Journal of Machine Learning Research 27 (2026) 1-28 Submitted 8/25; Revised 7/26; Published 9/26

SVarM: Linear Support Varifold Machines for Classification and Regression on Geometric Data

Emmanuel Hartman EHARTMA2@COUGARNET.UH.EDU
Department of Mathematics
University of Houston
Houston, TX 77204-3049, US

Nicolas Charon NCHARON@CENTRAL.UH.EDU
Department of Mathematics
University of Houston
Houston, TX 77204-3049, USA

Editor: Bharath Sriperumbudur

Abstract

Despite progress in the rapidly developing field of geometric deep learning, performing statistical analysis on geometric data—where each observation is a shape such as a curve, graph, or surface—remains challenging due to the non-Euclidean nature of shape spaces, which are defined as equivalence classes under invariance groups. Building machine learning frameworks that incorporate such invariances, notably to shape parametrization, is often crucial to ensure generalizability of the trained models to new observations. This work proposes *SVarM* to exploit varifold representations of shapes as measures and their duality with test functions $h : \mathbb{R}^n \times S^{n-1} \to \mathbb{R}$. This method provides a general framework akin to linear support vector machines but operating instead over the infinite-dimensional space of varifolds. We develop classification and regression models on shape data sets by introducing a neural network-based representation of the trainable test function h. This approach demonstrates strong performance and robustness across various shape graph and surface data sets, achieving results comparable to state-of-the-art methods while significantly reducing the number of trainable parameters.

Keywords: geometric deep learning, linear regression, classification, varifolds.

1. Introduction

Shape analysis, or geometric data science, is a field dedicated to building statistical and machine learning methods able to retrieve and analyze the morphological variability in geometric structures. This is a particularly central problem in applications such as computer vision or biomedical imaging where observations often come as segmented curves, surfaces, densities or other types of complex geometric data. Various approaches have been proposed, including Riemannian and elastic shape space models (Grenander, 1996; Michor and Mumford, 2007; Jermyn et al., 2012; Younes, 2019), topology based methods (Ghrist, 2008; Carlsson, 2009; Edelsbrunner and Morozov, 2017), and metric/functional matching frameworks (Bronstein et al., 2006; Mémoli, 2011; Ovsjanikov et al., 2012). These different methods have proved quite successful in tackling problems such as pairwise comparison, regression, classification or clustering for data sets of shapes. However, acquisition protocols have advanced significantly, leading to much larger and higher-resolution datasets. As a result, many existing methods do not scale well to modern applications involving tens of thousands of subjects, each with potentially hundreds of thousands of vertices. In view of the rapid development of new machine learning paradigms, in particular neural network models, and their impressive

©2026 Emmanuel Hartman and Nicolas Charon.

License: CC-BY 4.0, see https://creativecommons.org/licenses/by/4.0/. Attribution requirements are provided at http://jmlr.org/papers/v27/25-1908.html.

achievements in image processing and analysis tasks, one can reasonably expect similar tools to be able to address those challenges on geometric data. Yet, the very particular and intricate nature of shape spaces poses unique challenges when it comes to designing robust neural network models for shape analysis tasks. The importance of this problem within machine learning was recently recognized through the emergence of the research area known *geometric deep learning*, see the surveys (Bronstein et al., 2017; Cao et al., 2020; Gerken et al., 2023) and references therein for the current state-of-the-art.

In contrast with many works in geometric deep learning that focus on developing neural network architectures (in particular convolutional models) defined directly on e.g. surface meshes (Kostrikov et al., 2018; Jiang et al., 2019; Hanocka et al., 2019), the approach we propose in this paper instead relies on the embedding of shapes into a certain measure space. Specifically, our learning framework leverages the concept of *varifolds* from geometric measure theory (Almgren, 1966; Allard, 1972) along with the natural representation of a shape $q \in \mathcal{S}$—such as a curve or a surface in $\mathbb{R}^n$—as its corresponding varifold $\mu_q \in \mathcal{V}$. The mapping $q \mapsto \mu_q$ is invariant under reparameterization, thereby providing a (nonlinear) embedding of the shape space $\mathcal{S}$ into $\mathcal{V}$, the space of measures (or distributions) on $\mathbb{R}^n \times S^{n-1}$. This enables us to instead build the learning framework on the space of varifolds by exploiting the duality of these measures with continuous test functions on $\mathbb{R}^n \times S^{n-1}$.We simply propose to rely on affine functionals on $\mathcal{V}$ of the form $\mu \mapsto \langle \mu, h \rangle + \beta$. Here, $h \in C_0(\mathbb{R}^n \times S^{n-1})$ and $\beta \in \mathbb{R}$ become trainable model parameters that are to be estimated for a given statistical task. Due to its analogy with support vector machines (but considered in the infinite-dimensional varifold space), we coin our model **S**upport **Var**ifold **M**achines (SVarM) and derive formulations of the shape regression and classification problems based on this setting. This overall idea also shares certain similarities with recent works such as (Bhadra and Srivastava, 2025) which introduces a regression model for function data based on the square root velocity function (SRVF) representation. Our use of varifolds provides, arguably, a more convenient setting when dealing with geometric instead of functional data. Importantly, our approach allows for a principled handling of complex shapes without requiring prior regularization or registration. Additionally, by focusing on approximating continuous functions in a low-dimensional space, the model remains lightweight, requiring significantly fewer trainable parameters than traditional models operating on raw geometric data as inputs.

The presentation and contributions of the paper are organized into three main sections. Section 2 introduces the mathematical settings of shape spaces and of varifolds. We then proceed by deriving a few theoretical results related to the approximation properties of SVarM for regression and classification as well as its stability with respect to geometric perturbations. Section 3.2 details our proposed neural network architecture, which approximates the trainable test function h of the model. Finally, Section 4 demonstrates the practical effectiveness of the approach on various regression and classification problems for real surface and shape graph data sets.

2. Mathematical foundations and results

The concept of "shape" encompasses a wide variety of geometric structures (see e.g. (Younes, 2019; Bauer et al., 2014; Srivastava and Klassen, 2016) for general overviews on shape analysis). Although the framework we introduce can in principle apply to any type of shapes that can be viewed as measures or varifolds, our focus here will be on open/closed curves embedded in the plane or 3D space, surfaces in $\mathbb{R}^3$, and by extension structures known as shape graphs, see (Sukurdeep et al., 2022; Bal et al., 2024). The challenging aspect of dealing with such objects is that there is generally not a simple nor canonical way to represent them in a common Euclidean or even functional space. Indeed, a curve or surface is usually defined by a parametrization function $q : M \to \mathbb{R}^n$, where

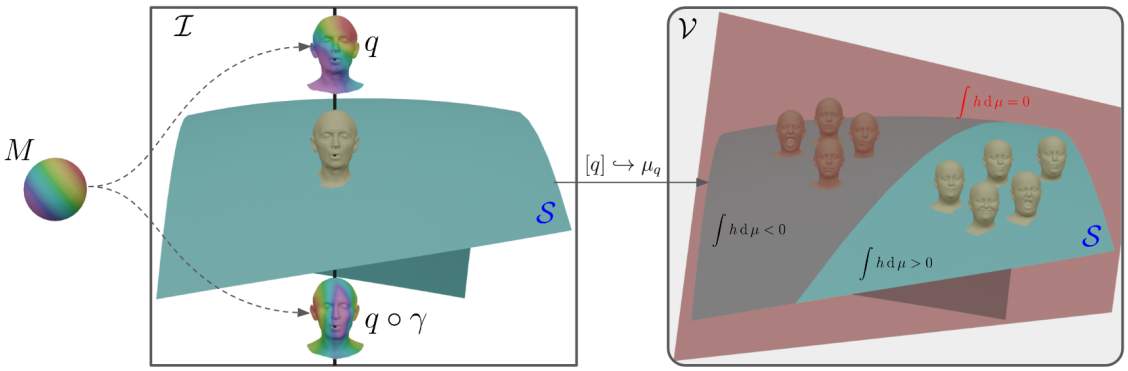


Figure 1: Representation of the quotient shape space $\mathcal{S}$ (left) and its embedding in varifold space $\mathcal{V}$ (right). A linear form $\mu \mapsto \langle \mu, h \rangle$ used by SVarM is visualized as the red hyperplane in $\mathcal{V}$.

$n = 2$ or 3 and M is the parameter manifold (for instance an interval or S^1 for open and closed curves, a disk or S^2 for open/closed surfaces...).

2.1 Shape Space

In this paper, we consider oriented rectifiable sets parameterized by bi-Lipschitz regular embeddings of of a compact, oriented, Riemannian manifold with or without boundary, M to $\mathbb{R}^n$. We denote such an embedding by $q \in \mathcal{I} \doteq \mathrm{Emb}_{\mathrm{Lip}}(M, \mathbb{R}^n)$. However, q itself does not uniquely identify the shape $q(M) \subseteq \mathbb{R}^n$. The action of any reparametrization $q \circ \gamma$, where $\gamma \in \mathcal{D} = \mathrm{Diff}_{\mathrm{Lip}}(M)$ is an orientation preserving bi-Lipschitz homeomorphism of M, leaves the shape $q(M)$ invariant. In other words, the actual shape associated to q is in essence the equivalence class $[q] = \{q \circ \gamma : \ \gamma \in \mathrm{Diff}_{\mathrm{Lip}}(M)\}$, and the shape space $\mathcal{S}$ can be thus identified with the quotient $\mathcal{S} = \mathcal{I}/\mathcal{D}$, see Figure 1 for an illustration. Note that quotienting out other shape invariances, such as to rigid motions of $\mathbb{R}^n$, can also be relevant in certain situations, but this work focuses on creating a parametrization-invariant model, as rigid alignment can usually be applied in a separate step. In fact, Section 4.1 and Section 4.2 illustrate how SVarM can itself be used for that very task.

We want to add that the above shape space setting also comprises discrete objects, e.g. polygonal curves and triangulated surfaces. This setting also includes discrete objects such as polygonal curves and triangulated surfaces. In these cases, M is subdivided into a 1- or 2- dimensional mesh, and the parametrization q is a Lipschitz piecewise affine map over that mesh. Consequently, $q(M)$ can be viewed as a union of adjacent cells in $\mathbb{R}^n$. The issue of parametrization invariance still remains very relevant in this situation. Namely, a proper learning model should exhibit some degree of robustness to operations such as mesh subdivision, simplification or resampling.

2.2 Varifolds

The concept of varifold goes back to early works in geometric measure theory, notably (Almgren, 1966; Allard, 1972), where it provided a natural and convenient setting to reframe geometric variational problems into problems on measure spaces. In more recent years, multiple works in applied mathematics have leveraged varifold representations in the context of shape registration (Charon and Trouvé, 2013; Kaltenmark et al., 2017; Lee et al., 2024; Stouffer et al., 2024), computational geometry (Hsieh and Charon, 2021; Buet et al., 2022; Paul et al., 2024), or computer vision (Pierson et al., 2022; Besnier et al., 2023; Hartman et al., 2025). One of the fundamental features explaining the interest in varifolds for these different problems is that it provides a representation of shapes,

both continuous and discrete, that is independent of parametrization/sampling, as we shall explain below. This suggests that it can be well-suited in the design of learning models on shape spaces.

Definition 1 *A varifold is a positive Radon measure on $\mathbb{R}^n \times S^{n-1}$ where S^{n-1} denotes the unit sphere of $\mathbb{R}^n$. We shall write, in short, $\mathcal{V} \doteq \mathcal{M}^+(\mathbb{R}^n \times S^{n-1})$ the space of oriented varifolds.*

To be exact, the above definition corresponds to dimension/codimension 1 oriented varifolds of $\mathbb{R}^n$. A more general construction can be obtained by replacing S^{n-1} with Grassmannians, c.f. (Allard, 1972) or (Charon and Trouvé, 2013). We will denote $\text{supp}(\mu) \subseteq \mathbb{R}^n \times S^{n-1}$ the *support* of a varifold $\mu \in \mathcal{V}$. Furthermore, by the usual Riesz-Markov-Kakutani representation theorem, any varifold μ can be viewed as an element of the dual to $C_0(\mathbb{R}^n \times S^{n-1})$, the space of continuous test functions on $\mathbb{R}^n \times S^{n-1}$ vanishing at infinity and we shall use the notation $\langle \mu, h \rangle = \int_{\mathbb{R}^n \times S^{n-1}} h(x,v) d\mu(x,v)$ for the duality bracket between $\mathcal{V}$ and $C_0(\mathbb{R}^n \times S^{n-1})$. In particular, a Dirac varifold $\delta_{(x_0,v_0)}$ for some $(x_0, v_0) \in \mathbb{R}^n \times S^{n-1}$ is such that $\langle \delta_{(x_0,v_0)}, h \rangle = h(x_0, v_0)$.

Now, going back to shape spaces, any Lipschitz embedded curve or surface $q : M \to \mathbb{R}^n$ (where M is of dimension 1 or $n-1$) can be canonically mapped to a varifold $\mu_q \in \mathcal{V}$ that is defined as the pushforward measure $\mu_q := (q, v_q)_* \, \text{vol}_q$, where v_q denotes the unit tangent/normal vector field and vol_q is the induced length/area measure on M. More explicitly, for any $h \in C_0(\mathbb{R}^n \times S^{n-1})$,

$$\langle \mu_q, h \rangle = \int_{\mathbb{R}^n \times S^{n-1}} h(x,v) \, \mathrm{d}\, \mu_q(x,v) = \int_M h(q(m), v_q(m)) \, \mathrm{d}\, \text{vol}_q(m).$$

Intuitively, μ_q represents the distribution of tangent/normal vectors at the different points of the shape $q(M)$ with local density given by the length/area measure along $q(M)$. A key property of the mapping $q \mapsto \mu_q$ is that it only depends on the oriented image set $q(M)$ and not on parametrizations:

Theorem 1 *For any $q \in \mathcal{I}$ and $\gamma \in \mathcal{D}$, it holds that $\mu_{q \circ \gamma} = \mu_q$. Consequently, one obtains a well-defined map $[q] \in \mathcal{S} \mapsto \mu_q \in \mathcal{V}$ which, in addition, is injective.*

Proof This result and its proof are similar to that of previous works such as (Charon and Trouvé, 2013; Bauer et al., 2019), albeit in a slightly different setting. We recap the main arguments, most of which rely on known results from geometric measure theory. Introducing all the related definitions and concepts in detail is out of the scope of this paper and we will thus refer the reader to (Simon, 2014) for a precise exposition.

