

Article

# Research on Clothing Image Retrieval Combining Topology Features with Color Texture Features

Xu Zhang <sup>1,\*</sup>, Huadong Sun <sup>1</sup> and Jian Ma <sup>2</sup>

<sup>1</sup> School of Computer and Information Engineering, Harbin University of Commerce, Harbin 150028, China; sunhuadong1209@163.com

<sup>2</sup> Shenzhen Comen Medical Instruments Co., Ltd., Shenzhen 518000, China; 15140167452@163.com

\* Correspondence: zhangxu\_hrbcu@163.com

**Abstract:** Topological data analysis (TDA) is a method of feature extraction based on data topological structure. Image feature extraction using TDA has been shown to be superior to other feature extraction techniques in some problems, so it has recently received the attention of researchers. In this paper, clothing image retrieval based on topology features and color texture features is studied. The main work is as follows: (1) Based on the analysis of image data by persistent homology, the feature construction method of a topology feature histogram is proposed, which can represent the ruler of image local topological data, and make up for the shortcomings of traditional feature extraction methods. (2) The improvement of Wasserstein distance is presented, while the similarity measure method named topology feature histogram distance is proposed. (3) Because the single feature has some problems such as the incomplete description of image information and poor robustness, the clothing image retrieval is realized by combining the topology feature with the color texture feature. The experimental results show that the proposed algorithm, namely topology feature histogram + corresponding distance, can effectively reduce the computation time while ensuring the accuracy. Compared with the method using only color texture, the retrieval rate of top5 is improved by 14.9%. Compared with the method using cubic complex + Wasserstein distance, the retrieval rate of top5 is improved by 3.8%, while saving 3.93 s computation time.

**Keywords:** clothing image retrieval; topological data analysis; topology feature histogram; topology feature histogram distance

**MSC:** 68U10



**Citation:** Zhang, X.; Sun, H.; Ma, J. Research on Clothing Image Retrieval Combining Topology Features with Color Texture Features. *Mathematics* **2024**, *12*, 2363. <https://doi.org/10.3390/math12152363>

Academic Editors: Silvia Liberata Ullo and Li Zhang

Received: 28 June 2024

Revised: 25 July 2024

Accepted: 26 July 2024

Published: 29 July 2024



**Copyright:** © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

At present, image retrieval technology has been widely used in daily life and commercial areas, and clothing image retrieval is a typical application in commercial areas. In the early application stage of clothing retrieval, the main way of clothing retrieval is text-based retrieval. Later, with the deepening of computer vision-related technology research and the continuous maturity of means, there are more and more methods for image analysis and image processing, and content-based image retrieval has become the focus of a large number of researchers and enterprises. At the beginning of the research on content-based clothing retrieval, almost all of the research work was realized by traditional image retrieval technology, that is, the features used to describe images are extracted manually.

For example, in the task of image retrieval, Y.J. Choi [1] and the others divide the image into six different blocks in a specific way, extract color features from each block and transform them into feature vectors, and finally measure the similarity of each block and combine them into total distance as the basis of similarity between images. M. Gupta et al. [2] proposed a method that includes four attributes of clothing at the same time, namely texture, contour, feature, and color. Different attributes are given different weights, and finally, the four metric distances are added as the total similarity distance, so as to match

clothing. But, because of many problems such as the background, posture, and angle of clothing images, researchers have made many improvements to traditional features, which greatly improve their retrieval accuracy.

In 2017, Huang Dongyan et al. [3] proposed a joint segmentation algorithm for clothing images, which is based on HOG (Histogram of Oriented Gradients) features and E-SVM (Exemplar Support Vector Machine) classifier, and improves the accuracy of clothing image segmentation to a certain extent. Jin Jie et al. [4] fuses three features of color, texture, and shape as total features for retrieval, and the experiments show that the retrieval performance of the multi-feature fusion method is better than that of single feature algorithm.

Li Zongmin et al. [5] aimed to study the problem of the existing clothing retrieval framework retrieving the limbs of clothing in different scenes, due to the differences in the different styles of the same category of clothing and the information interference in different shooting backgrounds, and a new clothing segmentation method and clothing style recognition based on cross-domain dictionary learning is proposed, thus improving the accuracy of clothing style recognition in different scenes. Based on clothing elements, Wang Mengmeng et al. [6] used different methods such as roughness, orientation, and contrast to express the different clothing fabrics according to their texture, fabric weaving, and color characteristics. At the same time, they constructed various decision trees to distinguish the membership principles of different characteristics, which met the requirements of the classification decisions of clothing fabrics. Tao Binjiao et al. [7] improved the traditional algorithm of the weighted color histogram by combining the Grabcut image to block the foreground area of the image, and calculated the color histogram, respectively, and then gave different weights, which significantly improved its efficiency.

In 2018, Ge Jun et al. [8] also adopted the idea that a single feature cannot clearly and completely express the image content containing a lot of information. By accumulating the histogram features extracted from the image and weighting them by Hu invariant moments, this finally fuses the local binary pattern (LBP) and applies it to image retrieval. Chen Qian et al. [9] put forward the main color extraction algorithm to extract the color features of clothing images using the perception of human vision. Experiments show that this method has a better effect than other existing methods.

In 2019, Miao Zhiwen et al. [10] extracted the shape, color, and LBP-GLCM texture features of clothing images, and assigned different weights according to the characteristics of each feature. This method can integrate the advantages of each single feature, thus improving the retrieval accuracy. Hu Ying et al. [11] firstly locates the key points of clothing images through deep learning, and extracts the local features around the key points, which are used as the local area information of clothing. At the same time, the global features of clothing images are extracted through a convolutional neural network, which is used as the overall information of clothing, and the features also contain advanced semantic information on the images. Finally, combining the extracted local information with the global information as the basis for judging the image similarity, it can improve the accuracy of image retrieval to a certain extent.

In 2020, Qin Hui [12] mainly improved the SIFT feature matching method of images to solve the problem that the feature vectors of SIFT at high latitudes are not conducive to calculation, and the matching speed was accelerated by converting feature points into binary coding, thus greatly improving the image retrieval accuracy and matching efficiency. Wu Zhixin et al. [13] established a fusion method of the two features by extracting HSV feature histogram as a color feature and a gray level co-occurrence matrix as texture feature, and then normalized the extracted features. The results show that this method has certain advantages over a single feature.

Among many techniques in the field of computer vision, with the great achievements of deep learning, more and more scholars began to deeply study the application of this technology to clothing image retrieval. Kiapour [14] et al. used a convolutional neural network to process the image and extract the depth features of the image as the measurement basis. Different parameters can effectively reduce the influence of unimportant information in the

background area of the image and improve the ability of image retrieval in different scenes. Chen [15] and others put forward a depth neural network model RCNN [16] to realize clothing target detection in order to accurately locate clothing information in complex images. By accurately locating clothing, more accurate feature extraction can be obtained, and the influence of background area can be reduced, thus reducing redundant information. In addition, Liu [17] trains clothing images from different angles and extracts features for retrieval by designing a depth neural network model so as to solve the problem of clothing similarity from different angles. Wang [18] and Verma [19] and others adjust the parameters of the network model by adding an attention mechanism to the network model to extract clothing image features better. Garcia [20] established a dynamic clothing retrieval system, which intercepts clothing images from a video and retrieves them.

Topology data analysis (TDA) is a mathematical method, which makes use of topology ideas to mine the structure of datasets by analyzing the intrinsic topology characteristics of data. TDA provides a general framework for data analysis. Its advantage is that it can extract information from a large number of high-dimensional data and has stable anti-noise performance.

Since 2004, Carlsson and other scholars [21,22] began to study the application and practice of the continuous coherence method based on their in-depth understanding of topology and geometry. For a three-dimensional model, mainly through the construction of three-dimensional point cloud data complex, the corresponding topology feature is calculated to describe the skeleton of the three-dimensional model.

For two-dimensional images, we can imitate the idea of a three-dimensional model to extract feature points and build a two-dimensional skeleton structure for topology feature retrieval. For example, Zhang Jingliang et al. [23] applied the method of persistent homology to represent the image space by transforming the image space and constructing a simple complex structure under different parameters. Finally, the topology feature of these complexes are calculated using persistent homology, and then the topology information is represented by barcode graph. Experiments show that, in this continuous approximation process, not only the topology invariant features of the image can be obtained, but also some geometric features related to the topology structure of the image space can be obtained. Through the experiments of simple geometric images and natural images, it is concluded that the similarity between images can be analyzed by the topology invariant features of images, and different images can be distinguished by the geometric features of images. In particular, by comparing the similarity and difference of topology structure and geometric structure between images taken from different angles, it can be concluded that the similarity between the original image and the deformed image can be reflected to a certain extent by topology invariance, in which the image deformation includes deflection, rotation, reversal, scaling, and occlusion due to shooting angle and distance.

From the research results of the above literature, the current image features include two aspects: one being the underlying visual features such as color, texture, and shape, and the other being the depth features extracted by neural network, which have their own shortcomings. In addition, the research on topology is rarely used in clothing image retrieval. Therefore, this paper comprehensively considers the characteristics of clothing images, and improves the performance of clothing image retrieval through the fusion of topology, color texture, and depth features.

## 2. Topology Feature Extraction

In the field of topology, there is an important property called symmetry. Generally speaking, “symmetry” means that the result of an object after transformation remains the same as before. Mathematicians call symmetry in topology “symmetries persistent homologies”, which is the key to solving problems such as network analysis, data mining, and understanding brain neural network connection graphs [24].

