Article

Multi-View Graph Fusion for Semi-Supervised Learning: Application to Image-Based Face Beauty Prediction

Fadi Dornaika [1,2,3,*] and Abdelmalik Moujahid [2]

1   Henan Key Laboratory of Big Data Analysis and Processing, Henan University, Kaifeng 475001, China
2   Department of Computer Science and Artificial Intelligence, Faculty of Computer Science, University of the Basque Country UPV/EHU, M. Lardizabal 1, 20018 Donostia-San Sebastián, Spain; abdelmalik.moujahid@ehu.eus
3   IKERBASQUE, Basque Foundation for Science, Plaza Euskadi, 5, 48009 Bilbao, Spain
*   Correspondence: fadi.dornaika@ehu.eus

Abstract: Facial Beauty Prediction (FBP) is an important visual recognition problem to evaluate the attractiveness of faces according to human perception. Most existing FBP methods are based on supervised solutions using geometric or deep features. Semi-supervised learning for FBP is an almost unexplored research area. In this work, we propose a graph-based semi-supervised method in which multiple graphs are constructed to find the appropriate graph representation of the face images (with and without scores). The proposed method combines both geometric and deep feature-based graphs to produce a high-level representation of face images instead of using a single face descriptor and also improves the discriminative ability of graph-based score propagation methods. In addition to the data graph, our proposed approach fuses an additional graph adaptively built on the predicted beauty values. Experimental results on the SCUTFBP-5500 facial beauty dataset demonstrate the superiority of the proposed algorithm compared to other state-of-the-art methods.

Keywords: face beauty prediction; graph-based semi-supervised learning; graph fusion; score propagation; label graph; flexible manifold embedding

1. Introduction

The concept of beauty has always attracted people to find its true meaning. Since ancient times, beauty has been identified with proportion, even though the common definition of beauty in the Greek and Roman world also included the notion that proportion must always be accompanied by the grace of colour (and light). When the philosophers of ancient Greece began to discuss the origin of all things, they clearly recognized the correspondence between form and beauty. Pythagoras, however, was the one who explicitly stated these things, reconciling cosmology, mathematics, natural science and aesthetics. Pythagoras marks the birth of an aesthetic-mathematical view of the universe: all things exist because they are ordered, and they are ordered because they are the realisation of mathematical laws, which are, at the same time, a condition of existence and of beauty [1].

The philosophers of the Middle Ages went further, explaining that, for beauty to exist, there must be not only proper proportion but also integrity or perfection as well as clarity or splendour—as things of clear colour are considered beautiful. They considered that beauty is associated with brightly coloured things [1].

Eighteenth-century aesthetics placed great emphasis on the subjective and indeterminable aspects of taste, claiming that the basis of aesthetic experience is the dispassionate pleasure derived from the contemplation of beauty. Beauty is that which pleases in an objective way, without originating in or being attributable to a concept. When we judge an object to be beautiful, we assume that our judgment has universal value and that everyone must (or should) share our judgement. However, since the universality of judgments of

taste does not presuppose the existence of a concept to conform to, the universality of beauty is subjective [2].

Additionally, the idea of beauty is relative not only to different historical periods but also, even in the same period and even in the same country, different aesthetic ideals can coexist [3]. Cross-cultural agreement on attractiveness is evidence against the notion that attractiveness standards are learned gradually by exposure to culturally portrayed ideals [3,4]. This suggests that there is something universal about attractive faces (and unattractive faces) that is recognized by both individuals and cultures.

Both beauty and ugliness are not properties of external phenomena but are psychological in nature. Ugliness arises when certain fantasies arise in the consciousness that alter human aesthetic perception, and many experiences of ugliness lead to a process of integrating feelings of disturbance and disorder into familiar forms of understanding and aesthetic order [5].

The human face has attracted great interest among psychologists and other scientists over time. It has been suggested that facial features associated with attractiveness and healthy appearance are valid cues to aspects of underlying physiological health. Symmetry, sexual dimorphism, skin colour, facial adiposity and skin homogeneity are identified as factors contributing to attractiveness or healthy appearance [6–10]. For a recent review, please see [11].

In today's society, people pay great attention to their appearance, which has led to an increased demand for plastic surgery. Economic development, changing social and cultural norms, globalisation and contact with cultural media are the main factors behind the rapid development of cosmetic surgery. Cosmetic plastic surgery includes surgical and non-surgical procedures that enhance and reshape the structures of the body to improve appearance and self-confidence. According to the American Society of Plastic Surgeons (ASPS) (https://www.plasticsurgery.org/documents/News/Statistics/2020/plastic-surgery-statistics-full-report-2020.pdf, accessed on 10 May 2022), 15.6 million cosmetic procedures were performed in the United States in 2020, with three of the top five cosmetic procedures involving the face, namely 352,555 rhinoplasties, 352,112 eyelid surgeries and 234,374 facelift procedures. The total cost was approximately $16.7 billion in the U.S. alone.

Information about an individual's facial morphology can have several important clinical and forensic applications: Information on patient-specific models, enhancement and reduction of the need for extensive surgery for craniofacial abnormalities/trauma, prediction/reconstruction of facial shape from skeletal remains and the identification of suspects from DNA [12].

The fast-growing beauty market urgently needs a more precise definition of the beauty standard for faces. In animation and computer game design, research into the beauty of faces can also be a reference for designers creating virtual characters [13].

Compared to other tasks in facial image analysis, such as face recognition, facial expression recognition, gender classification, face age estimation and ethnicity classification, evaluating the beauty of faces is more challenging because it is difficult to apply a well-defined concept to describe the beauty information of facial images, which is a key problem in this field. Moreover, the classification of facial shape, which is of great importance for many beauty and fashion purposes, appears to be relatively complex and requires many features to distinguish [14].

Attempts to quantify facial beauty have been made by investigators from different areas, such as psychology, the arts, plastic surgeons and, more recently, image analysis [15,16]. In general, the task of automatically evaluating the beauty of a face can be accomplished by three categories of procedures. The first category concerns classification, in which beauty is defined by levels or classes. The second category deals with the ordering of face images, where the task is to rank the images according to their attractiveness. The third category deals with the automatic scoring of the beauty of faces. Schemes belonging to this category are the most general ones, since the solution for the first two categories is solved implicitly.

The existing methods rely on machine learning and computer vision techniques and have proven to achieve promising results.

Likely the first attempt for an accurate predictor of facial attractiveness was the one proposed in [17], where numerous facial features (90 geometric features and eight related to face symmetry, hair and skin colour and skin smoothness) were used along with an average human attractiveness score. The facial attractiveness ratings provided by the final predictor correlated strongly with the human ratings.

However, one of the major challenges of research in this area is how to create accurate face representations, which can be either feature-based, holistic or hybrid. The first category includes geometric, colour, texture or other local representations. Geometric features are based on the use of landmarks in the face and take into account their positions, distances between them or the relationship of these distances to each other [18]. On the other hand, holistic approaches focus on global information of the face instead of local features, e.g., using eigenfaces [19].

Other approaches were based on a family of convolutional neural networks whose input is the raw image, and the output is the beauty score [20,21]. Deep-learning-based methods have also been used for holistic approaches [22,23]. Indeed, over the past decade, studies in this area have focused in part on constructing larger and more reliable benchmarks, with many works using different neural networks, such as [22,24–26].