First, we need to show that the Var map is well-defined in the sense that μ_q is independent of the parametrization. Denoting the embedded shape $S = q(M)$, S is a rectifiable subset of $\mathbb{R}^n$ of dimension d ($d = 1, 2$ being the dimension of the parameter space M). Then the area formula (Theorem 3.3 in (Simon, 2014)) implies in this case that:

$$\langle \mu_q, h \rangle = \int_M h(q(m), v_q(m)) d \, \text{vol}_q(m) = \int_S h(x, v_S(x)) d\mathcal{H}^d(x) \tag{1}$$

where $\mathcal{H}^d$ denotes the d-dimensional Hausdorff measure of $\mathbb{R}^n$, and $v_S(x)$ is the tangent/normal vector to S at x which is defined for $\mathcal{H}^d$-a.a. $x \in S$. This shows that $\langle \mu_q, h \rangle$ only depends on $S = q(M)$ and not the parametrization q itself.

Next, we need to prove the injectivity of the varifold map on $\mathcal{S}$. Let $[q]$, $[q']$ be two distinct shapes of $\mathcal{S}$. Then $S = q(M)$ and $S' = q'(M)$ are two rectifiable subsets of dimension d of $\mathbb{R}^n$. Let us first consider the case in which those two sets are distinct. Without loss of generality, we may then assume that $\mathcal{H}^d(S \setminus S') > 0$. By continuity of q and q' and the compactness of M, one can deduce that there exist $x_0 \in S$, $r > 0$ and $\epsilon > 0$ such that $\text{dist}((B(x_0, r), S') > \epsilon > 0$. We can thus

construct a nonnegative test function $h \in C_0(\mathbb{R}^n \times S^{n-1})$ such that $h(x,v) = 1$ for any $x \in B(x_0, r)$ and $h(x,v) = 0$ for $\|x - x_0\| \geq r + \epsilon$. From (1), we obtain that:

$$\langle \mu_q, h \rangle = \int_S h(x, v_S(x)) d\mathcal{H}^d(x) \geq \int_{B(x_0,r)\cap S} h(x, v_S(x)) d\mathcal{H}^d(x) = \mathcal{H}^d(B(x_0, r) \cap S) > 0$$

while

$$\langle \mu_{q'}, h \rangle = \int_{S'} h(x, v_S(x)) d\mathcal{H}^d(x) = 0$$

since the support of h does not intersect S' by construction. It follows that $\langle \mu_q, h \rangle \neq \langle \mu_{q'}, h \rangle$ and thus the two varifolds μ_q and $\mu_{q'}$ are distinct in $\mathcal{V}$. On the other hand, if the two sets S and S' coincide, then we have a well-defined bi-Lipschitz homeomorphism $\psi : M \to M$ with $\psi = q^{-1} \circ q'$ and where q^{-1} is the inverse of q on the image set $q(M) = q'(M)$. By definition, $q' = q \circ \psi$ and since we assumed that $[q] \neq [q']$, ψ cannot be orientation-preserving. This implies that one can find a subset $B \subseteq M$ with $\mathcal{H}^d(q'(B)) \neq 0$ such that for any $x \in B$, we have $v_{q\circ\psi}(x) = -v_{q'}(x)$. Since $\mathcal{H}^d(q'(B)) \neq 0$, there exists $\phi : \mathbb{R}^3 \to \mathbb{R}$ such that $\phi(x) = 0$ for $x \in q'(M - B)$ and $u := \int_B \phi(q'(x)) v_{q'}(x) \, d\,\mathrm{vol}_{q'}(x) \neq 0$. Therefore,

$$\int_B \phi(q')(v_{q'} \cdot u) \, d\,\mathrm{vol}_{q'} = |u|^2 > 0, \text{ and}$$
$$\int_B \phi(q \circ \psi)(v_{q\circ\psi} \cdot u) \, d\,\mathrm{vol}_{q\circ\psi} = \int_B \phi(q')(-v_{q'} \cdot u) \, d\,\mathrm{vol}_{q'} < 0$$

It follows that for $h \in C_0(\mathbb{R}^n \times S^{n-1})$ defined by $h(x,v) = \phi(x)(u \cdot v)$, we have $\langle \mu_q, h \rangle \neq \langle \mu_{q'}, h \rangle$ and so $\mu_q \neq \mu_{q'}$. ■

This property will be fundamental for what follows, as it allows us to embed any shape of $\mathcal{S}$ into $\mathcal{V}$ (see Figure 1) and exploit the dual nature of $\mathcal{V}$.

Another advantage of the varifold framework is that it applies very naturally to the discrete geometric setting as well. Specifically, when q is a piecewise affine embedding defined over a polyhedral mesh on M, the shape $q(M)$ is made of a reunion of polyhedral cells $\{F_i\}$ in $\mathbb{R}^n$ (segments or triangles) which can each be approximated by a single weighted Dirac varifold $w_i \delta_{(x_i, v_i)}$, with $x_i \in \mathbb{R}^n$, $v_i \in S^{n-1}$ and $w_i > 0$ being respectively the center of mass, unit tangent/normal vector and length/area of F_i. This leads to a discrete varifold approximation $\mu_q \approx \sum_i w_i \delta_{(x_i,v_i)}$ of q. This type of approximations in $\mathcal{V}$ has been routinely used and analyzed in works such as (Kaltenmark et al., 2017; Buet et al., 2022; Paul et al., 2024), to which we refer for more details.

2.3 Theoretical Properties of SVarM

As stated in the introduction, the core of the SVarM framework is to represent affine weak-* continuous forms on $\mathcal{V}$ which can then trained as regressors or classifiers on shape spaces via the identification given by Theorem 1. Before we delve into the training process itself in Section 3, we shall first discuss the precise mathematical class of affine functionals that will be considered in SVarM. Moreover, we will establish a few theoretical properties of those functionals that are directly relevant for the shape regression or classification problems considered later on.

2.3.1 Existence of Regression Functions and Separators

Let us first recall a few definitions related to the different topologies on varifolds. The space of signed measures $\mathcal{M}(\mathbb{R}^n \times S^{n-1})$ can be identified with the topological dual to the space of test functions $C_0(\mathbb{R}^n \times S^{n-1})$ equipped with the $\|\cdot\|_\infty$ norm. As such, different topologies are usually

considered on $\mathcal{M}(\mathbb{R}^n \times S^{n-1})$ and by restriction on $\mathcal{V}$. The first of which is the dual metric topology, i.e. the topology induced by the *total variation* norm:

$$\|\mu\|_{TV} = \sup_{\|h\|_\infty \leq 1} |\langle \mu, h \rangle| = |\mu|(\mathbb{R}^n \times S^{n-1})$$

In particular, for a varifold μ_q associated to $q \in \mathcal{I}$, one can see from (1) that $\|\mu_q\|_{TV}$ is the total length/area of the shape $q(M)$. The TV norm induces a rather strong topology on varifolds: for instance, the TV norm between two Dirac masses $\delta_{(x_0,v_0)}$ and $\delta_{(x_1,v_1)}$ is 2 whenever $(x_0, v_0) \neq (x_1, v_1)$ and 0 when the position and direction coincide. It is often more natural to consider the *weak-* topology* on $\mathcal{M}(\mathbb{R}^n \times S^{n-1})$. By definition, a sequence (μ_i) weak-* converges to μ, denoted $\mu_i \overset{*}{\rightharpoonup} \mu$, when $\langle \mu_i, h \rangle \xrightarrow[i\to\infty]{} \langle \mu, h \rangle$ for all $h \in C_0(\mathbb{R}^n \times S^{n-1})$. Note that the weak-* convergence can be also locally metrized by a variety of usual metrics on $\mathcal{M}(\mathbb{R}^n \times S^{n-1})$ such as the bounded Lipschitz norm or the Wasserstein metrics of optimal transport (for normalized measures), c.f. (Villani, 2008) Chap. 6.

As a consequence of standard results on Banach spaces and weak-* topology, specifically Proposition 3.14 in (Brézis, 2011), any linear function $\mathcal{V} \to \mathbb{R}$ that is continuous with respect to the weak-* convergence is necessarily of the form $\mu \mapsto \langle \mu, h \rangle$ for some test function $h \in C_0(\mathbb{R}^n \times S^{n-1})$. It follows that the set of affine weak-* continuous functions $\varphi : \mathcal{V} \to \mathbb{R}$ are of the form $\varphi(\mu) = \langle \mu, h \rangle + \beta$ for some $h \in C_0(\mathbb{R}^n \times S^{n-1})$ and $\beta \in \mathbb{R}$. Thus we have the following:

Proposition 2 *The set of functionals modeled by SVarM are exactly the affine weak-* continuous functionals on $\mathcal{V}$.*

However, this does not hold anymore when the continuity is with respect to the stronger TV metric topology due to the lack of reflexiveness of $\mathcal{V}$. In that case, one can only recover the following approximation bound over any finite set of varifolds:

Proposition 3 *Suppose $\varphi : \mathcal{V} \to \mathbb{R}$ is an affine function and continuous with respect to the TV norm topology and fix $C \subseteq \mathcal{V}$ any finite subset of the space of varifolds. For every $\epsilon > 0$, there exists $h \in C_0(\mathbb{R}^n \times S^{n-1}, \mathbb{R})$ and $\beta \in \mathbb{R}$ such that*

$$\sup_{\mu \in C} |\varphi(\mu) - (\langle \mu, h \rangle + \beta)| < \epsilon.$$

Proof Let us first consider the case of linear forms i.e. let $\varphi : \mathcal{M}(\mathbb{R}^n \times S^{n-1}) \to \mathbb{R}$ be a continuous linear functional, in other words φ belongs to the bidual space $C_0(\mathbb{R}^n \times S^{n-1}, \mathbb{R})^{**}$. By Goldstine theorem (Rudin, 1991), $C_0(\mathbb{R}^n \times S^{n-1}, \mathbb{R})$ is dense in its bi-dual with respect to the weak-* topology on $C_0(\mathbb{R}^n \times S^{n-1}, \mathbb{R})^{**}$. Specifically, for the canonical embedding J defined by:

$$\begin{aligned} J : C_0(\mathbb{R}^n \times S^{n-1}, \mathbb{R}) &\to C_0(\mathbb{R}^n \times S^{n-1}, \mathbb{R})^{**} \\ h &\mapsto (\mu \in \mathcal{M}(\mathbb{R}^n \times S^{n-1}) \mapsto \langle \mu, h \rangle) \end{aligned}$$

we have $J(C_0(\mathbb{R}^n \times S^{n-1}, \mathbb{R}))$ dense in $C_0(\mathbb{R}^n \times S^{n-1}, \mathbb{R})^{**}$. Consequently, we can find a sequence $(h_i) \subseteq C_0(\mathbb{R}^n \times S^{n-1}, \mathbb{R})$ so that $J(h_i)(\mu) = \langle \mu, h_i \rangle \to \varphi(\mu)$ as $i \to \infty$ for all $\mu \in \mathcal{M}(\mathbb{R}^n \times S^{n-1})$. Now , since C is assumed to be finite, choosing i large enough so that $|\varphi(\mu) - \langle \mu, h_i \rangle| < \epsilon$ for all $\mu \in C$ leads to the desired results. The extension to affine continuous functionals is then straightforward by incorporating the extra bias term $\beta \in \mathbb{R}$. ■

For classification problems, we are interested in determining when two given sets of varifolds (or shapes) can be separated by SVarM.

