




What's in a Face? Metric Learning for Face Characterization

O. Sendik¹  D. Lischinski²  D. Cohen-Or¹ 

¹Tel Aviv University, Israel

²The Hebrew University of Jerusalem, Israel



Figure 1: We analyze faces of different individuals to identify which facial parts constitute their most characteristic features. The process of characterization also enables us to select a most characteristic portrait of an individual, out of a set. By leveraging these characteristic portraits, we can synthesize an effective facial hybrid, which fuses together the characteristic facial parts of two different individuals.

Abstract

We present a method for determining which facial parts (mouth, nose, etc.) best characterize an individual, given a set of that individual's portraits. We introduce a novel distinctiveness analysis of a set of portraits, which leverages the deep features extracted by a pre-trained face recognition CNN and a hair segmentation FCN, in the context of a weakly supervised metric learning scheme. Our analysis enables the generation of a polarized class activation map (PCAM) for an individual's portrait via a transformation that localizes and amplifies the discriminative regions of the deep feature maps extracted by the aforementioned networks. A user study that we conducted shows that there is a surprisingly good agreement between the face parts that users indicate as characteristic and the face parts automatically selected by our method. We demonstrate a few applications of our method, including determining the most and the least representative portraits among a set of portraits of an individual, and the creation of facial hybrids: portraits that combine the characteristic recognizable facial features of two individuals. Our face characterization analysis is also effective for ranking portraits in order to find an individual's look-alikes (Doppelgängers).

Keywords: facial hybrids, face recognition, feature polarization, neural networks

CCS Concepts

• **Computing methodologies** → **Neural networks; Image processing;**

1. Introduction

Face recognition is an intricate cognitive process that humans excel at, with entire areas in the human brain dedicated to this task. Studies show that face recognition can take place in a holistic fash-

ion, from just a quick glance [TF93]; but the perceived identity of a face is also strongly affected by a more cognitive processing of specific facial features, such as lip thickness, eye shape, etc. [AY16]. This is particularly true when the faces are well familiar to the observer [JE09].

In this paper, we address the intriguing task of *face characterization*: identifying which facial parts of a particular individual constitute his/her most characteristic, distinctive and recognizable features. To the best of our knowledge, so far this task has only been addressed using classical methods, e.g., [ZLO10], while in the current work we seek to identify such parts by using deep neural networks, in conjunction with weakly supervised metric learning, to analyze a set of portraits of an individual of interest.

In recent years it has become apparent that deep neural networks are extremely successful in performing a wide variety of vision tasks, and, in particular, have surpassed human performance in face recognition [TYRW14]. Furthermore, recent studies in cognitive neuroscience [AY17] show that there is a strong correlation between facial features that human observers find perceptually sensitive and the features that CNNs rely on for the task of face recognition. These findings suggest that by tapping into the hidden layers, and making use of the deep features extracted by networks trained for face recognition, it is possible to determine the characteristic facial parts of an individual.

Specifically, one can leverage a network trained for face recognition, to extract Class Activation Maps (CAMs), using the technique proposed by Zhou et al. [ZKL*16]. CAMs highlight the image regions which most strongly contributed to the recognition task; however, they are not stable in the sense that they do not provide sufficiently consistent and localized activations. Moreover, Zhou et al. demonstrate that CAMs may be used for segmenting a recognized object in its entirety, since they tend to indicate most of the object as active. Thus, they are not well suited for our task of localizing only certain parts of the face.

To overcome these difficulties, we affix the network with an additional *distinctiveness* analysis layer that helps us detect and better localize the characteristic facial regions of individuals. This layer is trained in a weakly supervised manner, using two sets of portraits. In effect, this new layer localizes and amplifies the discriminative parts of the feature maps of analyzed portraits in the first set, when they are considered against the portraits in the second set, while attenuating the other regions. Since the resulting modified activation maps typically exhibit a polar nature, in the sense that face parts are either attenuated or not, we name them Polarized CAMs (PCAMs).

We quantitatively compare PCAMs with CAMs by defining measures which show that CAMs are not effective for our purpose, since they often indicate the same facial regions as active for both sets of portraits, and the active regions are not consistent across different images of the same individual. We also report the results of a user study that we conducted, showing that the facial parts selected by our method largely agree with selections made by the participants of our study.

We demonstrate a number of applications for which our face characterization analysis is useful.

By analyzing a set of portraits of a certain individual versus a set consisting of a mixture of many other individuals, we are able to identify the most and the least representative portraits of that individual.

Another fascinating application is the creation of *facial hybrids*.