In most applications, data are processed in the form of a point cloud, a large but limited set of points sampled from some underlying geometric objects in Euclidean space.

Persistent homology is a multi-scale method in a sense [24]. Its utility is illustrated by considering the following simple geometric cases. Suppose we first have a set of  $\epsilon$  distances between two points. Because the invariants to be calculated are highly dependent on the connectivity of the space being described, it is necessary to specify how small the  $\epsilon$  of the two points should be in order to become a connected component. To avoid making such a choice, this article uses the concept of persistence.

Given a topology space  $\mathbb{X}$  and a filter function  $f : \mathbb{X} \rightarrow \mathbb{R}$ , persistent homology studies the homology changes of subsets, where  $\mathbb{X}^t = f^{-1}(-\infty, t]$ . The algorithm captures the birth and death time of the homology class from the growth of subsets  $\mathbb{X}^{-\infty}$  to the time  $\mathbb{X}^{+\infty}$ . More persistent homological classes reveal information about the global structure of space  $\mathbb{X}$ , as described by functions  $f$ .

An important proof of using persistence is the stability theorem. Cohen-Steiner et al. [25] proved that, for the sum of any two filter functions  $f$  and  $g$ , their persistence difference is always equal to the norm upper of  $L^\infty$  which is bound to their difference:

$$\|f - g\|_\infty = \max_{x \in X} |f(x) - g(x)| \quad (1)$$

This theorem guarantees that persistence can be used as a feature.

Extracting the topology feature of images primarily causes images to be analyzed as a topology space, while similar images are considered as two homologous spaces. For two topology spaces  $X$  and  $Y$ , their persistent homology can be described as follows. If there is a continuous mapping  $H : X \times [0, 1] \rightarrow Y$  that makes the sum  $H(x, 0) = f(x)$  and  $H(x, 1) = g(x)$ , then the two continuous mappings  $f, g : X \rightarrow Y$  are considered homologous. For a mapping  $f : X \rightarrow Y$ , if there is a mapping  $g : Y \rightarrow X$ , then it is called an identical homology mapping of  $f$  and  $g$ ,  $g$  is an inverse homology of  $f$ . Two spaces  $X, Y$  with homologous equivalent mappings  $f : X \rightarrow Y$  are called the homologous equivalent. Therefore, it can be used as an important basis for image retrieval using topology.

### 2.1. Witness Complex

Witness complex can be regarded as an approximation of restricted Delaunay triangulation, but this construction avoids the dimension curse related to Delaunay calculation; Specifically, the external extension of the dataset has little influence on the complexity of the algorithm.

Firstly, every pixel of a color RGB image is transformed into a five-dimensional point in space, and its dimensions are  $x, y$  coordinates and  $r, g$ , and  $b$  components of pixels, respectively. In this way, the analysis of an image is transformed into the analysis of a five-dimensional point cloud data, which acts as the point cluster  $Z$  for constructing the witness complex flow.

There are two main methods to select landmark set  $L$ . One is the random selection method, and the other is sequential maximum–minimum method. The algorithm steps [26] are as follows:

- (1) Randomly select a data point  $L_1$  in the point cluster  $Z$ .
- (2) Inductively, if a set of landmark points  $i - 1$  has been selected, the selection of the first landmark point  $i$  needs to satisfy the maximization function  $z \mapsto d(z, L_{i-1})$ , where  $z \mapsto d(z, L_{i-1})$  is the distance from point  $z$  to point set  $(L_1, L_2, \dots, L_{i-1})$ .
- (3) Repeat this operation until the required number of landmark points are selected.

The method can tend to cover datasets and disperse each other. Figure 1 shows the comparison between the two methods, where the number of landmark points are  $n = N/100$ .

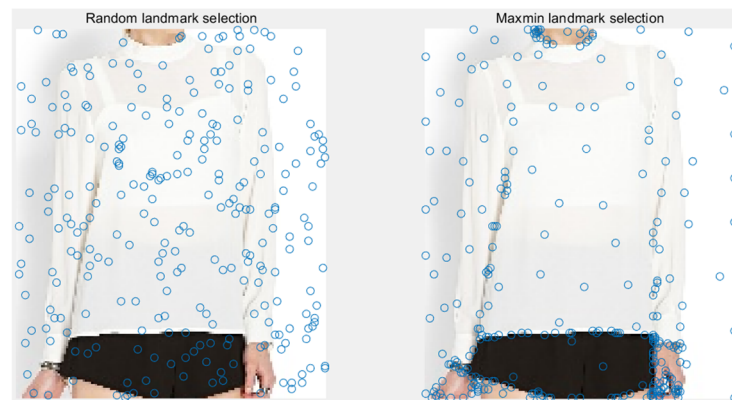


Figure 1. Comparison of landmark selection methods.

After selecting the landmark points, the distance matrix  $D$  of a dimension  $N \times n$  is obtained by calculation. For each non-negative integer  $v$ , a nested family of simple complexes  $W(D, R, v)$  is constructed, in which  $R \in [0, \infty]$ . The values are 0, 1, and 2. If the vertex set of  $W(D, R, v)$  is  $\{1, 2, \dots, n\}$ , then the different value  $v$  is defined as follows:

If  $v = 0$ , then for  $i = 1, 2, \dots, N$  defines  $m_i = 0$ .

If  $v > 0$ , then for  $i = 1, 2, \dots, N$  is defined  $m_i$  as the  $v$ -th minimum of the  $i$ -th column of  $D$ . Edge  $\sigma = [a, b]$  belongs  $W(D, R, v)$ , then there is a witness  $i \in (1, 2, \dots, N)$  satisfaction:

$$\max(D(a, i), D(b, i)) \leq R + m_i \tag{2}$$

$p$  - Simplex of  $\sigma = [a_0, a_1, \dots, a_p]$  belongs to  $W(D, R, v)$ , if all its edges belong to  $W(D, R, v)$ ; there is a witness satisfaction:

$$\max(D(a_0, i), D(a_1, i), \dots, D(a_p, i)) \leq R + m_i \tag{3}$$

The first task of data preprocessing is to generate a simple list (up to the dimension of  $p$ -dimensional homology  $p + 1$ ). For each simplex  $\sigma$ , it is necessary to identify its face and determine its appearance time, that is,  $R = R_\sigma$ , and the minimum value of  $\sigma \in [D, R]$ . By definition,  $R_\sigma = \max\{R_\tau\}$ , in which  $\tau$  is an edge of  $\sigma$ . Its steps [26] are as follows:

(1) A matrix  $E$  which is  $n \times n$  calculated with non-diagonal terms  $E(i, j) = R_{[ij]}$ , which records the occurrence time of each edge.

(2) Generate a simple list that appears in time.

(3) Calculate the appearance time of each simplex as the maximum value of the appearance time of its edges.

Step 1 can be expressed algebraically as a “minimum” matrix product:  $E = D \odot D^*$ . Here,  $\odot$  is the action:

$$[A \odot B](i, j) = \min_{k \in \{1, \dots, n\}} \max(A(i, k), B(k, j)) \tag{4}$$

For Step 2, the list of temporal edges which are born from time  $r$  can be used to generalize high-dimensional units: for example, simplex  $[a_0 \dots a_p]$  occurs in time  $r$  if th three simplexes  $[a_1 \dots a_p]$ ,  $[a_0 \dots a_{p-1}]$ , and  $[a_0 a_p]$  all occur at time  $r$ . Step 3 can be performed simultaneously with Step 2. Then, the persistent homologous group on the interval  $R \in [0, r]$  can be calculated using the algorithm in paper [27].

When  $v = 0$ , the complex structure  $W(D, R, v)$  will be very similar to the complex *Rips*, specifically meeting the following conditions:

$$W(D, R, 0) \subseteq Rips(L, 2R) \subseteq W(D, 2R, 0) \tag{5}$$

Their homology group structures are similar, so the calculated persistence graphs are similar.

When  $v = 1$ , in some cases, it is the most suitable parameter to witness the complex, and it can be interpreted as a series of coverages of space  $X$  by Voronoi [28]—like regions around each landmark point, which gradually overlap as  $R$  approaches infinity.

When  $v = 2$ , although the persistence of the complex was sometimes not so good, the following conditions were met when  $R = 0$ :

$$W(D, 0, 2) = W(D) \tag{6}$$

In practice, the complex can give a cleaner persistent interval diagram when  $v = 2$ , that is, there is less “noise”, which is proved in paper [29]. The calculation effect is shown in Figure 2. In this paper, one-dimensional feature information is mainly used for calculation. Although the complexity of the witness complex has been greatly reduced by selecting landmark points, it mainly aims at the calculation of spatial point cloud data, and there is a more suitable processing method for images, that is, cube complex.