Definition 4 *Two disjoint sets $C_1, C_2 \subseteq \mathcal{V}$ are strictly separated by some test function h when:*

$$\sup_{\mu \in C_1} \langle \mu, h \rangle \; < \; \inf_{\nu \in C_2} \langle \nu, h \rangle \tag{2}$$

Note that, by reintroducing the extra bias parameter $\beta \in \mathbb{R}$, the above is also equivalent to $\sup_{\mu \in C_1} \langle \mu, h \rangle + \beta < 0 < \inf \nu \in C_2 \langle \nu, h \rangle + \beta$, which is analogous to the separation by a hyperplane property in classical SVM. A first generic result, that follows from the Hahn-Banach separation theorem is:

Theorem 2 *Suppose $C_1, C_2 \subseteq \mathcal{V}$ are convex weak-* compact subsets such that $C_1 \cap C_2 = \emptyset$. Then there exists a test function $h \in C_0(\mathbb{R}^n \times S^{n-1}, \mathbb{R})$ that separates C_1 and C_2 in the sense of (2).*

Proof Let $C_1, C_2 \subseteq \mathcal{M}(\mathbb{R}^n \times S^{n-1})$ be convex weak-* compact subsets such that $C_1 \cap C_2 = \emptyset$. Since the space $\mathcal{M}(\mathbb{R}^n \times S^{n-1})$ with the weak-* topology is a locally convex space, the Hahn-Banach separation theorem (Theorem 1.7 in (Brézis, 2011)) applies and there exists a weak-* continuous linear form $\varphi : \mathcal{V} \to \mathbb{R}$ such that:

$$\sup_{\mu \in C_1} \varphi(\mu) < \inf_{\nu \in C_2} \varphi(\nu).$$

As already noted above, φ must be of the form $\varphi(\mu) = \langle \mu, h \rangle$ for some test function $h \in C_0(\mathbb{R}^n \times S^{n-1}, \mathbb{R})$, which leads to the result. ■

The weak-* compactness assumption is relatively natural as, by the Banach-Alaoglu theorem, closed sets of varifolds with bounded total variation are in particular weak-* compact. However, the condition of convexity in $\mathcal{V}$ in Theorem 2 is rather abstract and not immediate to connect to concrete situations of interest. We will thus look to derive more specific conditions for certain families of measure sets C_1 and C_2. First of all, we can focus on the situation where C_1 and C_2 are both finite subsets of $\mathcal{V}$, which is natural when considering classification and regression problems in practice. In that case, C_1 and C_2 are automatically weak-* compact. We can then state the following simpler condition:

Theorem 3 *Let $C_1, C_2 \subseteq \mathcal{V}$ be disjoint finite sets of varifolds which satisfy the property that either 1) for each $\nu \in C_2$, $supp(\nu) \not\subseteq \bigcup_{\mu \in C_1} supp(\mu)$ or 2) for each $\mu \in C_1$, $supp(\mu) \not\subseteq \bigcup_{\nu \in C_2} supp(\nu)$. Then there exists $h \in C_0(\mathbb{R}^n \times S^{n-1}, \mathbb{R})$ that separates C_1 and C_2.*

Proof Without loss of generality, let us assume that $C_1, C_2 \subseteq \mathcal{S}$ are two finite sets in $\mathcal{V}$ with $C_1 = \{\mu_1, \ldots, \mu_p\}$ and $C_2 = \{\nu_1, \ldots, \nu_q\}$, such that for each $\nu \in C_2$, $\text{supp}(\nu) \not\subseteq \bigcup_{\mu \in C_1} \text{supp}(\mu)$. Let us denote by $\text{conv}(C_1)$ and $\text{conv}(C_2)$ the convex hulls of the two sets C_1 and C_2 in $\mathcal{V}$. By contradiction, if $\text{conv}(C_1) \cap \text{conv}(C_2) \neq \emptyset$, there exist $(\alpha_1, \ldots, \alpha_p)$ and $(\beta_1, \ldots, \beta_q)$ in the probability simplexes of $\mathbb{R}^p_+$ and $\mathbb{R}^q_+$ respectively such that:

$$\sum_{i=1}^{p} \alpha_i \mu_i = \sum_{j=1}^{q} \beta_j \nu_j \implies \text{supp}\left(\sum_{j=1}^{q} \beta_j \nu_j\right) = \text{supp}\left(\sum_{i=1}^{p} \alpha_i \mu_i\right)$$

Now, on the one hand, $\text{supp}\left(\sum_{i=1}^{p} \alpha_i \mu_i\right) \subseteq \bigcup_{\mu \in C_1} \text{supp}(\mu)$. On the other hand, we have:

$$\text{supp}\left(\sum_{j=1}^{q} \beta_j \nu_j\right) = \bigcup_{j : \beta_j \neq 0} \text{supp}(\nu_j).$$

Since not all β_j's can be 0, we deduce that there exist j such that $\text{supp}(\nu_j) \subseteq \bigcup_{\mu \in C_1} \text{supp}(\mu)$ which contradicts the assumption. Therefore, $\text{conv}(C_1)$ and $\text{conv}(C_2)$ are two disjoint convex sets in $\mathcal{V}$.

They are also weak-* compact as the convex hulls of finite sets; indeed $\mathrm{conv}(C_1)$ is the image of the compact probability simplex by the continuous map $(\alpha_1, \dots, \alpha_p) \mapsto \sum_{i=1}^{p} \alpha_i \mu_i$, and similarly for $\mathrm{conv}(C_2)$. Then, by Theorem 2, we deduce that $\mathrm{conv}(C_1)$ and $\mathrm{conv}(C_2)$ and thus C_1 and C_2 can be separated by a test function $h \in C_0(\mathbb{R}^n \times S^{n-1})$. ■

Remark 5 *Note that this only provides a sufficient but not necessary condition of separability. For instance, if C_1 and C_2 are both made of varifolds all having the same compact support set $K \subseteq \mathbb{R}^n \times S^{n-1}$ but such that their total mass satisfy $\mu(K) < \nu(K)$ for all $\mu \in C_1$ and $\nu \in C_2$, then C_1 and C_2 are still separable by taking any test function h that is constant equal to 1 on K.*

The above condition on the support of the varifolds becomes even more explicit when considering varifolds that are associated to continuous shapes via the mapping $q \in \mathcal{I} \mapsto \mu_q$ introduced in Section 2.2. We get specifically:

Corollary 6 *Let Σ_1, Σ_2 be finite sets of continuous shapes in $\mathcal{S}$ such that either 1) for each $q' \in \Sigma_2$, $q'(M) \not\subseteq \bigcup_{q \in \Sigma_1} q(M)$ or 2) for each $q \in \Sigma_1$, $q(M) \not\subseteq \bigcup_{q' \in \Sigma_2} q'(M)$. Then the two sets Σ_1, Σ_2 are strictly separable by a test function $h \in C_0(\mathbb{R}^n \times S^{n-1}, \mathbb{R})$.*

This indeed follows directly from Theorem 3 since for $q \in \mathcal{I}$, $\mathrm{supp}(\mu_q) = \{(q(m), v_q(m)) : m \in M\}$ and thus the assumption of Corollary 6 implies the assumption of Theorem 3.

Remark 7 *When the condition of Corollary 6 is not satisfied, it is possible to find sets that are not separable within the SVarM setting. Figure 2 shows such an example in which the two sets (green and red), each made of two shapes, cannot be separated. This is because the two convex hulls of the varifolds associated to the shapes in those sets do intersect in $\mathcal{V}$; specifically, one can see that the varifold $\frac{1}{2}\mu_{q_1} + \frac{1}{2}\mu_{q_2}$, q_1 and q_2 being the two shapes of the first (green) set, is equal to $\frac{1}{2}\mu_{q'_1} + \frac{1}{2}\mu_{q'_2}$ where q'_1 and q'_2 are the two shapes in the second (red) set, and therefore belongs to the intersection of the two convex hulls. This example further shows that the weaker and somewhat more natural condition that the support of any varifold in one of the two sets is not included in the support of any varifold from the other set is not sufficient to guarantee separability by SVarM since the example of Figure 2 is not separable and does satisfy this condition (but does not satisfy that of Corollary 6).*

One can also obtain a corresponding result for discrete varifolds such as varifolds approximating discrete meshes. It is formulated below and is again a direct corollary of Theorem 3:

Corollary 8 *Let C_1 and C_2 be two finite sets of discrete varifolds that are given by $C_1 = \left\{\sum_{i=1}^{N_k} r_i^{(k)} \delta_{(x_i^{(k)}, v_i^{(k)})}\right\}_{k=1,\dots,K_1}$ and $C_2 = \left\{\sum_{i=1}^{\tilde{N}_k} \tilde{r}_i^{(k)} \delta_{(\tilde{x}_i^{(k)}, \tilde{v}_i^{(k)})}\right\}_{k=1,\dots,K_2}$. Assume that these sets satisfy either of the two following conditions:*

1. *For each $k = 1, \dots, K_1$, $\{(x_i^{(k)}, v_i^{(k)}), i = 1, \dots N_k\} \not\subseteq \bigcup_{l=1}^{K_2} \{(\tilde{x}_i^{(l)}, \tilde{v}_i^{(l)}), i = 1, \dots \tilde{N}_l\}$*
2. *For each $k = 1, \dots, K_2$, $\{(\tilde{x}_i^{(k)}, \tilde{v}_i^{(k)}), i = 1, \dots \tilde{N}_k\} \not\subseteq \bigcup_{l=1}^{K_1} \{(x_i^{(l)}, v_i^{(l)}), i = 1, \dots N_l\}$*

Then the two sets C_1, C_2 are separable by a test function $h \in C_0(\mathbb{R}^n \times S^{n-1}, \mathbb{R})$.

2.3.2 Stability with Respect to Shape Deformations and Perturbations

Another key theoretical question regarding the SVarM approach is the stability of the model under perturbations of the input data. We show below that this type of result can hold under the assumption that h is Lipchitz continuous. As described in Section 3.2, we will approximate h

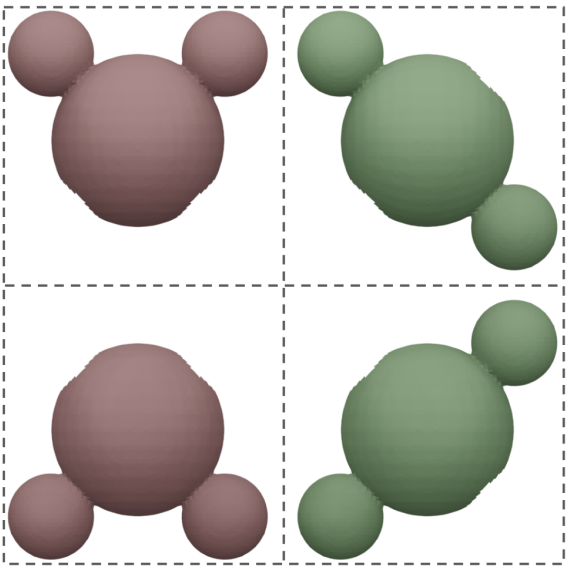


Figure 2: An example of two sets of shapes (in green and red) that cannot be separated within the SVarM framework. Note also that neither shape in one set is a subset of a shape from the other set, showing that this is not a sufficient condition of separability for SVarM.

using a multilayer perceptron (MLP) architecture. For fixed weights (i.e., after training), an MLP is a Lipchitz function, and one can efficiently compute an upper bound on its Lipchitz constant (Virmaux and Scaman, 2018). Let us first examine stability for general varifolds in $\mathcal{V}$:

Theorem 4 *Let $\mu, \nu \in \mathcal{V}$ and $h \in C_0(\mathbb{R}^n \times S^{n-1})$ such that h is Lipchitz with constant K. Then,*

$$|\langle \mu, h \rangle - \langle \nu, h \rangle| \leq \|h\|_\infty \cdot \|\mu - \nu\|_{TV}$$

and

$$|\langle \mu, h \rangle - \langle \nu, h \rangle| \leq |\|\mu\|_{TV} - \|\nu\|_{TV}| \cdot \|h\|_\infty + \|\mu\|_{TV} \cdot K \cdot W_1\left(\frac{\mu}{m_\mu}, \frac{\nu}{m_\nu}\right)$$

where W_1 denotes the 1-Wasserstein distance.

Proof Let $\mu, \nu \in \mathcal{V}$ and $h \in C_0(\mathbb{R}^n \times S^{n-1})$ such that h is Lipchitz with constant K.

Bound 1: Based on the definition of the TV norm on $\mathcal{M}(\mathbb{R}^n \times S^{n-1})$ given above, we can bound the difference in the evaluation of μ and ν on $h \neq 0$ as follows:

$$\begin{aligned}
|\langle \mu, h \rangle - \langle \nu, h \rangle| &= \left| \int_{\mathbb{R}^n \times S^{n-1}} h \,\mathrm{d}\,\mu - \int_{\mathbb{R}^n \times S^{n-1}} h \,\mathrm{d}\,\nu \right| \\
&\left| \int_{\mathbb{R}^n \times S^{n-1}} h \,\mathrm{d}(\mu - \nu) \right| \\
&\left| \int_{\mathbb{R}^n \times S^{n-1}} \frac{h}{\|h\|_\infty} \,\mathrm{d}(\mu - \nu) \right| \|h\|_\infty \\
&\leq \|\mu - \nu\|_{\mathrm{TV}} \|h\|_\infty.
\end{aligned}$$

Bound 2: For brevity, let us define the total masses $m_\mu = \|\mu\|_{\mathrm{TV}}, m_\nu = \|\nu\|_{\mathrm{TV}}$ and the corresponding normalized measures $\overline{\mu} = \frac{\mu}{m_\mu}, \overline{\nu} = \frac{\nu}{m_\nu}$. We begin by with a series of inequalities:

$$\begin{aligned}
|\langle \mu, h\rangle - \langle \nu, h\rangle| =& |m_\mu \langle \overline{\mu}, h\rangle - m_\nu \langle \overline{\nu}, h\rangle| \\
=& |m_\mu \langle \overline{\mu}, h\rangle - m_\mu \langle \overline{\nu}, h\rangle + m_\mu \langle \overline{\nu}, h\rangle - m_\nu \langle \overline{\nu}, h\rangle| \\
\leq& |m_\mu \langle \overline{\nu}, h\rangle - m_\nu \langle \overline{\nu}, h\rangle| + |m_\mu \langle \overline{\mu}, h\rangle - m_\mu \langle \overline{\nu}, h\rangle| \\
\leq& |m_\mu - m_\nu| \cdot \left| \int_{\mathbb{R}^n \times S^{n-1}} h \,\mathrm{d}\,\overline{\nu} \right| + m_\mu \left| \int_{\mathbb{R}^n \times S^{n-1}} h \,\mathrm{d}\,\overline{\mu} - \int_{\mathbb{R}^n \times S^{n-1}} h \,\mathrm{d}\,\overline{\nu} \right| \\
\leq& |m_\mu - m_\nu| \cdot \|h\|_\infty + m_\mu \left| \int_{\mathbb{R}^n \times S^{n-1}} h \,\mathrm{d}\,\overline{\mu} - \int_{\mathbb{R}^n \times S^{n-1}} h \,\mathrm{d}\,\overline{\nu} \right|.
\end{aligned}$$