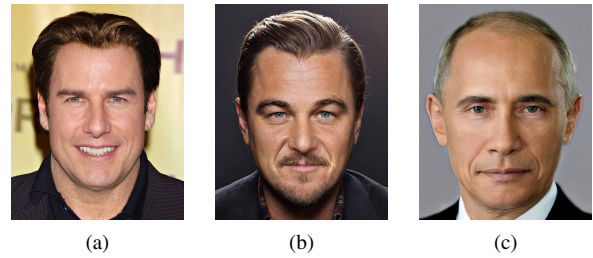


Figure 2: Three examples of GesichterMix's facial hybrids: (a) Tom Cruise and John Travolta, (b) Sean Penn and Leonardo Di-Caprio, and (c) Vladimir Putin and Barack Obama.

There are artists, e.g., the German Instagram artist GesichterMix [Ges17] and the Norwegian artist ThatNordicGuy [Tha18], who create striking facial hybrids of various pairs of celebrities by meticulously combining their most characteristic and distinctive face parts into a single portrait (see Figure 2 for a few examples). Inspired by their work, we employ our distinctiveness analysis for automatically creating such hybrids. More specifically, by analyzing two sets of celebrity portraits we are able to determine, given two specific portraits, which regions from each portrait should be fused together to form an effective facial hybrid.

Doppelgänger is a term from the German language referring to a look-alike or double of a living person (occasionally with some interesting paranormal connotations). The Doppelgänger Week, which occurs during the first week of February, involves searching for your own look-alike and sharing it. However, searching for one's Doppelgänger is not a simple task to automate, since one must take into account the variations in pose, illumination and various in-the-wild differences between portraits, which must be overcome. We show that face characterization analysis of an individual also improves the ability to find his/her Doppelgängers. Specifically, given a query image of an individual, whom we have analyzed using our method, we use the most distinctive facial features in order to define a more effective metric for comparing identities. Given a set of portraits of other individuals, our metric is then used to rank them according to their similarity to the queried individual.

The rest of this paper is organized as follows: We present background and related work in Section 2. In Sections 3 and 4 we present our algorithm with emphasis on our main contribution. Experimental results of our method are reported in Section 5; Section 6 demonstrates the aforementioned applications of our face characterization. Finally, Section 7 presents conclusions and discusses future research directions.

2. Related Work

Machine analysis and recognition of faces has been an active research area since the early 1990s [CWS95], with face detection probably being the most studied problem [HL01]. With the recent rise of neural networks, many face analysis tasks that were previously solved in various ad hoc manners are now solved within the same framework. Benchmarks of inference tasks such as face recognition, verification, expression and action analysis, facial landmarks, face-based age, gender and race estimation, and more,

have been increasingly pushed to the limits (and often exceeded those) of human performance [DTM14, TYRW14].

With the increase in data availability, algorithms which leverage the existence of sets of portraits have emerged, ranging from reconstructing 3D face models from large unstructured photo collections [KSS11], to computing optical flow between pairs of portraits [KSS12], and even reconstruction of a controllable model capturing a persona from a large photo collection [SSKS15]. Finally, a fair amount of effort has been directed towards synthesis challenges. Tasks such as face synthesis, face swapping, [NMaTa*17, BKD*08] and even turning static single still portraits into vivid facial animations [AECOKC17] are currently active areas of research.

In this work, our goal is to analyze a set of portraits of a given individual in order to automatically identify his/her most characteristic facial parts. Psychological studies [MLGM02, TAAMA11] support the hypothesis that the human visual system analyzes faces by quickly building holistic representations to extract useful second-order information provided by the variation between the faces of different individuals. In their review, Maurer et al. [MLGM02] conclude that there are three stages of processing associated with face recognition. The first is face detection (based on first-order information). The second is a holistic processing (the integration of facial features following detection), and finally, the third is face discrimination (based on second-order information extracted from the holistic representation).

From this perspective, the recognition process first attempts to recognize a face at once. A second attempt, relying only on individual face parts, kicks in only after the first, holistic attempt has failed. The ability to recognize faces using only parts of a face, serves as motivation to formalizing the characterization problem. Further motivation for our work is provided by a previous study by Abudarham and Yovel [AY17], which compared CNNs and humans showing that there is a strong correlation between facial features that humans consider perceptually sensitive and the features that CNNs rely on when performing face recognition.

3. Face Characterization: Overview

The goal of this work is to determine which face parts characterize an individual. The main insight of our approach is that we can infer this via a weakly supervised analysis of the activations of a pretrained neural network. This question is difficult to answer from just one pair of images, but, as we shall see, it can be answered robustly given two sets of portraits, one for the individual being characterized and another for arbitrary faces other than those of the individual. We denote these two sets as \mathcal{P}_{ind} and \mathcal{P}_{arb} respectively.

