AffectiveTDA: Using Topological Data Analysis to Improve Analysis and Explainability in Affective Computing

Hamza Elhamdadi, Shaun Canavan, and Paul Rosen

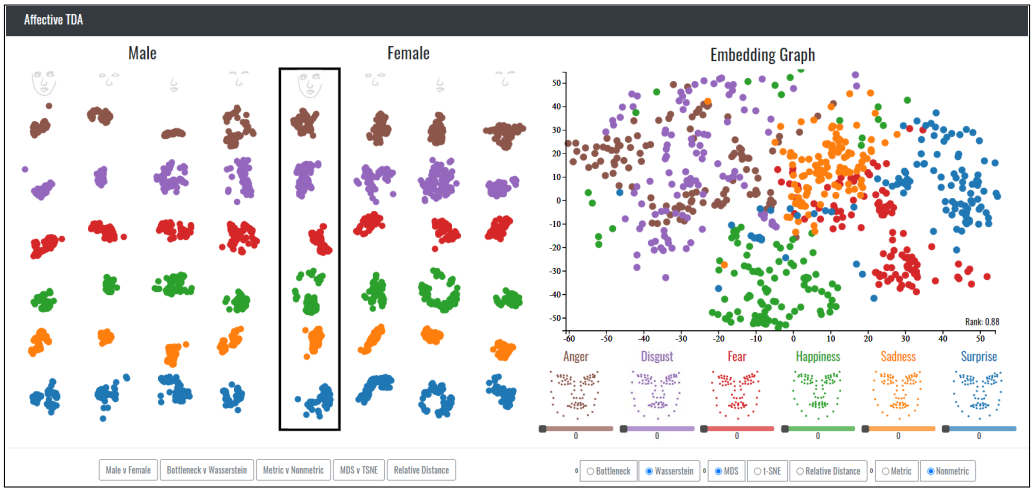


Fig. 1. Our affective computing visualization provides numerous options for comparing and contrasting the data. For example, the small multiples view (left) is comparing a male (left) subject to a female (right) subject considering different subsets of facial landmarks (columns), across emotions (rows) of *anger*, *disgust*, *fear*, *happiness*, *sadness*, and *surprise*. Each point represents one facial pose. The embedding graph (top right) compares all facial poses across all emotions, in this case showing the female full face using MDS on non-metric topology. The 3D landmarks (lower right) show a single facial pose per emotion. Settings are selected on the bottom.

Abstract—We present an approach utilizing Topological Data Analysis to study the structure of face poses used in affective computing, i.e., the process of recognizing human emotion. The approach uses a conditional comparison of different emotions, both respective and irrespective of time, with multiple topological distance metrics, dimension reduction techniques, and face subsections (e.g., eyes, nose, mouth, etc.). The results confirm that our topology-based approach captures known patterns, distinctions between emotions, and distinctions between individuals, which is an important step towards more robust and explainable emotion recognition by machines.

Index Terms—Affective computing, topological data analysis, explainability, visualization

1 Introduction

Affective computing, computer-based detection of human affects, has applications that span education (e.g., judging learners' confidence), healthcare (e.g., judging pain), and product marketing (e.g., measuring consumers' response to products). Early work in measuring affect began in the late 1960's spearheaded by Ekman and Friesen [23]. Their work culminated in a classification of six basic emotions: *anger*, *disgust*, *fear*, *happiness*, *sadness*, and *surprise* [24], which were later expanded [22]. The field of affective computing has seen significant growth since the seminal work from Rosalind Picard [65]. The vast majority of research in affective computing has been focused on machine learning algorithms trained on emotion data to classify affect. Like many machine learning solutions, these neural networks focused on classifying the input emotion and ignored data inspection and decision-making explainability.

- *Hamza Elhamdadi is with the University of Massachusetts Amherst. E-mail: helhamdadi@umass.edu.*
- *Shaun Canavan is with the University of South Florida. E-mail: scanavan@usf.edu.*
- *Paul Rosen is with the University of South Florida. E-mail: prosen@usf.edu.*

Manuscript received xx xxx. 201x; accepted xx xxx. 201x. Date of Publication xx xxx. 201x; date of current version xx xxx. 201x. For information on obtaining reprints of this article, please send e-mail to: reprints@ieee.org. Digital Object Identifier: xx.xxxx/TVCG.201x.xxxxxxx

To inspect the data, an effective visual representation of emotion data must address numerous challenges. First, the affect data are quite large, captured by multiple high-speed video cameras. Fortunately, previous affective computing research already partially addressed this issue by reducing the data to 83 landmark points tracked temporally. Nevertheless, the problem remains challenging because patterns in emotion occur over extended time periods, represented by a series of 83-landmark poses. Furthermore, patterns of interest may occur in different time sequences lasting for different lengths of time, making alignment and comparison non-trivial. Finally, changes in landmarks are simultaneously subtle and subject to noise from the extraction process, making them difficult to observe.

This paper presents a visual analytics approach utilizing Topological Data Analysis (TDA) to examine emotion data respective and irrespective of time. By using TDA to address this problem, our approach can capture and track the topological "shape" of facial landmarks over time in a manner robust to noise [20]. After analysis, the data are presented for investigation using familiar visualizations, e.g., timelines (see Fig. 5 top) and scatterplots (see Fig. 1 top right), and through landmark-based representations (see Fig. 1 bottom right or Fig. 10). These interfaces enable tracking facial movement, comparing emotions, and comparing individuals, while also providing the ability to derive precise explanations for features identified in the data.

A natural question at this point would be, why is TDA well-suited to this problem? Our approach utilizes one of the foundational tools of TDA, namely persistent homology. There are four main advantages

to this tool. (1) Persistent homology has a solid mathematical grounding, and its output is explainable. (2) Persistent homology extracts homology groups, which in our context are (connected) components and tunnels/cycles. These fundamental shapes match well with the shapes of a face. (3) The homology groups are extracted at multiple scales, without the need to specify any thresholds or other parameters. This means that persistent homology captures all of the topological structures without any user intervention[1]. (4) Finally, it classifies features by their importance with a measure called persistence, which automatically differentiates topological signal from noise [13].

The specific contributions of this paper are:

1. a mapping of affective computing data to TDA (see Sect. 4), including a novel non-metric formulation of geometry for faster and more accurate topology extraction (see Sect. 4.3);
2. a visual analytics interface that enables analyzing, comparing, and contrasting multiple data configurations (see Sect. 6);
3. an evaluation that uses our methodology to explain features in data that were extracted by state-of-the-art emotion detection machine learning algorithms (see Sect. 7.2); and
4. an evaluation of the ability of TDA to differentiate emotions within the same individual (see Sect. 7.3) and differentiate multiple individuals showing the same emotion (see Sect. 7.4).

Perhaps most importantly, our approach opens the door to explainability in a way that may help to unlock open questions in the affective computing community.

2 Background in Affective Computing

Affective computing has applications in fields as varied as medicine [85], entertainment [31], and security [45]. Most notably, the expression recognition sub-field focuses on detecting subjects' affective states automatically.

2.1 Expression Recognition

While successful 2D facial-expression image recognition exists [29, 43, 49], the approaches suffer from weaknesses, such as occlusion from, e.g., a rotating head. We focus our discussion instead on a few representative 3D facial recognition approaches. Zhen et al. [91] developed a model that localized points within each muscular region of the face and extracted features that include coordinate, normal, and shape index [46]. The features were then used to train a Support Vector Machine (SVM) [77] to recognize expressions. Xue et al. [82] proposed a method for 4D (3D + time) expression recognition, which showed promise differentiating difficult emotions, such as *anger* and *sadness*. The method extracted local patch sequences from consecutive 3D video frames and represented them with a 3D discrete cosine transform. Then, a nearest-neighbor classifier was used to recognize the expressions. Hariri et al. [36] proposed an approach to expression recognition using manifold-based classification. The approach sampled the face by extracting local geometry as covariance regions, which were used with an SVM to recognize expressions.

Some recent techniques showed that not all regions of the face carry the same importance in emotion recognition. Hernandez-Matamoros et al. [37] found that segmenting the face based on the eyes and mouth resulted in improved expression recognition. Fabiano et al. [28] further illustrated that different areas of the face carry different levels of importance for emotions, e.g., one subject *happiness* had more important features on the right eye and eyebrow, while embarrassment had more on the left eye and eyebrow. We utilize this information in our visualization design by targeting specific subsets of facial features.

2.2 Affective Computing in Visualization

There has been limited work in the visualization community on affective computing; what exists has been primarily focused on *visualizing affective states*, i.e., considering valence and arousal, not inspecting the landmarks used as input to affective computing algorithms.

Early work on visualizing affective states concerns the glyph-based Self-Assessment Manikin (SAM), which measures pleasure, arousal, and dominance of a person's affective state [7]. Cernea et al. [9] later described guidelines for conveying the user emotion through the use of widgets that depict the affective states of valence and arousal. The widgets employed *emotion scents*, hue-varied colormaps representing either valance or arousal, e.g., red and green represent negative and positive valance, respectively. Emotion-prints was an early system to provided real-time feedback of valance and arousal to users using touch-displays [10]. More recently, Kovacevik et al. [47] employed ideas from SAM and emotion scents to create a glyph for simultaneous representation of valence and arousal. Their research focused on video game players' and developers' awareness of emotions elicited from a particular gaming experience. For visualizing affect over extended periods, AffectAura provided an interface that enabled users to visualize emotional states over time for the purpose of reflection [56].

There has also been some work visualizing the affective state of multiple individuals using, e.g., virtual agents in collaborative work [11] or using a visual analytics interface to access the emotional state of students in a classroom [86]. Qin et al. [66] created HeartBees, which was an interface to demonstrate the affect of a crowd using physiological data. The interface used an abstract flocking behavior to demonstrate the collective emotional state.

In contrast to all of these prior approaches, our work focuses on using TDA and visualization to investigate the data used in classifying expression, i.e., the input data, not the emotional state itself. There has been some recent work that looked at the explainability of deep networks in expression recognition, e.g., [62]. These approaches focus on visualizing heatmaps that highlight *what* parts of the image most influenced decision making, not necessarily *why*.

2.3 Dataset

To evaluate our approach, we use the BU4DFE 3D facial expression dataset [84], which has been extensively used for expression recognition [12, 27, 63, 74], 3D shape reconstruction [32, 33, 53], face tracking [8, 64], and face recognition [3, 30, 42, 72]. The dataset contains 101 subjects (58 female and 43 male) from multiple ethnicities, including Caucasian, African American, Asian, and Hispanic, with an age range of 18-45 years old. Each modality has the six basic emotions: *anger*, *disgust*, *fear*, *happiness*, *sadness*, and *surprise*. For each sequence, the expression is the result of gradually building from neutral to low then high intensity and back again. Each of the video sequences is 3-4 seconds in length.

The data are captured using the Di3D dynamic face capturing system [17], which consists of three cameras, two to capture stereo and one to capture texture. Passive stereophotogrammetry is used on each pair of stereo images to create the 3D facial pose models with an RMS accuracy of 0.2 mm. Each 3D model contains 83 facial landmarks (see Fig. 2), which correspond to the key areas of the face that include the mouth, eyes, eyebrows, nose, and jawline. The landmarks are the result of using an active appearance model [14] that detects the landmarks on the 2D texture images, which are aligned and projected into the corresponding 3D models.

3 Overview of the Pipeline

TDA has received significant attention in the visualization community, e.g., [73]. We utilize a foundational tool of TDA, persistent homology, which has been studied in graph analysis [34, 35, 67, 71], high-dimensional data analysis [78], and multivariate analysis [68]. We utilize persistent homology to capture the topology of the landmarks of each facial pose into a structure known as a persistence diagram. We then compare the topology of different subsets of facial poses to reveal their relationships. Our processing pipeline contains three main stages, which are fed into the visualization (see Sect. 6).

Stage 1: The first stage is extracting the topology of a single facial pose. We offer two variations, a Euclidean metric-based approach (see Sect. 4.1) and a novel non-metric-based approach (see Sect. 4.3).

Stage 2: Once the topologies of individual poses are extracted, we compare the topology to that of other poses to determine their pairwise dissimilarity (see Sect. 5.1).

[1]Note that while persistent homology itself has no parameters, our pipeline does have some options available to users.

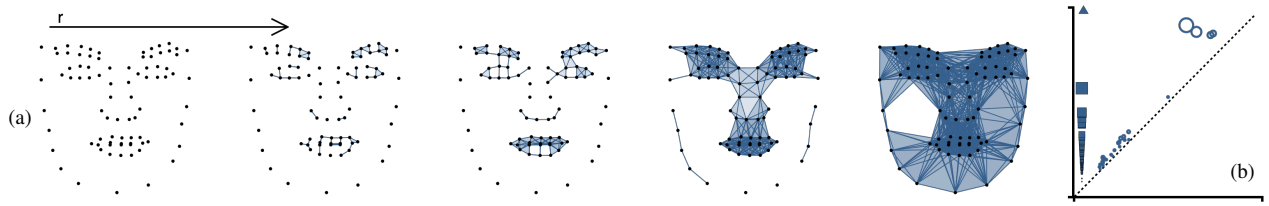


Fig. 2. An illustration of persistent homology on the 83 facial landmarks on the female subject F001. (a) The persistent homology is calculated by forming a Rips filtration and tracking/extracting the associated homology groups. Starting at $r = 0$, if the pairwise distance between any two or three points is less than r, an edge or triangle is formed, respectively. As r increases, components merge, and tunnels form and disappear. (b) The topology is visualized with a persistence diagram. Square points are H_0 components (the triangle indicates a single infinite H_0 component), and the hollow circles are the H_1 tunnels. The horizontal position of points is their birth r_{b_i} and their vertical position is their death r_{d_i}. Distance from the dotted diagonal, as well as object size, is proportional to its persistence.

Stage 3: Finally, using the topological dissimilarity, we utilize a variety of dimension reduction techniques to highlight different aspects of the dissimilarity between groups of facial poses (see Sect. 5.2).

4 Topological Data Analysis of Facial Landmarks

We consider two variations for extracting the topology of facial poses, a Euclidean metric approach, followed by a novel non-metric variant.

4.1 Euclidean Metric Persistent Homology on Landmarks

Homology deals with the topological features of a space. Given a topological space $\mathbb{X}$, we are interested in extracting the $\mathsf{H}_0(\mathbb{X})$ and $\mathsf{H}_1(\mathbb{X})$ homology groups, which correspond to (connected) components and tunnels/cycles of $\mathbb{X}$, respectively[2]. In practice, there may not exist a single scale that captures the topological structures of the data. Instead, we use a multi-scale notion of homology, called *persistent homology*, to describe the topological features of a space at different spatial resolutions. We briefly describe persistent homology in our limited context. Nevertheless, understanding persistent homology can be daunting for those who are unfamiliar with it. For a high-level overview, see [79], or for detailed background, see [19].

To calculate the persistent homology of a single facial pose, we first calculate the Euclidean distance between all 83 landmarks. We then apply a geometric construction, the Rips complex, $R(r)$, on the point set. In brief, for a given distance, r, the Rips complex has all 0-simplicies, i.e., points, for all values of r. A 1-simplex, i.e., an edge, between two points is formed *iff* r is greater than or equal to their distance. A 2-simplex, i.e., a triangle, is formed among three points *iff* r is greater than or equal to every pairwise distance between the points.