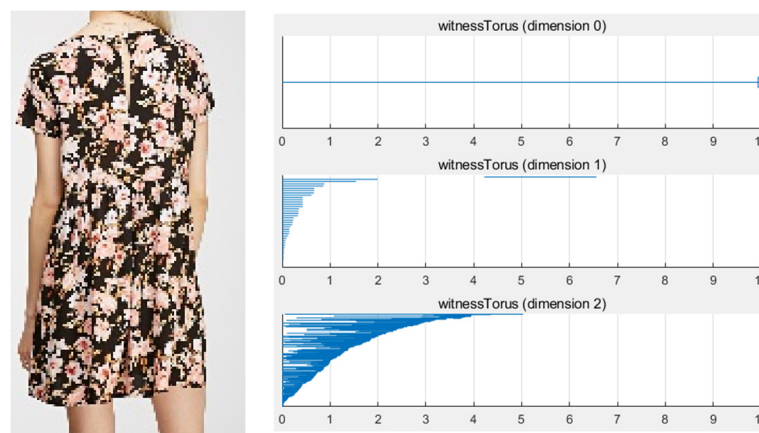


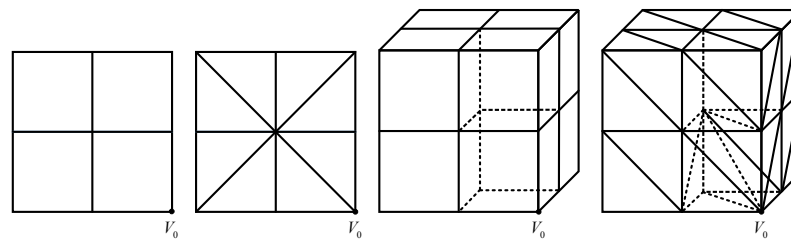
Figure 2. Witness complex calculation results.

### 2.2. Cube Complex

Persistence is underused in applications, in large part because of its high computational cost. The standard algorithm [30] requires three runs, which may be prohibited even for small-size data (for example,  $64 \times 64 \times 64$ ). In addition to the high time complexity, there are two problems: (1) the currently available memory consumption is very large, even for small data, so it is forbidden for commodity calculation; (2) several applications focus on higher dimensional data, such as 4D, 5D, or higher. There are few implementations of general dimensions, and the existing implementations cannot expand well with the increase in dimensions, thus introducing larger computing time and low memory efficiency. Hubert Wagner et al. [31] proposed an effective computing framework, which focuses on uniform or periodic sampling data common in visualization and data analysis, that is, image data include pixels (2D images), voxels (3D scanning, simulation), or their high-dimensional analogies, such as 4D time-varying data. In this work, the name “cube” is used to represent these data. This method uses a specific triangulation, namely Freudenthal triangulation [32]. This triangulation can be easily extended to general dimensions.

Next, the cube complex is introduced in detail. Firstly, a basic interval is defined as a unit interval  $[k, k+1]$  or a degenerate interval  $[k, k]$ . For  $d$ -dimensional space, a cube is the product of  $d$  basic intervals  $I: \prod_{i=1}^d I_i$ , and the number of non-degenerate intervals in the product is the dimension of the cube. Cube 0, Cube 1, Cube 2, and Cube 3 are vertices, edges, squares, and three-dimensional cubes (voxels), respectively. Given two cubes:  $b \subseteq R^d$ ,  $a$  is a face of  $b$  only if one  $a \subseteq b$ . A cube complex of dimension  $d$  is a set of cubes with the maximum dimension  $d$ . Similarly to the definition of a simple complex, it must be closed under the surface and intersection point. In this article, cube complexes will be used to describe data. In Figure 3, two-dimensional and three-dimensional cube complexes are shown, depicting a two-dimensional image with a size of  $3 \times 3$  and a three-dimensional

image with a size of  $3 \times 3 \times 3$ . And, the corresponding simple complex representation is given.



**Figure 3.** Simple complex of two-dimensional and three-dimensional cubes.

After constructing the simple complex of the data, it is necessary to calculate the boundary matrix of the structure. The information of the boundary matrix is hidden in the simple complex structure. The construction of the cube complex abandons the standard construction method, including triangulation space. There are two advantages. First of all, the size of the complex is significantly reduced, and the memory and runtime efficiency are significantly improved, especially for high-dimensional data. Second, cubic complexes allow the use of more compact data structures. For  $d$ -dimensional data, the correlation ratio is expressed as  $\rho_d = S_d/C_d$ , where  $C_d$  and  $S_d$  are the size of the cube complex and its triangulation, respectively. The exact formula  $\rho_d$  given is very important because the triangulation of the minimum cardinal cube is an open problem [33]. Here, by triangulating all the cubes of the cube complex in each dimension, given the lower bound  $\rho_d$  of  $d \leq 7$ . When triangulating a  $d$ -dimensional cube, only the  $d$ -simplex and their  $(d - 1)$ -dimensional intersections are calculated. Finally, considering that some simplex will be the common surface of many high-dimensional simplex, we can obtain the number  $\tau_d$  of the  $d$ -dimensional simplex in the triangulation of the  $d$ -dimensional cube.

$$\rho_d \geq \frac{\sum_{i=0}^d \binom{d}{i} \tau_i + \sum_{i=0}^{d-1} \binom{d}{i+1} (\tau_{i+1} - 1)}{2^d} \tag{7}$$

The specific operation of the cube complex for image data processing is as follows: firstly, the input of the cube complex requires a square image. Its input requirements are different from that of witness complex flow. It does not input point cloud data, but directly inputs pixel matrix. Take a 2D image with  $5 \times 5$  pixels as an example, as shown in Figure 4. Because of the regular structure, the relationships between cells can be immediately read from their coordinates, and the necessary information for each cell (i.e., order in filtering, function values) can be stored in a  $9 \times 9$  array. You can then immediately obtain the dimension of any cell (whether it is a vertex, an edge, or a square), as well as its face and coplanar, that is, the cell where it is a face. This is achieved by checking that the coordinates are modulus 2, because cubes can be defined as the product of intervals, and even coordinates correspond to the degenerate interval of a cube. This structure is called cube graph, and the main advantage of this data structure is improved memory efficiency. Boundary relationships are implicitly encoded in the coordinates of the cells, which themselves are implicit, and each cube persistence can be accessed randomly. The above properties are applicable to any dimension because of the inductive structure of the complex of cubic complexes and the fact that the cube is the product of intervals.

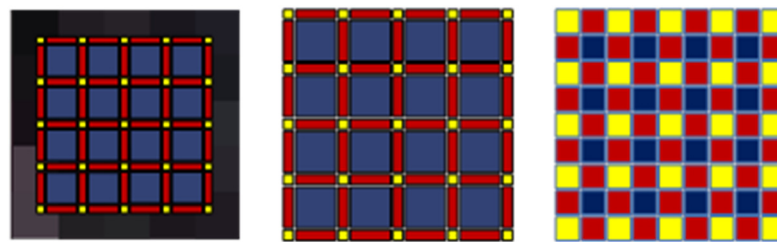


Figure 4. Cube complex of 2D gray image.

Each vertex (yellow) corresponds to one pixel: edge (blue); cube (red); and the cube complex itself. All the filter build information is encoded in a  $9 \times 9$  array, with each element corresponding to a cell. Next consider the input data of dimension  $d$  and size  $w^d$ , where  $w$  is the number of vertices in each dimension. You can append information to a two-dimensional array with elements  $(2w - 1)^d$ . This array consists of overlapping copies of arrays of size  $3^d$ . Therefore, for two-dimensional image data,  $w$  is the size of a square image. The main advantages of this data structure are the improved memory efficiency, the fact that the boundary relationship is implicitly encoded in the coordinates of cells, and that each cell can be accessed randomly and its boundary can be located quickly.

Now, an effective algorithm is used to calculate the filtering of cubic complex caused by a given function. This article uses the cube map data structure to store additional information (function values, filtering order) for each cell. The result of the algorithm is a sorted boundary matrix, which is used as the input of simplification step. Because there is only one non-zero element in each column of cubic data boundary matrix, sparse matrix representation is usually used. The intuition behind the algorithm is that, when all vertices are iterated in descending order, the remainder of the vertices that are not added to the filter is known and can be added to the filter. At the same time, you cannot create a boundary matrix in the same step, because the indexes of adjacent cells may not have been calculated. Figure 5 illustrates the data structure of the algorithm, in which the value of filter function  $f$  is assigned to vertices and extended to all cubes. Cells are assigned indexes during filtration. These indexes are separate for each dimension. Vertices are marked as V, edges as E, and squares as S.

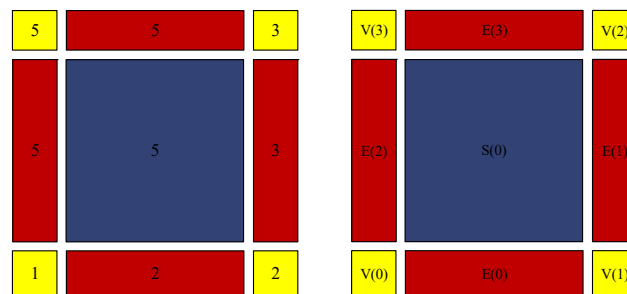


Figure 5. Cube complex of 2D gray image (cube map data structure).

Edelsbrunner et al. [27] designed an algorithm to calculate persistent homology, which is calculated in cube time (complex number size). For images, functions are defined on all pixels or voxels. First, these values are interpreted as the values of vertices of a complex number. Then, the filter of complex numbers is calculated, and the sorted boundary matrix is generated. This matrix is the input of the reduction algorithm. Filtering can be described as adding incremental cells one by one. To achieve this, a filter build algorithm extends the function to all complex units by assigning the maximum value of each unit of its vertices. Then, all cells are sorted in ascending order according to the function value, so that each cell is added to the filter after all faces. Such cell sequences are called low star filtering. After calculating the sort of cells, you can generate a sort boundary matrix.



In the reduction step, the algorithm reduces the columns of the sorted boundary matrix from left to right. Each new column is reduced by adding the column that has been reduced until its lowest non-zero entry is as high as possible. The simplified matrix encodes all persistent coherence information. Because the cube complex requires the input data to be square, and the image size of the dataset is  $256 \times 256$ , it can be directly input for calculation, and the calculation results are as shown in Figure 6.