Next we define the function $\overline{h} := \frac{h}{K}$. Since the Lipchitz constant of h is K, it follows that $\overline{h}$ is $1-$Lipchitz. Thus we can write

$$\left| \int_{\mathbb{R}^n \times S^{n-1}} h \,\mathrm{d}\,\overline{\mu} - \int_{\mathbb{R}^n \times S^{n-1}} h \,\mathrm{d}\,\overline{\nu} \right| = K \left| \int_{\mathbb{R}^n \times S^{n-1}} \overline{h} \,\mathrm{d}\,\overline{\mu} - \int_{\mathbb{R}^n \times S^{n-1}} \overline{h} \,\mathrm{d}\,\overline{\nu} \right|.$$

By the Kantorovich-Rubinstein formulation of the 1-Wasserstein distance,

$$W_1(\overline{\mu}, \overline{\nu}) = \sup_{\|f\|_{W^{1,\infty}} \leq 1} \left| \int_{\mathbb{R}^n \times S^{n-1}} f \,\mathrm{d}\,\overline{\mu} - \int_{\mathbb{R}^n \times S^{n-1}} f \,\mathrm{d}\,\overline{\nu} \right| \geq \left| \int_{\mathbb{R}^n \times S^{n-1}} \overline{h} \,\mathrm{d}\,\overline{\mu} - \int_{\mathbb{R}^n \times S^{n-1}} \overline{h} \,\mathrm{d}\,\overline{\nu} \right|$$

and we conclude that:

$$|\langle \mu, h\rangle - \langle \nu, h\rangle| \leq |m_\mu - m_\nu| \cdot \|h\|_\infty + m_\mu \cdot K \cdot W_1(\overline{\mu}, \overline{\nu}).$$

■

The first bound involving the total variation between μ and ν implies stability of $\langle \mu, h\rangle$ under partial occlusion of some parts of a shape (c.f. experiments of Section 4.5) while the second bound can be thought of as a control of the changes in $\langle \mu, h\rangle$ under variations of the total mass and support. Our second result demonstrates that the outputs of SVarM are stable in the space of shapes with respect to sufficiently regular perturbations in $\mathcal{S}$:

Theorem 5 *Let $h \in C_0(\mathbb{R}^n \times S^{n-1})$ such that h is Lipschitz with constant K and q, q' be bi-Lipschitz embeddings from a d-dimensional manifold M into $\mathbb{R}^n$ (with either $d = 1$ or $d = n-1$) such that $\|q - q'\|_{W^{1,\infty}} \leq C$. Then*

$$|\langle \mu_q, h\rangle - \langle \mu_{q'}, h\rangle| \leq C A_q (K\sqrt{2dL^{2d-1} + 1} + d^{d+1/2} L^{4d-1} \|h\|_\infty)$$

where $A_q = \int_M \mathrm{dvol}_q$ is the total length/area of $q(M)$, and L the maximum of the Lipschitz constants of q and q'.

Proof Let $h \in C_0(\mathbb{R}^n \times S^{n-1})$ be such that h is Lipschitz with constant K and q, q' be bi-Lipschitz embeddings from a d-dimensional manifold M into $\mathbb{R}^n$ such that $\|q - q'\|_{W^{1,\infty}} \leq C$. We estimate the difference between the actions of the varifolds associated with q and q' on h by:

$$\begin{aligned}
|\langle \mu_q, h\rangle - \langle \mu_{q'}, h\rangle| =& \left| \int_M h(q, v_q) \,\mathrm{d}\,\mathrm{vol}_q - \int_M h(q', v_{q'}) \,\mathrm{d}\,\mathrm{vol}_{q'} \right| \\
=& \left| \int_M h(q, v_q)\sqrt{|Jq|} - h(q', v_{q'})\sqrt{|Jq|} + h(q', v_{q'})\sqrt{|Jq|} - h(q', v_{q'})\sqrt{|Jq'|} \,\mathrm{d}\,m \right| \\
\leq& \int_M \left| h(q, v_q) - h(q', v_{q'}) \right| \sqrt{|Jq|} \,\mathrm{d}\,m + \int_M |h(q', v_{q'})| \frac{\left| \sqrt{|Jq|} - \sqrt{|Jq'|} \right|}{\sqrt{|Jq|}} \sqrt{|Jq|} \,\mathrm{d}\,m.
\end{aligned}$$

where $Jq(m) = Dq(m)^T Dq(m)$ and $|Jq(m)|$ is the absolute value of the determinant of $Jq(m)$. Since q, q' are bi-Lipchitz embeddings, there exists $L > 0$ such that the singular values of Dq and Dq' are in $[1/L, L]$ almost everywhere on M. Therefore, for almost every $m \in M$, $\frac{1}{L^d} < \sqrt{|Jq|}, \sqrt{|Jq'|} < L^d$ and $\frac{\sqrt{d}}{L} < \|Dq\|_F, \|Dq'\|_F < \sqrt{d}L$. Moreover, for almost every $m \in M$,

$$\frac{\left|\sqrt{|Jq|} - \sqrt{|Jq'|}\right|}{\sqrt{|Jq|}} \leq \frac{L^{2d}}{2} \left||Jq| - |Jq'|\right|.$$

To control the differences in the Jacobian determinants we leverage the following perturbation bound for determinants using the Frobenius norm.

Lemma 9 (Theorem 2.12 of (Ipsen and Rehman, 2008)) *Let A and E be $d \times d$ matrices. Then $||A| - |A + E|| \leq d\|E\|_F \max(\|A\|_F, \|A + E\|_F)^{d-1}$.*

Substituting $A = Jq$ and $E = Jq - Jq'$ we have directly that

$$||Jq| - |Jq'|| \leq d\|Jq - Jq'\|_F \max(\|Jq\|_F, \|Jq'\|_F)^{d-1}$$

Since $\|Jq\|_F \leq \|Dq\|_F^2 \leq dL^2$ and $\|Jq'\|_F \leq \|Dq'\|_F^2 \leq dL^2$, it follows that

$$||Jq| - |Jq'|| \leq d\|Jq - Jq'\|_F (dL^2)^{d-1} = d^d \|Jq - Jq'\|_F L^{2(d-1)}$$

Finally, to control the Frobenius norm of the difference of the Jacobians of q and q' we use,

$$\|Jq - Jq'\|_F \leq (\|Dq\|_F + \|Dq'\|_F)\|Dq - Dq'\|_F \leq 2\sqrt{d}LC.$$

So, combining these inequalities we obtain

$$\int_M |h(q, v_q)| \frac{\left|\sqrt{|Jq|} - \sqrt{|Jq'|}\right|}{\sqrt{|Jq|}} \sqrt{|Jq|} \,\mathrm{d}\, m \leq d^{d+1/2} C L^{4d-1} \int_M |h(q, v_q)| \sqrt{|Jq|} \,\mathrm{d}\, m \leq d^{d+1/2} C L^{4d-1} \|h\|_\infty A_q.$$

where $A_q = \int_M \sqrt{|Jq|} \,\mathrm{d}\, m = \int_M \mathrm{dvol}_q$. Moreover, v_q is the unit tangent (or normal) vector of q computed by $v_q = \frac{\partial_1 q \wedge \ldots \wedge \partial_d q}{\sqrt{|J_q|}}$. Therefore, for almost every $m \in M$,

$$\|v_q - v_{q'}\| \leq 2L^d \|(\partial_1 q \wedge \ldots \partial_d q) - (\partial_1 q' \wedge \ldots \partial_d q')\| \leq 2L^d \sum_{i=1}^{d} \|(\partial_1 q \wedge \ldots \wedge (\partial_i q - \partial_i q') \wedge \ldots \wedge \partial_d q')\|$$

$$\leq 2L^d \sum_{i=1}^{d} \|\partial_1 q\| \cdot \ldots \cdot \|\partial_i q - \partial_i q'\| \cdot \ldots \cdot \|\partial_d q'\| \leq 2dL^{2d-1} \|Dq - Dq'\| \leq 2dL^d C.$$

where we have again used Hadamard's inequality to bound the norm of the exterior product. So, $\sqrt{\|q - q'\|^2 + \|v_q - v_{q'}\|^2} \leq C\sqrt{2dL^{2d-1} + 1}$ almost everywhere. Since h is Lipschitz with constant K, we have for almost every $m \in M$,

$$|h(q, v_q) - h(q', v_{q'})| \leq K\sqrt{\|q - q'\|^2 + \|v_q - v_{q'}\|^2} \leq CK\sqrt{2dL^{2d-1} + 1}.$$

Therefore,

$$\int_M \left|h(q, v_q) - h(q', v_{q'})\right| \sqrt{|Jq|} \,\mathrm{d}\, m \leq CK\sqrt{2dL^{2d-1} + 1} \int_M \sqrt{|Jq|} \,\mathrm{d}\, m = CKA_q\sqrt{2dL^{2d-1} + 1}$$

Thus, we have obtained that:

$$|\langle \mu_q, h\rangle - \langle \mu_{q'}, h\rangle| \leq C A_q (K\sqrt{2dL^{2d-1}+1} + d^{d+1/2} L^{4d-1} \|h\|_\infty).$$

■

3. SVarM regression and classification models

The SVarM framework for regression or classification consists in estimating a map $h : \mathbb{R}^n \times S^{n-1} \to \mathbb{R}$ (for $n = 2$ or 3) which, together with the bias β, represents any weak-* continuous affine map on $\mathcal{V}$ as we have previously shown in Proposition 2. In the absence of any clear finite-dimensional family of functions to represent h, we focus here on non-parametric models; we first examine the situation in which h is taken in some reproducing kernel Hilbert space of functions on $\mathbb{R}^n \times S^{n-1}$ leading to a formulation akin to kernel regression/classification, before relying instead on the representation of h via a neural network model which we argue is a more effective approach when dealing with large data sets. From now on, we shall assume that a training set $\{\mu^{(k)}, y^{(k)}\}_{k=1,...,R}$ is given, where $\{\mu^{(1)}, ..., \mu^{(R)}\}$ is a set of varifolds (computed from the corresponding input shapes) and $\{y^{(1)}, ..., y^{(R)}\}$ are the corresponding labels (either real numbers in regression problems or integers giving class membership for classification).

3.1 Reproducing kernel Hilbert space approach

Reproducing kernel Hilbert spaces (RKHS) in the context of varifolds have been considered in several previous works (Charon and Trouvé, 2013; Kaltenmark et al., 2017; Hsieh and Charon, 2021; Paul et al., 2024), primarily as a way to build explicit metrics on the space of curves and surfaces. One can rely on a similar construction in order to specify the space of test functions h, in which case the problem is related to *kernel ridge regression* with a theoretically explicit form for the solution. Before discussing this approach, we begin with a brief recap of the main background material on RKHS, referring the reader to e.g. (Aronszajn, 1950) or (Younes, 2019) Chap. 8 for more detailed presentations.

Let us start by introducing a positive definite kernel function K on $\mathbb{R}^n \times S^{n-1}$, i.e. a function $K : (\mathbb{R}^n \times S^{n-1}) \times (\mathbb{R}^n \times S^{n-1}) \to \mathbb{R}$ which is symmetric and satisfies the property that for any $(x_i, v_i)_{i=1,...,m} \in (\mathbb{R}^n \times S^{n-1})^m$ and $(w_i) \in \mathbb{R}^m$, it holds that:

$$\sum_{i,j=1}^{m} w_i K((x_i, v_i), (x_j, v_j)) w_j \geq 0$$

with strict inequality as long as all points are distinct and $(w_i)_{i=1,...,m} \neq 0$, in other words the matrix $[K((x_i, v_i), (x_j, v_j))]_{i,j}$ is a symmetric positive definite $m \times m$ matrix. Such kernel functions can in particular be constructed, as in (Charon and Trouvé, 2013; Kaltenmark et al., 2017), by taking the tensor product of a positive definite kernel on $\mathbb{R}^n$ and S^{n-1}, for which there are many known kernel families. If one makes the additional assumption that K is a continuous function vanishing at infinity, it follows from the classical theory of kernels, c.f. (Younes, 2019), that there exists a unique RKHS W of functions on $\mathbb{R}^n \times S^{n-1}$ such that W is continuously embedded into $C_0(\mathbb{R}^n \times S^{n-1})$ which is "associated" to K. The previous statement means specifically that the Hilbert inner product on W between any two functions $f(\cdot) = K((x, v), \cdot)$ and $g(\cdot) = K((x', v'), \cdot)$ is explicitly given by $\langle f, g\rangle_W = K((x, v), (x', v'))$. More general continuous embeddings of W into

spaces $C_0^s(\mathbb{R}^n \times S^{n-1})$ for $s \in \mathbb{N}$ can be also recovered under additional assumptions on the regularity of the kernel K.