To extract the persistent homology (see Fig. 2(a)), we consider a finite sequence of increasing distances, $0 = r_0 \leq r_1 \leq \cdots \leq r_m = \infty$. A sequence of Rips complexes, known as a Rips filtration, is connected by inclusions, $R(r_0) \to R(r_1) \to \cdots \to R(r_m)$, and the homology of each is calculated, tracking the homomorphisms induced by the inclusions, $\mathsf{H}(R(r_0)) \to \mathsf{H}(R(r_1)) \to \cdots \to \mathsf{H}(R(r_m))$. As the distance increases, topological features, i.e., components and tunnels, appear and disappear. The appearance is known as a *birth* event, r_{b_i}, and the disappearance is known as a *death* event, r_{d_i}. The birth and death of all features are stored as a multi-set of points in the plane, (r_{b_i}, r_{d_i}), known as the *persistence diagram*, which is often visualized in the scatterplot display (see Fig. 2(b)). From the points, we devise an importance measure, called *persistence*, which helps to differentiate signal from noise. The persistence is simply the difference between the birth and death of a feature, i.e., $r_{d_i} - r_{b_i}$. Furthermore, in visualizations of the persistence diagram, such as Fig. 2(b), distance from the diagonal dotted line represents the persistence of a feature.

In addition to considering all the topology of all landmarks, we provide the user the functionality to consider only related subsets of features. In particular, they have the option of including/excluding jawline, mouth, nose, left/right eyes, and left/right eyebrows in the calculation of the topology.

[2]We do not consider H_2 (voids) because despite our data being 3D, it is nearly flat. Thus, voids rarely occur, and when they do, they have low persistence.

4.2 Interpolating Known Geometry

Our computation using facial landmarks ignores an important aspect of the data, namely the known connectivity between landmarks. In other words, landmarks of, e.g., the mouth, have known connectivity to their neighboring landmarks. Fig. 3(f) shows this connectivity. This raises two questions. First, does our failure to consider this connectivity impact the features we extract, and second, how do we efficiently consider the connectivity?

We first consider using interpolation of the connectivity to supersample additional landmarks. For our experiment, we take the known connectivity and interpolate across each edge, such that points are no further than a user-defined ε apart. Fig. 3(b) through Fig. 3(e) show four examples with ever-smaller ε values. As expected, as ε gets smaller, the data looks increasingly similar to the known connectivity in Fig. 3(f).

We now consider the impact of the connectivity by comparing the persistence diagrams of H_1 features in the original data in Fig. 3(a) to the lowest ε data in Fig. 3(e). The persistence diagrams are clearly different (the H_0 features are also different but more difficult to observe pictorially). The difference is exceedingly important because it means *using the 83 landmark points alone is insufficient to capture the topological structure of the data*.

To overcome this limitation, we considered using the supersampled landmarks for calculations. However, there are three interrelated problems to this approach. (1) The first is the challenge of selecting an appropriate ε value. The smaller the value, the closer the representation is to the geometric structure. For example, Fig. 3(c) appears sufficient for this example, but it is unclear if this is sufficient for all of the data, leading one to perhaps select an even smaller ε. (2) However, the second challenge is that the smaller the ε, the longer the computation time for detecting the topological features. Fig. 4(a) shows that as ε is divided in half, the compute time grows exponentially. (3) The third related challenge is that the smaller ε, the greater the number of topological features generated. Fig. 4(b) shows this extreme growth. To make matters worse, the vast majority of these features are *topological noise* with very low persistence. In other words, they do not contribute to our understanding of the shape of the face.

4.3 A Non-metric Variant of Persistent Homology

We instead use a novel modification to the persistent homology calculation to utilize this connectivity as follows. Instead of considering 83 landmark points, we consider the relationship between 81 landmark edges formed by the known connectivity of the landmark points (see Fig. 3(f)). We calculate a distance matrix representation of the landmark edges, where the distance is the *shortest Euclidean distance between line segments*. Finally, we run persistent homology calculations on this distance matrix.

One immediate question should be the appropriateness of this configuration for persistent homology calculations, particularly considering that this representation breaks two important axioms of a metric space, namely the identity of indiscernibles and the triangle inequality. Fortunately, persistent homology calculations themselves do not explicitly require a metric space—they have a weaker requirement of inclusion [19]. In other words, as long as in the filtration $R(r_i) \subset R(r_{i+1})$, the calculation can proceed.

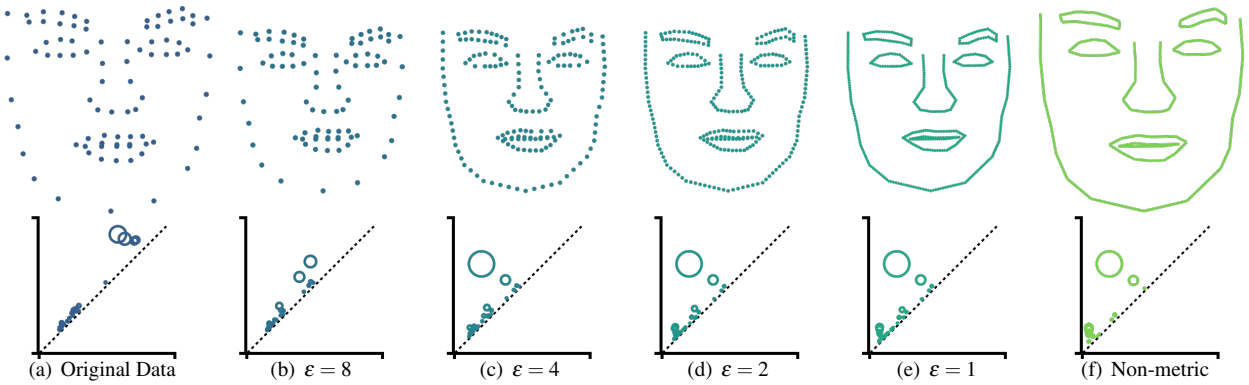


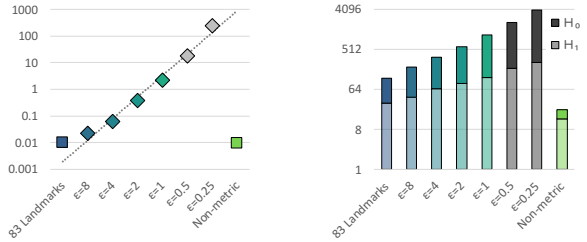
Fig. 3. Illustration of supersampling and non-metric persistent homology shows the data and the persistence diagram of H_1 features. (a) The original 83 landmarks from a single pose on the female subject F001. (b-e) Supersampling the landmarks with different ε values shows a significant difference in the persistence diagram between the (a) original and (e) supersampled data. (f) Our non-metric representation of the data requires significantly less data and produces a persistence diagram similar to that of (e) supersampling.

The challenge is that the Rips complex does require that the underlying space is metric. We define a new non-metric Rips complex that satisfies the inclusion property, where: 0-simplicies, representing landmark edges, are present for all values of r; 1-simplicies appear when r is *strictly greater than* the non-metric distance between a pair of 0-simplicies; and 2-simplices appear when r is *strictly greater than* all of the non-metric distances of the three related 1-simplicies. Fortunately, this definition is similar enough to the standard Rips complex that careful ordering of inclusions (i.e., observing the strictly greater than cases) in the filtration allows us to utilize conventional persistent homology tools on our non-metric distances.

Fig. 3(f) shows the landmark edges and the persistence diagram of the associated H_1 features. Our non-metric approach overcomes all three limitations of supersampling. (1) The result is very similar to the output of the supersampling in Fig. 3(e) without the need for specifying any ε parameter. (2) Furthermore, Fig. 4(a) shows that the compute time for our non-metric approach is approximately the same as that of the original 83 landmark points. (3) Finally, Fig. 4(b) shows the number of topological features output is small (i.e., we avoid outputting extraneous topological noise).

5 Comparing Facial Pose Topology

Thus far, we have introduced a method for extracting the topological features from a single facial pose. We now describe how we compare the topology of multiple facial poses. We start by describing the notion of topological distance between persistence diagrams, which serves as a pairwise dissimilarity between them (see Sect. 5.1). Next, we discuss how dimension reduction is used on all pairwise dissimilarities to cluster, compare, and summarize changes in topology (see Sect. 5.2).



(a) Calculation Time in Seconds (b) Number of Features Generated

Fig. 4. Plots of (a) the compute time and (b) the number of topological features generated for the original 83 landmarks, six levels of supersampling ($\varepsilon = 8$, 4, 2, 1, 0.5, and 0.25), and our non-metric representation. The regression line in (a) only considers the supersampling data points.

5.1 Dissimilarity Between Poses

Once persistence diagrams are calculated, we wish to explore the relationship between them by performing pairwise comparisons of features of the persistence diagrams. This type of pairwise comparison is commonly performed using bottleneck or Wasserstein distance [20]. *Intuitively speaking, these measures find the best match between the features of two persistence diagrams and report the topological feature of the largest distortion, in the case of bottleneck distance, or the average topological distortion, in the case of Wasserstein distance.*

Technically speaking, consider two persistence diagrams, X and Y, let η be a bijection, with all diagonal points, (x,x), added for infinite cardinality [44]. The bottleneck distance is $W_\infty(X,Y) = \inf_{\eta:X\to Y} \sup_{x\in X} \|x-\eta(x)\|_\infty$, and the 1-Wasserstein distance, which we use, is $W_1(X,Y) = \inf_{\eta:X\to Y} \Sigma_{x\in X} \|x-\eta(x)\|_\infty$. Our implementation computes the bottleneck and 1-Wasserstein distance for H_0 and H_1 features separately, and combines the results. In other words, for bottleneck, $\overline{W_\infty}(X,Y) = max(W_\infty(X_{H_0},Y_{H_0}),W_\infty(X_{H_1},Y_{H_1}))$, and for 1-Wasserstein, $\overline{W_1}(X,Y) = W_1(X_{H_0},Y_{H_0}) + W_1(X_{H_1},Y_{H_1})$.

5.2 Summarizing Topological Dissimilarity

Once the set of all persistence diagrams is calculated, we explore the relationship between them by calculating all pairwise dissimilarities between poses, forming a dissimilarity matrix representing all of the topological variations between facial poses. However, a dissimilarity matrix, such as this, is difficult to explore directly. We investigated several options to represent and evaluate the relationship between different facial poses and emotions. Importantly, each technique preserves a different aspect of the dissimilarity matrix, providing different perspectives on the data.

The first approach we used is 1D relative distance. In this approach, a keyframe or focal pose is selected by the user. All other facial poses are positioned by their relative distance (i.e., pairwise distance) to that keyframe. *Relative distance perfectly preserves the relationship between the keyframe and all other frames. It does not, however, provide information about the relationship between other pairs of frames.*

Next, we consider two dimension reduction techniques, with each using the pairwise dissimilarity matrix directly. We first consider Multidimensional Scaling (MDS) [48], which *tries to preserve pairwise distances between the topology of poses*. Second, we use t-SNE [76] and UMAP [57], which attempt to *preserve the clustering structure by considering a local neighbor*. Both t-SNE and UMAP contain hyperparameters that can impact the structures visible to the user. We have performed a structured evaluation of various hyperparameters and found that the structures visible in our results are, by-and-large, stable across a wide variety of parameter values (see our supplement for an example). Therefore, we use the default parameters in our evaluation. To measure the dimension reduction quality for all methods, we cal-

culate the goodness-of-fit using the Spearman rank correlation of the Shepard diagram (denoted in the lower right of images as *Rank*).

Note that *none of these approaches directly consider time*. Nevertheless, the temporal components of the data are presented in the visualization when relevant.

6 Visualization

To examine the topological structure of facial landmark data, we built a visualization (see Fig. 1 and Fig. 10) with the following design criteria:

[D1] provide multiple ways to evaluate temporal and non-temporal aspects of the data (e.g., animated, static, and non-temporal visualizations);

[D2] provide multiple conditional perspectives (e.g., bottleneck vs. Wasserstein, MDS vs. t-SNE vs. UMAP, etc.) on the topology;

[D3] allow comparison of data between two or more emotions;

[D4] allow for investigating subsets of landmarks; and

[D5] provide direct explanations for the topological differences between facial poses.

Small Multiples (Fig. 1 left) Our interface features a small multiples display for comparing different data conditions. The interface features a comparison between two conditions, including comparing two subjects, bottleneck vs. Wasserstein distance, metric vs. non-metric topology, t-SNE vs. MDS, etc. [D2]. The interface further divides each column into comparisons of different subsets of facial features, including full face, eyes+nose, mouth+nose, and eyebrows+nose [D4]. Finally, each row represents one of the six main emotions, *anger*, *disgust*, *fear*, *happiness*, *sadness*, and *surprise*. The user selects the data for the embedding by selecting a column and enabling/disabling specific emotions of interest by selecting rows [D3]. Each small multiple is shown using the visualization modality chosen in the settings at the bottom.

Embedding Graph (Fig. 1 top right) The primary visualization tool in our approach is the embedding graph, which is either a line chart or scatterplot representation of time-varying topological data for the selected emotions [D1]. For the line chart, time is plotted horizontally, while the 1D *relative distance* is plotted vertically (see Fig. 5 top) [D2]. Each selected emotion is overlaid for time-dependent comparison [D3]. The keyframe is user-selectable, and the visualization updates as the keyframe is modified. The scatterplot representation uses 2D dimension reduction for both the horizontal and vertical axes. The choice of MDS, t-SNE, or UMAP is provided for the user. The data are either shown as points or connected via a path (see Fig. 6(a)) if the user wants a temporal context [D1].

The plot is also interactive—selecting a point updates the time-index used in other visualizations, e.g., the 3D landmarks.

3D Landmarks (Fig. 1 lower right) The 3D landmarks represent the data of the current time-index for the respective emotion [D1]. The faces are placed side-by-side for comparison [D3][D5]. The sliders beneath each can be used to animate or adjust their time-index [D1].

Persistence Diagrams (Fig. 10) When additional details about a given facial pose are desired, the persistence diagram captured by persistent homology is represented by a scatterplot [D5]. The persistence diagram plots feature birth horizontally and death vertically. In this context, H_0 features are represented as solid squares, and H_1 features are represented as rings. The size of each element is proportional to its persistence (i.e., importance). Furthermore, the distance from the dashed diagonal to a feature is also a measure of persistence.

Representative Components and Cycles (Fig. 10) A byproduct of the calculation of persistent homology is a structure known as generators, which are the landmark elements that generated a particular topological feature. For H_0, the generators are the 0-simplices representing the joining of two components. For H_1 features, the data are output in the form of a representative cycle[3]. Each topological feature is associated with a generator that we use to *identify what input data generated that topological feature*, with a general focus on the high persistence features in the data [D5].

[3]For reasons outside the scope of this paper, there is no single generator for a cycle but instead a class of generators. A representative cycle, which is a byproduct of a process called boundary matrix reduction, is output instead [21].