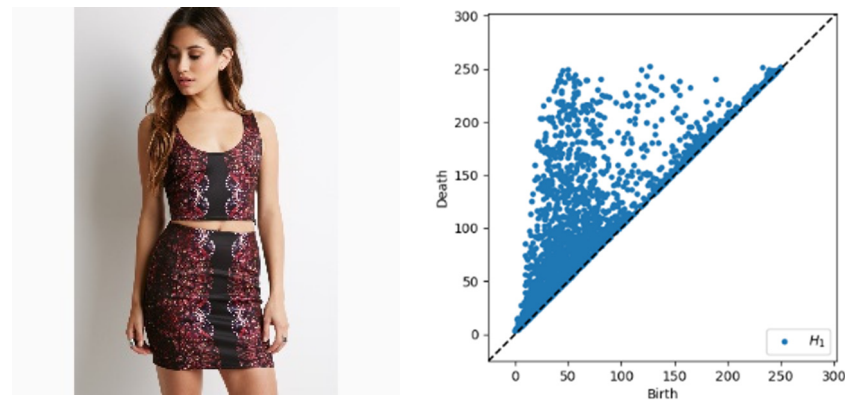


Figure 6. Calculation results of cube complex.

### 2.3. Topology Feature Histogram

In this paper, we construct witness complex and cube complex to extract the topology features of images, which are mainly biased towards the extraction of overall features of images. On this basis, we propose a new local feature extraction method, that is, using Rips complex to construct topology histogram to extract image features.

Both witness complex and cube complex calculate the whole image, and extract the whole features. In this paper, the sliding window method is used, only a limited number of pixels in the window are calculated each time, and finally, all the output features of the window are counted. The purpose of this method is to extract the repetitive and regular local topology feature in the two-dimensional image plane as the basis for retrieval.

Topology feature histogram uses Rips complex, which is the simplest simple complex structure construction method, and it has good performance when there are a few data points. However, due to the particularity of digital images, each image can be regarded as a dense point cloud. And, with the change of image resolution, the number of point clouds increases rapidly, and in the region with little color change, the distance between adjacent data points is too small, which leads to the complexity of constructing Rips complex, which not only requires a lot of calculation, but also cannot extract effective features. Among them, the best window size is  $3 \times 3$ . Because of the calculation method of the Rips complex, an excessively large window will lead to a long time consumption of feature extraction calculation, while a small window will lead to insufficient feature types, resulting in poor calculation accuracy. Firstly, the whole image is scanned sequentially using the window, and the point cloud data containing nine data points is extracted. Then, the Rips complex is constructed by the extracted data, the homology information is calculated and written into the persistence graph, and the topology eigenvalue of the window is calculated, as shown in Figure 7.

After calculating the persistence graph for each  $3 \times 3$  window, it is necessary to calculate an eigenvalue according to the persistence graph. If the central pixel coordinate of each window is  $(x_0, y_0)$ , the eigenvalue of the window is

$$Rips_{(x_0, y_0)} = \sum_{i=0} \frac{(dx_i + dy_i)}{m} \tag{8}$$

where  $dx_i$  and  $dy_i$  are the value of x and y of the i-th coordinate in the one-dimensional information in the persistence graph calculated using the complex, that is, the birth time and death time of the hole, where m is equal to 4, which means that, when taking nine pixels for calculation, there will be four holes connected by four points in most cases, so in order to avoid having an excessive value for the final calculated eigenvalue, scaling processing is carried out. Finally, the topology histogram is obtained by counting all the window eigenvalues of the whole image, in which the abscissa of the histogram represents the eigenvalues and the ordinate represents the number of eigenvalues in the image. The calculation results are shown in Figure 8.

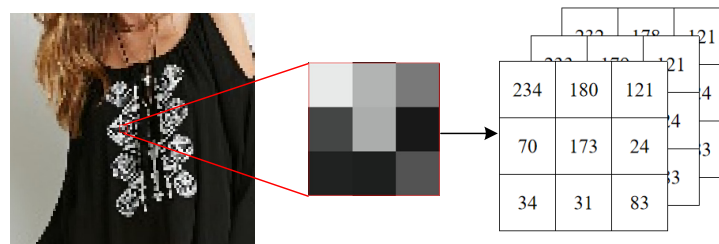


Figure 7. Image extraction  $3 \times 3$  window point cloud data.

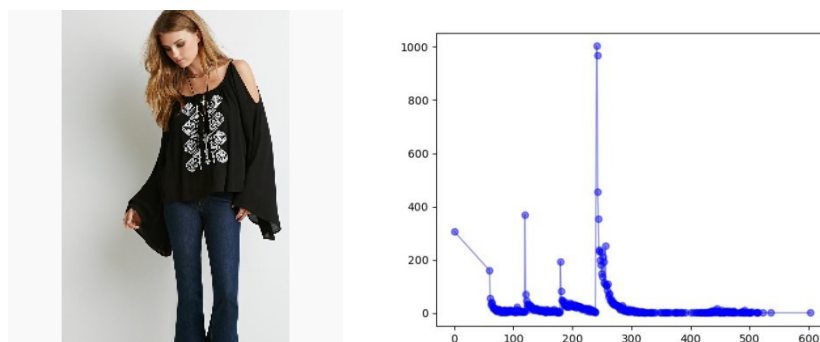


Figure 8. Calculation results of topology feature histogram.

### 3. Distance Measure of Topology Feature

The distance measure of the topology feature are mainly divided into the measure of persistent graph and the measure of the vector graph, in which the vector graph is transformed by mapping a persistent graph. Euclidean distance can be used as the measurement method for a vector graph. However, there are two main measure methods for two-dimensional persistence graphs with many points and discrete points: one is the bottleneck distance algorithm [34], and the other is the Wasserstein distance algorithm [35]. Finally, focusing on the topology feature histogram proposed in this paper, we propose a topology histogram distance algorithm, which is improved by the Wasserstein distance.

#### 3.1. Bottleneck Distance

The bottleneck distance between two persistence graphs measures the cost between their corresponding relationship points. Simply put, two persistence graphs are transformed into the same scale coordinate system, and then every point in one persistence graph finds the nearest match in the other persistence graph, and calculates the distance between the two points. If the point is closer to the diagonal or  $y$  axis, no point matching is required. The matching effect is shown in Figure 9.

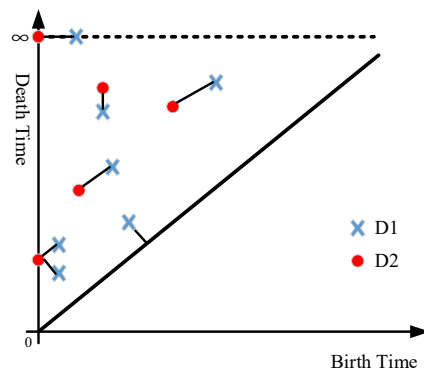


Figure 9. Matching results of the bottleneck distance.

If  $D$  and  $D'$  are two persistence graphs, their bottleneck distance is

$$d_B(D, D') = \min_{\sigma} \max_{p \in D} d(p, \sigma(p)) \tag{9}$$

where  $\sigma$  differs in all bijections between  $D$  and  $D'$  and satisfies

$$d((u, v), (u', v')) = \min \left\{ \max\{|u - u'|, |v - v'|\}, \max\left\{\frac{v - u}{2}, \frac{v' - u'}{2}\right\} \right\} \tag{10}$$

where  $(u, v)$  and  $(u', v')$  satisfy  $(u, v) \in \mathbb{R}^2, u \leq v$ .

### 3.2. Wasserstein Distance

Wasserstein distance was originally used to calculate the distance between two histograms, that is, the distance between two discrete distributions. Later, it can also be used to calculate the distance between two probability distributions, that is, the distance of continuous distribution. It can be popularly understood as the minimum transportation cost required to move objects in several positions to several other positions. Its definition formula is as follows:

$$W_p(\mu, \nu) = \inf_{\begin{matrix} X \sim \mu \\ Y \sim \nu \end{matrix}} (E\|X - Y\|^p)^{\frac{1}{p}}, p \geq 1 \tag{11}$$

where  $X$  and  $Y$  are  $d$ -dimensional space sets satisfying the sum probability distribution, respectively, which  $p$  represents the norm of  $L_p$  and  $\inf$  represents lower bound. Compared with other distance measures, the Wasserstein distance has the following advantages: (1) it can calculate the distance between discrete distribution and continuous distribution; (2) it cannot only calculate the distance between two distributions, but also gives a clear transition scheme from one distribution to another; and (3) it does not lose its original geometric characteristics when transferring from one distribution to another. This method is more accurate than the bottleneck distance measurement method, but the calculation cost is also higher.

### 3.3. Topology Feature Histogram Distance

For the measurement method of the topology feature histogram distance, the Wasserstein distance can be directly used for the calculation. However, due to the large amount of histogram data, the data amount of two histograms is usually different, according to the calculation rules of Wasserstein distance, and it is necessary to allocate weights in advance to ensure that the total weights of the two histograms are equal, which makes the calculation complexity rise and the calculation results are of too large an order of magnitude. Therefore, in order to solve the above problems, this paper proposes an improved measurement algorithm.

First, the two histograms to be calculated are simplified, that is, with the abscissa  $x_n$  for each point of each histogram, the simplified formula is

$$H_{1,2}(x_n) = H_{1,2}(x_n) - \min(H_1(x_n), H_2(x_n)) \quad (12)$$

where  $x_n$  represents the same eigenvalue in the two histograms and  $H_i(x_n)$  represents the number of features at the eigenvalue  $x_n$ . The simplification is based on simplifying the same eigenvalue part of two histograms, and only different eigenvalues remain after simplification, so as to facilitate subsequent calculation.