More recently, efforts to automatically assess the beauty of faces have shifted to manifold-based semi-supervised learning. In the field of machine learning, semi-supervised learning has indeed proven useful when relatively few labelled training samples are available; however, a large number of unlabelled samples are available [27–31].

Flexible Manifold Embedding (FME) [31], as one of the semi-supervised learning methods, can effectively utilize the information of labelled data as well as a manifold structure of labelled and unlabelled data. Much recent work in this area has focused on combining deep facial features with manifold learning methods to learn a better representation for predicting the beauty prediction [32].

The main contributions of this work can be summarized as follows: Our approach fuses two different deep face features and geometric features, which allows us to achieve better results compared with using a single descriptor. We integrated the information in the label space with the feature space, which improves the learning performance and increases the understandability of the obtained results. Moreover, since the relevance of all graphs was the same, our proposed algorithm did not need to search for the weights of the different graph matrices, which saves us a huge amount of time.

The remainder of this paper is organized as follows. In Section 2, we briefly explain the multimetric methods and the flexible embedding of manifolds. The details of the algorithm of the proposed method are presented in Section 3. The results of facial beauty prediction on the SCUTFBP-5500 dataset are presented in Section 4, followed by our concluding remarks.

2. Related Work

2.1. Multi-View Graph-Based Semi-Supervised Learning

Graph-based approaches have proven to be particularly successful in semi-supervised learning, especially in solving the label propagation problem [33–35]. However, their performance strongly depends on whether the underlying manifold of the given data is contained in the input graph. On the other hand, in many current real-world applications, the same instances are represented by multiple heterogeneous data sources. This may require tools that automatically integrate different data representations to achieve better predictive performance.

Recently, many efforts have been made to solve the problem of learning similarity metrics, which involves predicting whether pairs of samples are similar or not. Most metric learning methods share the common property that they all attempt to learn a single distance metric by finding a discriminative threshold. Deciding whether two samples are similar or not by relying only on this score has been shown to be unstable and less robust to variations

in the data structure. To overcome these weaknesses, multimetric and multiview learning approaches have recently been proposed [36–41].

Some semi-supervised learning methods rely on merging multiple graphs to maximize the variance between classes and minimize the variance within classes. These strategies exploit the different information of each view (or graph) and have been shown to improve classification performance. For example, the authors in reference [36] took a simple approach in which a set of *k*-NN graphs with different values of the neighbourhood parameter *K* are linearly integrated with some noisy graphs. Each graph was assigned the same weight value, and the relevant graphs were selected from the noisy graphs. Only one type of similarity graph was used in their algorithm.

Other authors went further and attempted to merge information from both the feature space and the label space. In this case, dynamic unequal weights were assigned to each constructed graph from the feature space, while only one fixed weight was used for the information from the label space [37].

The same goal was achieved by other authors [38] by integrating correlations between labels and similarities between instances into a new way of propagating labels. The intuition behind this approach is to dynamically update the similarity measures by merging information from multiple labels and multiple classes, which can be understood in a probabilistic framework. The K-nearest neighbour (KNN) matrix is used to obtain the intrinsic structure of the input data.

By integrating multiple graphs in the label propagation framework, many works succeeded in improving the performance of these algorithms [28,32,37,42]. The main notations used in our paper are shown in Table 1.

Table 1. The primary notations used in the paper.

Notation	Description
$\mathbf{X}$	Training data samples $\in \mathbb{R}^{d \times n}$
	$\mathbf{X} = [\mathbf{x}_1, \ldots, \mathbf{x}_\ell, \mathbf{x}_{\ell+1}, \ldots, \mathbf{x}_n]$
ℓ	Number of labelled samples
u	Number of unlabelled samples ($u = n - \ell$)
d	Dimensionality of data
n	Number of training samples/images
C	Total number of classes
$\mathbf{F}$	Prediction label matrix $\in \mathbb{R}^{n \times C}$ or Prediction score vector $\in \mathbb{R}^n$
$\mathbf{Y}$	Binary label matrix $\in \mathbb{R}^{n \times C}$ or ground-truth score vector $\in \mathbb{R}^n$
$\mathbf{U}$	Indicator diagonal Matrix $\in \mathbb{R}^{n \times n}$
$\mathbf{Q}$	Projection Matrix $\in \mathbb{R}^{d \times d}$ or Projection vector $\in \mathbb{R}^d$
$\mathbf{b}$	Bias vector $\in \mathbb{R}^c$ or bias scalar

2.2. Flexible Manifold Embedding (FME) for Semi-Supervised Classification

Let $\mathbf{X} \in \mathbb{R}^{d \times n}$ denote the data matrix containing n samples, each of dimensionality d, where l samples are labelled and $u = n - l$ samples have no labels. Assume that the data matrix is ordered such that the first l samples have labels and the unlabelled samples range from $l + 1$ to $l + u$. In addition, in classification problems with C classes, the corresponding labels are ordered in a binary label matrix $\mathbf{Y} = [\mathbf{Y}_l, \mathbf{Y}_u] = [\mathbf{y}_1; \mathbf{y}_2; \ldots; \mathbf{y}_l; \mathbf{y}_{l+1}; \ldots; \mathbf{y}_{l+u}] \in \mathbb{R}^{n \times C}$ such that $\mathbf{y}_i$ is a C dimensional vector that determines the class for the samples $\mathbf{x}_i$. For any sample $\mathbf{x}_i$ belonging to class c, the cth element of $\mathbf{y}_i$ is one, and the remainder is zero.

The Flexible Manifold Embedding (FME) method was introduced in [31] and consists of a unified framework for semi-supervised and unsupervised manifold learning that can handle data from nonlinear manifolds. For semi-supervised learning, FME enables mappings for unseen data points through a linear regression function that is able to handle the data from the nonlinear manifold by modeling the regression residual. It is useful for both semi-supervised settings: (i) transductive and (ii) inductive.

In the inductive setting, the framework is able to predict the label of unseen test samples. Many semi-supervised methods are transductive, meaning that they cannot handle unseen test samples. The FME method attempts to simultaneously find the optimal prediction label matrix **F** by minimizing the following function:

$$\min_{\mathbf{F},\mathbf{W},\mathbf{b}} Tr\left(\mathbf{F}^T \mathbf{L}\, \mathbf{F}\right) + Tr\left(\mathbf{F} - \mathbf{Y}\right)^T \mathbf{U}\left(\mathbf{F} - \mathbf{Y}\right) + \mu\left(||\mathbf{Q}||^2 + \gamma\,||\mathbf{X}^T\mathbf{Q} + \mathbf{1}\mathbf{b}^T - \mathbf{F}||^2\right) \quad (1)$$

where the first and second terms are the label smoothness and label fitness terms, respectively. **U** is the diagonal matrix, where U_{ii} is zero for the unlabelled samples and a non-zero value for the labelled samples. The third term consists of the regularization of the projection matrix (**Q**) and the linear regression model (provided by the projection matrix **Q** and the shift vector **b**). By seeking a small regression error (last term), FME seeks a nonlinear representation that is close to a linear model. This makes the model flexible.