7 Evaluation

We evaluate our approach by first performing a detailed evaluation of two individuals—one female ('F001') and one male ('M001') from the BU4DFE expression dataset [84]. We then evaluate the ability of our approach to differentiate individuals using the entire dataset of 101 subjects. Each of these individuals has approximately 600 facial poses (6 emotions $\times \sim 100$ frames per emotion). Since the data provided are large and time-varying, our approach allows conditional observation of the topology of emotions based upon individuals, emotions, selected subset of facial features (full face, eyes+nose, mouth+nose, eyebrows+nose), topological dissimilarity, and dimension reduction technique. Our evaluation looks at how these conditional comparisons can be matched to known phenomena in affective computing. We note that one of the coauthors of this paper, Shaun Canavan, is a researcher in affective computing and provided detailed feedback at every stage of the design.

7.1 Implementation and Performance

We implemented our approach using Python for data management, non-metric distance, and dimension reduction calculations, ripser [6] for persistent homology calculations, Hera [44] for topological distance, and D3.js for the user interface. Persistent homology and topological dissimilarity are pre-calculated for all combinations of landmark subsets. Dimension reduction is performed at run-time, taking at most a few seconds; as this data is calculated, it is also stored in a short-term cache to improve performance. The user interface is interactive. Our source code is available at `https://github.com/USFDataVisualization/AffectiveTDA`.

We evaluated the computational performance of the persistent homology and bottleneck and Wasserstein dissimilarity matrix calculations for F001 and M001 in Table 1. The calculations were performed on a Linux workstation with a 3.40GHz Intel i7-6700 CPU and 48 GB of RAM. In this table, we compare the metric landmark point-based approach and our novel non-metric landmark edge-based approach. Comparing persistent homology calculations, our non-metric approach took approximately twice as long as the metric approach. This is entirely attributable to the extra cost of calculating segment-segment distance (instead of point-point distance). The performance benefit of the non-metric approach comes with the calculation of the dissimilarity matrices, which saw a $10x - 15x$ speedup over the metric approach. This is attributable to the reduced number of noise features created by the non-metric approach, as described in Sect. 4.3. Overall, our approach saw a speedup of $\sim 7.5x$.

Table 1. Computation time for metric (M) and non-metric (NM) approaches to extract persistent homology features and calculate the dissimilarity matrix for all frames from each subject, including subsets of facial landmarks (full face, eyes+nose, mouth+nose, eyebrows+nose). The number of landmarks input to each method is similar, 83 for full face metric and 81 for non-metric. In addition, we show the average number of H_0 and H_1 features generated per frame.

		$\lvert H_0 \rvert$ (avg)	$\lvert H_1 \rvert$ (avg)	Persistent Homology	Topological Distance		Total
					Bottle.	Wasser.	
Female (F001)	M	43.7	12.8	84.3 *s*	1833.8 *s*	1851.7 *s*	3769.8 *s*
	NM	4.3	6.3	188.5 *s*	189.8 *s*	133.7 *s*	512.1 *s*
Male (M001)	M	43.7	13.4	91.6 *s*	1877.6 *s*	2039.2 *s*	4008.4 *s*
	NM	4.3	6.5	188.1 *s*	207.5 *s*	136.4 *s*	532.0 *s*

7.2 Relative Distance Topology and Action Units (AUs)

In affective computing, there are various approaches for recognizing expressions, as detailed in Sect. 2.1. One promising approach is the use of action units (AUs) [25], which are facial muscle movements linked to expression. AUs are represented as an intensity from $[0, 5]$, where 0 is

inactive, and > 0 is an active AU, with higher values representing more intense movement. Specific configurations of active AUs have been shown to be useful for recognizing facial expressions [51,52,54,69,81].

AUs are generally created in one of two ways. Either an expert manually annotates video frames, or a machine learning algorithm extracts them from the data. While the former is a slow and tedious process, the latter is fast but lacks any explainability in measuring the activity of AUs. We automatically detect 17 AUs (see Table 2) using the publicly available OpenFace toolkit [4], which is commonly used in affective computing literature [75]. However, coming from a machine learning model, the extracted AUs lack specific explainability.

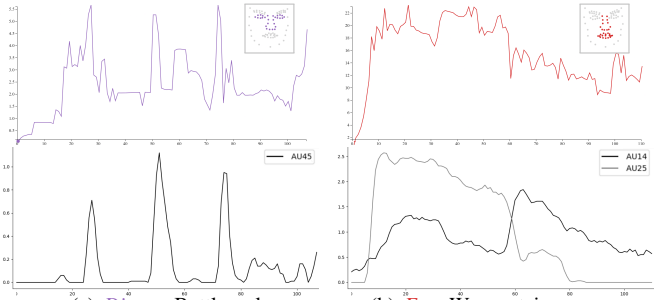
We now demonstrate how our topology-based system can explain certain AU features detected by OpenFace by comparing the output of each. Our approach is as follows, since all sequences begin with a neutral pose, we consider relative distance with respect to the first frame of the sequence. We hypothesized that we would observe similar signals in the AUs and their associated facial features using our topology-based approach. Fig. 5 shows two examples of this relationship. In Fig. 5(a), we compare the activity of the nose and eyes to AU45 (blink) for the F001 *disgust* emotion. The relative distance shows three clear spikes at the same frames as AU45 (approximately frames 28, 52, and 75). However, AU45 does not tell the entire story of the activity that the topology is capturing.

Instead, we hypothesize that the topology is a combination of multiple AUs. Fig. 5(b) shows an example comparing the mouth+nose to AU14 (dimple) and AU25 (lips part) for the F001 *fear* emotion. In this case, a linear combination of both AUs seems to capture a more complete picture of the activity represented in the topology. A broader analysis of both subjects, multiple emotions, and multiple facial features, as seen in Fig. 7 and Fig. 8, revealed these relationships are widely observable. This confirms our hypothesis of a strong similarity between topology features and AUs.

Nevertheless, there is still not a perfect one-to-one relationship between the topology and AUs. The AUs go through further contextual processing than the topology does, e.g., to separate the activity of AU7 (eyelid tightening) from AU45 (blinking) and other related movements. One challenge with the contextual processing in the state-of-the-art in affective computing is the lack of explainability, which our topology-based approach provides (see Sect. 8).

Table 2. Action Units (AUs) and their corresponding facial muscle movements as used in our evaluation.

Action Unit	Facial Muscles	Description
AU1	Frontalis, pars medialis	Inner eyebrow raise
AU2	Frontalis, pars lateralis	Outer eyebrow raise
AU4	Depressor Glabellae, Depressor Supercilli, Currugator	Eyebrow lower
AU5	Levator palpebrae superioris	Upper eyelid raise
AU6	Orbicularis oculi, pars orbitalis	Cheek raise
AU7	Orbicularis oculi, pars palpebralis	Eyelid tighten
AU9	Levator labii superioris alaquae nasi	Nose wrinkle
AU10	Levator Labii Superioris, Caput infraorbitalis	Upper lip raise
AU12	Zygomatic Major	Lip corner pull
AU14	Buccinator	Dimple
AU15	Depressor anguli oris	Lip corner depress
AU17	Mentalis	Chin raise
AU20	Risorius	Lip stretch
AU23	Orbicularis oris	Lip tighten
AU25	Depressor Labii, Relaxation of Mentalis, Orbicularis Oris	Lips part
AU26	Masetter, Temporal/Internal Pterygoid	Jaw drop
AU45	Levator Palpebrae, Orbicularis Oculi, Pars Palpebralis	Blink



(a) *Disgust* Bottleneck Eyes+Nose / AU45

(b) *Fear* Wasserstein Mouth+Nose / AU14+25

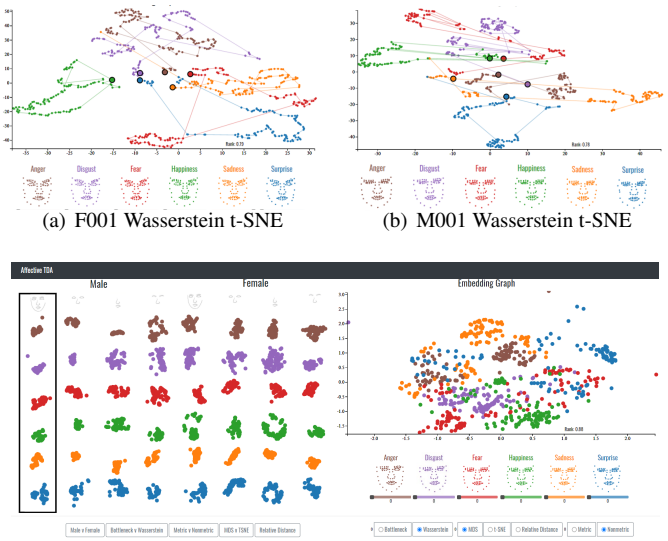
Fig. 5. A comparison of relative distance on non-metric topology (top) to Action Units (AUs) (bottom) on F001. The results demonstrate the similarity between the features extracted by the topology and AUs, which are commonly used in affective computing.

7.3 Comparing and Differentiating Expressions

Next, we consider whether the topological features are sufficient for differentiating the six different emotions present in the data. To perform this evaluation, we look at the full face topology for the female subject using t-SNE (Fig. 6(a), Shepard fitness: 0.79) and MDS (Fig. 1, Shepard fitness: 0.88), and male subject using t-SNE (Fig. 6(b), Shepard fitness: 0.78) and MDS (Fig. 6(c), Shepard fitness: 0.88).

We begin by examining the female and male subjects using t-SNE, as seen in Fig. 6(a) and Fig. 6(b), respectively. We make three important observations about the resulting images. (1) First, for both subjects, the emotional states tend to form separate clusters, indicating that they are indeed differentiable. This is particularly important if the topology were to be used for predicting unknown emotional states. (2) Second, most of the emotions begin and end towards the centers of the plots. This colocation is caused by the neutral facial pose that subjects were asked to begin and end with each sequence. (3) The final observation is that the facial poses form temporally coherent 'strings.' This observation is particularly poignant, considering that *nowhere in calculating the topological dissimilarity does it utilize temporal information*.

We next consider the MDS projections for the female subject and male subject in Fig. 1 and Fig. 6(c), respectively. With the female

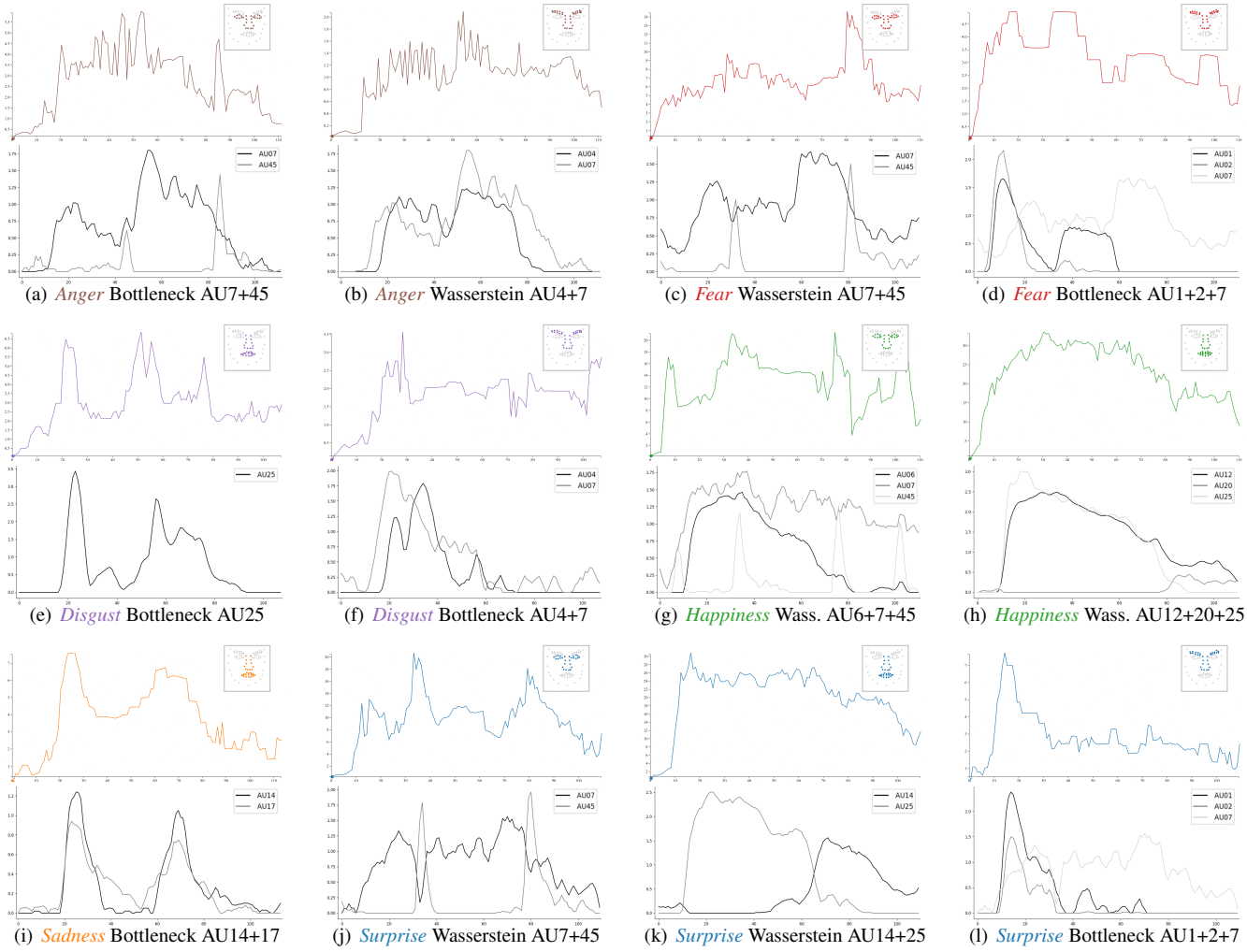


(a) F001 Wasserstein t-SNE

(b) M001 Wasserstein t-SNE

(c) M001 Wasserstein MDS

Fig. 6. Evaluation using (a-b) t-SNE and (c) MDS to determine how effective our non-metric topology-based approach is at differentiating emotions. F001 MDS can be found in Fig. 1.



(a) *Anger* Bottleneck AU7+45 (b) *Anger* Wasserstein AU4+7 (c) *Fear* Wasserstein AU7+45 (d) *Fear* Bottleneck AU1+2+7

(e) *Disgust* Bottleneck AU25 (f) *Disgust* Bottleneck AU4+7 (g) *Happiness* Wass. AU6+7+45 (h) *Happiness* Wass. AU12+20+25

(i) *Sadness* Bottleneck AU14+17 (j) *Surprise* Wasserstein AU7+45 (k) *Surprise* Wasserstein AU14+25 (l) *Surprise* Bottleneck AU1+2+7

Fig. 7. Comparison of F001 non-metric topology (top) and AUs (bottom) shows a similarity between eyes+nose (column 1), mouth+nose (column 2), and eyebrows+nose (column 3) and AUs associated with those facial regions.

subject, we observe that *happiness*, *surprise*, *fear*, and *sadness* largely cluster into separate regions of the plot with limited overlap or mixing. The *anger* and *disgust* emotions, on the other hand, overlap significantly, which corresponds with the recent literature in the affective computing community that considers them to be similar expressions [59]. Another interesting finding is that the emotional states of the male are less differentiated than those of the female. Interestingly, it is commonly accepted in affective computing that, in general, the expressiveness of females is more differentiable than that of males [15]. Given that we are only observing two subjects, no broad gender-based conclusions can be made in our case. Nevertheless, *this* female's expressions are more differentiated than *this* male's expressions.