After simplification, the calculation time of histogram distance can be greatly shortened. The traditional Wasserstein distance needs to assign different weights to both calculation parties to ensure the same total weights. However, due to the large amount of data in the histogram, the computational magnitude will increase. Therefore, this paper uses the improved Wasserstein distance to calculate the distance of the simplified histogram. Its basic principle is that the histogram with more data is divided into two parts  $H_{1a}$  and  $H_{1b}$ , in which the amount of data in the first part is equal to that in the histogram  $H_2$ , so that the weights of the two parts are the same without allocation, and the distance of  $L_1$  is calculated by Wasserstein. The distance of each data pair  $H_2$  in the histogram  $H_1$  of the other part is the distance from the abscissa of the data to the mean value of the abscissa of  $H_2$ , and the distance  $L_2$  is obtained by sum. The calculation formula is as follows:

$$L_2 = \sum_{i=0}^m \text{abs}(x_i - \text{mean}(H_2)) \times H_{1b}(x_i) \quad (13)$$

This method is equivalent to distributing more data of  $H_1$  than  $H_2$  to  $H_2$  evenly, and finally taking the sum of  $L_1$  and  $L_2$  as the final distance. If the data amount of  $H_1$  and  $H_2$  is the same, it can be calculated directly.

It can be seen from the above analysis that the calculation of bottleneck distance is relatively simple, while the Wasserstein distance is relatively accurate, but the calculation cost is relatively high. The calculation speed of a vector graph is fast, but the corresponding accuracy is low. According to the topology feature histogram proposed in this paper, the improved Wasserstein distance is used to calculate it. The experimental results show that the improved Wasserstein distance has achieved a good retrieval performance.

#### 4. Clothing Image Retrieval

The typical CBIR (content-based image retrieval) system firstly needs to select the appropriate feature extraction methods according to different fields, application scenarios, and systems to analyze and process all the images in the image database, and extract the different forms of image features in various ways. Then, the features are quantified and saved in the database, and finally, an index mechanism is established to facilitate users' query. In the field of clothing image retrieval, there are many ways to measure the similarity level of two images, one of which is Metric learning [36]. The accuracy of image retrieval has a great relationship with the selected similarity measurement method. In order to make image retrieval more accurate and efficient, it is necessary to select a suitable distance calculation method between images. The measurement function [37] mainly includes the Manhattan distance, Euclidean distance, histogram intersection distance, and cosine similarity. In the stage of image retrieval, different kinds and different forms of features will be extracted from the different kinds of images. In actual operation, it is necessary to select appropriate similarity calculation methods according to the characteristics of different images, so that the required images can be retrieved more accurately and efficiently.

In the clothing image retrieval technology in this paper, feature extraction is mainly divided into three types. Among them, image color texture features are mainly aimed at the underlying features of the image. Image depth features are extracted from the model trained by neural network. The topology feature is to construct a complex filter flow through a variety of methods, and finally calculate the persistence graph, in which a

series of two-dimensional coordinates are taken as the topology feature of the image. After multi-feature fusion, it cannot only solve the shortcoming of the insufficient expression information of single feature, but also extract more abundant image information, which improves the retrieval efficiency.

#### 4.1. Color Histogram

Color is one of the most widely used underlying visual features. It is one of the most basic visual bases for human beings to distinguish different things, so it is also one of the most representative features to describe image content, and it is widely used in the field of image retrieval.

In the CBIR system, the color histogram is easy to extract, so it is also one of the most widely used color features. Firstly, the color space of the image is quantized, and then the color distribution is counted to form a color histogram. RGB, LUV, HSV, YCrCb are commonly used color space models. Among them, RGB color space is the most common and simplest in the field of computer vision. After choosing the appropriate color space, we need to quantify the color space properly, and finally calculate the corresponding color histogram features according to the statistical information of color distribution.

When extracting color histogram features, in order to transform color features into vector form, a certain quantization strategy is needed. The main idea is that, the richer the color information, the smaller the quantization interval, while the larger the quantization interval, the lesser the color information. At the same time, the selection of quantization granularity should also be appropriate. An excessively fine quantization granularity will lead to excessively high feature dimension, which will lead to too low computational efficiency when the image database is too large, while excessively coarse quantization granularity will lead to obviously different colors being quantized into the same interval, which will lead to wrong feature information and affect retrieval accuracy. Therefore, it is very important to select appropriate quantization granularity.

#### 4.2. SURF

SURF (Speeded-Up Robust Feature) is a fast and stable local invariant feature extraction algorithm. Similar to the SIFT feature calculation method, the salient feature points in the linear scale space of the image need to be selected as key points at first, and then the gradient direction distribution of pixels around the key points of the image is counted and taken as local features in turn. In contrast to the SIFT feature, the Gaussian filter is replaced by an integral image and frame filter, and an approximate differential operator is calculated in different scale spaces, so the calculation speed is faster than the SIFT feature. The SURF algorithm consists of the five following steps:

- (1) Calculate the integral image of the input image;
- (2) Using frame filters to calculate discrete operators on the multiple scales of the original image;
- (3) The image is divided into different scale spaces, the determinant of Hessian matrix is calculated, respectively, and the maximum value is selected to accurately locate the detected key points.
- (4) The main direction of each key point is calculated by window sliding method.
- (5) Using the parameters of the scale space in which the feature points are located and their main azimuth angles, the eigenvalues of their nearest neighbors are obtained, and a new feature descriptor is generated.

On this basis, a rectangular window area of  $20s \times 20s$  is constructed with each key point as the center, and it is divided into  $25 \ 4 \times 4$  sub-areas, and given Gaussian weight coefficients. By summing the transverse wavelet response  $d_x$ , longitudinal wavelet response  $d_y$  and the absolute value of wavelet response of the key points, the four-dimensional vector  $(\sum d_x, \sum d_y, \sum |d_x|, \sum |d_y|)$  of the region at each key point is obtained, and the final  $(4 \times 4) \times 4 = 64$ -dimensional vector  $V_s = (i_1, i_2, \dots, i_{64})$  is obtained as the feature descriptor. For details of SURF calculation, please refer to [38].

### 4.3. Convergence Strategy

In order to make the computer program better implement the multi-feature fusion image retrieval method, it is necessary to use appropriate rules to fuse the feature vectors. A direct and effective method is to connect the features of various patterns into a new feature vector to represent them. However, this method will lead to a rapid increase in feature dimensions and a decrease in time efficiency, which is not suitable for large-scale clothing retrieval. In addition, there are many hierarchical structures among various clothing categories, and the method of direct series connection ignores the statistical relationship between different pattern features.

Therefore, this paper adopts the fusion strategy at the level of similarity measure. Generally, there are linear fusion and nonlinear fusion methods, and linear fusion is adopted in this paper.

$$S_{Fusion} = \alpha_1 S_1(F_1) + \alpha_2 S_2(F_2) + \dots \quad (14)$$

where  $S_{Fusion}$  represents the similarity measure after fusion,  $S_1$  represents the similarity calculated between two images based on the first feature  $F_1$ ,  $\alpha_1$  denotes the proportion of similarity measure  $S_1$  in the fused measure  $S_{Fusion}$ , and so on.

In this paper, the similarity measure level is used for linear fusion. At present, there are mainly two kinds of image feature extraction, one is for image visual feature extraction, and the other is for image depth feature extraction through trained neural network. Therefore, in the following experiments, we fuse the new topology feature with these two methods to achieve clothing image retrieval.

## 5. Experiments and Analysis

### 5.1. Datasets

In order to ensure the accuracy and effectiveness, the dataset used here is DeepFashion, an open clothing image database of Chinese University of Hong Kong, which is commonly used in scientific research fields such as clothing image retrieval and classification. It contains 800,000 pictures, including pictures from different angles, different scenes, buyer shows, and so on. The WOMEN dataset in In-shop Clothes Retrieval Benchmark was selected in this paper, which includes 7982 items and 52,712 images, which are divided into 14 categories, and each image size is  $256 \times 256$ . The image database is shown in Figure 10.



Figure 10. Example of commodity image.

## 5.2. Evaluation of Image Retrieval

Top-K index [39] is the most commonly used in image retrieval tasks. In the task of image retrieval, it is necessary to calculate the similarity degree between different images through appropriate similarity measurement methods, and arrange the calculated results in turn, which can be used to evaluate the ability of the algorithm for image retrieval. In addition, the common evaluation indexes also include precision, which refers to the proportion of images with the same kind of images to be retrieved in all retrieval results in an image retrieval task. The average precision rate is the sum of P divided by the number of searches, and the calculation formula of precision rate is shown in (15),

$$P = \frac{TP}{TP + FP} \quad (15)$$

where  $TP$  represents the number of positive cases correctly predicted as positive cases, and  $FP$  represents the number of negative cases incorrectly predicted as positive cases.