Let us now consider the regression problem (the case of classification can be treated in a similar fashion). It can be framed, in this setting, similarly to a kernel ridge regression problem:

$$\min_{h \in W, \beta \in \mathbb{R}} \lambda \|h\|_W^2 + \frac{1}{R} \sum_{k=1}^{R} \left(\langle \mu^{(k)}, h \rangle + \beta - y^{(k)} \right)^2 \tag{3}$$

where $\lambda > 0$ weighs the regularization term given by the RKHS norm $\|\cdot\|_W$. By direct differentiation with respect to β, one finds the linear equation satisfied by the optimal (h_*, β_*):

$$\beta_* + \frac{1}{R} \sum_{k=1}^{R} \langle \mu^{(k)}, h_* \rangle = \bar{y}. \tag{4}$$

where $\bar{y} = \sum_{k=1}^{R} y^{(k)}/R$ is the average value of the labels. We now derive an explicit optimality condition on h by focusing on the practical situation in which each $\mu^{(k)}$ is a discrete varifold that we write $\mu^{(k)} = \sum_{i=1}^{m_k} w_i^{(k)} \delta_{(x_i^{(k)}, v_i^{(k)})}$. In that case, the second term of (3) only depends on the values of h at the points $(x_i^{(k)}, v_i^{(k)})$. In order to remove repetitions, we shall denote by $(\tilde{x}_i, \tilde{v}_i)$ for $i = 1, \ldots, \tilde{m}$ (with $\tilde{m} \leq m \doteq m_1 + \ldots + m_R$) the subset of distinct points in $\mathbb{R}^n \times S^{n-1}$ among the $(x_i^{(k)}, v_i^{(k)})$'s. This implies that the optimal h^* must lie in the orthogonal subspace $W_0^\perp$ where:

$$W_0 = \{h \in W : h(\tilde{x}_i, \tilde{v}_i) = 0 \text{ for all } i = 1, \ldots, \tilde{m}\}$$

Indeed, if $h^* \notin W_0^\perp$, then, denoting $h_\perp^*$ its orthogonal projection onto $W_0^\perp$, we would have that $\|h_\perp^*\|_W^2 < \|h^*\|_W^2$ while the second term in (3) remains the same for $h_\perp^*$ due to the definition of W_0. Thus h^* would not be optimal. From standard properties of RKHS, specifically Lemma 8.1 in (Younes, 2019), it follows that h_* must take the form:

$$h_*(\cdot) = \sum_{i=1}^{\tilde{m}} \alpha_i K((\tilde{x}_i, \tilde{v}_i), \cdot)$$

where $\alpha = (\alpha_i) \in \mathbb{R}^{\tilde{m}}$ is to be determined. Introducing $\tilde{\mathbf{K}} = [K((\tilde{x}_i, \tilde{v}_i), (\tilde{x}_j, \tilde{v}_j))]$ the matrix of the kernel evaluations between all pairs of points, which is positive definite, one has $\|h_*\|_W^2 = \alpha^T \tilde{\mathbf{K}} \alpha$. On the other hand, we may define a second kernel matrix $\mathbf{K} = [K((x_i^{(k)}, v_i^{(k)}), (\tilde{x}_j, \tilde{v}_j))] \in \mathbb{R}^{m \times \tilde{m}}$ so that the energy to minimize in (3) becomes:

$$J(\alpha) = \lambda \alpha^T \tilde{\mathbf{K}} \alpha + \frac{1}{R} \sum_{k=1}^{R} \left((w^{(k)})^T (\Pi^{(k)})^T \mathbf{K} \alpha + \beta - y^{(k)} \right)^2$$

in which $w^{(k)} \in \mathbb{R}^{m_k}$ denotes the k-th block vector of w and $\Pi^{(k)} \in \mathbb{R}^{m_k \times m}$ is the orthogonal projection that extracts the k-th block entries of a vector in $\mathbb{R}^m$. This amounts to the minimization of a quadratic function of α for which the optimality condition is:

$$\left[\lambda \tilde{\mathbf{K}} + \mathbf{K}^T \left(\frac{1}{R} \sum_{k=1}^{R} (\Pi^{(k)})^T w^{(k)} (w^{(k)})^T \Pi^{(k)} \right) \mathbf{K} \right] \alpha_* + \mathbf{K}^T (w/R) \beta = \mathbf{K}^T (w \otimes y/R)$$

where we use the notation $w \otimes y$ for the vector made of the R blocks $y^{(k)}w^{(k)}$, $k = 1, \dots, R$. Ultimately, combining the above with (4), we see that the optimal α_*, β_* (and thus the solution of the kernel regression problem) are obtained by solving the linear system:

$$\begin{bmatrix} \lambda\tilde{\mathbf{K}} + \mathbf{K}^T \left(\frac{1}{R} \sum_{k=1}^{R} (\Pi^{(k)})^T w^{(k)} (w^{(k)})^T \Pi^{(k)} \right) \mathbf{K} & \mathbf{K}^T(w/R) \\ w^T\mathbf{K}/R & 1 \end{bmatrix} \begin{bmatrix} \alpha_* \\ \beta_* \end{bmatrix} = \begin{bmatrix} \mathbf{K}^T(w \otimes y/R) \\ \bar{y} \end{bmatrix} \tag{5}$$

We note that the above block matrix is invertible for any $\lambda > 0$, as can be checked using the Schur complement criterion for block matrices. Despite the nice and simple mathematical structure of the problem, it should be pointed that the resulting linear system (5) has a size essentially of the order of the total number of Diracs in the whole training set, which can be over 100 millions for some of the experiments of Section 4. Furthermore, this linear system may be very ill-conditioned if λ is not properly calibrated. Thus, this direct strategy is likely not practical except for small datasets of shapes with limited number of vertices. Although this is out of the scope of the present paper, we suspect that such a kernel-based approach can stil be made practical if combined with some sparse approximation schemes allowing to reduce the dimensionality of the solution space, such as the quantization approach of (Hsieh and Charon, 2021) or the Nystrom method considered in (Paul et al., 2024). In addition, one could leverage recently proposed techniques for this class of linear systems, such as methods that rely on preconditioning (Avron et al., 2017; Zhao et al., 2024) or multilevel Newton schemes (Zhang et al., 2023). Yet, in order to avoid any such preprocessing steps, we will consider, in the next section, another representation for the test function space.

3.2 Neural network model

As an alternative to the above kernel strategy, we shall introduce a more direct approach that consists in modeling the test function via a neural network, which is now ubiquitous in the modern machine learning literature. We thus approximate h using a multilevel perceptron (MLP), namely consider a set of trainable functions $h_\theta : \mathbb{R}^{2n} \to \mathbb{R}$ where θ denotes the trainable parameters of the MLP. Note that, based on the many existing universal approximation and expressivity theorems for MLPs (Hornik et al., 1989; Cybenko, 1989; Haykin, 1994; Lu et al., 2017), one can easily adapt the separability properties from Section 2.3 for functions h expressible by a sufficiently wide and/or deep MLP. Informally, if two sets of varifolds (resp. shapes) satisfy the assumptions of Theorem 2 (resp. Theorem 3), then one can find a separating test function for those two sets that can be expressed with a sufficiently wide/deep MLP. Given a discrete varifold $\mu = \sum_{i=1}^{m} w_i \delta_{(x_i, v_i)}$, we pass each support $(x_i, v_i) \in \mathbb{R}^n \times S^{n-1} \subseteq \mathbb{R}^{2n}$ through our MLP to compute a vector $H = [h_\theta(x_1, v_1), ..., h_\theta(x_m, v_m)]$. We then view the weights of μ as a vector $W = [w_1, ..., w_m]$ and compute

$$H^T W + \beta = \sum_{i=1}^{m} w_i h_\theta(x_i, v_i) + \beta = \langle \mu, h_\theta \rangle + \beta$$

where β is a trainable bias term. As an important note, the number of support Diracs m has no influence on the network architecture and the number of trainable parameters in (θ, β). Those are determined by the dimension n of the ambient space. The number m only affects the evaluation of the network on the varifold μ. For the linear regression problem in $\mathcal{V}$, we train this model in a supervised manner. Similar to the previous kernel approach, we minimize the mean squared error loss to the known labels:

$$\mathrm{MSE}(\theta, \beta) = \frac{1}{R} \sum_{j=1}^{R} \left(\langle \mu_j, h_\theta \rangle + \beta - y_j \right)^2 .$$

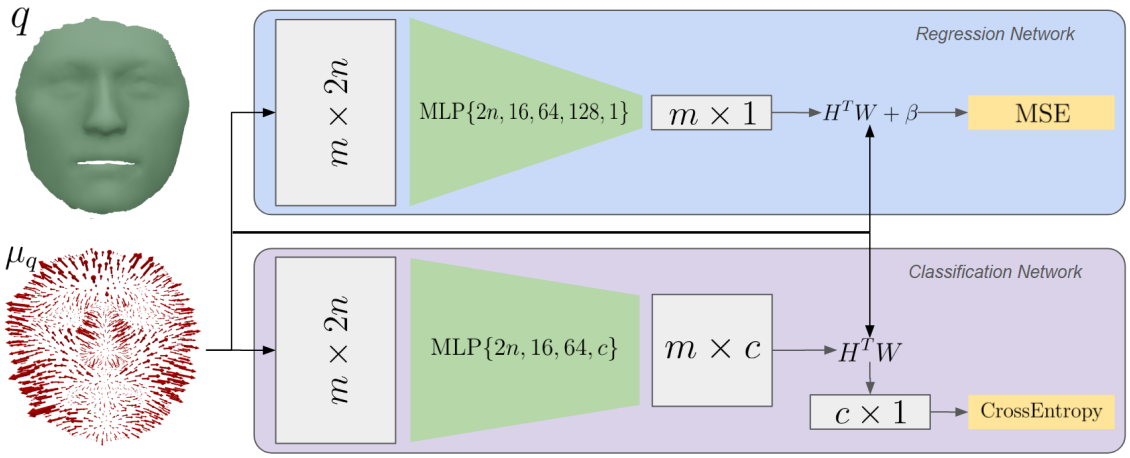


Figure 3: Model Summary: A mesh q is mapped to its varifold representation μ_q. The supports of the varifold are passed through an MLP and the resulting values are integrated with respect to the masses of μ_q. For regression, a trainable bias term is added to integral. For classification, the integral produces a c-dimensional vector viewed as logits of the probability that q belongs to each class.

To perform multi-class classification of varifolds, we slightly modify the architecture to learn a vector valued functions $h : \mathbb{R}^n \times S^{n-1} \to \mathbb{R}^c$ where $\int_{\mathbb{R}^n \times S^{n-1}} h \, \mathrm{d}\, \mu$ is a c-dimensional vector where each entry is interpreted as a logit of the probability that μ belongs to the corresponding class. The function h is again modeled via a multi-layer perceptron (MLP) $h_\theta : \mathbb{R}^{2n} \to \mathbb{R}^c$. For $\mu = \sum_{i=1}^m w_i \delta_{(x_i, u_i)}$, we evaluate h_θ on the supports to obtain an $m \times c$ matrix H with each row being the vector $h_\theta(x_i, u_i)$. Therefore, $H^T W = \langle \mu, h_\theta \rangle$ yields a c-dimensional logit vector. These logits are passed through a softmax and used with the standard cross-entropy loss for classification. A schematic of both models is outlined in Figure 3.

Remark 10 (Invariance to Orientation) *It is important to note that the model presented so far relies on the orientation of the tangent/normals to the shapes, meaning that one assumes a consistent orientation of the different shapes in the data set. For certain applications that involve non-oriented data, it is necessary to make the SVarM approach invariant to the orientation of v_i. This can easily be achieved by taking $\widehat{h_\theta} = h_\theta(x, v) + h_\theta(x, -v)$ and computing $\langle \mu, \widehat{h_\theta} \rangle + \beta$. In particular, in the applications to shape graphs of Section 4.4, we rely on this modification to remain robust to reorientation of the edges of the shape graphs.*

4. Numerical Experiments

In this section, we present a series of numerical experiments designed to evaluate the performance, generalizability, and efficiency of our proposed model across a range of tasks involving 3D data. We consider three distinct scenarios: rotational alignment of 3D human body scans, classification of remeshed surface representations of MNIST digits and classification of facial graph and surface data extracted from the COMA data set. These experiments demonstrate the flexibility of our model across diverse input types and provide quantitative and qualitative results with comparisons to baseline methods. For each task, we outline the data preprocessing steps, describe the experimental setup, and report results using standard performance metrics. In the final section, we present a series of ablation and robustness experiments to highlight the stability of our model under noise and

occlusions. These analyses provide deeper insight into the factors driving the model's performance and demonstrate its robustness across a range of challenging conditions. All of the models utilize sigmoid activation functions and are trained using the Adam optimizer with a fixed learning rate of 0.005 for 100 epochs to ensure consistency in the presented results. All experiments were conducted on a standard home PC equipped with an Intel 3.2 GHz CPU and a GeForce GTX 2070 GPU (1620 MHz). A Pytorch implementation of the SVarM framework is publicly available on GitHub at `https://github.com/SVarM25/SVarM`.

4.1 Regression for Single Axis Rotation Alignment of Human Bodies

Our first experiment evaluates the effectiveness of our regression model on a toy data set consisting of 3D human body scans from the FAUST data set (Bogo et al., 2014) rotated along a fixed axis. In many real-world applications involving 3D data—such as human pose estimation, animation, or medical imaging—scans are often misaligned due to variations in sensor orientation, subject posture, or scanning artifacts. Such rotational misalignment can significantly degrade the performance of downstream tasks. Therefore, estimating and correcting the rotation of 3D scans is a crucial preprocessing step.

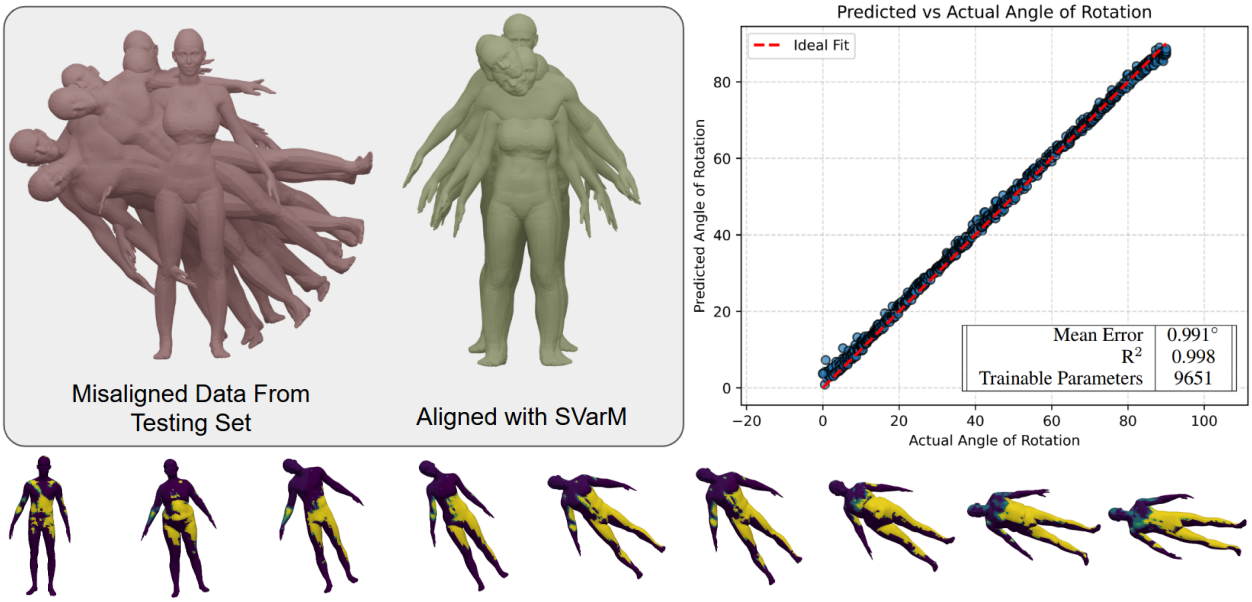


Figure 4: Regression on rotation angle around a fixed axis. We display a selection of human scans from our testing set of misaligned data (red) and present same data rotated by the angle predicted by the SVarM regression model (yellow). Additionally, we plot of the predicted angles against the ground truth on the entire testing set (right) reporting the mean error and correlation coefficient on the entire test set. Lastly, we display the learned function h evaluated on several meshes from the testing data set; the color of each face corresponds to $h(c_f, v_f)$ ranging from blue (low values) to yellow (higher values).