7.4 Comparing and Differentiating Individuals

Finally, we consider how topological features allow the differentiation of each of the 101 subjects in the BU4DFE dataset.

To perform this evaluation, we look at the full face topology of a subset of 10 subjects (F001-F010) using t-SNE (Fig. 9(a)-9(f)). We notice that, for all six emotions, all ten subjects form relatively independent clusters; this is particularly true of the *anger* (Fig. 9(a)) and *sadness* (Fig. 9(e)) emotions, while some minor overlap occurs for a few of the individuals in the other emotions.

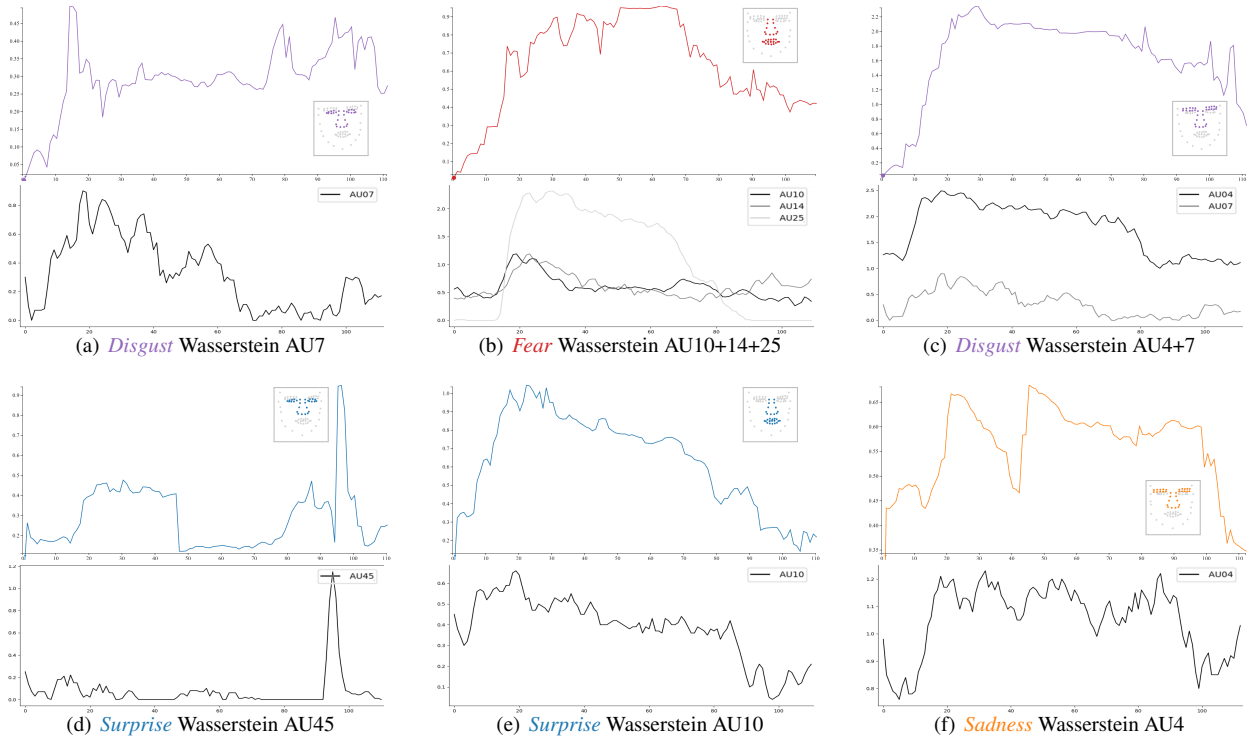
We performed a similar evaluation for all 101 subjects (58 female and 43 male) (Fig. 9(g)-9(l)). We can see that the clustering behavior seen with only 10 subjects scales to all the subjects of the dataset. To test the robustness of this t-SNE result to variations in hyperparameters, we ran the tests with four different perplexities (30, 40, 50, and 100) and found that the clusters remained roughly constant throughout (see supplemental material). The clustering phenomenon present in the t-SNE dimension reduction images was also present when we used UMAP (see supplemental materials).

8 Discussion

8.1 Contribution to Affective Computing

Due to the challenging nature of detecting AUs and determining emotion from expression, we hypothesize that our TDA-based approach can be used to provide new insights into these challenges. As it has been shown that temporal AU information can make recognizing emotions easier, our approach evaluates temporal facial expressions (i.e., AUs), which allows us to visualize a new representation of this data. As shown in Sect. 7, this representation shows the similarity between the topological signals and the AU signals over time, which provides the following insight, as validated by the coauthor on this paper, who is a researcher in affective computing.

Validation That AUs Are Correctly Detected A limitation of current machine learning AU detection models is their accuracy, as little improvement has been made compared to previous models [39]. This is mainly due to the models detecting AUs that are not active, as well as not detecting AUs that are active. Our TDA-based approach can validate that the detected AUs are correctly capturing the muscle movement of the face. More specifically, the proposed approach will ensure that the AUs that have been detected are correct. As shown in Fig. 5(a), AU45 has high-intensity values three times during the sequence. This



(a) *Disgust* Wasserstein AU7 (b) *Fear* Wasserstein AU10+14+25 (c) *Disgust* Wasserstein AU4+7

(d) *Surprise* Wasserstein AU45 (e) *Surprise* Wasserstein AU10 (f) *Sadness* Wasserstein AU4

Fig. 8. Comparison of M001 non-metric topology (top) and AUs (bottom) shows a similarity between eyes+nose (column 1), mouth+nose (column 2), and eyebrows+nose (column 3) and AUs associated with those facial regions.

correctly corresponds to the three blinks that occur in the data, which are also captured in the topological signal. If the blinks were not captured in the topological signal, then the spikes in the AU intensity signal could be attributed to mislabelling or noise. This could facilitate more intelligent active learning [1] that would improve the machine learning detection models.

Relationship Between Which AUs Occur Together AU co-occurrences [70] and patterns [39] can have a significant impact on the accuracy of machine learning models. Our approach can also give insight into multiple AUs that occur together, including specific AUs that occur when an emotion is expressed. This is detailed in Fig. 5(b), where AU14 and AU25 are active at different intensities at different times during the sequence. The topological signal shows a combination of the two AU signals corresponding to the most intense segments of the active AUs. AU14 is a dimple, and AU25 is active when the lips part, which are common muscle movements that could occur during a wide smile. When a smile occurs, AU6 and AU12 are commonly found together, according to the Facial Action Coding System (FACS) [25]. However, according to Barrett et al. [5], expressions vary across cultures and situations meaning, AU6 and AU12 may not be active in all smiles. The proposed approach can provide insight into this phenomenon, allowing investigations of the relationships of new AUs, over time, for different expressions.

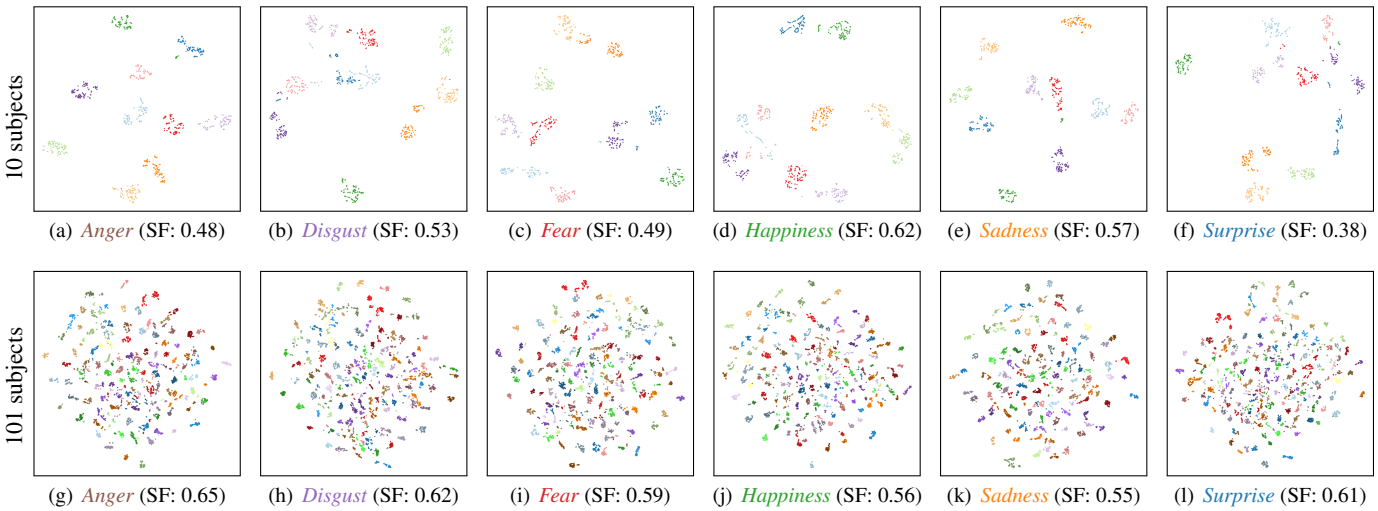
Detecting Facial Expressions There are many successful approaches to detecting facial expressions in affective computing [49,50,58,83,87]. Considering this, the purpose of the proposed approach is *not* to detect facial expressions but to provide greater analysis and explainability of the data. The insight provided by our visualization will allow new insight not previously seen in affective computing, as we can analyze the movement of the face using the proposed TDA-based approach, which directly corresponds to AU movements. This will result in new, more accurate ways of building facial expression detection models.

Explainability of Machine Learning Models Machine learning has given us many advancements in fields as diverse as medicine [80], security [2], and education [16]. However, one of the main limitations is the lack of explainability [18]. Considering this, one of the key advantages of TDA over machine learning is the explainability of the features identified in the process. We demonstrate this using an example of four facial poses from the female *surprise* data, as shown in Fig. 10. These examples focus on the opening and closing of the eyes and mouth, which is commonly associated with a surprised expression. Given a machine learning model that successfully detected the AUs associated with this expression, with the long list of possible muscle movements (e.g., AU1, AU2, AU5, AU25, and AU26), it is difficult to understand *why* such a model detected them. This is especially true given the black-box nature of neural networks [61]. In Fig. 10, we can directly see the features that change the data. In the persistence diagrams, the number and persistence of the most important features are clearly different. Furthermore, when evaluating the representative cycles, we can further associate the landmark geometry of each high persistence feature in the data. This explainability will facilitate more accurate emotion recognition systems. This is due to the insight that the explainability will give affective computing researchers to better understand and tune their models, resulting in more accurate models.

8.2 Topology Doesn't Capture Everything

Limits of Topology Some of the advantages of topology over geometry are also its biggest weaknesses. There are certain shapes of the data that may not be captured by topology alone. For example, smiles and frowns *may* have the same topological shape. That said, the relationship between the smile, nose, and jawline may be sufficient enough to disambiguate between smiles and frowns. Furthermore, smiles and frowns will also be associated with other changes in facial features, e.g., changes in eye or eyebrow shape. At the very least, our evaluation showed that smile emotions, e.g., *happiness*, and frown emotions, e.g., *sadness*, were differentiable in Sect. 7.3.

Differences Between TDA and Machine Learning Topology does not capture all features of AUs. The AU extraction may utilize other data, nonlinearities, knowledge of physiological relationships of AUs, etc., to determine the extent of the activation of the AUs. That said, it is also important to understand that *the AU information is not ground*



(a) *Anger* (SF: 0.48) (b) *Disgust* (SF: 0.53) (c) *Fear* (SF: 0.49) (d) *Happiness* (SF: 0.62) (e) *Sadness* (SF: 0.57) (f) *Surprise* (SF: 0.38)

(g) *Anger* (SF: 0.65) (h) *Disgust* (SF: 0.62) (i) *Fear* (SF: 0.59) (j) *Happiness* (SF: 0.56) (k) *Sadness* (SF: 0.55) (l) *Surprise* (SF: 0.61)

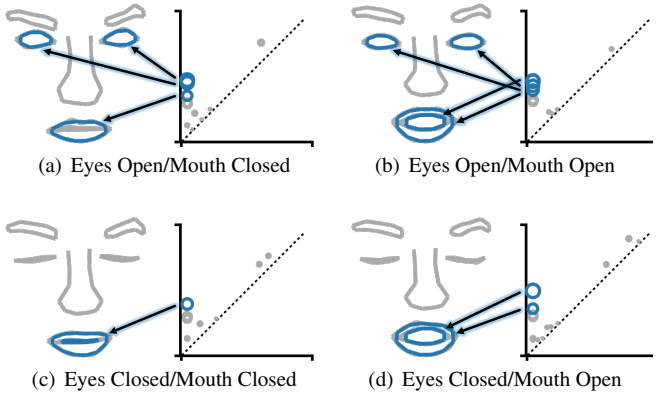
Fig. 9. Clustering of 10 subjects (F001-F010) on the top and all 101 subjects on the bottom. For 10 subjects, each subject is colored differently. For 101 subject tests, 12 colors were mapped to 101 by creating roughly 10 shades per color. Each plot includes the associated Shepard fitness (SF).

truth. It is the output of a machine learning technique, and it may, in fact, not be showing genuine AU activation. On the other hand, all of the topological features we observed are in the data, and for any matching feature, TDA provides evidence as to why the machine learning algorithm classified AU activation as it did.

8.3 Challenges and Limitations of Automatic AU Detection

The use of AUs for expression recognition is a promising approach. However, there are significant challenges in the detection of them [26]. Many works that have developed approaches for automatically detecting AUs using machine learning have focused on learning single AUs. However, it has been shown that patterns of AUs can have a significant impact on detection [39]. This leads to a bigger limitation of current machine learning approaches to AU detection, namely the data. State-of-the-art machine learning models require a large amount of good data to be accurate. Current models are trained on data that has biases in ethnicity [55], as well as significant imbalances in the distribution of AUs [88, 89]. Along with these data biases and imbalances, the ground truth data is often manually annotated, which is subjective and can lead to errors [60]. This results in machine learning models that are not learning how to represent an AU but learning the distribution of data [39]. In machine learning, many times, the solution to the problem is to collect more data. In the case of AUs, it has the additional challenge that multiple AUs occur simultaneously [70], resulting in an unresolvable imbalance of data for the AUs that occur more often (e.g., AU6 or AU12) [39]. These challenges lead to machine learning models that often fail to recognize AUs that are active, as well as recognize AUs that are not active [90].

Along with challenges in detecting AUs, there is a larger discussion of how humans learn and express emotions [41]. Broadly, this discussion can be categorized into two hypotheses. The first hypothesis states that emotions can be recognized from facial expressions (AUs) [24]. This hypothesis is the basis for the Facial Action Coding System (FACS) AUs [25] and is a significant motivation for the field of affective computing. The second hypothesis contradicts that and states that expressions vary across cultures, situations, and people in the same situation. Because of this, emotions cannot be recognized from facial expressions alone [5]. Instead, other factors such as context, physiology, age, and gender should be considered. While these two hypotheses contradict one another, recent work has shown validity in both hypotheses. More specifically, while it is difficult to determine emotion from AUs given a single facial image when temporal AU information and context are considered, the task becomes easier [38]. Along with this, it has also been shown that the fusion of physiological signals, such as heart rate and respiration, along with AUs can be used to accurately recognize pain in subjects [40]. Although it is a challenging problem to automatically detect AUs using machine learning, these works show that AUs are still a promising approach to solve the challenging problem of emotion from expression.