The Laplacian graph $\mathbf{L} = \mathbf{D} - \mathbf{W}$, where **D** is the diagonal degree matrix, and **W** the similarity matrix. The parameters μ and γ refer to balancing terms and $||.||^2$ denotes the ℓ_2 norm of a matrix. **U** is the diagonal matrix such that the diagonal elements are 1 for labelled nodes and 0 for unlabelled nodes.

The closed-form solution of Equation (1) giving **F**, **Q** and **b** is as follows.

$$\mathbf{Q} = \gamma\left[\gamma\,\mathbf{X}\,\mathbf{H}_c\,\mathbf{X}^T + \mathbf{I}\right]^{-1}\mathbf{X}\,\mathbf{H}_c\,\mathbf{F} \quad (2)$$

$$\mathbf{b} = \frac{1}{n}[\mathbf{F}^T\mathbf{1} - \mathbf{Q}^T\mathbf{X}\mathbf{1}] \quad (3)$$

$$\mathbf{F} = (\mathbf{U} + \mathbf{L} + \mu\,\gamma\,\mathbf{H}_c - \mu\,\gamma^2\,\mathbf{N})^{-1}\mathbf{U}\,\mathbf{Y} \quad (4)$$

where $\mathbf{N} = \mathbf{X}_c^T\,\mathbf{X}_c\,(\gamma\,\mathbf{X}_c^T\,\mathbf{X}_c + \mathbf{I})^{-1}$. $\mathbf{X}_c$ and $\mathbf{H}_c = \mathbf{I} - \frac{1}{n}\mathbf{1}\,\mathbf{1}^T$ are the centred data matrix and the centring matrix, respectively. The mathematical proof of Equations (2)–(4) are described in [31].

In a previous work [43], we extended the FME algorithm with a multi-metric fusion approach by integrating multiple graphs and including the label space in the fusion process. The proposed method provides a unified framework that combines two phases: Graph fusion and label propagation. Starting from an existing view, several simple graphs were efficiently created and used as input for our proposed fusion approach. Moreover, the information from the label space was integrated into other similarity graphs as a new form of graph, namely the correlation graph. We emphasize that the FME framework and the method proposed in [43] are designed for classification problems.

In this work, we extend these approaches to the face beauty prediction problem, which is a regression problem. In other words, the label of a given image is determined by a real score instead of a vector of class probabilities.

3. Proposed Method

In this section, we present our Multi Similarity Metric Fusion Manifold Embedding (MSMFME) for Face Beauty Prediction. Since our knowledge of face beauty data is limited, using a single descriptor to build the similarity graph may result in an incomplete representation of the beauty information of a face image. The main goal is twofold: first, we construct different similarity graphs based on different descriptors; then, we construct the label space information graph, which we call a correlation graph, and merge both the label space graph and the data space graph into a new graph, which is used in the FME algorithm to predict the beauty value of the unknown face images.

A graphical representation of the main idea of this approach can be seen in Figure 1. We emphasize that the unlabelled images play two roles in our proposed scheme. First, they help to capture the global and local structure of the data via the estimation of the graphs built on the facial image features/descriptors.

Second, during the iterative process, a coarse prediction of the beauty score is produced for the unlabelled images, which will provide additional information through the

score graph (correlation graph). Since this graph is merged with the data graph, a better prediction of the scores is expected.
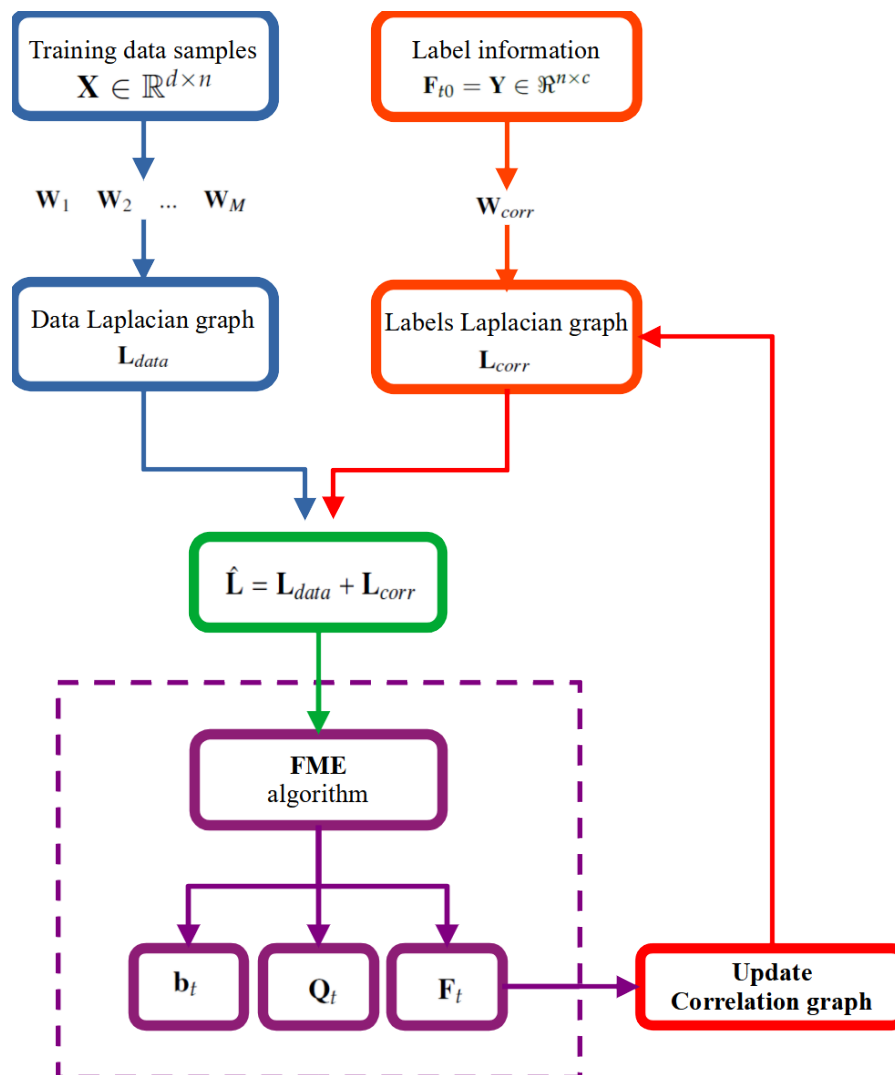

Figure 1. The general flowchart of the multi-view graph fusion that integrates label graph for semi-supervised learning. It is used for multi-class classification problems. Given the training dataset, we construct different similarity graphs based on different descriptors. Then, from the available labelling data, we construct the label space information graph, which we call a correlation graph, and merge both the label space graph and the data space graph into a new graph that is used in the FME algorithm to predict the beauty value of the unknown face images.

The graph construction technique used to construct the different graphs representing both the data space and the label space mainly uses the *K*-nearest neighbour (K-NNG) graph construction approach [44]. The *K*-NN graph of a set of vertices $V = \{x_1, x_2, \ldots, x_n\} \in \mathbb{R}^d$ is a undirected graph where V is the set of vertices and E is the set of edges from each $x_i \in V$ to its K most similar objects under a certain similarity measure. In this work, the Gaussian kernel function was used to encode the similarity between neighbours. Let that graph be described by the similarity matrix $\mathbf{W}$.