When all the image results obtained at the end of the retrieval are numbered from small to large according to the order of the distance between them and the image to be retrieved,  $N$  represents the average value of the image numbers of the same things as the image to be retrieved in the retrieval results, and the calculation formula is shown in (16):

$$N = \frac{1}{M} \sum_{x=1}^M x \quad (16)$$

$M$  denotes the number of images similar to the image to be retrieved in the returned results, and  $x$  denotes the serial number of the returned images in the total results. Generally, the Top-K index is used to evaluate the quality of the retrieval results, which indicates the first K images with the smallest calculation distance as the retrieval results. Image retrieval tasks not only need to consider all the above factors, but also have strict rules on retrieval time, which often reflects the quality of model performance and whether the similarity measurement algorithm is suitable to a certain extent. For an image retrieval task, the calculation of the retrieval time is shown as (17), and when evaluating an image retrieval algorithm, it is necessary to comprehensively consider accuracy, ranking and retrieval time.

$$t = \frac{\text{one retrieval time}}{\text{total number of pictures in one retrieval}} \quad (17)$$

## 5.3. Experimental Results of Topology Feature

### 5.3.1. Results of Witness Complex

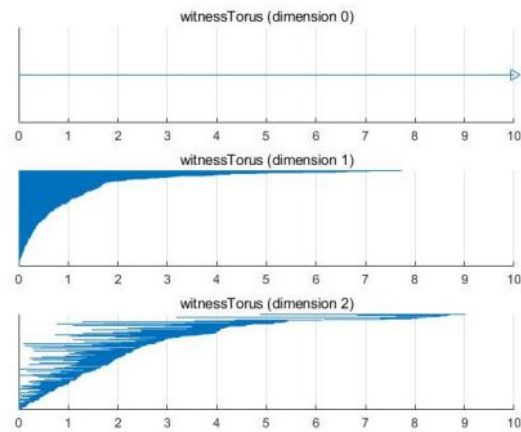
The difference between the witness complex and cube complex is that the cube complex directly inputs images for operation, and the calculation of the witness complex is not only based on the image itself, but also affected by the selection of data landmark points. Then, the calculation results are analyzed for different data quantities and different landmark points.

The selection number of landmark points only needs to satisfy:  $N/n \geq 20$  [26].  $N$  and  $n$  are the number of dot clusters  $Z$  and the number of landmark points, respectively. The smaller the ratio, the more landmark points are selected, and the calculation is finer, but the calculation time is longer. Therefore, the selection of landmark points should not be excessive.

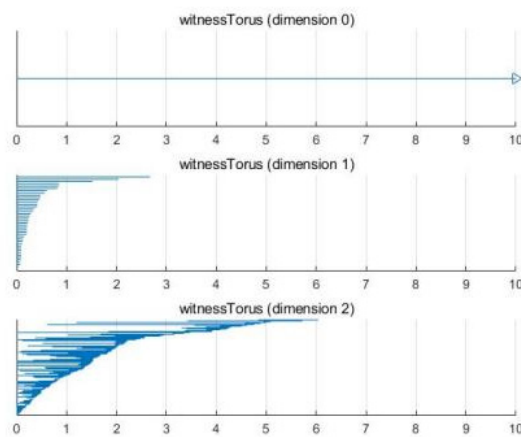
There is no requirement for the horizontal–vertical ratio of the image size in the calculation of the witness complex. Therefore, because both sides of the image in this dataset are invalid information, images are clipped and calculated. The specific clipping is to clip 40 pixels at the left and right ends, and the clipped image size is  $176 \times 256$ , which can reduce the calculation time of the witness complex without losing image information. Table 1 shows the running time in different sizes and ratios after cutting. Figure 11 shows the calculation results in the different sizes and ratios after cutting.

**Table 1.** Running time at different ratios after cutting.

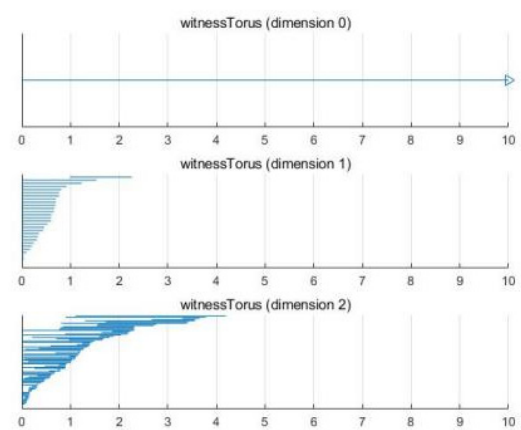
Image Size	Ratio	Running Time
176 × 256	128	67.14 s
176 × 256	192	31.14 s
176 × 256	256	18.41 s



(a) Ratio is 128



(b) Ratio is 192



(c) Ratio is 256

**Figure 11.** Results in different ratios after cropping.



### 5.3.2. Results of Topology Feature Histogram

During the calculation processing of the topology feature histogram, because of the particularity of the triangle structure and the calculation method of the Rips complex, the hole composed of three points cannot die out after filling, and the whole structure composed of three points is very common, especially in the area with strong pixel change. Therefore, in order to avoid this situation, the  $x$  and  $y$  coordinates of the pixels need to be scaled first.

This zoom mode has little influence on the area with gentle pixel gray change, but has an obvious effect on extracting the topology features from areas with strong pixel gray changes. As shown in Figure 12, it can be clearly seen that the appropriate multiple of  $x$  and  $y$  coordinates can make certain topology features appear. Of course, different scales of scaling will extract the different scales of topology features. If the scaling factor is too small, the influence of the pixel coordinates on distance will be too small, which leads to the inability to extract most features. When the RGB influence on the distance is too small, this results in similar features for most pixel blocks, which cannot be effectively distinguished.

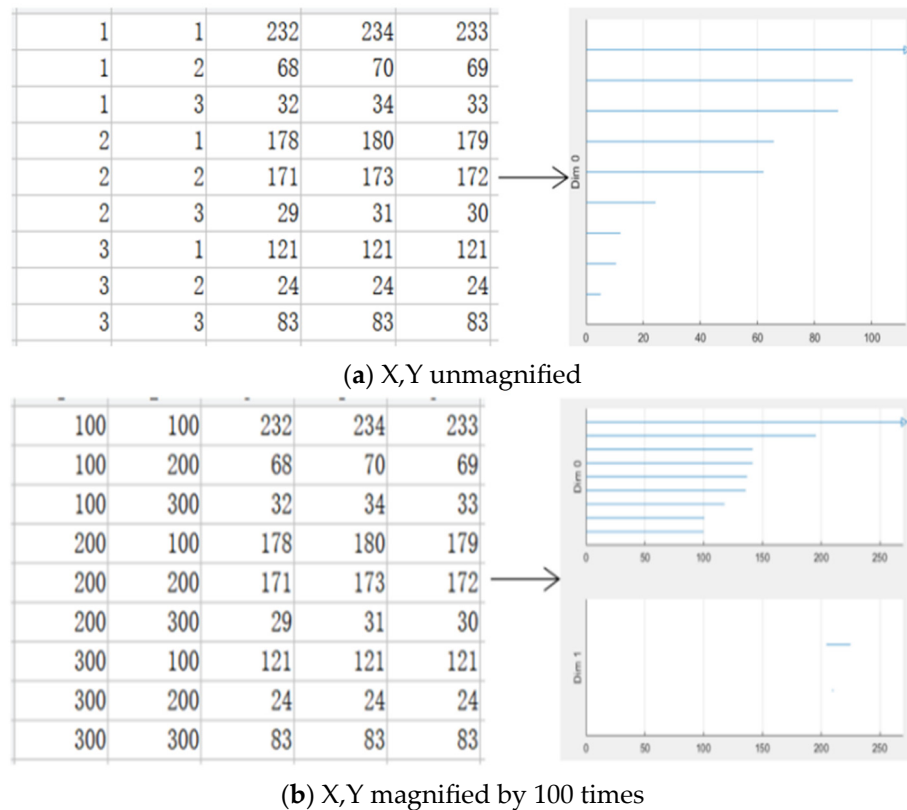
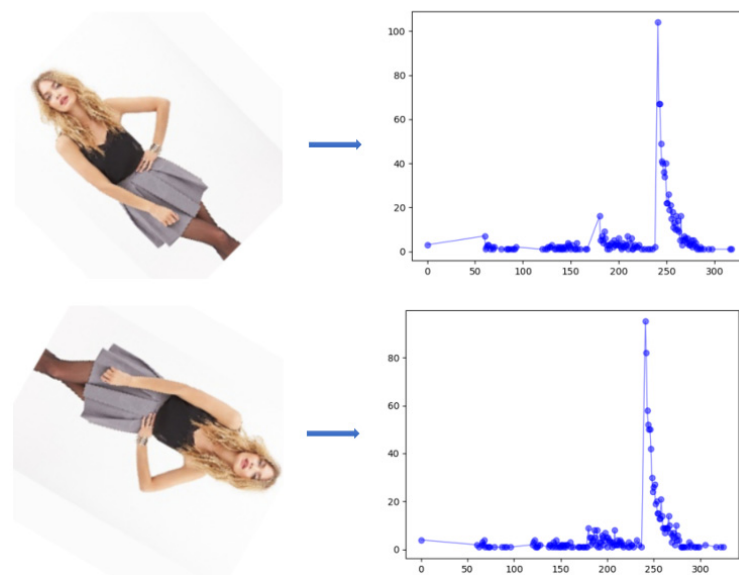


Figure 12. Result of zoomed-in 100 compared with the result without zoom.

In this paper, the  $256 \times 256$  size image is transformed into a  $254 \times 254$  size vector after  $3 \times 3$  windowing with one step size, and then normalized to 0–255 and turned into gray image.