(a) Eyes Open/Mouth Closed (b) Eyes Open/Mouth Open

(c) Eyes Closed/Mouth Closed (d) Eyes Closed/Mouth Open

Fig. 10. Illustration of the explainability of our TDA-based approach using four poses from F001 *surprise*. For each, the persistence diagrams are shown (right), along with the highest persistence representative cycles (left). The number and persistence of features explains the difference between poses, while the representative cycles explain which landmarks generated those topological features.

9 Conclusions

In this paper, we have demonstrated that TDA can be used to discern and better understand patterns that exist between emotions. Paired with machine learning approaches to affective computing, TDA provides a means to evaluate particular aspects of the data to discern what parts of the face may be causing the machine learning algorithm to recognize the data as a particular emotion or explain the shortcomings or misdiagnoses that the machine learning algorithm provides, e.g., if the algorithm recognizes a *happiness* emotion as *anger*, TDA may help to discern what led to this misdiagnosis. The next phase of this work is to begin evaluating these affective computing hypotheses, particularly those discussed in Sect. 8, using our TDA-based analysis and interface.

Acknowledgments

We thank the anonymous reviewers for their feedback. This project was supported in part by the National Science Foundation (IIS-1845204).

REFERENCES

[1] M. U. Ahmed, K. J. Woo, K. Y. Hyeon, M. R. Bashar, and P. K. Rhee. Wild facial expression recognition based on incremental active learning. *Cognitive Systems Research*, 52:212–222, 2018.

[2] V. Albiero, K. KS, K. Vangara, K. Zhang, M. C. King, and K. W. Bowyer. Analysis of gender inequality in face recognition accuracy. In *Proceedings of the IEEE Winter Conference on Applications of Computer Vision Workshops*, pp. 81–89, 2020.

[3] D. Aouada, K. Cherenkova, G. Gusev, B. Ottersten, et al. 3D Deformation Signature for Dynamic Face Recognition. In *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 2138–2142, 2020.

[4] T. Baltrušaitis et al. Openface: an open source facial behavior analysis toolkit. In *IEEE Winter Conference on Applications of Computer Vision (WACV)*, pp. 1–10, 2016.

[5] L. F. Barrett, R. Adolphs, S. Marsella, A. M. Martinez, and S. D. Pollak. Emotional expressions reconsidered: Challenges to inferring emotion from human facial movements. *Psychological Science in the Public Interest*, 20(1):1–68, 2019.

[6] U. Bauer. Ripser: Efficient Computation of Vietoris-Rips Persistence Barcodes. *arXiv preprint arXiv:1908.02518*, 2019.

[7] M. Bradley and P. Lang. Measuring Emotion: The Self-Assessment Manikin and the Semantic Differential. *Journal of Behavior Therapy and Experimental Psychiatry*, 1994.

[8] S. Canavan, P. Liu, X. Zhang, and L. Yin. Landmark Localization on 3D/4D Range Data Using a Shape Index-Based Statistical Shape Model with Global and Local Constraints. *Computer Vision and Image Understanding*, 139:136–148, 2015.

[9] D. Cernea, C. Weber, A. Ebert, and A. Kerren. Emotion Scents - A Method of Representing User Emotions on GUI Widgets. *SPIE Digital Library*, 2013.

[10] D. Cernea, C. Weber, A. Ebert, and A. Kerren. Emotion-prints: Interaction-driven emotion visualization on multi-touch interfaces. In *Visualization and Data Analysis 2015*, vol. 9397, p. 93970A. International Society for Optics and Photonics, 2015.

[11] D. Cernea, C. Weber, A. Kerren, and A. Ebert. Group affective tone awareness and regulation through virtual agents. In *Proceeding of IVA 2014 Workshop on Affective Agents, Boston, MA, USA, 27-29 August*, pp. 9–16, 2014.

[12] N. Chinaev, A. Chigorin, and I. Laptev. Mobileface: 3D Face Reconstruction with Efficient CNN Regression. In *European Conference on Computer Vision (ECCV)*, pp. 0–0, 2018.

[13] D. Cohen-Steiner, H. Edelsbrunner, and J. Harer. Stability of persistence diagrams. *Discrete & computational geometry*, 37(1):103–120, 2007.

[14] T. Cootes, G. Edwards, and C. Taylor. Active Appearance Models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(6):681–685, 2001.

[15] Y. Deng, L. Chang, M. Yang, M. Huo, and R. Zhou. Gender differences in emotional response: Inconsistency between experience and expressivity. *PloS one*, 11(6):e0158666, 2016.

[16] A. Dhall, G. Sharma, R. Goecke, and T. Gedeon. Emotiw 2020: Driver gaze, group emotion, student engagement and physiological signal based challenges. In *International Conference on Multimodal Interaction*, pp. 784–789, 2020.

[17] *Di3D Inc.* `http://www.di3d.com`.

[18] F. K. Došilović, M. Brčić, and N. Hlupić. Explainable artificial intelligence: A survey. In *International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO)*, pp. 0210–0215, 2018.

[19] H. Edelsbrunner and J. Harer. Persistent Homology - A Survey. *Contemporary Mathematics*, 453:257–282, 2008.

[20] H. Edelsbrunner and J. Harer. *Computational Topology: An Introduction*. American Mathematical Society, 2010.

[21] H. Edelsbrunner, D. Letscher, and A. Zomorodian. Topological persistence and simplification. In *Proceedings 41st Annual Symposium on Foundations of Computer Science*, pp. 454–463, 2000.

[22] P. Ekman. Basic Emotions. *Handbook of Cognition and Emotion*, 98(45-60):16, 1999.

[23] P. Ekman and W. Friesen. The Repertoire of Nonverbal Behavior: Categories, Origins, Usage, and Coding. *Semiotica*, 1(1):49–98, 1969.

[24] P. Ekman, W. Friesen, M. O'Sullivan, A. Chan, I. Diacoyanni-Tarlatzis, K. Heider, R. Krause, W. A. LeCompte, T. Pitcairn, P. Ricci-Bitti, et al. Universals and Cultural Differences in the Judgments of Facial Expressions of Emotion. *Journal of Personality and Social Psychology*, 53(4):712, 1987.

[25] R. Ekman. *What the face reveals: Basic and applied studies of spontaneous expression using the Facial Action Coding System (FACS)*. Oxford University Press, USA, 1997.

[26] I. O. Ertugrul, J. F. Cohn, L. A. Jeni, Z. Zhang, L. Yin, and Q. Ji. Crossing domains for au coding: Perspectives, approaches, and measures. *IEEE Transactions on Biometrics, Behavior, and Identity Science*, 2(2):158–171, 2020.

[27] D. Fabiano and S. Canavan. Deformable Synthesis Model for Emotion Recognition. In *IEEE International Conference on Automatic Face and Gesture Recognition*, pp. 1–5, 2019.

[28] D. Fabiano, M. Jaishanker, and S. Canavan. Impact of Multiple Modalities on Emotion Recognition: Investigation into 3d Facial Landmarks, Action Units, and Physiological Data. *arXiv preprint arXiv:2005.08341*, 2020.

[29] Y. Fan, V. Li, and J. C. Lam. Facial Expression Recognition with Deeply-Supervised Attention Network. *IEEE Transactions on Affective Computing*, 2020.

[30] S. L. Fernandes and G. J. Bala. 3D and 4D Face Recognition: A Comprehensive Review. *Recent Patents on Engineering*, 8(2), 2014.

[31] J. Fleureau, P. Guillotel, and Q. Huynh-Thu. Physiological-Based Affect Event Detector for Entertainment Video Applications. *IEEE Transactions on Affective Computing*, 3(3):379–385, 2012.

[32] W. Gao, X. Zhao, Z. Gao, J. Zou, P. Dou, and I. Kakadiaris. 3D Face Reconstruction From Volumes of Videos Using a MapReduce Framework. *IEEE Access*, 7:165559–165570, 2019.

[33] S. Grasshof, H. Ackermann, S. Brandt, and J. Ostermann. Multilinear Modelling of Faces and Expressions. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2020.

[34] M. Hajij, E. Munch, and P. Rosen. Fast and scalable complex network descriptor using pagerank and persistent homology. In *2020 International Conference on Intelligent Data Science Technologies and Applications (IDSTA)*, pp. 110–114. IEEE, 2020.

[35] M. Hajij, B. Wang, C. Scheidegger, and P. Rosen. Visual Detection of Structural Changes in Time-Varying Graphs Using Persistent Homology. In *IEEE Pacific Visualization*, pp. 125–134, 2018.

[36] W. Hariri, H. Tabia, N. Farah, A. Benouareth, and D. Declercq. 3D Facial Expression Recognition Using Kernel Methods on Riemannian Manifold. *Engineering Applications of Artificial Intelligence*, 64:25–32, 2017.

[37] A. Hernandez-Matamoros, A. Bonarini, E. Escamilla-Hernandez, M. Nakano-Miyatake, and H. Perez-Meana. Facial Expression Recognition with Automatic Segmentation of Face Regions Using a Fuzzy-Based Classification Approach. *Knowledge-Based Systems*, 110:1–14, 2016.

[38] S. Hinduja, S. Aathreya, S. Canavan, J. Cohn, and L. Yin. Recognizing context using facial expression dynamics from action unit patterns. *IEEE Transactions on Affective Computing (Under Review)*, 2020.

[39] S. Hinduja, S. Canavan, and S. Aathreya. Impact of action unit occurrence patterns on detection. *arXiv preprint arXiv:2010.07982*, 2020.

[40] S. Hinduja, S. Canavan, and G. Kaur. Multimodal fusion of physiological signals and facial action units for pain recognition. In *IEEE International Conference on Automatic Face and Gesture Recognition*, pp. 387–391, 2020.

[41] K. Hoemann, R. Wu, V. LoBue, L. M. Oakes, F. Xu, and L. F. Barrett. Developing an understanding of emotion categories: Lessons from objects. *Trends in Cognitive Sciences*, 24(1):39–51, 2020.

[42] S. Jannat, D. Fabiano, S. Canavan, and T. Neal. Subject Identification across Large Expression Variations Using 3D Facial Landmarks. *arXiv preprint arXiv:2005.08339*, 2020.

[43] A. Kalam, M. Haque, M. Jashem, M. Hasan, M. Ibrahim, and T. Jabid. Facial Expression Recognition Using Local Composition Pattern. In *International Conference on Computer and Communications Management*, pp. 63–67, 2019.

[44] M. Kerber, D. Morozov, and A. Nigmetov. Geometry Helps to Compare Persistence Diagrams. *Journal of Experimental Algorithmics*, 22:1–4, 2017.

[45] H. Kim, J. Ben-Othman, L. Mokdad, and K. Lim. CONTVERB: Continuous Virtual Emotion Recognition Using Replaceable Barriers for Intelligent Emotion-based IoT Services and Applications. *IEEE Network*, 2020.

[46] J. Koenderink and A. V. Doorn. Surface Shape and Curvature Scales. *Image and Vision Computing*, 10(8):557–564, 1992.

[47] N. Kovacevic, R. Wampfler, B. Solenthaler, M. Gross, and T. Günther.

Glyph-Based Visualization of Affective States. *EuroVis: Short Papers*, 2020.

[48] J. Kruskal. Multidimensional Scaling by Optimizing Goodness of Fit to a Non-Metric Hypothesis. *Psychometrika*, 29(1):1–27, 1964.

[49] S. Li and W. Deng. Deep Facial Expression Recognition: A Survey. *IEEE Transactions on Affective Computing*, 2020.

[50] Y. Li, J. Zeng, S. Shan, and X. Chen. Occlusion aware facial expression recognition using cnn with attention mechanism. *IEEE Transactions on Image Processing*, 28(5):2439–2450, 2018.

[51] J. J. Lien, T. Kanade, J. F. Cohn, and C.-C. Li. Automated facial expression recognition based on facs action units. In *IEEE International Conference on Automatic Face and Gesture Recognition*, pp. 390–395, 1998.

[52] D. Liu, X. Ouyang, S. Xu, P. Zhou, K. He, and S. Wen. Saanet: Siamese action-units attention network for improving dynamic facial expression recognition. *Neurocomputing*, 413:145–157, 2020.

[53] F. Liu, L. Tran, and X. Liu. 3D Face Modeling from Diverse Raw Scan Data. In *IEEE International Conference on Computer Vision*, pp. 9408–9418, 2019.

[54] P. Lucey, J. F. Cohn, I. Matthews, S. Lucey, S. Sridharan, J. Howlett, and K. M. Prkachin. Automatically detecting pain in video through facial action units. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, 41(3):664–674, 2010.

[55] S. M. Mavadati, M. H. Mahoor, K. Bartlett, P. Trinh, and J. F. Cohn. Disfa: A spontaneous facial action intensity database. *IEEE Transactions on Affective Computing*, 4(2):151–160, 2013.

[56] D. McDuff, A. Karlson, A. Kapoor, A. Roseway, and M. Czerwinski. Affectaura: an intelligent system for emotional memory. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 849–858, 2012.

[57] L. McInnes, J. Healy, and J. Melville. Umap: Uniform manifold approximation and projection for dimension reduction. *arXiv preprint arXiv:1802.03426*, 2018.

[58] S. Minaee, M. Minaei, and A. Abdolrashidi. Deep-emotion: Facial expression recognition using attentional convolutional network. *Sensors*, 21(9):3046, 2021.

[59] C. Molho, J. M. Tybur, E. Güler, D. Balliet, and W. Hofmann. Disgust and anger relate to different aggressive responses to moral violations. *Psychological Science*, 28(5):609–619, 2017.

[60] K. O'Connor, A. Sarker, J. Perrone, and G. G. Hernandez. Promoting reproducible research for characterizing nonmedical use of medications through data annotation: Description of a twitter corpus and guidelines. *Journal of Medical Internet Research*, 22(2):e15861, 2020.

[61] S. J. Oh, B. Schiele, and M. Fritz. Towards reverse-engineering black-box neural networks. In *Explainable AI: Interpreting, Explaining and Visualizing Deep Learning*, pp. 121–144. Springer, 2019.

[62] N. Otberdout, A. Kacem, M. Daoudi, L. Ballihi, and S. Berretti. Automatic analysis of facial expressions based on deep covariance trajectories. *IEEE transactions on neural networks and learning systems*, 31(10):3892–3905, 2019.

[63] G. Patil and P. Suja. Emotion Recognition from 3D Videos Using Optical Flow Method. In *IEEE International Conference On Smart Technologies For Smart Nation (SmartTechCon)*, pp. 825–829, 2017.

[64] H. Pham, V. Pavlovic, J. Cai, and T. jen Cham. Robust Real-Time Performance-Driven 3D Face Tracking. In *IEEE International Conference on Pattern Recognition (ICPR)*, pp. 1851–1856, 2016.

[65] R. Picard. *Affective Computing*. MIT press, 2000.