Note that the data graph quantifies the pairwise similarities based on the image descriptors, while the score graph (correlation graph) quantifies the pairwise similarities based on the image scores.

3.1. Multi Graphs for Manifold Smoothness

We construct M distinct similarity graphs, $\mathbf{W}_k, k = 1, \ldots, M$, with M distinct data matrices. Each data matrix corresponds to a particular descriptor. In this work, we considered two similarity graphs for deep features and one similarity graph for geometric features. For each graph, we compute its Laplacian matrix ($\mathbf{L}_k$). In the context of semi-supervised beauty scoring, the soft label matrix reduces to a vector of n scores.

Without loss of generality, we assume that this vector is given by $\mathbf{F} = [f_1, f_2, \ldots, f_l, f_{l+1}, \ldots, f_n]^T$ where the first l images have known scores and the remaining $n - l$ images have unknown scores. In addition to this vector, we have a vector storing the ground-truth scores of the images. The latter is named $\mathbf{Y}$ and given by $\mathbf{Y} = [y_1, y_2, \ldots, y_l, 0, \ldots, 0]^T$.

We assume that the manifold structure of the data is obtained by linear integration of the different similarity graphs. Based on this multi-view approach, we transform the label smoothness term of the FME into Equation (1) into the following form:

$$\sum_{k=1}^{M} Tr\left(\mathbf{F}^T \mathbf{L}_k \mathbf{F}\right) = \sum_{k=1}^{M} \mathbf{F}^T \mathbf{L}_k \mathbf{F} = \mathbf{F}^T \left[\sum_{k=1}^{M} \mathbf{L}_k\right] \mathbf{F} = \mathbf{F}^T \mathbf{L}_{data} \mathbf{F} \tag{5}$$

where $\mathbf{L}_{data} = \sum_{k=1}^{M} \mathbf{L}_k$ is the fused Laplacian matrix of the graphs built by different feature similarity matrices. The cost function given by Equation (1) is transformed into the following form by substituting the first two terms of the FME criterion. The trace operator is removed since $\mathbf{F}$ and $\mathbf{Y}$ are vectors.

$$g(\mathbf{F}, \mathbf{Q}, b) = \mathbf{F}^T \mathbf{L}_{data} \mathbf{F} + (\mathbf{F} - \mathbf{Y})^T \mathbf{U} (\mathbf{F} - \mathbf{Y}) + \mu \left(||\mathbf{Q}||^2 + \gamma ||\mathbf{X}^T \mathbf{Q} + \mathbf{1}\, b - \mathbf{F}||^2\right) \tag{6}$$

where the first term fuses M graphs associated with M available feature descriptors. The last two terms are the same as the last two terms in the FME algorithm.

3.2. Graph-Based Label Space Information

In semi-supervised learning, the information contained in the data space is generally well used, while the information that may be encoded in the label space is rarely used to obtain a general representation of the data. For example, in the face beauty problem, it is easy to understand that if two face images have similar beauty scores in the label space, they are likely to have strong similarity in the data space, i.e., the label space contains similarity information between two samples (cluster hypothesis [45]). In the complex face beauty prediction problem, two images with different image descriptors may have similar scores. Therefore, in this particular case, a graph based on the similarity of the scores can complement the data graph and contribute to a better score propagation among all images.

Indeed, we considered the label space as a new view of the data that could be described by a new similarity graph. The nodes of the graph were the samples and the weight of each edge was calculated by the Pearson correlation coefficient involving scores only. Since the face beauty score is not a vector, the Pearson correlation coefficient between the labels of two images cannot be calculated. To overcome this issue, for each image I_i, we define a new label vector $\mathbf{s}_i = [s_1; s_2; \ldots; s_K]$ with the purpose of making use of the label (score) information. Here, $s_1, s_2, \ldots,$ and s_K are the K nearest scores to the score of the image I_i. For two images I_i and I_j with label vectors $\mathbf{s}_i$ and $\mathbf{s}_j$, the similarity between them is calculated using the Pearson correlation coefficient, yielding the correlation similarity matrix $\mathbf{W}_{corr}$ with elements given by:

$$W_{corr}(i, j) = corr(\mathbf{s}_i, \mathbf{s}_j) = \frac{\sum_{m=1}^{K} (s_{i,m} - \bar{s}_i)(s_{j,m} - \bar{s}_j)}{\sqrt{\sum_{m=1}^{K} (s_{i,m} - \bar{s}_i)^2} \sqrt{\sum_{m=1}^{K} (s_{j,m} - \bar{s}_j)^2}} \tag{7}$$

where $\bar{s}_i$ is the average of the label vector $\mathbf{s}_i$ over the $K-$nearest neighbours of the image I_i.

The computation of the score graph is based on the scores s_j. If the image I_j is in the labelled part of the data, then s_j is set to the ground-truth score of that image. In the case

that the image does not belong to the labelled part of the data, then s_j is set to the soft score f_j estimated in the iterative process.

Afterward, we assign zero to the negative value obtained through Equation (7), and finally a sparse graph is generated by retaining the K' highest correlations values. In our work, K and K' are set to 10.

We add the obtained correlation graph to the previously mentioned data space graph by linearly adding it to the data graphs with the same weights as follows:

$$\hat{\mathbf{L}} = \mathbf{L}_{data} + \mathbf{L}_{corr} = \sum_{k=1}^{M} \mathbf{L}_k + \mathbf{D}_{corr} - \mathbf{W}_{corr} \tag{8}$$

where $\mathbf{L}_{corr} = \mathbf{D}_{corr} - \mathbf{W}_{corr}$ is the Laplacian of the correlation graph.

3.3. Proposed Algorithm

In multi-view score propagation, the criterion of FME becomes:

$$g(\mathbf{F}, \mathbf{Q}, b) = \mathbf{F}^T \,\hat{\mathbf{L}}\, \mathbf{F} + (\mathbf{F} - \mathbf{Y})^T \,\mathbf{U}\, (\mathbf{F} - \mathbf{Y}) + \mu \left(||\mathbf{Q}||^2 + \gamma \,||\mathbf{X}^T \mathbf{Q} + \mathbf{1}\, b - \mathbf{F}||^2 \right) \tag{9}$$

In the above equation, $\hat{\mathbf{L}}$ is given by Equation (8), and the data matrix $\mathbf{X}$ is simply the concatenation of the data matrices in all views. Thus, $\mathbf{X} = [\mathbf{X}_1; \ldots; \mathbf{X}_M]$

If $\hat{\mathbf{L}}$ is known (i.e., $\mathbf{L}_{corr}$ is known) then the closed-form solution of Equation (9) giving the optimal $\mathbf{F}$, $\mathbf{Q}$ and b is as follows.

$$\mathbf{Q} = \gamma \left[\gamma \, \mathbf{X}\, \mathbf{H}_c \, \mathbf{X}^T + \mathbf{I} \right]^{-1} \mathbf{X} \mathbf{H}_c \mathbf{F} \tag{10}$$

$$b = \frac{1}{n}[\mathbf{F}^T \mathbf{1} - \mathbf{Q}^T \mathbf{X} \mathbf{1}] \tag{11}$$

$$\mathbf{F} = (\mathbf{U} + \hat{\mathbf{L}} + \mu\,\gamma\,\mathbf{H}_c - \mu\,\gamma^2\,\mathbf{N})^{-1} \mathbf{U}\,\mathbf{Y} \tag{12}$$

where $\mathbf{N} = \mathbf{X}_c^T \mathbf{X}_c \,(\gamma\,\mathbf{X}_c^T \mathbf{X}_c + \mathbf{I})^{-1}$. $\mathbf{X}_c$ and $\mathbf{H}_c = \mathbf{I} - \frac{1}{n}\mathbf{1}\mathbf{1}^T$ are the centred data matrix and the centring matrix, respectively.