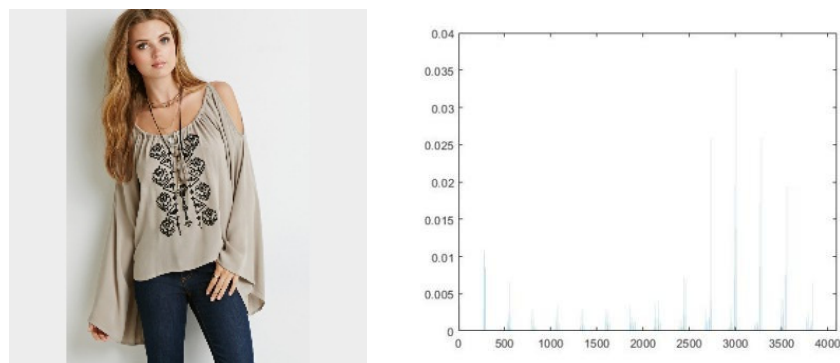
After the above processing, some local features of the image can be restored, and more features of image can be extracted with appropriate magnification. The disadvantage is that the calculation time is long, so the step size is three here. The feature is sensitive to the distance between pixels and insensitive to its spatial position information, so the feature has certain rotation invariance. Figure 13 shows topology feature histograms in different rotation angles.



**Figure 13.** Topology feature histogram in different rotation angles.

*5.4. Retrieval Result Combine Color Texture Feature with Topology Feature*

In this paper, color histogram, SURF features, and topology feature are used for clothing image retrieval. In the color histogram, after being divided into 16 intervals on the three components of R, G, and B, the colors are quantized, and finally  $16 \times 16 \times 16 = 4096$  color bins are formed. Then, count the number of these 4096 colors, that is, the color histogram. Due to the background of the image, the calculation effect after removing the histogram peak is shown in Figure 14. The similarity of two color histograms can be measured by Euclidean distance.



**Figure 14.** The image and its color histogram.

In the experiments, we extract SURF features from all images, and generate a  $64 \times n$ -dimensional feature from each image. For the measure of SURF features, we use the matching vectors closest to the first 30 distances, and calculate the average value as SURF distance, while the topology features are converted into two-dimensional vectors for calculation. In order to avoid the influence of color feature distance on the different color images of the same style, adaptive threshold processing is adopted because there are many different colors in the same style of clothing in this dataset. That is to say, for each image to be retrieved, when the color feature distance is too large in the calculation with the database, the color distance will not be assigned weight, that is, for clothing images with an excessively large color difference, the distance between them is only determined by the

SURF feature and topology feature. We adopt linear fusion, that is, the distances calculated by different features are linear fused together, as shown in (18).

$$S = \alpha_1 L_{color}(F_1, F_2) + \alpha_2 L_{SURF}(F_1, F_2) + \alpha_3 L_t(F_1, F_2) \tag{18}$$

where  $F_1$  represents the image to be retrieved,  $F_2$  represents the image in the database,  $\alpha_1$  represents the color feature distance weight,  $L_{color}$  represents the color feature distance,  $\alpha_2$  represents the SURF feature distance weight,  $L_{SURF}$  represents the SURF feature distance,  $\alpha_3$  represents the topology feature distance weight, and  $L_t$  represents the topology feature distance. There are seven ways to calculate the topology feature distance due to the different complex construction forms and measurement methods, and all the calculated distances need to be normalized.  $\alpha_1$  is 0 when the distance between the image to be retrieved and the color feature of the image in the database is large, and in the rest of cases, the weights of the three features are the same, all of which are 1.

In the experiment, the images to be retrieved are styles with more than 15 images in the dataset. Table 2 shows the retrieval results combined with the topology feature with color and SURF feature, where the witness complex is adopted in different distance measures and different landmark selection ratios, in which the image size is  $176 \times 256$ . It can be seen that, with the increase in ratio, the decrease in the number of landmark points will lead to the decrease in calculation, and the calculation time will decrease correspondingly but the accuracy will decrease. The retrieval results of the cube complex and topology feature histogram are shown in Table 3, and the time is the average calculation time of each image.

**Table 2.** Comparison of witness Complex retrieval results (different distance measures).

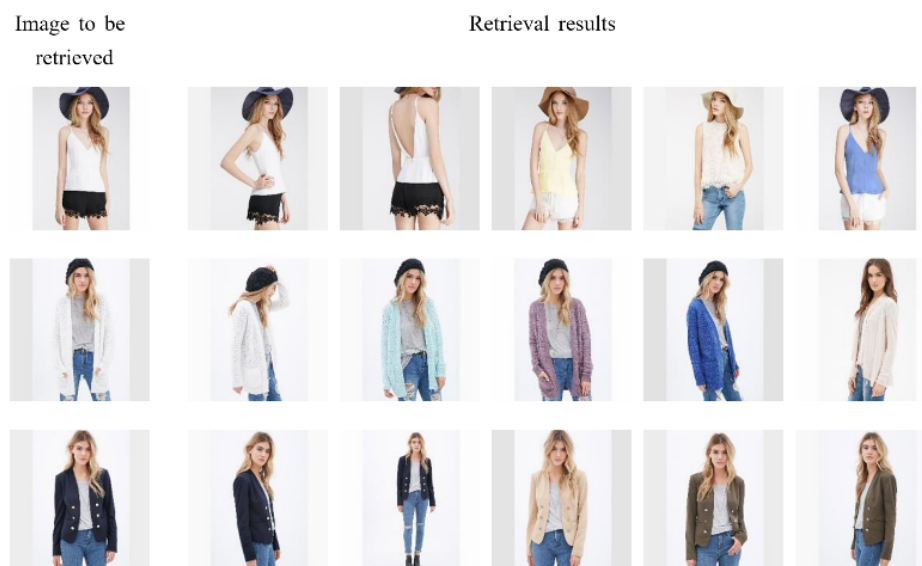
Distance Measure	Ratio	Top5 (%)	Top10 (%)	Top15 (%)	Time
Bottleneck distance	128	72.3	70.5	65.9	11.98
Bottleneck distance	192	71.4	69.8	64.1	10.65
Bottleneck distance	256	69.2	66.6	62.8	9.41
Wasserstein distance	128	81.5	78.6	74.5	12.96
Wasserstein distance	192	80.4	78.1	74.3	11.52
Wasserstein distance	256	79.2	77.4	73.8	10.07
Vector graph	128	69.6	66.3	64.8	9.02
Vector graph	192	68.8	65.5	62.1	9.02
Vector graph	256	68.7	65.1	60.8	9.02

**Table 3.** Comparison of retrieval results.

Method	Top5 (%)	Top10 (%)	Top15 (%)	Time
Only color surf	68.7	64.5	61.2	8.26
Color surf + cubic complex + bottleneck distance	74.5	71.5	69.8	11.78
Color Surf + cubic complex + Wasserstein distance	79.8	76.5	72.4	12.86
Color surf + cubic complex + vector graph	66.4	60.5	56.9	9.02
Color surf + topology feature Histogram + topology feature Histogram distance (our method)	83.6	80.9	78.8	8.93

As can be observed from Tables 2 and 3, for the measurement methods of the three topology features, the calculation speed of the vector graph is relatively fast, but the accuracy is poor. While the Wasserstein distance has high accuracy, but its calculation speed is the slowest. As can be seen from Table 3, after adding the topology feature, the retrieval effect is basically improved compared with only using the color and texture information. Our proposed method, i.e., the topology feature histogram + its corresponding distance, improves the retrieval rate of top5 by 14.9% compared with the method using color texture alone. Compared with the method of the cubic complex + Wasserstein distance, it is

improved by 3.8%, and the calculation time is saved by 3.93 s (accounting for 44%). It can be seen that the algorithm proposed retains the accuracy of Wasserstein distance, and can shorten some calculation time. Figure 15 shows some examples of top-5 retrieval results using the proposed algorithm, where the first column represents the clothing images to be retrieved, and the retrieval results show the first five similar images retrieved.



**Figure 15.** Some examples of retrieval results.

Experiments indicate that the three methods have good calculation results for images with different resolutions. Among the three methods, the witness complex is the slowest for feature extraction, which is affected by the image size and landmark selection ratio, and the amount of feature extracted directly affects the retrieval results and retrieval time. However, it has no requirement on the horizontal and vertical proportion and shape of the image, and it has little influence on the clothing images with different colors. The cube complex has the fastest calculation speed, but the cube complex has low retrieval accuracy and a large error for different colors. On the other hand, the algorithm proposed has good results under different color conditions, and the calculation speed is fast, so the comprehensive retrieval performance is the best.

## 6. Conclusions

This paper studies the technology of clothing image retrieval based on topology features. As for topology features, firstly, the different construction methods of image complex and their characteristics are studied, and the construction mode of topology feature histogram is proposed. This feature calculates the information of each pixel block of the image separately, and takes its statistical characteristics as the topology feature of the image, which has good scale and rotation invariance. According to the similarity measure of topology feature, this paper improves the Wasserstein distance and proposes the topology feature histogram distance. Experimental results show that the proposed topology feature measurement method can effectively reduce the operation time while ensuring the accuracy. In this paper, the color texture feature are combined with the topology feature, and then, the image retrieval is carried out. The experiments show that, after fusing the topology feature, the accuracy of clothing image retrieval has been improved.

Although persistent homology in algebraic topology has the advantages of mathematical principles, it is still difficult to give full play to the advantages of continuous homology method in image analysis and understanding technology, and there is room for further research and development. The shortcomings of this paper and the further work are as follows.

(1) In the related research, most studies use the one-dimensional Betty number as homology information for analysis and calculation, and 1-dimensional information can better represent the topology characteristics of data than 0-dimensional information, while the research on using higher-dimensional homology information for clothing image retrieval needs further discussion.

(2) The dataset studied in this paper is clothing images with a simple background, but in real life, the images and scenes shot are very complex, such as shadows, masks, deformation and so on. Therefore, the complex background image retrieval needs further discussion.