[66] C. Y. Qin, M. Constantinides, L. M. Aiello, and D. Quercia. Heartbees: Visualizing crowd affects. *arXiv preprint arXiv:2010.07209*, 2020.

[67] B. Rieck, U. Fugacci, J. Lukasczyk, and H. Leitte. Clique Community Persistence: A Topological Visual Analysis Approach for Complex Networks. *IEEE Transactions on Visualization and Computer Graphics*, 24(1):822–831, 2017.

[68] B. Rieck, H. Mara, and H. Leitte. Multivariate Data Analysis Using Persistence-Based Filtering and Topological Signatures. *IEEE Transactions on Visualization and Computer Graphics*, 18(12):2382–2391, 2012.

[69] G. Sikander and S. Anwar. A novel machine vision-based 3d facial action unit identification for fatigue detection. *IEEE Transactions on Intelligent Transportation Systems*, 2020.

[70] Y. Song, D. McDuff, D. Vasisht, and A. Kapoor. Exploiting sparsity and co-occurrence structure for action unit recognition. In *IEEE International Conference on Automatic Face and Gesture Recognition*, vol. 1, pp. 1–8, 2015.

[71] A. Suh, M. Hajij, B. Wang, C. Scheidegger, and P. Rosen. Persistent Homology Guided Force-Directed Graph Layouts. *IEEE Transactions on Visualization and Computer Graphics*, 26(1):697–707, 2019.

[72] Y. Sun and L. Yin. 3D Spatio-Temporal Face Recognition Using Dynamic Range Model Sequences. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, pp. 1–7, 2008.

[73] J. Tierny, G. Favelier, J. Levine, C. Gueunet, and M. Michaux. The Topology Toolkit. *IEEE Transactions on Visualization and Computer Graphics*, 24(1):832–842, 2017.

[74] S. Tornincasa, E. Vezzetti, S. Moos, M. G. Violante, F. Marcolin, N. Dagnes, L. Ulrich, and G. F. Tregnaghi. 3D Facial Action Units and Expression Recognition Using a Crisp Logic. *Computer Aided Design and Applications*, 16:256–268, 2019.

[75] M. T. Uddin and S. Canavan. Quantified facial temporal-expressiveness dynamics for affect analysis. *International Conference on Pattern Recognition*, 2020.

[76] L. van der Maaten and G. Hinton. Visualizing Data Using t-SNE. *Journal of Machine Learning Research*, 9:2579–2605, 2008.

[77] V. Vapnik. *The Nature of Statistical Learning Theory*. Springer science & business media, 2013.

[78] B. Wang, B. Summa, V. Pascucci, and M. Vejdemo-Johansson. Branching and Circular Features in High Dimensional Data. *IEEE Transactions on Visualization and Computer Graphics*, 17(12):1902–1911, 2011.

[79] S. Weinberger. What is... persistent homology? *Notices of the AMS*, 58(1):36–39, 2011.

[80] M. J. Willemink, W. A. Koszek, C. Hardell, J. Wu, D. Fleischmann, H. Harvey, L. R. Folio, R. M. Summers, D. L. Rubin, and M. P. Lungren. Preparing medical imaging data for machine learning. *Radiology*, 295(1):4–15, 2020.

[81] X. Xu and V. R. de Sa. Exploring multidimensional measurements for pain evaluation using facial action units. In *IEEE International Conference on Automatic Face and Gesture Recognition*, pp. 559–565, 2020.

[82] M. Xue, A. Mian, W. Liu, and L. Li. Automatic 4D Facial Expression Recognition Using DCT Features. In *IEEE Winter Conference on Applications of Computer Vision*, pp. 199–206, 2015.

[83] H. Yang, U. Ciftci, and L. Yin. Facial expression recognition by de-expression residue learning. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 2168–2177, 2018.

[84] L. Yin, X. Chen, Y. Sun, T. Worm, and M. Reale. A High-Resolution 3D Dynamic Facial Expression Database. In *IEEE International Conference on Automatic Face and Gesture Recognition*, vol. 126, p. 6, 2008.

[85] G. Zamzmi, C.-Y. Pai, D. Goldgof, R. Kasturi, T. Ashmeade, and Y. Sun. An Approach for Automated Multimodal Analysis of Infants' Pain. In *IEEE International Conference on Pattern Recognition (ICPR)*, 2016.

[86] H. Zeng, X. Shu, Y. Wang, Y. Wang, L. Zhang, T.-C. Pong, and H. Qu. Emotioncues: Emotion-oriented visual summarization of classroom videos. *IEEE transactions on visualization and computer graphics*, 2020.

[87] N. Zeng, H. Zhang, B. Song, W. Liu, Y. Li, and A. M. Dobaie. Facial expression recognition via learning deep sparse autoencoders. *Neurocomputing*, 273:643–649, 2018.

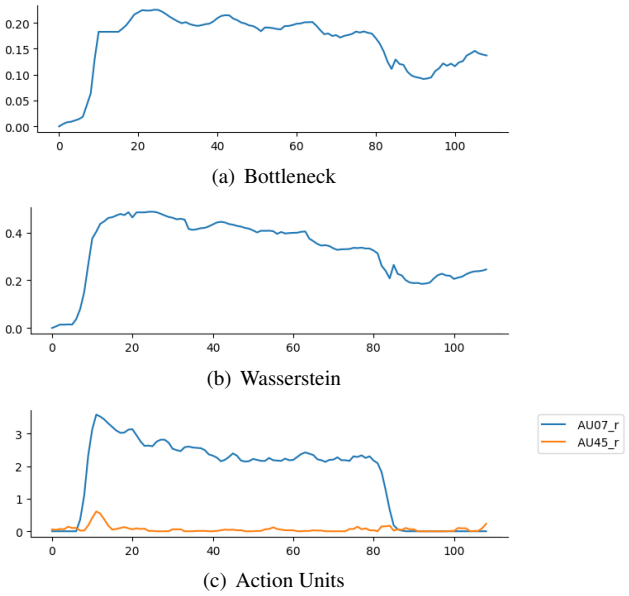
[88] X. Zhang, L. Yin, J. F. Cohn, S. Canavan, M. Reale, A. Horowitz, P. Liu, and J. M. Girard. Bp4d-spontaneous: a high-resolution spontaneous 3d dynamic facial expression database. *Image and Vision Computing*, 32(10):692–706, 2014.

[89] Z. Zhang, J. M. Girard, Y. Wu, X. Zhang, P. Liu, U. Ciftci, S. Canavan, M. Reale, A. Horowitz, H. Yang, et al. Multimodal spontaneous emotion corpus for human behavior analysis. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 3438–3446, 2016.

[90] Z. Zhang, T. Wang, and L. Yin. Region of interest based graph convolution: A heatmap regression approach for action unit detection. In *ACM International Conference on Multimedia*, pp. 2890–2898, 2020.

[91] Q. Zhen, D. Huang, Y. Wang, and L. Chen. Muscular Movement Model-Based Automatic 3D/4D Facial Expression Recognition. *IEEE Transactions on Multimedia*, 18(7):1438–1450, 2016.

ADDITIONAL EXAMPLES OF RELATIVE DISTANCE TOPOLOGY AND ACTION UNITS (AUS)