We build upon the method of Zhou et al. [ZKL*16] who proposed a procedure for generating Class Activation Maps (CAMs) that indicate the discriminative image regions used by a CNN to identify specific categories. However, as we demonstrate in Section 4, CAMs are not effective for our purpose. Hence, we apply metric learning with an implicit requirement of polarity between activation maps of different individuals. Informally, we seek a metric that would minimize the distance between the deep features of the individual of interest, while maximizing their distance to the features of

the other individuals. Using the learned metric, we are able to generate Polarized CAMs (PCAMs) from the transformed deep feature maps.

A drawback of employing off-the-shelf face recognition CNNs for face characterization is that they are insensitive to hair, although hair is well known to be a rather characteristic trait [BHK97, SBOR05]. In order to take hair into account, we train a Fully Convolutional network (FCN) to perform hair segmentation, and apply metric learning on its features as well. Both of the learned metrics (for the internal face parts and for hair) are then used to perform the face characterization.

Finally, having obtained the activation polarizing metrics for a particular individual, we analyze the entire set of his/her portraits to select a single, most representative, portrait. We then determine the individual's most characteristic face parts by comparing the PCAMs of the representative portrait to those of a set of portraits of other arbitrary individuals from \mathcal{P}_{arb} . The entire process is depicted in Figure 5, and explained in more detail in the next section.

4. Characterization Analysis

In order to determine the characteristic facial parts we use deep features extracted by a pre-trained face recognition CNN, in our case VGG-Face [PVZ15]. Our analysis requires the portrait sets \mathcal{P}_{ind} and \mathcal{P}_{arb} , to first be aligned to the same canonical pose. We use the method of Zhu and Ramanan [ZR12] for joint face detection, pose estimation, and landmark estimation. Once the landmarks are extracted, we align the portrait sets with an affine transformation between the landmarks detected for each portrait and those of our target canonical face. We denote the cropped and aligned portrait sets with \mathcal{P}_{ind}^A and \mathcal{P}_{arb}^A . We denote a single aligned portrait from these sets by P_{ind}^A and P_{arb}^A .

4.1. Class Activation Maps

Zhou et al. [ZKL*16] proposed a procedure for generating Class Activation Maps (CAMs) that indicate the discriminative image regions used by a CNN to identify specific categories. Their method consists of replacing the penultimate fully connected CNN layers with a Global Average Pooling (GAP) layer, followed by a fully-connected softmax layer, which is retrained for the classes at hand.

Formally, let us denote the CNN's feature map before the GAP layer by $f \in \mathbb{R}^{Z \times W \times H}$, where Z is the number of feature channels, and W and H are the width and height of the map. The GAP layer computes a global spatial average for each of the Z channels:

$$F_z = \frac{1}{WH} \sum_{w,h} f_{z,w,h} \quad (1)$$

The resulting vector of global averages $F = [F_1, \dots, F_Z]$ is fed into the final fully connected layer, trained to infer the appropriate class c . In other words, the final layer's output, $\text{softmax}(\mathbf{A}F)$, is a vector of class probabilities, where \mathbf{A} is a matrix $\mathbf{A} \in \mathbb{R}^{C \times Z}$. The CAM for class c is a two dimensional map depicting the active regions,

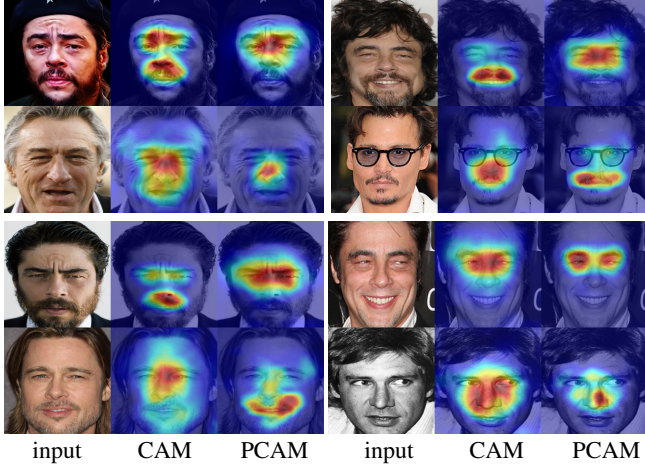


Figure 3: Class Activation Maps (middle columns) vs. our Polarized Class Activation Maps (right columns) for extracting activation maps, applied on portraits. The CAMs seem to spread to wide regions serving the face recognition process while the PCAMs select only the characteristic face part.