Finally, our proposed algorithm proceeds as follows. After initialization of the soft scores of all images $\mathbf{F}$, Equations (7), (8) and (12) are iterated until the solution stops to change or the maximum number of iterations is reached. In each iteration, the correlation graph $\mathbf{W}_{corr}$ changes with the predicted scores and the associated score similarities. After each iteration, the updated $\hat{\mathbf{L}}$ is used in the cost function of the proposed multi-view FME of Equation (9).

The final solution is the predicted scores as well as a linear mapping model that allow to recover the score of unseen images.

The flowchart of our proposed approach is shown in Figure 2. In the specific case of this work, Figure 2 shows the pipeline of this algorithm considering two deep feature similarity graphs and one geometric feature similarity graph. In each iteration, the prediction matrix F obtained from FME is used to update the correlation graph. The projection matrix $\mathbf{Q}$ obtained after t iterations, and the bias vector $\mathbf{b}$ can be used to estimate the beauty scores of the unseen samples using the following:

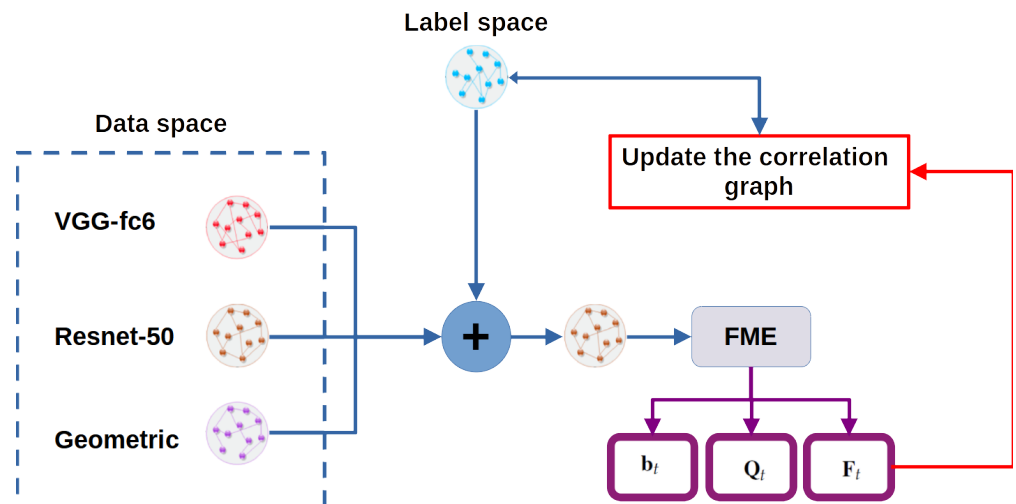$$f_{unseen} = \mathbf{x}_{unseen}^T \,\mathbf{Q} + b \tag{13}$$

Figure 2. The framework of the proposed method. Different similarity graphs corresponding to the different descriptors are constructed. The correlation graph is constructed in the score space. All graphs are merged into a single graph. The FME method is repeatedly used to process the resulting fused graph and to refine the face beauty prediction for the unseen images in each iteration.

The proposed MSMFME algorithm is summarized in Algorithm 1.

Algorithm 1: Multi similarity metric fusion manifold embedding for face beauty prediction

---

Input:

1. Vgg-fc6 feature matrix $\mathbf{X}_{fc6} = [\mathbf{x}^1_{fc6}, \mathbf{x}^2_{fc6}, \dots, \mathbf{x}^n_{fc6}]$
2. Resnet-50 feature matrix $\mathbf{X}_{resnet} = [\mathbf{x}^1_{resnet}, \mathbf{x}^2_{resnet}, \dots, \mathbf{x}^n_{resnet}]$
3. Geometric feature matrix $\mathbf{X}_{geo} = [\mathbf{x}^1_{geo}, \mathbf{x}^2_{geo}, \dots, \mathbf{x}^n_{geo}]$
4. Score beauty vector $\mathbf{Y} = [y_1, y_2, \dots, y_l, 0, 0, 0, \dots]$ (*l* images are scored)
5. Parameters μ, γ

Output: Prediction scores for all images $\mathbf{F} = [f_1, f_2, \dots, f_l, f_{l+1}, f_{l+2}, \dots, f_n]$.

Process: Construct three different data graphs $\mathbf{W}_{fc6}, \mathbf{W}_{resnet}, \mathbf{W}_{geo}$ ($M = 3$);
Calculate their corresponding Laplacian matrices $\mathbf{L}_{fc6}, \mathbf{L}_{resnet}$ and $\mathbf{L}_{geo}$
Set the data matrix to $\mathbf{X} = [\mathbf{X}_{fc6}; \mathbf{X}_{resnet}; \mathbf{X}_{geo}]$
Initialize the soft score vector $\mathbf{F} = \mathbf{Y}$
for t = 1: Maxiteration
Generate the correlation graph $\mathbf{W}_{corr}$ based on $\mathbf{F}$ using Equation (7).
Obtain the global Laplacian $\hat{\mathbf{L}}$ by fusing the $M + 1$ Laplacian matrices (Equation (8)).
Feed $\hat{\mathbf{L}}$ to the FME framework and calculate a new soft score vector $\mathbf{F}$ using Equations (10)–(12).
end for:
Use the projection vector $\mathbf{Q}$ and the bias b to estimate the beauty scores of unseen samples (Equation (13)).

---

4. Experimental Results

4.1. Databases

We evaluated the proposed multi-view method using the SCUT-FBP5500 benchmark dataset and performed a quantitative comparison of the results obtained with those obtained using a single-view approach only. The comparison is based on three measures: the Pearson correlation coefficient (PC), the mean absolute error (MAE) and the root mean square error (RMSE).

The SCUT-FBP5500 contains 5500 frontal, unobstructed faces between the ages of 15 and 60 with neutral expressions. It can be divided into four racial and gender subgroups, including 2000 Asian females, 2000 Asian males, 750 Caucasian females and 750 Caucasian

males. Most of the SCUT-FBP5500 images were collected from the Internet. All images were labelled with beauty values in the range of [1, 5]. Some examples can be found in Figure 3.

Figure 3. Some face images from the SCUT-FBP5500 dataset.