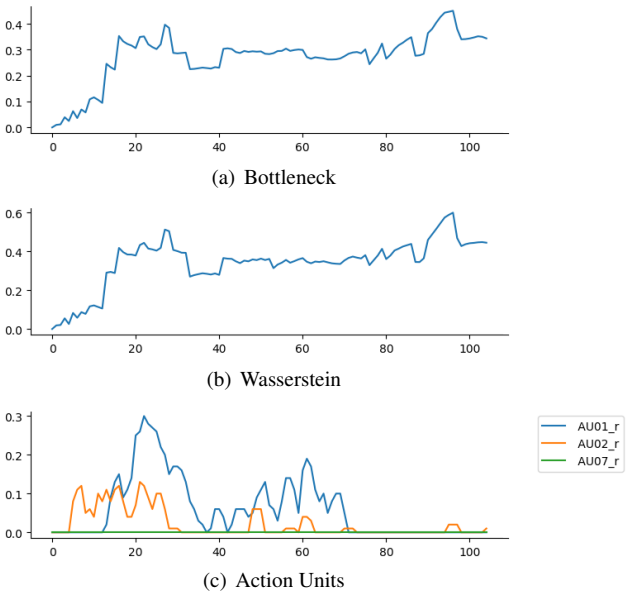


(a) Bottleneck

(b) Wasserstein

(c) Action Units

Fig. 11. F002 *Anger* using eyes, eyebrows, nose, and mouth

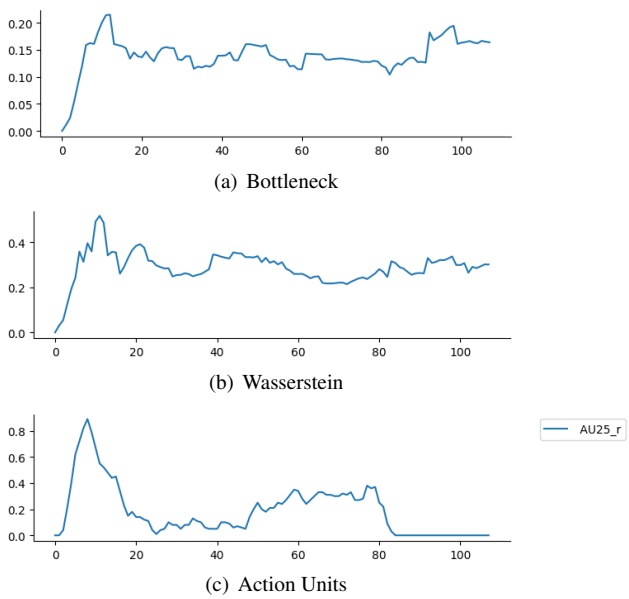


(a) Bottleneck

(b) Wasserstein

(c) Action Units

Fig. 12. F002 *Fear* using eyes, eyebrows, nose, and mouth

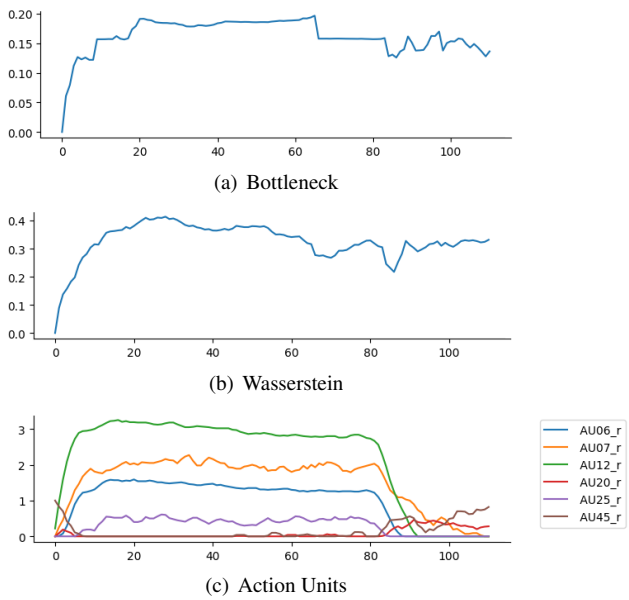


(a) Bottleneck

(b) Wasserstein

(c) Action Units

Fig. 13. F002 *Disgust* using eyes, eyebrows, nose, and mouth



(a) Bottleneck

(b) Wasserstein

(c) Action Units

Fig. 14. F002 *Happiness* using eyes, eyebrows, nose, and mouth

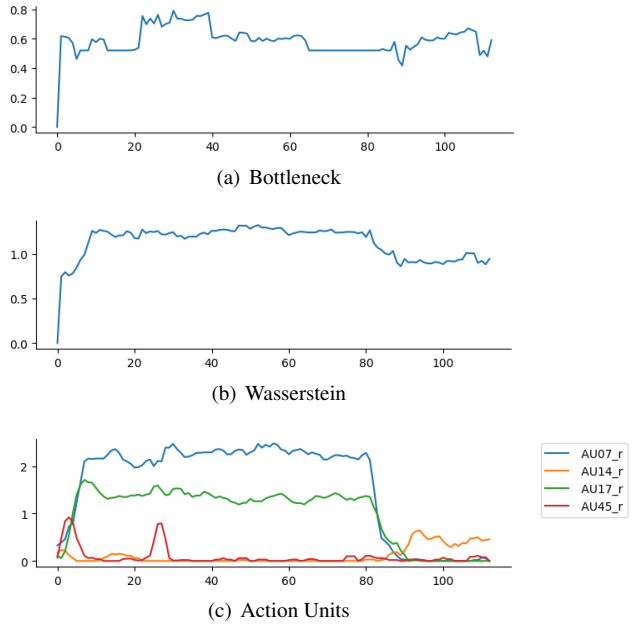


Fig. 15. F002 *Sadness* using eyes, eyebrows, nose, and mouth

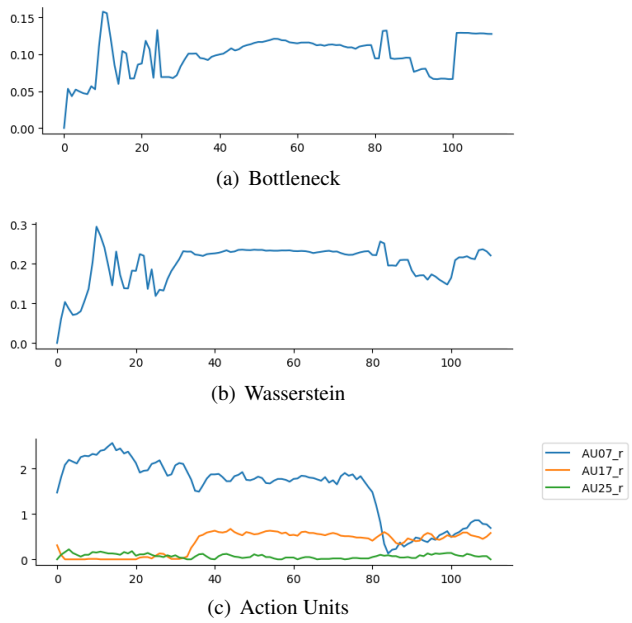


Fig. 16. M002 *Anger* using eyes, eyebrows, nose, and mouth

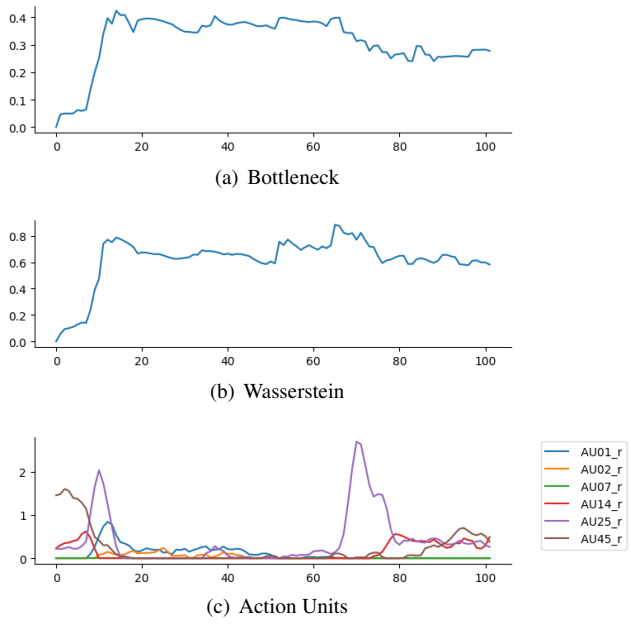


Fig. 17. F002 *Surprise* using eyes, eyebrows, nose, and mouth

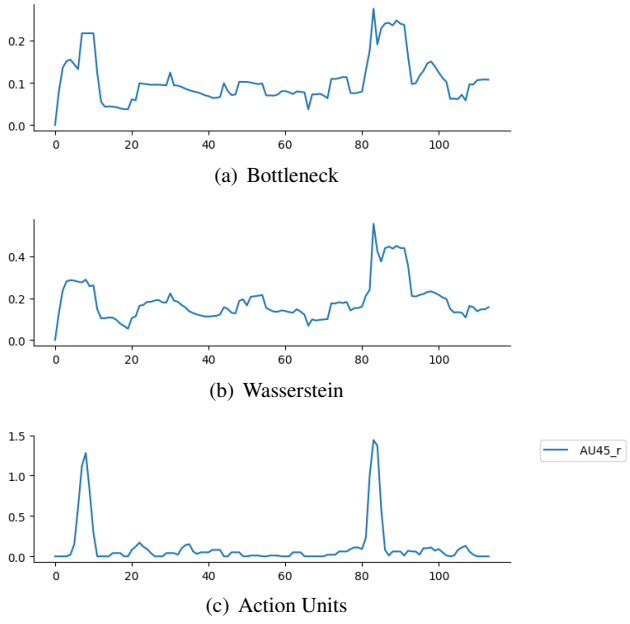
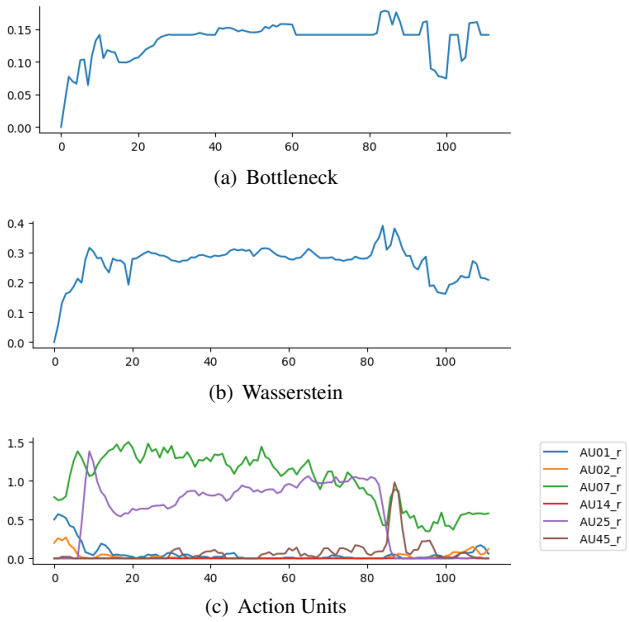


Fig. 18. M002 *Disgust* using eyes, eyebrows, nose, and mouth

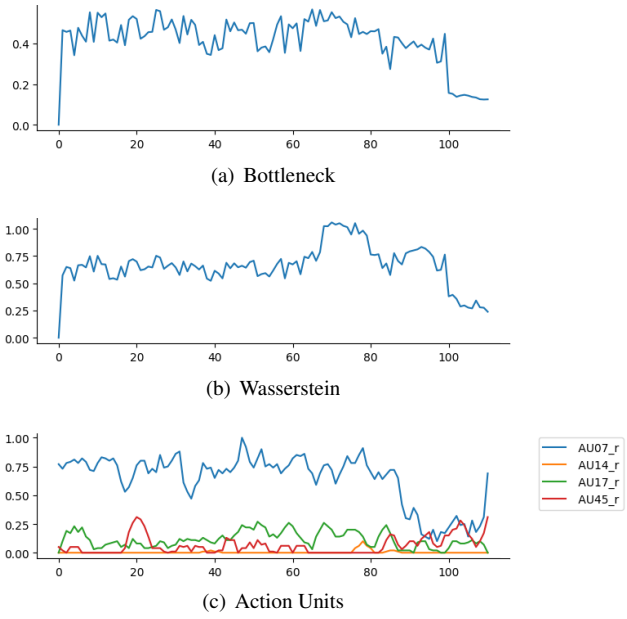


(a) Bottleneck

(b) Wasserstein

(c) Action Units

Fig. 19. M002 *Fear* using eyes, eyebrows, nose, and mouth

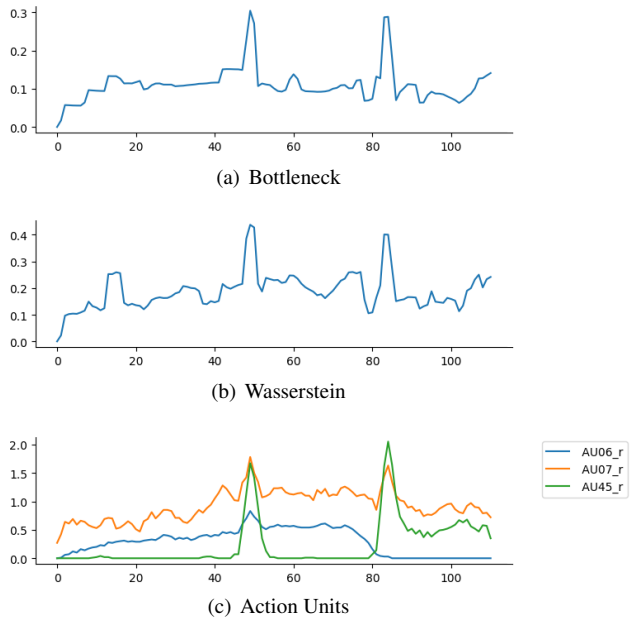


(a) Bottleneck

(b) Wasserstein

(c) Action Units

Fig. 20. M002 *Sadness* using eyes, eyebrows, nose, and mouth

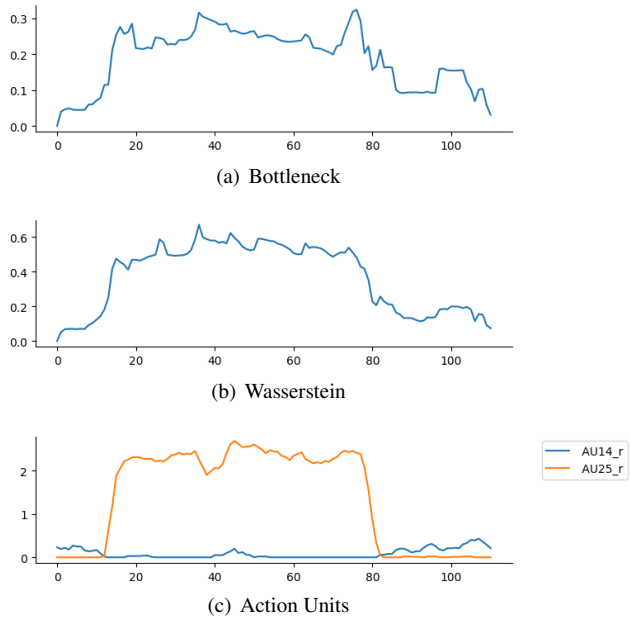


(a) Bottleneck

(b) Wasserstein

(c) Action Units

Fig. 21. M002 *Happiness* using eyes, eyebrows, nose, and mouth



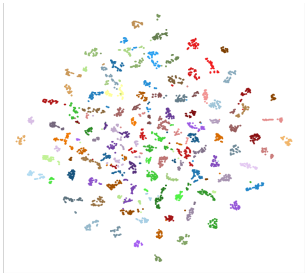
(a) Bottleneck

(b) Wasserstein

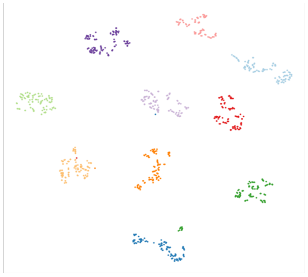
(c) Action Units

Fig. 22. M002 *Surprise* using eyes, eyebrows, nose, and mouth

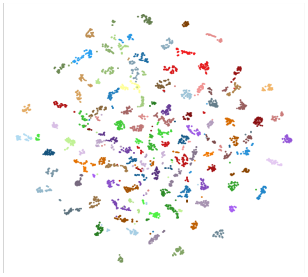
ADDITIONAL EXAMPLES COMPARING AND DIFFERENTIATING INDIVIDUALS



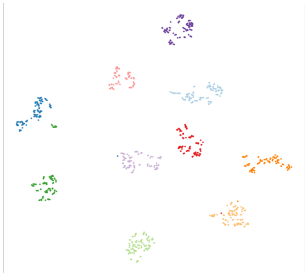
(a) *Anger* perplexity: 30 (full)



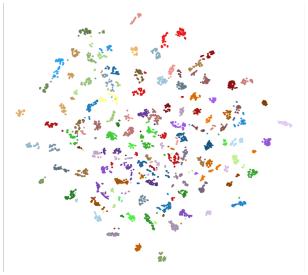
(b) *Anger* perplexity: 30 (10 subj.)



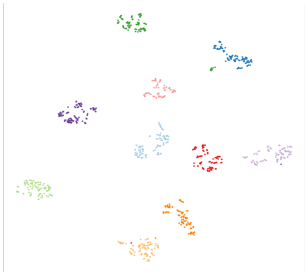
(c) *Anger* perplexity: 40 (full)



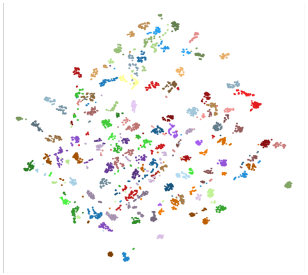
(d) *Anger* (10 subj.) Perplexity=40



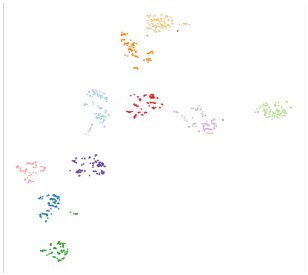
(e) *Anger* perplexity: 50 (full)



(f) *Anger* perplexity: 50 (10 subj.)

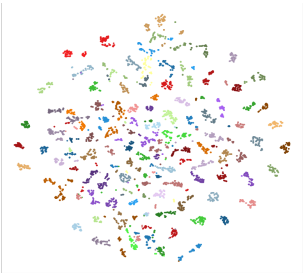


(g) *Anger* perplexity: 100 (full)

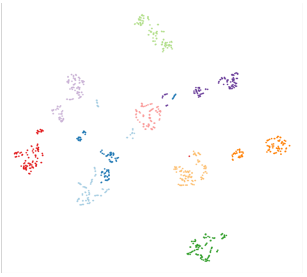


(h) *Anger* perplexity: 100 (10 subj.)

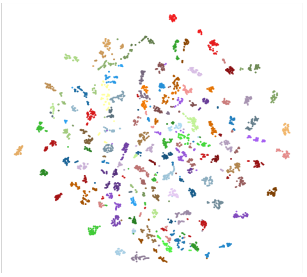
Fig. 23. t-SNE clustering of individual topological data for *Anger* emotion at different perplexities.



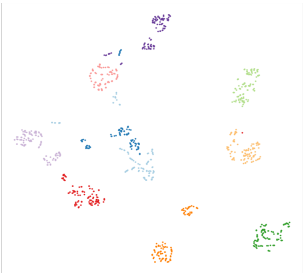
(a) *Disgust* perplexity: 30 (full)



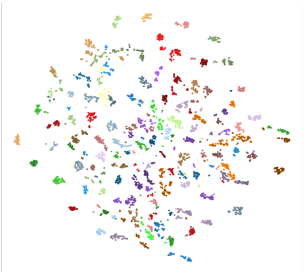
(b) *Disgust* perplexity: 30 (10 subj.)



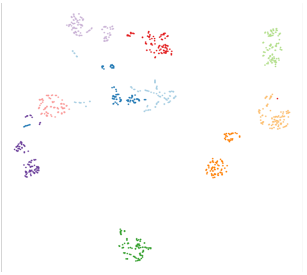
(c) *Disgust* perplexity: 40 (full)



(d) *Disgust* perplexity: 40 (10 subj.)

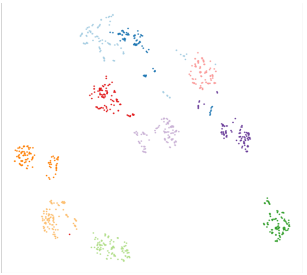


(e) *Disgust* perplexity: 50 (full)



(f) *Disgust* perplexity: 50 (10 subj.)

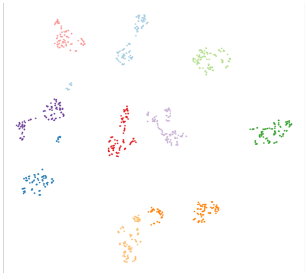
(g) *Disgust* perplexity: 100 (full)



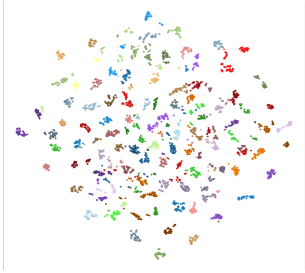
(h) *Disgust* perplexity: 100 (10 subj.)

Fig. 24. t-SNE clustering of individual topological data for *Disgust* emotion at different perplexities.

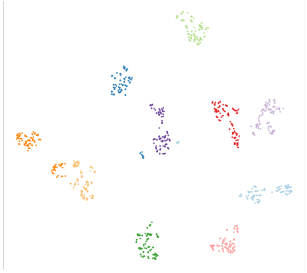
(a) *Fear* perplexity: 30 (full)



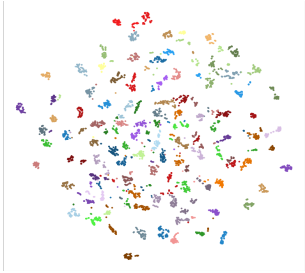
(b) *Fear* perplexity: 30 (10 subj.)



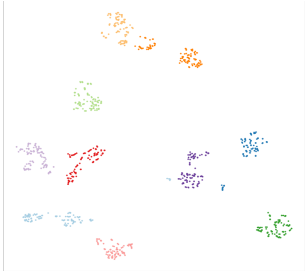
(c) *Fear* perplexity: 40 (full)



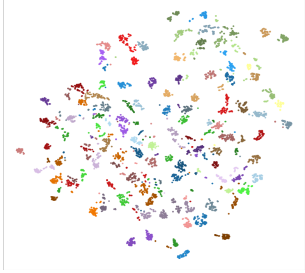
(d) *Fear* perplexity: 40 (10 subj.)



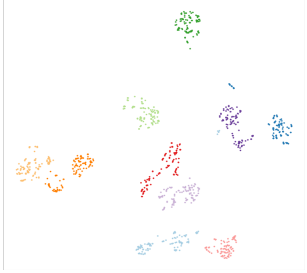
(e) *Fear* perplexity: 50 (full)



(f) *Fear* perplexity: 50 (10 subj.)