**Author Contributions:** Conceptualization, X.Z., H.S. and J.M.; methodology, H.S. and J.M.; software, X.Z.; validation, X.Z. and J.M.; formal analysis, H.S. and J.M.; investigation, X.Z.; resources, H.S.; data curation, J.M.; writing—original draft preparation, X.Z. and H.S.; writing—review and editing, X.Z. and H.S.; visualization, J.M.; supervision, H.S.; project administration, X.Z.; funding acquisition, X.Z. and H.S. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by Harbin City Science and Technology Plan Projects grant number 2022ZCZJCG006, Basic Research Support Program for Excellent Young Teachers in Provincial Undergraduate Universities in Heilongjiang Province grant number YQJH2023240.

**Data Availability Statement:** The original contributions presented in the study are included in the article, further inquiries can be directed to the corresponding author.

**Conflicts of Interest:** Author Jian Ma was employed by the Shenzhen Comen Medical Instruments Co., Ltd. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## References

1. Choi, Y.J.; Kim, K.J.; Nam, Y.; Cho, W.D. Retrieval of identical clothing images based on local color histograms. In Proceedings of the 2008 Third International Conference on Convergence and Hybrid Information Technology, Busan, Republic of Korea, 11–13 November 2008; Volume 1, pp. 818–823.
2. Gupta, M.; Bhatnagar, C.; Jalal, A.S. Clothing image retrieval based on multiple features for smarter shopping. *Procedia Comput. Sci.* **2018**, *125*, 143–148. [[CrossRef](#)]
3. Huang, D. *Research on Clothing Image Retrieval Technology Based on Joint Segmentation and Feature Matching*; Kunming University of Science and Technology: Kunming, China, 2018.
4. Jin, J. *Research on Key Technologies of Clothing Image Retrieval Based on Multi-Feature Fusion*; Kunming University of Science and Technology: Kunming, China, 2018.
5. Li, Z.M.; Li, Y.T.; Liu, Y.J.; Li, H. Clothing retrieval combining hierarchical over-segmentation and cross-domain dictionary learning. *J. Image Graph.* **2017**, *22*, 0358–0365.
6. Wang, M. *Research of Image Retrieval Technology Based on Clothing Elements*. Doctoral Dissertation, Nanjing Normal University, Nanjing, China, 2018.
7. Tao, B.; Chen, Q.; Pan, Z. Improvement of block weighted color histogram algorithms in clothing image retrieval system. *Laser J.* **2017**, *38*, 97–101.
8. Ge, J.; Yu, W. A Method of Clothing Image Retrieval Based on Color Shape Weighted Feature and LBP. *Mod. Comput. (Prof. Ed.)* **2018**, *19*, 33–38.
9. Chen, Q.; Tao, B.; Pan, Z.; Li, P.; Liu, S. Main Color Extraction Algorithm and its Application to Clothing Image Retrieval. *J. South China Norm. Univ. (Nat. Sci. Ed.)* **2019**, *51*, 111–119.
10. Miao, Z.; He, L.; Liu, D. A Clothing Image Retrieval Method Based on Weighted Color Shape Feature and LBP-GLCM Texture Feature Extraction. *Text. Rep.* **2019**, *4*, 4–7.
11. Hu, Y. *Research on Garment Image Retrieval Based on Landmark Feature*. Doctoral Dissertation, Shanghai Jiao Tong University, Shanghai, China, 2020. [[CrossRef](#)]
12. Qin, H. Construction and System Design of Clothing Retrieval Matching Model Based on Content. *Tech. Autom. Appl.* **2020**, *39*, 154–158+176.
13. Wu, Z.; Li, L.; Wang, J.; Jiang, H. An Image Retrieval Method of Clothing Fabric Based on Feature Fusion. *J. Cloth. Res.* **2021**, *6*, 42–47.
14. Hadi Kiapour, M.; Han, X.; Lazebnik, S.; Berg, A.C.; Berg, T.L. Where to buy it: Matching street clothing photos in online shops. In Proceedings of the IEEE International Conference on Computer Vision, Santiago, Chile, 7–13 December 2015; pp. 3343–3351.

15. Chen, Q.; Huang, J.; Feris, R.; Brown, L.M.; Dong, J.; Yan, S. Deep domain adaptation for describing people based on fine-grained clothing attributes. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015; pp. 5315–5324.
16. Wang, R.N.; Luo, M.; Feng, Q.; Peng, C.; He, D.B. Multi-Party Privacy-Preserving Faster R-CNN Framework for Object Detection. *IEEE Trans. Emerg. Top. Comput. Intell.* **2024**, *8*, 956–967. [[CrossRef](#)]
17. Liu, K.H.; Chen, T.Y.; Chen, C.S. Mvc: A dataset for view-invariant clothing retrieval and attribute prediction. In Proceedings of the 2016 ACM on International Conference on Multimedia Retrieval, New York, NY, USA, 6–9 June 2016; ACM: New York, NY, USA, 2016; pp. 313–316.
18. Wang, Z.; Gu, Y.; Zhang, Y.; Zhou, J.; Gu, X. Clothing retrieval with visual attention model. In Proceedings of the 2017 IEEE Visual Communications and Image Processing (VCIP), St. Petersburg, FL, USA, 10–13 December 2017; pp. 1–4. [[CrossRef](#)]
19. Verma, S.; Anand, S.; Arora, C.; Rai, A. Diversity in Fashion Recommendation using Semantic Parsing. In Proceedings of the 2018 25th IEEE International Conference on Image Processing (ICIP), Athens, Greece, 7–10 October 2018.
20. Garcia, N.; Vogiatzis, G. Dress like a star: Retrieving fashion products from videos. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 2293–2299.
21. Carlsson, G. Topology and data. *Bull. Am. Math. Soc.* **2009**, *46*, 255–308. [[CrossRef](#)]
22. Singh, G.; Mémoli, F.; Carlsson, G.E. Topology Methods for the Analysis of High Dimensional Data Sets and 3D Object Recognition. In *Eurographics Symposium on Point-Based Graphics*; Eurographics Association: Prague, Czech Republic, 2007; pp. 91–100.
23. Zhang, J.L.; Ju, X.M. Application of persistent homology to image classification and recognition. *Commun. Appl. Math. Comput.* **2017**, *31*, 494–508.
24. Cufar, M.; Virk, Z. Fast computation of persistent homology representatives with involuted persistent homology. *Found. Data Sci.* **2023**, *5*, 466–479. [[CrossRef](#)]
25. Cohen-Steiner, D.; Edelsbrunner, H.; Harer, J. Stability of persistence diagrams. *Discret. Comput. Geom.* **2007**, *37*, 103–120. [[CrossRef](#)]
26. De Silva, V.; Carlsson, G.E. Topology estimation using witness complexes. In Proceedings of the First Eurographics Conference on Point-Based Graphics (SPBG'04), Zurich, Switzerland, 2–4 June 2004; Eurographics Association: Goslar, Germany, 2004; pp. 157–166.
27. Edelsbrunner; Letscher; Zomorodian. Topology Persistence and Simplification. *Discret. Comput. Geom.* **2002**, *28*, 511–533. [[CrossRef](#)]
28. Amenta, N.; Bern, M. Surface Reconstruction by Voronoi Filtering. *Discret. Comput. Geom.* **1999**, *22*, 481–504. [[CrossRef](#)]
29. Ghrist, R. Barcodes: The persistent topology of data. *Bull. Am. Math. Soc.* **2007**, *45*, 61–75. [[CrossRef](#)]
30. Genevois, A. Special cube complexes revisited: A quasi-median generalization. *Can. J. Math.-J. Can. De Math.* **2023**, *75*, 743–777. [[CrossRef](#)]
31. Wagner, H.; Chen, C.; Vućini, E. Efficient Computation of Persistent Homology for Cubical Data. In *Mathematics and Visualization*; Springer: Berlin/Heidelberg, Germany, 2012.
32. Freudenthal, H. Simplicialzerlegungen von beschränkter Flachheit. *Ann. Math.* **1942**, *43*, 580–582. [[CrossRef](#)]
33. Zong, C. What is known about unit cubes. *Bull. Am. Math. Soc.* **2005**, *42*, 181–211. [[CrossRef](#)]
34. Babu, A.; John, S.J. Persistent homology based Bottleneck distance in hypergraph products. *Appl. Netw. Sci.* **2024**, *9*, 10. [[CrossRef](#)]
35. Wiesel, J.C.W. Measuring association with Wasserstein distances. *Bernoulli* **2022**, *28*, 2816–2832. [[CrossRef](#)]
36. Chen, S.; Gong, C.; Li, X.; Yang, J.; Niu, G.; Sugiyama, M. Boundary-restricted metric learning. *Mach. Learn.* **2023**, *112*, 4723–4762. [[CrossRef](#)]
37. Gao, J. *Learning Image Retrieval Algorithm Based on Deep Research and Implementation of Clothing*; Hunan University of Technology: Zhuzhou, China, 2022. [[CrossRef](#)]
38. Bay, H.; Tuytelaars, T.; Van Gool, L. SURF: Speeded Up Robust Features. In Proceedings of the Computer Vision—ECCV 2006, 9th European Conference on Computer Vision, Graz, Austria, 7–13 May 2006; Proceedings, Part I; Springer: Berlin/Heidelberg, Germany, 2006.
39. Hines, J.W.; Uhrig, R.E.; Wrest, D.J. Use of Auto associative Neural Networks for Signal Validation. *J. Intell. Robot. Syst.* **1998**, *21*, 143–154. [[CrossRef](#)]

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.