4.2. Features

Geometric features have been widely used to quantitatively describe human attraction [17,18]. The distance of the facial features and the proportions of a beautiful face have the same influence in Asians and Caucasians. The geometric methods used by the predecessors do not consider so much the direct use of facial features as a characteristic to predict a beautiful face. In this work, 81 points of facial contour are used as the geometric feature matrix as Figure 4 shows.
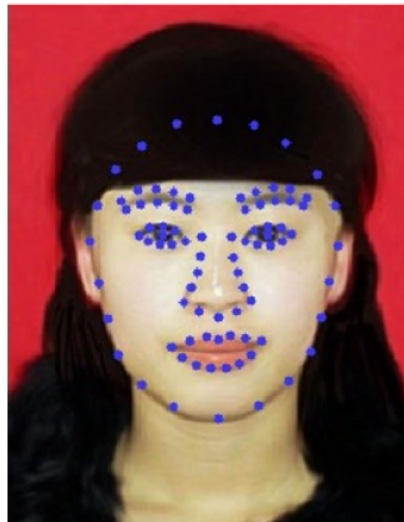

Figure 4. The 81 facial points detected in a face image.

The geometric features are given by the 2D locations of 81 facial points (a vector of 162 elements) corresponding to the significant facial components of each images, such as the eyes, eyebrows, nose and mouth. For each image, these points are provided as part of the dataset SCUT-FBP5500. The providers of this dataset also provided the set of 81 points for each image in the dataset. A GUI landmark location system is developed, where the original location of the landmarks are initialized by the active shape model (ASM) trained

by the SCUT-FBP dataset. Then, the detected landmarks by ASM are modified manually by volunteers to ensure the accuracy.

On the other hand, our approach also fuses two different deep face features extracted from two different CNN networks (see [46] for a recent review). The ResNet-50, which is a convolutional neural network that is 50 layers deep [47] and the VGG-face network [48]. The dimensions of the resulting descriptors are shown in Table 2. Note that the two CNNs were pretrained on the face dataset VGG-Face (https://www.robots.ox.ac.uk/~vgg/data/vgg_face/, accessed on 10 May 2022) to solve a classic face recognition problem.

Table 2. Face beauty dataset and image descriptors.

Dataset	Descriptor	Dimension	# of Instance
SCUT5500-AF	VGG-face+fc6 Resnet50 Geometric	4096 2048 162	2000
SCUT5500-AM	VGG-face+fc6 Resnet50 Geometric	4096 2048 162	2000
SCUT5500-CF	VGG-face+fc6 Resnet50 Geometric	4096 2048 162	750
SCUT5500-CM	VGG-face+fc6 Resnet50 Geometric	4096 2048 162	750

4.3. Training Setup and Experimental Results

Our proposed method was trained and tested using five-fold cross-validation. We conducted experiments to predict the beauty of faces using the four subsets or the entire set SCUT-FBP5500. The results of the average of all five-fold cross validations for the different subsets are shown in Table 3. In this setting, 80% of the images have a score, and the remaining 20% are unlabelled. The results show that merging multiple similarity graphs from the data space and the label space achieved the best performance compared to other methods using only a single graph. Moreover, using the Caucasian and Asian dataset, we can see that the more samples in the dataset, the better the performance in predicting the beauty of a face.

Although it is not fair to compare a semi-supervised method with supervised methods, Table 4 shows such a comparison. The reason for this is twofold. First, there are almost only supervised methods for predicting the beauty of faces in the literature. Second, the supervised approaches can give an indication of the actual performance of the semi-supervised approach, especially when the number of labelled images is exactly the same.

This table includes the MAE, the RMSE and the PC obtained with different supervised techniques. The same table also illustrates the performance with a recent semi-supervised approach. We can see that the proposed method performs well compared to many supervised techniques that use end-to-end deep learning paradigms. Moreover, our semi-supervised method outperformed a recent semi-supervised approach based on graph-based score propagation.

Finally, Figure 5 shows the predicted scores for three face images. In this figure, we compare the predictions obtained by two semi-supervised techniques: the NFME method presented in [29] and the proposed method. The top row shows the ground truth scores. We can see that the predictions of our proposed method are closer to the ground truth than those of the competing method. We can also see that both methods accurately predicted the score associated with the average degree of beauty.

Table 3. Semi-supervised face beauty prediction.

Dataset	Method	MAE ↓	RMSE ↓	PC (%) ↑
SCUT5500-AF	VGG-face+fc6	0.237	0.307	89.4
	ResNet-50	0.225	0.300	90.5
	Proposed method	**0.220**	**0.277**	**91.3**
SCUT5500-AM	VGG-face+fc6	0.232	0.301	89.9
	ResNet-50	0.224	0.283	91.5
	Proposed method	**0.218**	**0.276**	**92.2**
SCUT5500-CF	VGG-face+fc6	0.257	0.337	88.6
	ResNet-50	0.241	0.324	89.5
	Proposed method	**0.231**	**0.302**	**90.3**
SCUT5500-CM	VGG-face+fc6	0.234	0.318	88.7
	ResNet-50	0.232	0.317	89.9
	Proposed method	**0.230**	**0.300**	**90.5**
SCUT5500	VGG-face+fc6	0.242	0.317	89.0
	ResNet-50	0.229	0.302	90.2
	Proposed method	**0.221**	**0.2870**	**91.1**

Table 4. Comparison with the state-of-the-art supervised methods using the five-fold cross-validation scenario. Only Reference [29] is a semi-supervised approach.

Method	MAE ↓	RMSE ↓	PC (%) ↑
Alexnet [49]	0.2651	0.3481	86.34
Resnet-18 [49]	0.2419	0.3166	89.00
ResneXt-50 [49]	0.2291	0.3017	89.97
CNN with SCA [50]	0.2287	0.3014	90.03
PI-CNN [51]	0.2267	0.3016	89.78
CNN + LDL [52]	0.2201	0.2940	90.31
ResNet-18 based AaNet [53]	0.2236	0.2954	90.55
ResneXt-50-R^{3}CNN [54]	0.2120	0.2800	91.42
Semi-supervised [29]	0.2675	0.3455	86.60
Semi-supervised **(Ours)**	0.2210	0.2870	91.13

Ground truth score	**1.955**	**2.715**	**4.084**
Predicted score [29]	**2.509**	**2.675**	**3.923**
Predicted score (Ours)	**2.221**	**2.735**	**4.035**

Figure 5. Three prediction examples using two semi-supervised methods. The upper row shows the ground truth scores.

5. Conclusions

As mentioned in the introduction to this work, people largely agree on which faces are attractive, both within and across cultures. Our judgments about the attractiveness of others often happen unconsciously and influence us in ways of which we are unaware [55,56].

Psychologists have observed that citizens vote for more attractive political candidates, that judges give more lenient sentences to more attractive defendants and that teachers grade better-looking students more positively [57].

These observations have been difficult to explain; however, today, using novel technologies, such as digital face morphing and brain imaging, psychologists and neuroscientists are beginning to identify the different features that people find attractive in faces as well as the complex networks in the brain that respond to beautiful features. This work not only uncovers neural connections between evaluations of attractiveness and social traits, such as trustworthiness but also offers insights into our appreciation of works of art, such as the Nefertiti bust [58].

In this paper, we presented a general semi-supervised approach to the problem of facial beauty prediction. The main point behind this approach is the use of information available in both the data space and the label space (score space) by using multiple graphs to construct an efficient graph-based representation of the input face images. This graph is used as input to the FME algorithm in an iterative framework where the graph matrix associated with all scores is updated during the optimization process.