computed as a linear combination of the feature channels, each weighted by $\mathbf{A}_{c,z}$, denoting the (c, z) entry of \mathbf{A} :

$$\text{CAM}_c(w, h) = \sum_z \mathbf{A}_{c,z} f_{z,w,h}. \quad (2)$$

4.2. Polarized Class Activation Maps

Figure 3 shows the CAMs computed as described above for the sets \mathcal{P}_{ind}^A (portraits of Benicio Del Toro) and \mathcal{P}_{arb}^A (portraits of other random individuals). We configured the final classification layer to classify between these two classes. The figure shows four pairs of portraits (each pair consists of one portrait from \mathcal{P}_{ind}^A and another portrait from \mathcal{P}_{arb}^A). The corresponding CAMs appear next to the portraits in the 2nd and 5th columns.

These four pairs of portraits demonstrate that one cannot rely on CAMs to indicate which face parts of an individual are characteristic by examining the spatial distribution of the activation values or by comparing their magnitudes. First, the active regions are not consistent across different portraits of the same individual: in some of his portraits Del Toro's eye region is highly activated, but in others it is not. Second, there are significant overlaps between the CAMs of Del Toro's portraits and those of the other individuals. Thus, Del Toro's characteristic face parts cannot be easily determined by comparing his CAM values to those of others.

We believe that this unstable and inconsistent behavior of CAMs is a consequence of training the fully connected classification layer using a standard softmax loss, which does not encourage the activations of the two portraits to be mutually exclusive. Rather, the softmax loss achieves facial recognition by taking into account the combined appearance of all of the face parts. Hence, CAMs of faces

are likely to overlap. Our objective, on the other hand, is to minimize the overlap between the two activation maps, so that an active region in \mathcal{P}_{ind}^A is unlikely to be active in \mathcal{P}_{arb}^A . At the same time, however, we wish to preserve the ability to discriminate between the two portraits, correctly determining whether a portrait belongs to \mathcal{P}_{ind}^A or \mathcal{P}_{arb}^A . In other words, we cannot simply attenuate the activations of portraits from \mathcal{P}_{arb}^A completely. Our goal is then to select the most distinctive activations for the portraits from \mathcal{P}_{ind}^A while attenuating the less distinctive ones and letting them remain active for portraits from \mathcal{P}_{arb}^A .

To achieve this goal, we propose a scheme that applies Metric Learning to the deep feature maps f . Specifically, by learning a Mahalanobis distance between pairs of feature maps (f^i, f^j) ,

$$d_{\mathbf{M}}^2(f^i, f^j) = \left\| \mathbf{M}f^i - \mathbf{M}f^j \right\|_2^2, \quad (3)$$

where $\mathbf{M} \in \mathbb{R}^{2 \times Z \times W \times H}$, we construct a linear embedding of the deep feature maps, such that the intra-class distances are minimized, while the inter-class distances are maximized. Note that \mathbf{M} and f are reshaped into a matrix and a vector, such that their multiplication is well defined.

Using \mathbf{M} we redefine the GAP values from Eq. (1) as:

$$F_{c,z} = \frac{1}{WH} \sum_{w,h} \mathbf{M}_{c,z,w,h} \odot f_{z,w,h}, \quad (4)$$

where \odot denotes elementwise multiplication. In other words, we obtain two global average values for each channel of the feature map, representing the relative contribution of this channel to the classification of the portrait as belonging to \mathcal{P}_{ind}^A or \mathcal{P}_{arb}^A .

Finally, given an image, its PCAM is obtained by elementwise multiplication of the feature channels with the learned weights, and their weighted summation using the modified GAP values:

$$\text{PCAM}_c(w, h) = \sum_z F_{c,z} (\mathbf{M}_{c,z,w,h} \odot f_{z,w,h}). \quad (5)$$

Figure 3 shows the PCAMs, obtained as described above, in the 3rd and 6th columns. Note that for each pair of portraits, the overlap between the active regions in the corresponding PCAMs has been reduced. Moreover, Del Toro's eye region is now consistently indicated as active, and for each of the portraits from \mathcal{P}_{arb}^A , the PCAM activations over the eye regions are attenuated, which helps determine that the eye region is characteristic for Del Toro. In contrast, Del Toro's mouth region is consistently attenuated in his PCAMs, indicating that it is not one of his characteristic parts. We quantitatively compare the overlap and consistency of CAMs vs. PCAMs on a larger set of individuals in Section 5.