To address this, we frame rotational alignment as a regression problem, where the model learns to predict the angle by which a given mesh has been rotated. The goal is to realign each mesh to a canonical upright position. We generate our data set by applying known rotations to a subset of the FAUST 3D human body scans along a single axis, storing the corresponding angles as labels resulting in 8000 labeled meshes. Our model is trained on 90% of this data and evaluated on

the remaining 10%. In Figure 4, we present a selection of data set samples and demonstrate the model's ability to accurately predict rotation angles and realign the meshes. Moreover, to highlight the interpretability of our model, we display the evaluation of the learned test function h on samples from the data set where the color of each face of the mesh corresponds to the evaluation of the test function at (c_f, n_f). We note that the meshes that are misaligned by a larger angle have a higher density of faces with larger values of h.

4.2 Regression for Full Rotation Alignment of Human Data

Building on our initial experiment with scalar angle regression, we next evaluate the ability of our model to predict full 3D rotation matrices. While predicting rotation around a fixed axis simplifies the problem, many real-world scenarios involve arbitrary rotations in 3D space. Consequently, modeling the full rotation is essential for general applicability to object alignment.

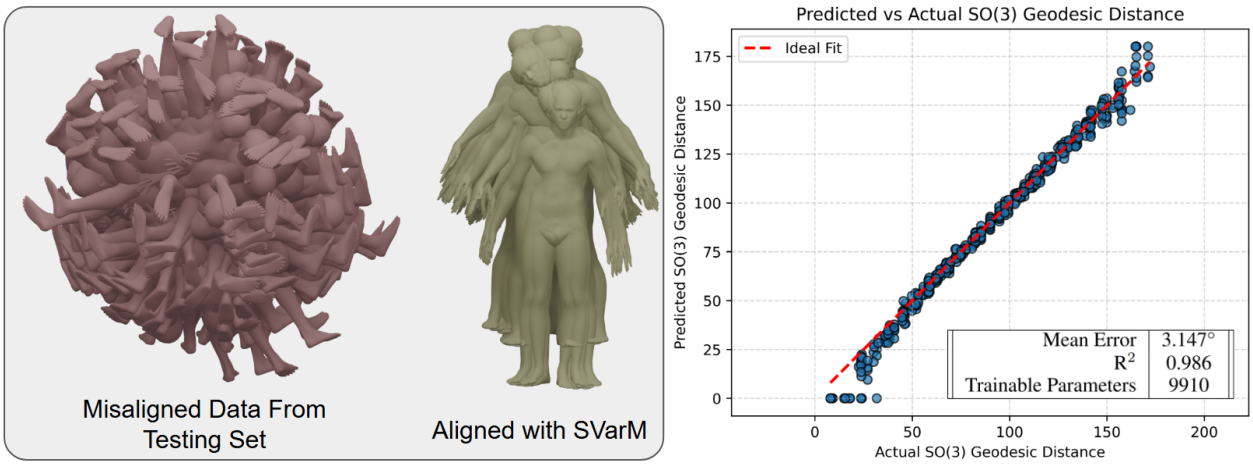


Figure 5: Multidimensional Regression to predict rotation matrices. We display a selection of misaligned data (red) and present same data rotated by the matrix predicted by the SVarM regression model (yellow). Additionally, we plot the geodesic distance in $SO(3)$ from the identity for the predicted matrices against the ground truth matrices (right) reporting the mean error and correlation coefficient on the entire test set.

The training procedure mirrors our previous setup: we generate synthetic data by applying random 3D rotations to FAUST meshes and use the known rotation matrices as supervision. This results in a set of 12840 meshes and train the model on 80% of the data and test it on the remaining 20%. In Figure 5, we display test examples of meshes before and after applying the predicted rotation matrix. In addition, we report quantitative metrics, including the average error and the coefficient of determination for the geodesic distances in $SO(3)$ of the rotation matrices. This experiment demonstrates that our regression framework can be extended from constrained single-axis alignment to general 3D rotational correction, paving the way for integration into more complex pipelines involving arbitrary pose variations and sensor noise.

To further highlight our method's effectiveness at solving the rotation alignment problem, we compare with results obtained by directly minimizing, over the rotation group $SO(3)$, the distance from the rotated surface to a target surface in the canonical upright position. Specifically, as a measure of surface discrepancy, we leverage the varifold kernel metric that has been routinely used for such purpose in several past works (Kaltenmark et al., 2017; Hartman et al., 2022). Note that other choices with similar properties, such as the Chamfer distance from the computer vision

literature, could be considered as well. The resulting discrepancy function is then minimized based on the quasi-Newton L-BFGS scheme applied to the 3D rotation parameters. This approach produces a mean error of 8.0558° with an R^2 value of 0.799. Additionally, direct minimization over $SO(3)$ involves no training time, but requires 15-30s per mesh to solve the alignment problem. By comparison, our trained model requires less than 15s to align the entire testing set and produces a mean error of 3.147° with an R^2 value of 0.986. Moreover, it involves a total of only 9910 trainable parameters and requires less than 40 minutes in total for training.

4.3 Classification of MNIST Digits

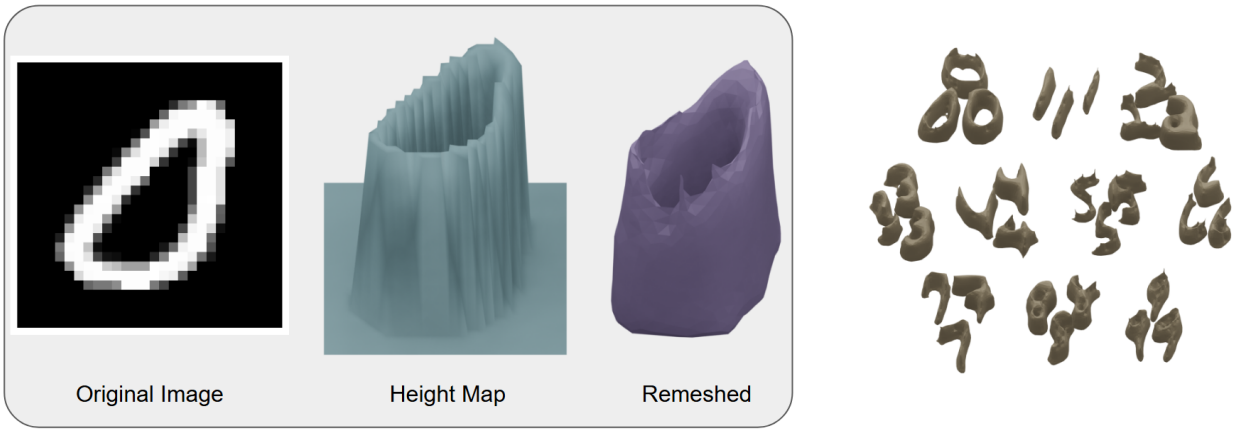


	Image Models				3D Models	
	Eff. CapsNet	DNN-5	ConvPMM	PMM	PointNet	SVarM
Accuracy	**99.84%**	97.20%	98.99%	97.38%	90.28%	**97.67%**
Trainable Parameters	161824	575051	129416	**4990**	1604883	**1850**

Figure 6: MNIST Surface Construction and Results. An image from the MNIST data set is converted to a 3D height map surface and remeshed, with examples of this preprocessing shown for each class of the MNIST data set. We compare several methods operating either on the images or surfaces, reporting accuracy and trainable parameters to highlight our method's efficiency.

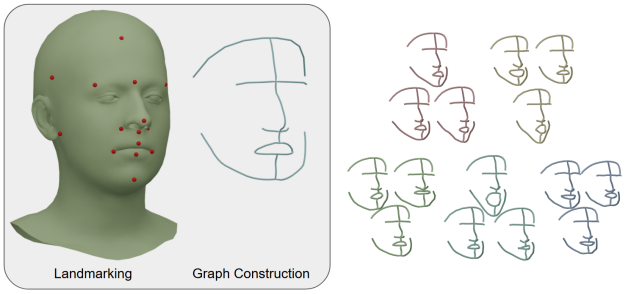
We next evaluate SVarM on the task of classifying handwritten digits from the well-known MNIST data set, which contains 70,000 greyscale images of handwritten digits with 60,000 training images and 10,000 images for testing. To adapt this image data set to our framework, following a similar setup to (Kostrikov et al., 2018), we convert each digit into a 3D surface representation by treating the grayscale pixel values as a height map. We then apply a remeshing operation to generate closed surface meshes corresponding to each digit. A schematic of the conversion process, along with representative examples of the resulting surface meshes, is shown in Figure 6. This pre-processing is not computationally burdensome and requires <20 minutes for the entire MNIST data set.

We compare our model's performance to that of several baseline methods (Mazzia et al., 2021; Pishchik, 2023; Cook et al., 2025) that operate directly on the original image data and are designed to be parameter-efficient. While two of these models (Eff. CapsNet (Mazzia et al., 2021) and ConvPMM (Cook et al., 2025)) achieve slightly higher accuracy, they involve significantly more trainable parameters. In contrast, SVarM achieves competitive accuracy while maintaining a much

smaller parameter footprint, underscoring its efficiency and suitability for low-resource learning settings.

4.4 Classification of Shape Graphs

To further evaluate the effectiveness of the SVarM model for classification tasks, we train the model on two separate data sets of shape graphs. First, we train SVarM on a data set of shape graphs derived from the CoMA data set (Ranjan et al., 2018), which contains 3D surface scans of human faces. This data set is comprised of approximately 21,000 scans of 12 distinct individuals. From each scan, we extract a shape graph by identifying a consistent set of facial landmarks and connecting them via geodesic paths along the surface of the mesh. A schematic illustration of this shape graph construction process is provided in Figure 7. We should note that shape graphs, just as curves, are naturally embedded in the space of varifolds by simple addition in $\mathcal{V}$ of the varifolds associated to each separate branch. We split the data into training and test sets, using 80% of



Method	Accuracy	Parameters
PointNet	64.32%	1,605,397
DGCNN	83.56%	1,014,924
VariGrad	91.28%	586,652
SVarM	**99.98%**	**1,980**

Figure 7: COMA Graph Construction and Results. A schematic illustrating the process of extracting a shape graph from a surface in the COMA data set with example graphs from 5 of the 12 classes. We compare the classification accuracy and model size for several models evaluated on this data set.

the scans for training and the remaining 20% for testing. To assess performance, we compare our proposed model against three baseline models commonly used for 3D shape classification. Our model achieves significantly higher classification accuracy across all metrics. A quantitative summary of these results is presented in Figure 7.

The second data set consists of 3D shape graphs representing neurons from five distinct regions of the mouse nervous system. This data set includes 443 neuron graphs, categorized into the Basal Ganglia, Cerebellum, Hippocampus, Main Olfactory Bulb, and Retina. Representative examples from each class are shown in Figure 8. We train the model on 80% neuron graphs and evaluate its performance on the remaining 20%. The trained model correctly classifies 97.73% of the shape graphs in the testing set. In Figure 8, for the test functions associated to the different classes, we visualize the marginal spatial and directional profiles defined (respectively) by

$$H_v(x) := \int_{S^{n-1}} h(x,v)d\mu(x,v) \qquad \text{and} \qquad H_x(v) := \int_{\mathbb{R}^n} h(x,v)d\mu(x,v). \tag{6}$$

This experiment highlights the ability of the SVarM model to be trained on smaller data sets without overfitting. In particular, this model is over-parameterized with 1525 trainable parameters and achieves similar accuracy on the training data as on the unseen data.