The experimental results on the dataset SCUT-5500 show that our multimetric fusion method achieved the best performance compared to using a single feature. It also performed well compared to many fully supervised methods (see Table 4). Our main task was to find the best graph that represented face beauty information.

In the future, we plan to explore non-linear representations of the data space and incorporate them into the regression model. This can be conducted either by using kernel representations of the data or by using explicit nonlinear mapping functions. In addition, adopting combinations of weighted similarity matrices associated with hand-crafted features and deep features would be another research direction to explore.

Author Contributions: Conceptualization, A.M and F.D.; methodology, F.D. and A.M.; validation, F.D. and A.M.; writing—original draft preparation, F.D. and A.M.; writing—review and editing, F.D. and A.M.; funding acquisition, F.D. and A.M. All authors have read and agreed to the published version of the manuscript.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The 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. Eco, U.; McEwen, A. *History of Beauty*; Rizzoli: New York, NY, USA 2005.
2. Kant, I. *Critique of the Power of Judgment*; The Cambridge Edition of the Works of Immanuel Kant, Cambridge University Press: Cambridge, UK, 2000. [CrossRef]
3. Wolf, N. *The Beauty Myth: How Images of Beauty Are Used against Women*; Random House: New York, NY, USA 2013.
4. Chernorizov, A.M.; Zinchenko, Y.P.; Qing, J.Z.; Petrakova, A. Face cognition in humans: Psychophysiological, developmental, and cross-cultural aspects. *Psychol. Russ. State Art* **2016**, *9*, 37–50. [CrossRef]
5. Eco, U. *On Ugliness*; Rizzoli: New York, NY, USA, 2007.
6. Grammer, K.; Thornhill, R. Human (Homo sapiens) facial attractiveness and sexual selection: The role of symmetry and averageness. *J. Comp. Psychol.* **1994**, *108*, 233–242. [CrossRef] [PubMed]
7. Rhodes, G.; Zebrowitz, L.A.; Clark, A.; Kalick, S.; Hightower, A.; McKay, R. Do facial averageness and symmetry signal health? *Evol. Hum. Behav.* **2001**, *22*, 31–46. [CrossRef]
8. Perrett, D.I.; Lee, K.J.; Penton-Voak, I.; Rowland, D.; Yoshikawa, S.; Burt, D.M.; Henzi, S.P.; Castles, D.L.; Akamatsu, S. Effects of sexual dimorphism on facial attractiveness. *Nature* **1998**, *394*, 884–887. [CrossRef] [PubMed]
9. Coetzee, V.; Perrett, D.I.; Stephen, I.D. Facial Adiposity: A Cue to Health? *Perception* **2009**, *38*, 1700–1711. [CrossRef]
10. Matts, P.J.; Fink, B.; Grammer, K.; Burquest, M. Color homogeneity and visual perception of age, health, and attractiveness of female facial skin. *J. Am. Acad. Dermatol.* **2007**, *57*, 977–984. [CrossRef] [PubMed]

11. de Jager, S.; Coetzee, N.; Coetzee, V. Facial Adiposity, Attractiveness, and Health: A Review. *Front. Psychol.* **2018**, *9*, 2562. [CrossRef]
12. Richmond, S.; Howe, L.J.; Lewis, S.; Stergiakouli, E.; Zhurov, A. Facial Genetics: A Brief Overview. *Front. Genet.* **2018**, *9*, 462. [CrossRef]
13. Zendle, D.; Meyer, R.; Ballou, N. The changing face of desktop video game monetisation: An exploration of exposure to loot boxes, pay to win, and cosmetic microtransactions in the most-played Steam games of 2010–2019. *PLoS ONE* **2020**, *15*, e0232780. [CrossRef]
14. Hossam, M.; Afify, A.A.; Rady, M.; Nabil, M.; Moussa, K.; Yousri, R.; Darweesh, M.S. A Comparative Study of Different Face Shape Classification Techniques. In Proceedings of the 2021 International Conference on Electronic Engineering (ICEEM), Menouf, Egypt, 3–4 July 2021; pp. 1–6. [CrossRef]
15. Gunes, H.; Piccardi, M. Assessing facial beauty through proportion analysis by image processing and supervised learning. *Int. J. Hum.-Comput. Stud.* **2006**, *64*, 1184–1199. [CrossRef]
16. Langlois, J.H.; Roggman, L.A. Attractive Faces Are Only Average. *Psychol. Sci.* **1990**, *1*, 115–121. [CrossRef]
17. Schölkopf, B.; Platt, J.; Hofmann, T. A Humanlike Predictor of Facial Attractiveness. In *Advances in Neural Information Processing Systems 19: Proceedings of the 2006 Conference*; MIT Press: Cambridge, MA, USA, 2007; pp. 649–656.
18. Zhang, D.; Zhao, Q.; Chen, F. Quantitative analysis of human facial beauty using geometric features. *Pattern Recognit.* **2011**, *44*, 940–950. [CrossRef]
19. Eisenthal, Y.; Dror, G.; Ruppin, E. Facial attractiveness: Beauty and the machine. *Neural Comput.* **2006**, *18*, 119–142. [CrossRef] [PubMed]
20. Gray, D.; Yu, K.; Xu, W.; Gong, Y. Predicting facial beauty without landmarks. In Proceedings of the Computer Vision–ECCV 2010, Crete, Greece, 5–11 September 2010; pp. 434–447.
21. Liu, X.; Li, T.; Peng, H.; Ouyang, I.C.; Kim, T.; Wang, R. Understanding Beauty via Deep Facial Features. In Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Long Beach, CA, USA, 16–20 June 2019; pp. 246–256. [CrossRef]
22. Gan, J.; Li, L.; Zhai, Y.; Liu, Y. Deep self-taught learning for facial beauty prediction. *Neurocomputing* **2014**, *144*, 295–303. [CrossRef]
23. Wang, S.; Shao, M.; Fu, Y. Attractive or not?: Beauty prediction with attractiveness-aware encoders and robust late fusion. In Proceedings of the 22nd ACM International Conference on Multimedia, Orlando, FL, USA, 3–7 November 2014; ACM: New York, NY, USA, 2014; pp. 805–808.
24. Nguyen, T.V.; Liu, S.; Ni, B.; Tan, J.; Rui, Y.; Yan, S. Towards decrypting attractiveness via multi-modality cues. *ACM Trans. Multimed. Comput. Commun. Appl. (TOMM)* **2013**, *9*, 28. [CrossRef]
25. Xie, D.; Liang, L.; Jin, L.; Xu, J.; Li, M. SCUT-FBP: A benchmark dataset for facial beauty perception. In Proceedings of the 2015 IEEE International Conference on Systems, Man, and Cybernetics (SMC), Hong Kong, China, 9–12 October 2015; pp. 1821–1826.
26. Xu, J.; Jin, L.; Liang, L.; Feng, Z.; Xie, D. A new humanlike facial attractiveness predictor with cascaded fine-tuning deep learning model. *arXiv* **2015**, arXiv:1511.02465.
27. Dornaika, F.; Bosaghzadeh, A. Exponential Local Discriminant Embedding and Its Application to Face Recognition. *IEEE Trans. Cybern.* **2013**, *43*, 921–934. [CrossRef]
28. Dornaika, F.; Elorza, A.; Wang, K.; Arganda-Carreras, I. Nonlinear, flexible, semisupervised learning scheme for face beauty scoring. *J. Electron. Imaging* **2019**, *28*, 1. [CrossRef]
29. Dornaika, F.; Wang, K.; Arganda-Carreras, I.; Elorza, A.; Moujahid, A. Toward graph-based semi-supervised face beauty prediction. *Expert Syst. Appl.* **2020**, *142*, 112990. [CrossRef]
30. El Traboulsi, Y.; Dornaika, F.; Assoum, A. Kernel flexible manifold embedding for pattern classification. *Neurocomputing* **2015**, *167*, 517–527. [CrossRef]
31. Nie, F.; Xu, D.; Tsang, I.W.H.; Zhang, C. Flexible Manifold Embedding: A Framework for Semi-Supervised and Unsupervised Dimension Reduction. *IEEE Trans. Image Process.* **2010**, *19*, 1921–1932. [CrossRef] [PubMed]
32. An, L.; Chen, X.; Yang, S. Multi-graph feature level fusion for person re-identification. *Neurocomputing* **2017**, *259*, 39–45. [CrossRef]
33. Ziraki, N.; Dornaika, F.; Bosaghzadeh, A. Multiple-view flexible semi-supervised classification through consistent graph construction and label propagation. *Neural Netw.* **2022**, *146*, 174–180. [CrossRef] [PubMed]
34. Namjoy, A.; Bosaghzadeh, A. A Sample Dependent Decision Fusion Algorithm for Graph-based Semi-supervised Learning. *Int. J. Eng.* **2020**, *33*, 1010–1019. [CrossRef]
35. Dornaika, F.; Dahbi, R.; Bosaghzadeh, A.; Ruichek, Y. Efficient dynamic graph construction for inductive semi-supervised learning. *Neural Netw.* **2017**, *94*, 192–203. [CrossRef]
36. Karasuyama, M.; Mamitsuka, H. Multiple Graph Label Propagation by Sparse Integration. *IEEE Trans. Neural Netw. Learn. Syst.* **2013**, *24*, 1999–2012. [CrossRef]
37. Lin, G.; Liao, K.; Sun, B.; Chen, Y.; Zhao, F. Dynamic graph fusion label propagation for semi-supervised multi-modality classification. *Pattern Recognit.* **2017**, *68*, 14–23. [CrossRef]
38. Wang, B.; Tsotsos, J. Dynamic label propagation for semi-supervised multi-class multi-label classification. *Pattern Recognit.* **2016**, *52*, 75–84. [CrossRef]
39. Zhang, L.; Zhang, D. MetricFusion: Generalized Metric Swarm Learning for Similarity Measure. *Inf. Fusion* **2016**, *30*, 80–90. [CrossRef]