4.3. Metric Learning

We employ weakly-supervised metric learning to obtain \mathbf{M} . Specifically, given two sets of feature maps (for \mathcal{P}_{ind}^A and for \mathcal{P}_{arb}^A), we learn \mathbf{M} by maximizing the classification margin via a hinge loss [SPVZ13]:

$$(\mathbf{M}, b) = \arg \min_{\mathbf{M}, b} \sum_{i,j} \max \left\{ r - \rho_{ij} \left(b - d_{\mathbf{M}}^2(f^i, f^j) \right), 0 \right\}, \quad (6)$$

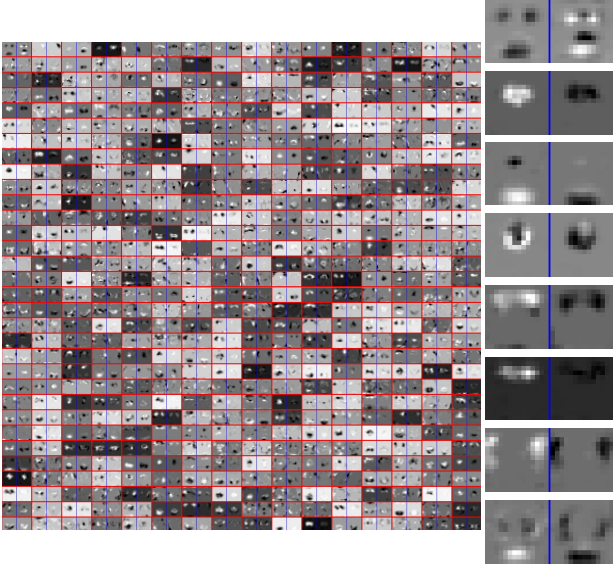


Figure 4: Results of our weakly supervised polarizing metric learning. Weights are grouped in pairs in order to depict the polar nature of the metric. In the right part, we zoom into specific pairs, showing the polar nature of the metric. Structured patterns, resembling face parts tend to emerge.

where b is the mean distance of all face features, and r specifies the classification margin with respect to b , such that the classification margin between positive and negative feature pairs is $2r$. A pair is positive when f^i and f^j are from portraits of the same set, and negative otherwise. The labels $\rho_{ij} = 1$ and $\rho_{ij} = -1$ denote the positive and negative pairs, respectively.

Eq. (6) is solved using stochastic gradient descent, where at each iteration t a positive or a negative pair is randomly drawn from the two training sets. The updates for \mathbf{M} and b are given by

$$\mathbf{M}_{t+1} = \begin{cases} \mathbf{M}_t, \rho_{ij} (b_t - (f^i - f^j)^T \mathbf{M}_t^T \mathbf{M}_t (f^i - f^j)) > r \\ \mathbf{M}_t - \gamma_M \rho_{ij} \mathbf{M}_t (f^i - f^j)(f^i - f^j)^T, \text{ otherwise} \end{cases} \quad (7)$$

$$b_{t+1} = \begin{cases} b_t, \rho_{ij} (b_t - (f^i - f^j)^T \mathbf{M}_t^T \mathbf{M}_t (f^i - f^j)) > r \\ b_t + \gamma_b \rho_{ij} b_t, \text{ otherwise} \end{cases} \quad (8)$$

where γ_M and γ_b are hyper parameters controlling the learning rate.

We observe that by optimizing the max margin loss in Eq. (6), we are able to get a more consistent set of activations which tend to overlap less. In Figure 4 we visualize the two rows of the learned metric \mathbf{M} , denoted as \mathbf{M}_{ind} and \mathbf{M}_{arb} , by reshaping each row to 512 weight maps, each corresponding to one of the 512 feature channels of f . We show the entire learned metric in the left part of the figure, and zoom on the right into eight pairs of weight maps. It is easy to see that structure appears in the learned metric, and that the pairs of weight maps tend to have inverse polarity. In other words, a region attenuated in one map is typically amplified in its counterpart.

4.4. Hair Distinctiveness

It is well known that hair is a facial feature that frequently changes in different portraits and that (partly) due to this reason face classification tasks (e.g., face detection or recognition) have been trained after cropping out the hair [TKB12]. VGG-Face [SZ14], for example, requires aligning the input portrait to a canonical pose which excludes hair. In order to enable our method to determine whether an individual's hair is characteristic, we seek to train a NN on a task which imposes activations on hair. We achieve this by training a Fully Convolutional Network (FCN) on the task of hair segmentation, following the architecture of Long et al. [LSD15]. We trained our network on the Figaro-1K dataset [MSLB18] which consists of unconstrained images containing various hair textures and styles, with manually labeled binary masks indicating hair pixels.