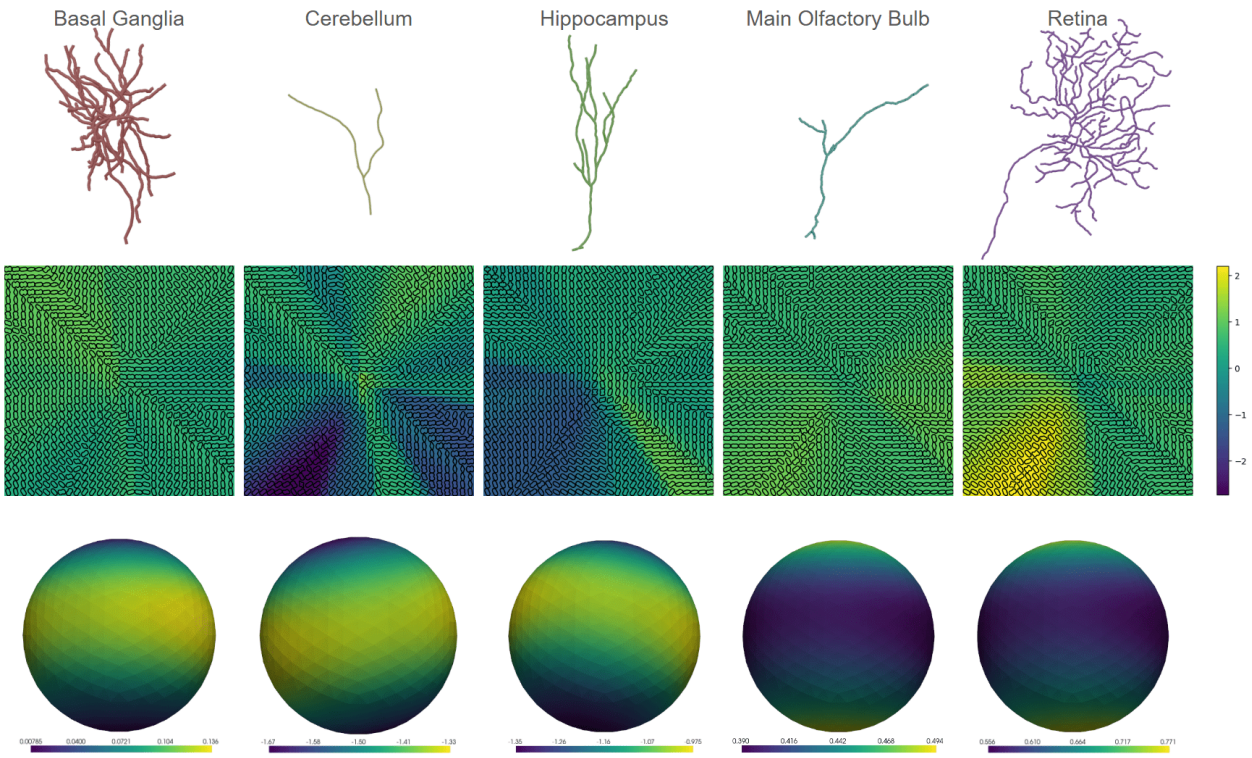
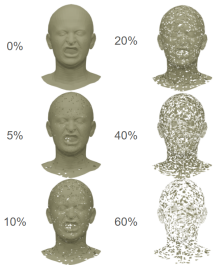


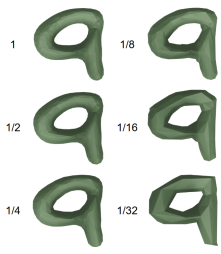
Figure 8: Neuron Shape Graphs. Row 1: Example of a neuron shape graph from five different regions in the mouse nervous system. Row 2: The marginal spatial profile $H_v(x_1, x_2, 0)$ (see Equation (6)) of the test function of each class is plotted as a color map. The black curves on each pixel illustrate the local direction profile of the function at that point. Row 3: we plot the marginal directional profile $H_x(v)$ of each test function as a color map on S^2.

4.5 Robustness Studies

In our final set of experiments, we seek to highlight SVarM's robustness to imaging noise and missing data. We train our classification model on 80% of the human face surface data from the CoMA data set. On the testing data, we remove faces from the meshes and report the results on the altered meshes. A visual representation of the meshes with missing faces is shown in Figure 9. We repeat this process, increasing the rate of missing faces and report the classification accuracy in Figure 9.



	SVarM
0%	99.9%
5%	99.8%
10%	98.9%
20%	96.3%
40%	85.6%
60%	67.3%



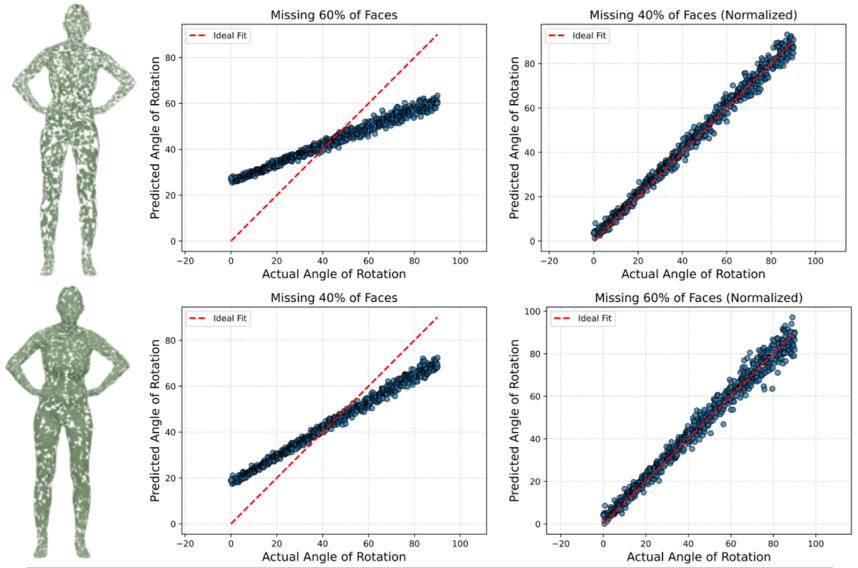
	SVarM	PointNet
1	97.7%	90.3%
1/2	97.7%	89.6%
1/4	97.4%	87.7%
1/8	97.3%	79.0%
1/16	91.5%	64.4%
1/32	86.2%	54.6%

Figure 9: Robustness experiments under different levels of missing data and mesh decimation.

Furthermore, we consider SVarM's robustness to parameterization by testing the classification model trained on the MNIST digits (c.f. Section 4.3) on down-sampled versions of the testing data.

For each mesh in the testing set we apply a mesh decimation operation to reduce the number of faces of the mesh to a fraction of the original and report the results on this down-sampled data. We repeat this process for smaller ratios of total faces and report the classification accuracy in Figure 9. A visual representation of the down-sampled meshes for different ratios of total faces Figure 9.

Finally, we explore the robustness of the SVarM model for a regression problem. We consider the trained regression model discussed in Section 4.1, we apply the face removal process to the testing set and report the results across varying rates of missing faces. As the model's predictions are sensitive to the total mass of the input varifolds, its performance deteriorates significantly when a large proportion of faces is removed. To mitigate this issue, we introduce a simple yet effective correction: we normalize the total mass of each varifold to a fixed value. This normalization ensures consistency in the model's input representations, regardless of the number of removed faces. In Figure 10, we illustrate the impact of missing faces on the regression performance, both with and without normalization, demonstrating the robustness of the SVarM regression model, as well as, the improvement gained by re-normalizing the masses.



	Regression		Regression Rescaled	
Missing	Mean Error (°)	R^2	Mean Error (°)	R^2
0%	0.991°	0.998	0.991°	0.998
5%	1.852°	0.993	1.120°	0.997
10%	2.875°	0.983	1.199°	0.996
20%	5.080°	0.949	1.417°	0.995
40%	9.355°	0.827	1.835°	0.992
60%	13.841°	0.6245	2.582°	0.984

Figure 10: Robustness experiments for regression on single axis angle of rotation under different proportions of missing faces.

5. Towards nonlinear regressors and classifiers in $\mathcal{V}$

In spite of the mathematical and numerical results presented in this paper, one of the obvious limitations of the SVarM model is that it is restricted to regression and classification functionals that are affine on the space of varifolds. Although we showed that these may be enough to separate

disjoint sets of shapes under relatively mild assumptions, the counterexample of Figure 2 still illustrates that some simple configurations are not separable within the SVarM framework. A natural way of generalizing the approach would be thus to consider nonlinear functionals on $\mathcal{V}$, by analogy with the machine learning models based on neural network representations that are commonly used for data living on Euclidean spaces.

In the context of the infinite-dimensional space of varifolds $\mathcal{V}$, we may build nonlinear maps $\psi : \mathcal{V} \to \mathbb{R}$ via the composition of multiple affine forms on $\mathcal{V}$ with some nonlinear function on $\mathbb{R}$. More specifically, a simple single hidden layer architecture would consist in modeling ψ as follows:

$$\psi(\mu) = \sum_{k=1}^{r} c_k \sigma(\langle \mu, h_k \rangle + \beta_k) \tag{7}$$

where $h_k \in C_0(\mathbb{R}^n \times S^{n-1})$ are a set of different test functions (each modeled for instance via an MLP as in Section 3.2), $\beta_k \in \mathbb{R}$ the associated biases, $\sigma : \mathbb{R} \to \mathbb{R}$ is a nonlinear function (such as RELU, tanh, sigmoid...) and $c_k \in \mathbb{R}$ are the coefficients of the output linear function. In the infinite width limit $r \to \infty$, one can approximate any weak-* continuous function $\mathcal{V} \to \mathbb{R}$:

Theorem 6 *Let $\sigma : \mathbb{R} \to \mathbb{R}$ be a continuous non-polynomial function, C a weak-* compact subset of $\mathcal{V}$ and $\psi : C \to \mathbb{R}$ a weak-* continuous function on C. Then for any $\varepsilon > 0$, there exist $r \in \mathbb{N}$, $(h_k)_{k=1,\dots,r}$ in $C_0(\mathbb{R}^n \times S^{n-1})$, $(\beta_k) \in \mathbb{R}^r$ and $(c_k) \in \mathbb{R}^r$ such that for all $\mu \in C$:*

$$\left| \psi(\mu) - \sum_{k=1}^{r} c_k \sigma(\langle \mu, h_k \rangle + \beta_k) \right| \leq \varepsilon$$

The statement follows directly from the more general result of Theorem 2.1 in (Ismailov, 2024) since the space $\mathcal{V}$ equipped with the weak-* convergence topology is a locally convex space which continuous linear forms are exactly of the form $\mu \mapsto \langle \mu, h \rangle$ (c.f. Proposition 2). Combined with our earlier remarks in Section 3.2, this implies that it is possible to approximate any weak-* continuous (nonlinear) function on $\mathcal{V}$ by considering sufficiently many test functions each represented with sufficiently wide (or deep) MLPs.

Using such more general architectures should theoretically enable to address, in particular, the nonseparability issues of SVarM. For instance, it is easy to derive a function ψ of the form (7) with only one test function h that can separate the pairs of shapes in the example of Figure 2. Namely, if h is defined so that $\langle \mu, h \rangle$ corresponds to the number of balls on the upper half of the object and $\sigma(x) = 2^{-|x-1|^2}$, then $\psi = 1$ for the green surfaces while $\psi = 1/2$ for the red ones. We leave it to future work, however, to implement and experiment these types of architectures on shape regression and classification problems.

6. Conclusion and future perspectives

In this work, we propose a learning framework on data sets of curves, surfaces and shape graphs that leverages the (nonlinear) embedding of shapes into varifold space. We focus on this particular measure space as it provides a parameterization-invariant representation of embedded shapes, making the model particularly well-suited for geometric data sets. Analogous to support vector machines—but in the infinite-dimensional setting of varifolds—the SVarM model then learns affine forms $\mu \mapsto \langle \mu, h \rangle + \beta$ to solve regression and classification problems involving 3D geometric data, by approximating a test function $h \in C_0(\mathbb{R}^n \times S^{n-1}, \mathbb{R})$ and bias β. We derive some theoretical guarantees on SVarM's capacity to approximate arbitrary affine forms, identify conditions under which sets of varifolds or embedded shapes can be linearly separated, and establish stability results

for the model. In addition, we present a series of numerical experiments demonstrating the model's effectiveness in both regression and classification tasks across various geometric data sets. These experiments highlight the model's computational efficiency and its robustness to imaging noise, including missing faces and sampling changes. The results also showcase the fact that SVarM can achieve comparable (and in some cases superior) accuracy to geometric deep learning models but with an overall much smaller number of trainable parameters, making it in line with recent efforts towards developing lightweight deep learning architectures (Liu et al., 2024).

While we focused on learning affine functions, we also outlined a more general approach to build and learn nonlinear operators over the space of varifolds, which we plan to implement and experiment for various types of supervised or unsupervised problems involving shape data sets in the future. Furthermore, although the present work was primarily concerned with objects such as curves and surfaces, the versatility of the varifold representation makes it possible, in principle, to extend SVarM to other types of geometric data structures, with two cases of interest being high-angular resolution diffusion images (Tuch et al., 2002) and spatial transcriptomics data (Stouffer et al., 2024) as these naturally fit the varifold setting. Lastly, an interesting aspect that has not yet been explored in this paper is the use, during training, of regularization penalties on the weights of the neural network defining h. Such penalties could enforce certain constraints on h so as to potentially increase the qualitative interpretability of the learned test function.

Acknowledgments

This work was supported by the National Science Foundation through the grants DMS-2438562 and DMS-2426550. The authors would also like to thank Murad Hossen for his assistance with the mouse neuron traces data set used in the numerical experiments section.

Appendix A. Ablation Studies

A.1 Choice of measure representation

In this work, we have focused on learning methods that are based on the varifold representations of shapes, i.e. by learning functions on the space $\mathcal{V} = \mathcal{M}(\mathbb{R}^n \times S^{n-1})$, with a shape q being mapped to the varifold $(q, v_q)_* \operatorname{vol} q$. Varifolds combine the information of point position together with the direction of the tangent or normal vector. It is then interesting to compare SVarM against similar approaches but that rely on other possible measure representations. In particular, by marginalizing a varifold with respect to either of the two components S^{n-1} or $\mathbb{R}^n$, one can quantify the advantage of using both the position and direction information for regression and classification problems. We thus compare the SVarM model to these two alternative measure representations, focusing on surface data. First we consider the space of usual *spatial measures*, $\mathcal{M}(\mathbb{R}^n)$, where $q \mapsto q_* \operatorname{vol} q$. This representation captures only the distribution of points in space (with respect to the local area measure), discarding all directional information. Conversely, we also consider *area measures* in $\mathcal{M}(S^{n-1})$, where $q \mapsto (v_q)_* \operatorname{vol}_q$, which only encodes the distribution of the normal vectors v_q on the sphere independently of the position. Note that area measures obviously do not allow to uniquely identify a shape but are nevertheless objects central to the field of directional statistics (Mardia and Jupp, 2009).