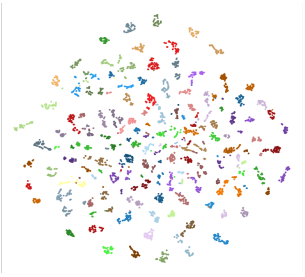


(g) *Fear* perplexity: 100 (full)

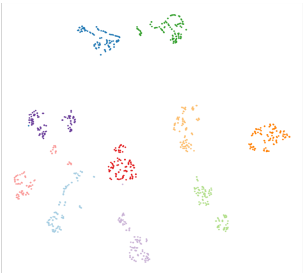


(h) *Fear* perplexity: 100 (10 subj.)

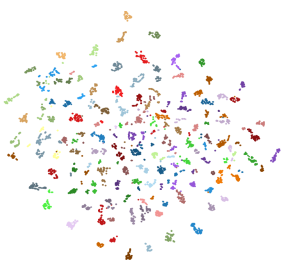
Fig. 25. t-SNE clustering of individual topological data for *Fear* emotion at different perplexities.



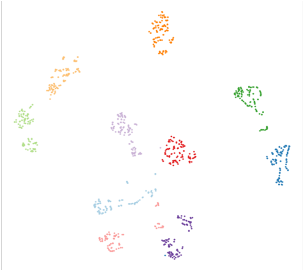
(a) *Happiness* perp.: 30 (full)



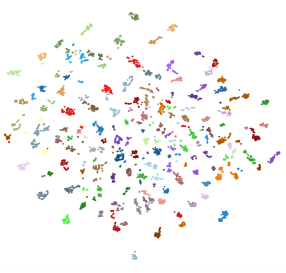
(b) *Happiness* perp.: 30 (10 subj.)



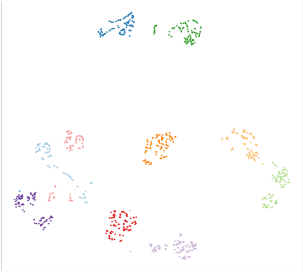
(c) *Happiness* perp.: 40 (full)



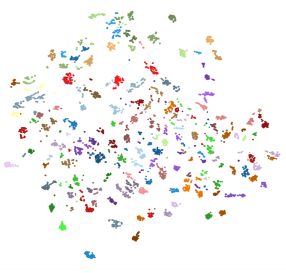
(d) *Happiness* perp.: 40 (10 subj.)



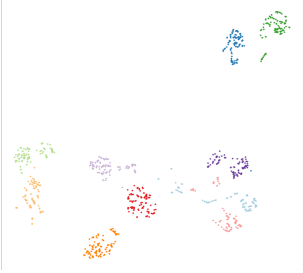
(e) *Happiness* perp.: 50 (full)



(f) *Happiness* perp.: 50 (10 subj.)

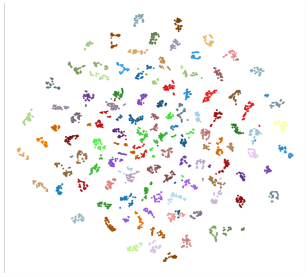


(g) *Happiness* perp.: 100 (full)

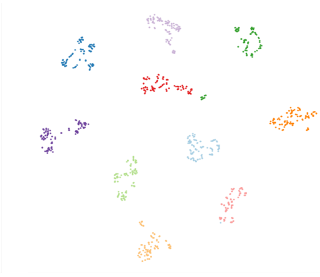


(h) *Happiness* perp.: 100 (10 subj.)

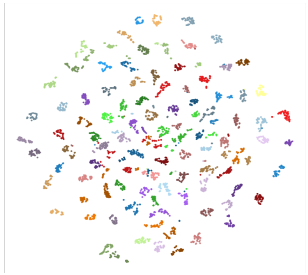
Fig. 26. t-SNE clustering of individual topological data for *Happiness* emotion at different perplexities.



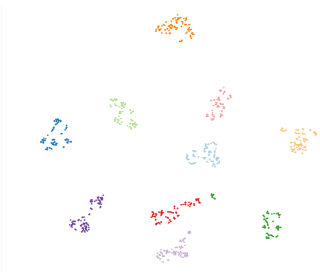
(a) *Sadness* perplexity: 30 (full)



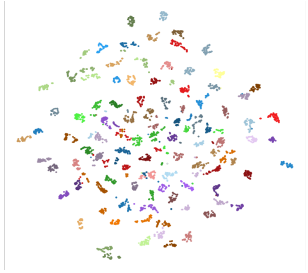
(b) *Sadness* perplexity: 30 (10 subj.)



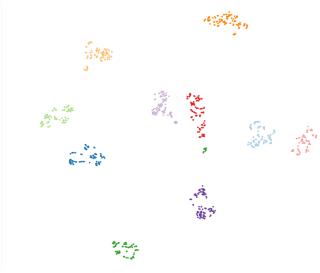
(c) *Sadness* perplexity: 40 (full)



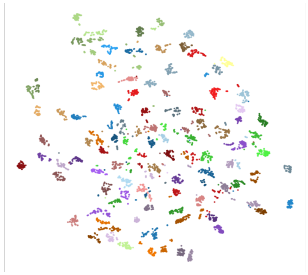
(d) *Sadness* perplexity: 40 (10 subj.)



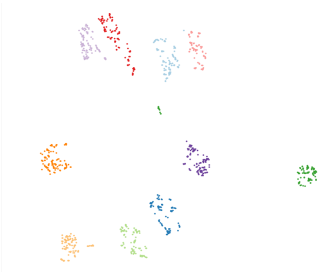
(e) *Sadness* perplexity: 50 (full)



(f) *Sadness* perplexity: 50 (10 subj.)

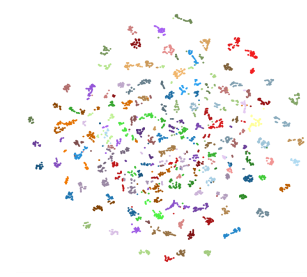


(g) *Sadness* perplexity: 100 (full)

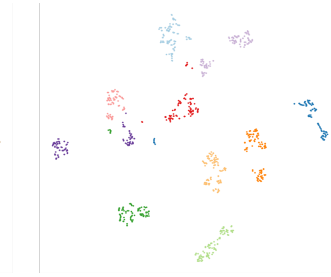


(h) *Sadness* perplexity: 100 (10 subj.)

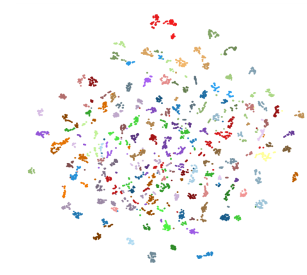
Fig. 27. t-SNE clustering of individual topological data for *Sadness* emotion at different perplexities.



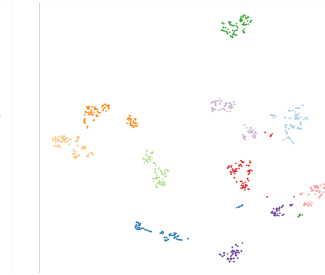
(a) *Surprise* perplexity: 30 (full)



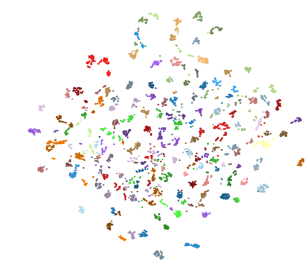
(b) *Surprise* perplexity: 30 (10 subj.)



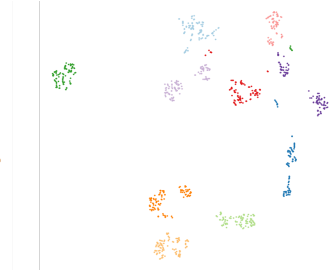
(c) *Surprise* perplexity: 40 (full)



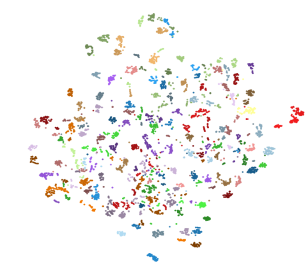
(d) *Surprise* perplexity: 40 (10 subj.)



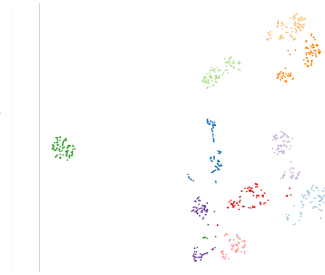
(e) *Surprise* perplexity: 50 (full)



(f) *Surprise* perplexity: 50 (10 subj.)

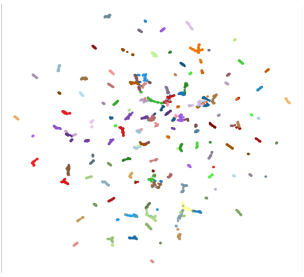


(g) *Surprise* perplexity: 100 (full)



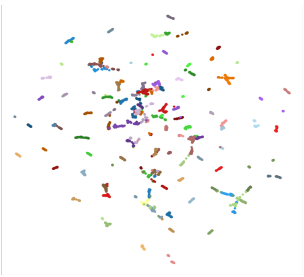
(h) *Surprise* perplexity: 100 (10 subj.)

Fig. 28. t-SNE clustering of individual topological data for *Surprise* emotion at different perplexities.



(a) *Anger* perplexity: 30 (full)

(b) *Anger* perplexity: 30 (10 subj.)



(c) *Anger* perplexity: 40 (full)

(d) *Anger* (10 subj.) Perplexity=40

(e) *Anger* perplexity: 50 (full)

(f) *Anger* perplexity: 50 (10 subj.)

(g) *Anger* perplexity: 100 (full)

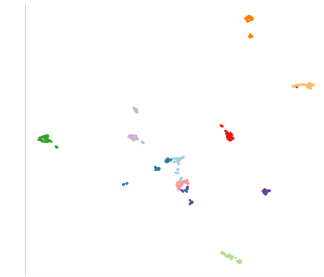
(h) *Anger* perplexity: 100 (10 subj.)

Fig. 29. UMAP clustering of individual topological data for *Anger* emotion at different perplexities.

(a) *Disgust* perplexity: 30 (full)

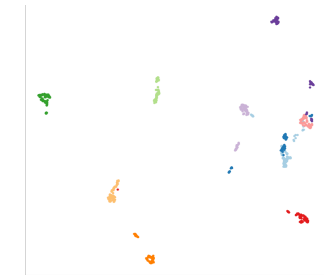
(b) *Disgust* perplexity: 30 (10 subj.)

(c) *Disgust* perplexity: 40 (full)

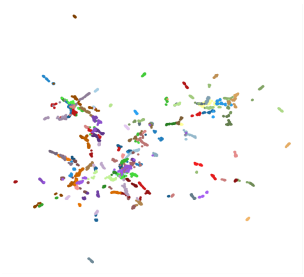


(d) *Disgust* perplexity: 40 (10 subj.)

(e) *Disgust* perplexity: 50 (full)



(f) *Disgust* perplexity: 50 (10 subj.)

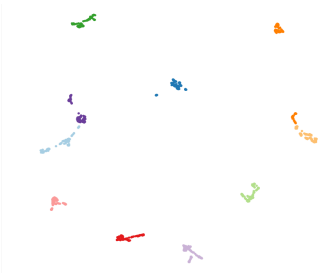


(g) *Disgust* perplexity: 100 (full)

(h) *Disgust* perplexity: 100 (10 subj.)

Fig. 30. UMAP clustering of individual topological data for *Disgust* emotion at different perplexities.

(a) *Fear* perplexity: 30 (full)



(b) *Fear* perplexity: 30 (10 subj.)

(c) *Fear* perplexity: 40 (full)

(d) *Fear* perplexity: 40 (10 subj.)

(e) *Fear* perplexity: 50 (full)

(f) *Fear* perplexity: 50 (10 subj.)

(g) *Fear* perplexity: 100 (full)

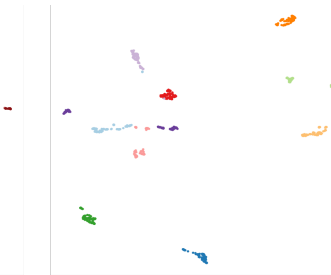
(h) *Fear* perplexity: 100 (10 subj.)

Fig. 31. UMAP clustering of individual topological data for *Fear* emotion at different perplexities.

(a) *Happiness* perp.: 30 (full)

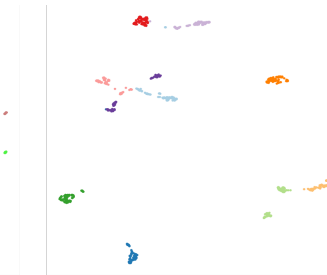
(b) *Happiness* perp.: 30 (10 subj.)

(c) *Happiness* perp.: 40 (full)

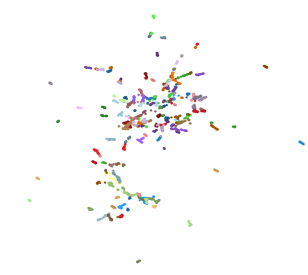


(d) *Happiness* perp.: 40 (10 subj.)

(e) *Happiness* perp.: 50 (full)



(f) *Happiness* perp.: 50 (10 subj.)

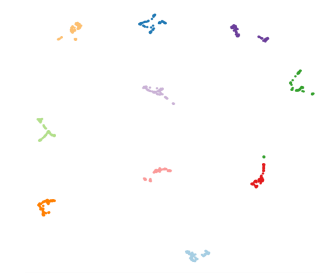


(g) *Happiness* perp.: 100 (full)

(h) *Happiness* perp.: 100 (10 subj.)

Fig. 32. UMAP clustering of individual topological data for *Happiness* emotion at different perplexities.

(a) *Sadness* perplexity: 30 (full)



(b) *Sadness* perplexity: 30 (10 subj.)

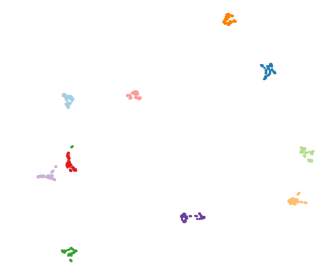
(c) *Sadness* perplexity: 40 (full)

(d) *Sadness* perplexity: 40 (10 subj.)

(e) *Sadness* perplexity: 50 (full)

(f) *Sadness* perplexity: 50 (10 subj.)

(g) *Sadness* perplexity: 100 (full)



(h) *Sadness* perplexity: 100 (10 subj.)

Fig. 33. UMAP clustering of individual topological data for *Sadness* emotion at different perplexities.

(a) *Surprise* perplexity: 30 (full)

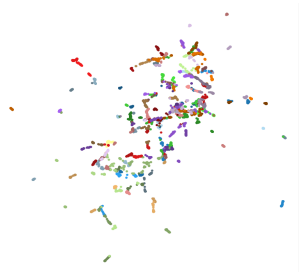
(b) *Surprise* perplexity: 30 (10 subj.)

(c) *Surprise* perplexity: 40 (full)

(d) *Surprise* perplexity: 40 (10 subj.)

(e) *Surprise* perplexity: 50 (full)

(f) *Surprise* perplexity: 50 (10 subj.)



(g) *Surprise* perplexity: 100 (full)

(h) *Surprise* perplexity: 100 (10 subj.)

Fig. 34. UMAP clustering of individual topological data for *Surprise* emotion at different perplexities.