Armed with the pretrained hair segmentation network, we apply the exact same metric learning procedure as described in the previous sections, on the hair-sensitive feature. We denote the hair-sensitive metric by \mathbf{M}^H and its first and second rows by \mathbf{M}_{ind}^H and \mathbf{M}_{arb}^H , respectively.

4.5. Face Part Selection

Given \mathbf{M} and \mathbf{M}^H , we can finally determine an individual's characteristic face parts. We first select the most representative portrait of the individual from \mathcal{P}_{ind}^A by searching for the portrait whose deep feature map is altered the least after weighing it by the learned metrics. Put formally,

$$k_{rep} = \arg \min_k \left\{ \left\| \mathbf{M}_{ind} f_{ind}^k - f_{ind}^k \right\| + \nu \left\| \mathbf{M}_{ind}^H f_{ind}^{H,k} - f_{ind}^{H,k} \right\| \right\}. \quad (9)$$

We denote the features provided by feeding a portrait from \mathcal{P}_{ind}^A into VGG-Face and our hair segmentation FCN by f_{ind}^k and $f_{ind}^{H,k}$ respectively. Similarly, by searching for the maximum instead of a minimum, we may determine the individual's least representative portrait. The hyper parameter ν enables balancing between the magnitudes of the face and hair activation values.

Once we select the most representative aligned portrait, which we denote by $P_{krep}^A \in \mathcal{P}_{ind}^A$, we determine its characteristic face parts by comparing the PCAM values inside each face part to those of the portraits of other individuals in \mathcal{P}_{arb}^A . In other words, a face part of P_{krep}^A is determined as characteristic if its total PCAM activation is stronger than that in the majority of portraits from \mathcal{P}_{arb}^A . Hence, each comparison requires segmenting the face part in order to sum the activations across the containing segment. All of the face parts other than the hair are segmented by using facial landmarks, with the method proposed by [ZR12], while the hair is segmented using our own hair segmentation FCN from Section 4.4. These face parts are shown in yellow in the inset. The entire process is depicted in Figure 5.



5. Results and Evaluation

Our entire pipeline was implemented in Matlab with the process of Metric Learning requiring about 15 minutes per individual (for

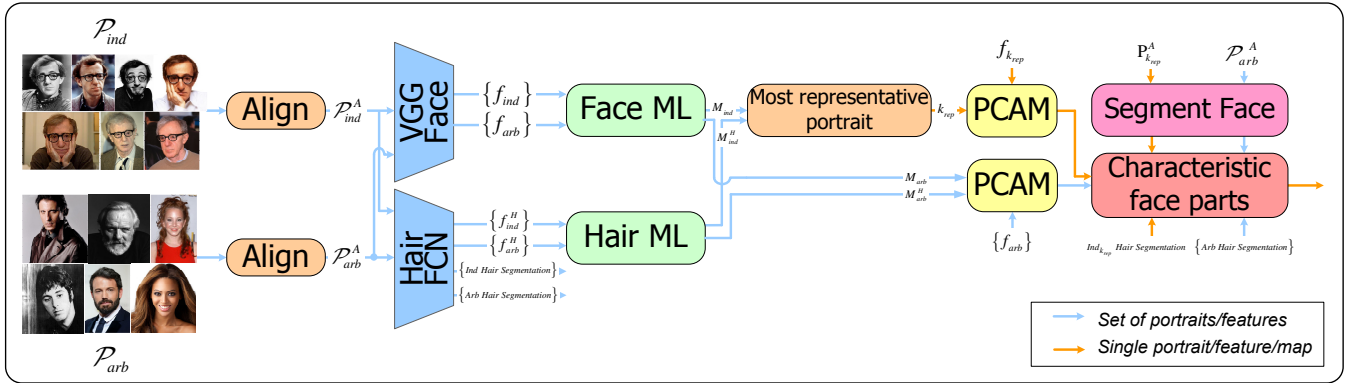


Figure 5: Our face characterization pipeline begins by aligning two sets of portraits, followed by feature extraction using a face recognition CNN and a hair segmentation FCN. Metric learning is applied on the features, enabling determining the most representative portrait of the analyzed individual. Given the latter portrait, we compare its PCAM to the PCAMs of portraits in \mathcal{P}_{arb}^A . We segment each portrait and detect the individual's characteristic face parts by comparing the total PCAM activation across each segment.

acquiring both \mathbf{M} and \mathbf{M}^H). Note that using a CAM involves training only a single penultimate layer, which is faster and converges within less than a minute. However, at test time, the run-time of both methods is dominated by the forward pass of the entire CNN.