Both alternatives encode strictly less geometric information about the shape q compared to the varifold representation. We retrain similar models for each of these measure representations, namely by learning a function on $\mathbb{R}^n$ or S^{n-1} instead of the product space. The results of this comparison

	1 Axis Rotation Regression		Full Rotation Regression		MNIST Classification
	Mean Error	R^2	Mean Error	R^2	Accuracy
$\mathcal{M}(\mathbb{R}^n)$	3.733°	0.967	4.302°	0.954	90.69%
$\mathcal{M}(S^{n-1})$	6.534°	0.908	7.847°	0.895	79.86%
$\mathcal{V}$	0.991°	0.998	3.147°	0.986	97.67%

Figure 11: Ablation on Measure Representations of Shapes

are reported in Table 11. They indicate that neither of these simplified models achieves performance comparable to the full SVarM approach.

A.2 Width and depth of the MLP representation of h_θ

We here present a second ablation study related to the architectural choices of the MLP parameterization of the test function h_θ in the SVarM architecture, specifically on the width and depth of the neural network. These results justify the choice of architecture used in the paper.

	MNIST Classification		
	Training Accuracy	Testing Accuracy	Trainable Parameters
$\text{mlp}(6, 8, 32, 10)$	95.51%	95.12%	674
$\text{mlp}(6, 16, 64, 10)$	98.12%	97.67%	1850
$\text{mlp}(6, 32, 128, 10)$	97.28%	97.47%	5738

Figure 12: Ablation on Width of MLP Parameterizations of Test Functions

Increasing the width of the hidden layers generally improves both training and testing accuracy. Notably, the jump from $\text{mlp}(6, 8, 32, 10)$ to $\text{mlp}(6, 16, 64, 10)$ significantly boosts performance, suggesting that a minimum representational capacity is needed for the model to generalize well. However, further increasing width to $\text{mlp}(6, 32, 128, 10)$ yields only marginal improvement in test accuracy, despite more than tripling the number of parameters. This indicates diminishing returns and potential over-parameterization. Adding more hidden layers beyond a moderate depth ap-

	MNIST Classification		
	Training Accuracy	Testing Accuracy	Trainable Parameters
$\text{mlp}(6, 64, 10)$	94.75%	95.39%	1098
$\text{mlp}(6, 16, 64, 10)$	98.12%	97.67%	1850
$\text{mlp}(6, 16, 64, 64, 10)$	97.64%	96.18%	6010

Figure 13: Ablation on Depth of MLP Parameterizations of Test Functions

pears to reduce generalization performance. While the two-hidden-layer model $\text{mlp}(6, 16, 64, 10)$ achieves the highest test accuracy, the deeper model $\text{mlp}(6, 16, 64, 64, 10)$ over-fits slightly, achieving higher training accuracy but worse test accuracy. Conversely, the shallow model $\text{mlp}(6, 64, 10)$ under-performs.

References

W. Allard. On the first variation of a varifold. *Annals of mathematics*, 95(3), 1972.

F. Almgren. *Plateau's Problem: An Invitation to Varifold Geometry.* Student Mathematical Library, 1966.

N. Aronszajn. Theory of reproducing kernels. *Trans. Amer. Math. Soc.*, 68:337–404, 1950.

H. Avron, K. L. Clarkson, and D. P. Woodruff. Faster kernel ridge regression using sketching and preconditioning. *SIAM Journal on Matrix Analysis and Applications*, 38(4):1116–1138, 2017.

A. B. Bal, X. Guo, T. Needham, and A. Srivastava. Statistical analysis of complex shape graphs. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2024.

M. Bauer, M. Bruveris, and P. W. Michor. Overview of the Geometries of Shape Spaces and Diffeomorphism Groups. *Journal of Mathematical Imaging and Vision*, 50(1):60–97, 2014.

M. Bauer, M. Bruveris, N. Charon, and J. Møller-Andersen. A relaxed approach for curve matching with elastic metrics. *ESAIM: Control, Optimisation and Calculus of Variations*, 25:72, 2019.

T. Besnier, S. Arguillère, E. Pierson, and M. Daoudi. Toward mesh-invariant 3D generative deep learning with geometric measures. *Computers & Graphics*, 115:309–320, 2023.

S. Bhadra and A. Srivastava. Scalar on shape regression using function data. In *International Workshop on Functional and Operatorial Statistics*, pages 43–50. Springer, 2025.

F. Bogo, J. Romero, M. Loper, and M. J. Black. FAUST: Dataset and evaluation for 3D mesh registration. In *Proceedings IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, Piscataway, NJ, USA, June 2014. IEEE.

H. Brézis. *Functional analysis, Sobolev spaces and partial differential equations*, volume 2. Springer, 2011.

A. M. Bronstein, M. M. Bronstein, and R. Kimmel. Generalized multidimensional scaling: a framework for isometry-invariant partial surface matching. *Proceedings of the National Academy of Sciences*, 103(5):1168–1172, 2006.

M. M. Bronstein, J. Bruna, Y. LeCun, A. Szlam, and P. Vandergheynst. Geometric deep learning: going beyond euclidean data. *IEEE Signal Processing Magazine*, 34(4):18–42, 2017.

B. Buet, G. P. Leonardi, and S. Masnou. Weak and approximate curvatures of a measure: a varifold perspective. *Nonlinear Analysis*, 222:112983, 2022.

W. Cao, Z. Yan, Z. He, and Z. He. A comprehensive survey on geometric deep learning. *IEEE Access*, 8:35929–35949, 2020.

G. Carlsson. Topology and data. *Bulletin of the American Mathematical Society*, 46(2):255–308, 2009.

N. Charon and A. Trouvé. The varifold representation of nonoriented shapes for diffeomorphic registration. *SIAM journal on Imaging Sciences*, 6(4):2547–2580, 2013.

P. Cook, D. Jammooa, M. Hjorth-Jensen, D. D. Lee, and D. Lee. Parametric matrix models, 2025. URL `https://arxiv.org/abs/2401.11694`.

G. Cybenko. Approximation by superpositions of a sigmoidal function. *Mathematics of control, signals and systems*, 2(4):303–314, 1989.

H. Edelsbrunner and D. Morozov. Persistent homology. In *Handbook of Discrete and Computational Geometry*, pages 637–661. Chapman and Hall/CRC, 2017.

J. E. Gerken, J. Aronsson, O. Carlsson, H. Linander, F. Ohlsson, C. Petersson, and D. Persson. Geometric deep learning and equivariant neural networks. *Artificial Intelligence Review*, 56(12): 14605–14662, 2023.

R. Ghrist. Barcodes: the persistent topology of data. *Bulletin of the American Mathematical Society*, 45(1):61–75, 2008.

U. Grenander. *Elements of pattern theory*. JHU Press, 1996.

R. Hanocka, A. Hertz, N. Fish, R. Giryes, S. Fleishman, and D. Cohen-Or. Meshcnn: a network with an edge. *ACM Transactions on Graphics (ToG)*, 38(4):1–12, 2019.

E. Hartman, Y. Sukurdeep, E. Klassen, N. Charon, and M. Bauer. Elastic Shape Analysis of Surfaces with Second-Order Sobolev Metrics: A Comprehensive Numerical Framework. *To Appear in IJCV*, 2022.

E. Hartman, E. Pierson, M. Bauer, M. Daoudi, and N. Charon. Basis restricted elastic shape analysis on the space of unregistered surfaces. *International Journal of Computer Vision*, 133 (4):1999–2024, 2025.

S. Haykin. *Neural networks: a comprehensive foundation*. Prentice Hall PTR, 1994.

K. Hornik, M. Stinchcombe, and H. White. Multilayer feedforward networks are universal approximators. *Neural networks*, 2(5):359–366, 1989.

H.-W. Hsieh and N. Charon. Metrics, quantization and registration in varifold spaces. *Foundations of Computational Mathematics*, 21(5):1317–1361, 2021.

I. C. Ipsen and R. Rehman. Perturbation bounds for determinants and characteristic polynomials. *SIAM Journal on Matrix Analysis and Applications*, 30(2):762–776, 2008.

V. Ismailov. Universal approximation theorem for neural networks with inputs from a topological vector space, 2024. URL `https://arxiv.org/abs/2409.12913`.

I. H. Jermyn, S. Kurtek, E. Klassen, and A. Srivastava. Elastic Shape Matching of Parameterized Surfaces Using Square Root Normal Fields. In *European Conference on Computer Vision*, page 804–817. Springer, 2012.

B. Jiang, Z. Zhang, D. Lin, J. Tang, and B. Luo. Semi-supervised learning with graph learning-convolutional networks. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 11313–11320, 2019.

I. Kaltenmark, B. Charlier, and N. Charon. A general framework for curve and surface comparison and registration with oriented varifolds. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3346–3355, 2017.

I. Kostrikov, Z. Jiang, D. Panozzo, D. Zorin, and J. Bruna. Surface networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2540–2548, 2018.

J. Lee, X. Cai, C.-B. Schönlieb, and S. Masnou. Neural varifolds: an aggregate representation for quantifying the geometry of point clouds. *arXiv preprint arXiv:2407.04844*, 2024.

H.-I. Liu, M. Galindo, H. Xie, L.-K. Wong, H.-H. Shuai, Y.-H. Li, and W.-H. Cheng. Lightweight deep learning for resource-constrained environments: A survey. *ACM Computing Surveys*, 56 (10):1–42, 2024.

Z. Lu, H. Pu, F. Wang, Z. Hu, and L. Wang. The expressive power of neural networks: A view from the width. *Advances in neural information processing systems*, 30, 2017.

K. V. Mardia and P. E. Jupp. *Directional statistics*. John Wiley & Sons, 2009.

V. Mazzia, F. Salvetti, and M. Chiaberge. Efficient-capsnet: Capsule network with self-attention routing. *Scientific Reports*, 11:14634, 2021. doi: 10.1038/s41598-021-93977-0. URL `https://doi.org/10.1038/s41598-021-93977-0`.

F. Mémoli. Gromov–Wasserstein distances and the metric approach to object matching. *Foundations of computational mathematics*, 11:417–487, 2011.

P. W. Michor and D. Mumford. An Overview of the Riemannian Metrics on Spaces of Curves Using the Hamiltonian Approach. *Applied and Computational Harmonic Analysis*, 23(1):74–113, 2007.

M. Ovsjanikov, M. Ben-Chen, J. Solomon, A. Butscher, and L. Guibas. Functional maps: a flexible representation of maps between shapes. *ACM Transactions on Graphics (ToG)*, 31(4):1–11, 2012.

A. Paul, N. Campbell, and T. Shardlow. Sparse Nystrom Approximation of Currents and Varifolds. *arXiv preprint arXiv:2406.09932*, 2024.

E. Pierson, M. Daoudi, and S. Arguillere. 3d shape sequence of human comparison and classification using current and varifolds. In *European Conference on Computer Vision*, pages 523–539. Springer, 2022.

E. Pishchik. Trainable activations for image classification. *Preprints*, January 2023. doi: 10.20944/preprints202301.0463.v1. URL `https://doi.org/10.20944/preprints202301.0463.v1`.

A. Ranjan, T. Bolkart, S. Sanyal, and M. J. Black. Generating 3D faces using convolutional mesh autoencoders. In *European Conference on Computer Vision (ECCV)*, pages 725–741, 2018. URL `http://coma.is.tue.mpg.de/`.

W. Rudin. Functional analysis 2nd ed. *International Series in Pure and Applied Mathematics. McGraw-Hill, Inc., New York*, 1991.

L. Simon. Introduction to geometric measure theory. *Tsinghua Lectures*, 2(2):3–1, 2014.

A. Srivastava and E. P. Klassen. *Functional and Shape Data Analysis*, volume 1. Springer, 2016.

K. M. Stouffer, A. Trouvé, L. Younes, M. Kunst, L. Ng, H. Zeng, M. Anant, J. Fan, Y. Kim, X. Chen, et al. Cross-modality mapping using image varifolds to align tissue-scale atlases to molecular-scale measures with application to 2D brain sections. *Nature communications*, 15(1): 3530, 2024.

Y. Sukurdeep, M. Bauer, and N. Charon. A new variational model for shape graph registration with partial matching constraints. *SIAM Journal on Imaging Sciences*, 15(1):261–292, 2022.

D. S. Tuch, T. G. Reese, M. R. Wiegell, N. Makris, J. W. Belliveau, and V. J. Wedeen. High angular resolution diffusion imaging reveals intravoxel white matter fiber heterogeneity. *Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine*, 48(4):577–582, 2002.

C. Villani. *Optimal Transport: Old and New*, volume 338. Springer Science & Business Media, 2008.

A. Virmaux and K. Scaman. Lipschitz regularity of deep neural networks: analysis and efficient estimation. In S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 31. Curran Associates, Inc., 2018. URL https://proceedings.neurips.cc/paper_files/paper/2018/file/d54e99a6c03704e95e6965532dec148b-Paper.pdf.

L. Younes. *Shapes and Diffeomorphisms*, volume 171. Springer, 2019.

J. Zhang, S. Zhou, C. Fu, and F. Ye. Fast newton method to solve klr based on multilevel circulant matrix with log-linear complexity: J. zhang et al. *Applied Intelligence*, 53(18):21407–21421, 2023.

S. Zhao, T. Xu, H. Huang, E. Chow, and Y. Xi. An adaptive factorized nyström preconditioner for regularized kernel matrices. *SIAM Journal on Scientific Computing*, 46(4):A2351–A2376, 2024.