40. Cao, Q.; Ying, Y.; Li, P. Similarity Metric Learning for Face Recognition. In Proceedings of the IEEE International Conference on Computer Vision (ICCV), Sydney, Australia, 1–8 December 2013.

41. Zhang, Y.; Zhang, H.; Nasrabadi, N.M.; Huang, T.S. Multi-metric learning for multi-sensor fusion based classification. *Inf. Fusion* **2013**, *14*, 431–440. [CrossRef]

42. Zhou, D.; Bousquet, O.; Lal, T.; Weston, J.; Schölkopf, B. Learning with Local and Global Consistency. In *Advances in Neural Information Processing Systems*; Thrun, S., Saul, L., Schölkopf, B., Eds.; MIT Press: Cambridge, MA, USA, 2003; Volume 16.

43. Bahrami, S.; Bosaghzadeh, A.; Dornaika, F. Multi Similarity Metric Fusion in Graph-Based Semi-Supervised Learning. *Computation* **2019**, *7*, 15. [CrossRef]

44. Eppstein, D.; Paterson, M.; Yao, F. On Nearest-Neighbor Graphs. *Comput. Geom.* **1997**, *17*, 263–282. [CrossRef]

45. Manning, C.; Prabhakar, R.; Hinrich, S. *"16. Flat Clustering". Introduction to Information Retrieval*; Cambridge University Press: New York, NY, USA, 2008.

46. Saeed, J.N.; Abdulazeez, A.M. Facial Beauty Prediction and Analysis Based on Deep Convolutional Neural Network: A Review. *J. Soft Comput. Data Min.* **2021**, *2*, 1–12. [CrossRef]

47. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep Residual Learning for Image Recognition. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778.

48. Parkhi, O.M.; Vedaldi, A.; Zisserman, A. Deep Face Recognition. In Proceedings of the BMVC 2015, Swansea, UK, 7–10 September 2015; Volume 1, p. 6.

49. Liang, L.; Lin, L.; Jin, L.; Xie, D.; Li, M. SCUT-FBP5500: A diverse benchmark dataset for multi-paradigm facial beauty prediction. In Proceedings of the 2018 24th International Conference on Pattern Recognition (ICPR), Beijing, China, 20–24 August 2018; pp. 1598–1603.

50. Cao, K.; Choi, K.n.; Jung, H.; Duan, L. Deep Learning for Facial Beauty Prediction. *Information* **2020**, *11*, 391. [CrossRef]

51. Xu, J.; Jin, L.; Liang, L.; Feng, Z.; Xie, D.; Mao, H. Facial attractiveness prediction using psychologically inspired convolutional neural network (PI-CNN). In Proceedings of the 2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), New Orleans, LA, USA, 5–9 March 2017; pp. 1657–1661.

52. Fan, Y.Y.; Liu, S.; Li, B.; Guo, Z.; Samal, A.; Wan, J.; Li, S.Z. Label distribution-based facial attractiveness computation by deep residual learning. *IEEE Trans. Multimed.* **2017**, *20*, 2196–2208. [CrossRef]

53. Lin, L.; Liang, L.; Jin, L.; Chen, W. Attribute-Aware Convolutional Neural Networks for Facial Beauty Prediction. In Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI), Macao, China,10–16 August 2019; pp. 847–853.

54. Lin, L.; Liang, L.; Jin, L. Regression guided by relative ranking using convolutional neural network (R3CNN) for facial beauty prediction. *IEEE Trans. Affect. Comput.* **2019**, *13*, 122–134. [CrossRef]

55. Langlois, J.H.; Kalakanis, L.; Rubenstein, A.J.; Larson, A.; Hallam, M.; Smoot, M. Maxims or myths of beauty? A meta-analytic and theoretical review. *Psychol. Bull.* **2000**, *126*, 390–423. [CrossRef]

56. Cunningham, M.R.; Roberts, A.R.; Barbee, A.P.; Druen, P.B.; Wu, C.H. "Their ideas of beauty are, on the whole, the same as ours": Consistency and variability in the cross-cultural perception of female physical attractiveness. *J. Personal. Soc. Psychol.* **1995**, *68*, 261. [CrossRef]

57. Wald, C. Beauty: 4 Big questions. *Nature* **2015**, *526*, S17. [CrossRef]

58. Wald, C. Neuroscience: The aesthetic brain. *Nature* **2015**, *526*, S2–S3. [CrossRef]