We summarize certain technical details and choices which are crucial for reproducing the results that follow, in the supplementary material. We plan to make our full implementation and data available upon publication of this work.

5.1. PCAM vs. CAM face characterization

In Section 4.2, we argued that CAMs are ill-suited for face characterization due to their tendency to overlap and their inconsistency. In order to quantify the measure of overlap, we normalize the PCAM and CAM activation maps and binarize them by thresholding their values at 0.1. As a measure of *overlap*, we calculate the Jaccard index, also known as the Intersection over Union, $J(A, B) = 100 \frac{|A \cap B|}{|A \cup B|}$ for random pairs of portraits (A, B) , where $A \in \mathcal{P}_{ind}$ and $B \in \mathcal{P}_{arb}$.

In Figure 6, we plot the distribution of Jaccard indices for four individuals under characterization. In the upper left subplot of each quartile, we show that the Jaccard indices for PCAM pairs are typically smaller than those of CAM pairs. We can also see that the distribution modes (denoted by μ) are significantly smaller for PCAM pairs, indicating their lower tendency to overlap.

In order to quantify the increased *consistency* of face characterization of PCAMs vs. CAM, we assign each one of the possible selections a number between 1 and 16. There are 16 options since for this comparison, we use 4 parts (nose, mouth, eyes and eyebrows). Note that in this analysis we do not consider hair, hence reproducing the exact same method as proposed by [ZKL*16]. For each pair of portraits, one from \mathcal{P}_{ind} and the other from \mathcal{P}_{arb} , we record the number of the selection produced using CAMs and using PCAMs. The histograms of these selections, shown in the top right subplot of each quartile in Figure 6, show that PCAMs have a higher probability of yielding the same selection. For example,

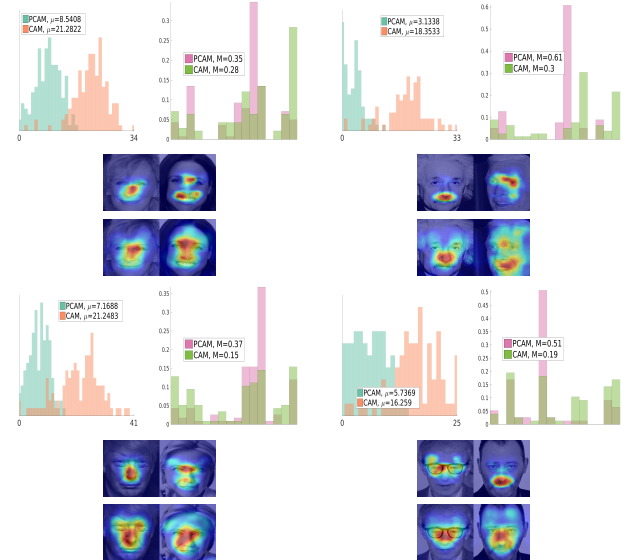


Figure 6: PCAM vs. CAM: The top left subplots within each quartile depict the distribution of Jaccard indices, which for PCAM (green) exhibit lower values than those of CAM (red), indicating a smaller tendency of activations to overlap. The top right subplots present the histogram of face part selections, which demonstrate that PCAM produces a distribution with a higher mode. The bottom subplots show portraits with PCAM (upper heat map pair) and CAM (bottom heat map pair) overlays respectively. Again, it is apparent that PCAMs are more concentrated and overlap less, compared with CAMs.

for Woody Allen, the PCAM and CAM modes are found to consist of 51% of the selections and 19% of the selections respectively. This difference demonstrates the significantly more stable nature of PCAM based selections.

We also show an example of the PCAM vs. CAM based activations in the bottom subplot of each quartile. In all four examples, CAM activations are more spread, while PCAM activations are more concentrated. More such results are presented in the supplementary material.

5.2. User Study

In order to quantitatively evaluate our PCAM-based part selection strategy, we conducted a user study where users were requested to indicate the characteristic face parts for different individuals. For each individual, the participants were shown a gallery of six randomly chosen portraits of that individual, and were asked to indicate all of the parts that they judged to be characteristic, choosing from Hair, Eyes/Eyeglasses, Eyebrows, Nose, and Mouth. The participants were asked to indicate as many parts as they felt appropriate, or none, for each individual. We collected a total of 4000 such characterizations for 20 individuals (10 males and 10 females), from 200 participants. Volunteers were recruited through social media posts and were able to answer the study only on a desktop (and not on a mobile phone).

The user study participants are shown six portraits per individual. We selected 10 out of the 20 individuals randomly from the Face-Scrub dataset [NW14], while the remaining 10 individuals were chosen from the WebCaricature dataset [HLS*17]. We assume that the prior familiarity of the study participants with the celebrities compensates for the small size of the gallery of presented images for those individuals.

For each individual in our study, we record and report two types of results, corresponding to joint and independent face part selection. The first type assumes that users select the most characteristic face parts in a joint, inter-dependent manner. Since users are requested to select the characteristic face parts out of 5 options (Hair, Eyes/Eyeglasses, Eyebrows, Nose and Mouth), there are 2^5 possible selections, and we record the most frequent one. The second type assumes that users select each face part independently, disregarding the other face part selections. In this case, we record the face parts that were indicated as characteristic by the majority of the users (more than 50%). In either case, the result for each individual may be represented as a binary vector of length 5.

In order to compare the selections predicted by our method to those made by the users, we calculate and report the Hamming distances between the selection vector produced by our method and the two types of vectors representing the users' selections described above, for each individual. Note that the Hamming distance measures the disagreement between our method and the selections made by humans, accounting both for parts that were indicated as characteristic and for those that were not.

As another measure, we also fit a probability mass function to the selections of the users for each of the 20 individuals. We use it to measure the likelihood of our PCAM-based part selections and report the ratio between the resulting likelihood and a naïve guess.

We model the joint face part selection with a generalized Bernoulli distribution (also known as Multinoulli) with a random variable ranging between 1 and 2^5 . For the independent face part selection, we fit an independent Bernoulli distribution for each face part. We note that without any prior knowledge, the likelihood of a random guess is $1/2^5 = 0.03125$.

Table 1 reports the selection made by our method for each of the 20 individuals, along with the joint and the independent selections made by the users, accompanied by the Hamming distances (denoted by Δ) and likelihood ratios (denoted by ρ) mentioned above. We denote the face parts by H, E, B, N and M, standing for Hair, Eyes, Eyebrows, Nose and Mouth, respectively. The results indicate a strong agreement between our method and the study participants. Most of the Hamming distances between the selections of our method and the most common joint and independent user selections are either 0 (full agreement) or 0.2 (agreement on 4 out of 5 parts). We stress that this holds across male and female faces. In addition, for 17 out of the 20 individuals, $\rho > 1$, meaning that we improved upon a naïve guess, while for 11 of the individuals, we improve upon a naïve guess by a factor of three or more.

Averages of the Hamming distances and the likelihood ratios are reported in Table 5.2. Note that on average, our method improves upon a random guess almost by a factor of 4.

It is interesting to examine the cases for which $\rho < 1$. This seems to occur when there's a good consensus among the study participants on a selection that differs from our method's. For example, 91% of the users agree that Melina Kanakaredes's (bottom left in Table 1) eyebrows are not characteristic, in contrast to our algorithm.

In general, we observe that users tend to select fewer face parts as characteristic, compared to our method. The overall percent of users who selected more than 2 characteristic face parts for an individual, stands on 14%, while our algorithm selected 3 or more characteristic parts for 8 out of 20 (40%) individuals. This is a significant difference, which obviously increases the Hamming distances and reduces the likelihoods of our method's selections. However, in all but one case, our method's selection is a strict superset of the common selections of the users.

Another noticeable difference between human selections and our algorithm is that, for the 20 individuals in the study, eyebrows were *never* chosen as characteristic by the majority of the users. It should be noted that eyebrows were also the least selected face part by our method (in 6 out of 20 individuals).

For each of the 20 individuals, we also show the part-wise selection consensus across users. For each face part, we show the percentage of the majority vote. Agreement on a characteristic face part is shown as a magenta bar, while agreement on a non characteristic one is shown as a cyan bar. For example, 95 percent of the users believed that Woody Allen's (top left) eyebrows are not characteristic and 90 percent indicated that Bob Marley's (top right) hair is.

Through these percentages we see that users often disagree about what characterizes face parts. However, in most cases, the selection of our algorithm agrees with the selection of the majority of users.

Individuals from WebCaricature	Most representative					
	PCAM	E	N,M	E,N	H,B,M	H,M
	Joint	$E(\Delta=0, \rho=9.19)$	$M(\Delta=0.2, \rho=4.59)$	$E(\Delta=0.2, \rho=4.92)$	$H,M(\Delta=0.2, \rho=1.29)$	$H(\Delta=0.2, \rho=4.77)$
	Independent	$E(\Delta=0, \rho=8.54)$	$M(\Delta=0.2, \rho=4.71)$	$E(\Delta=0.2, \rho=4.35)$	$H,M(\Delta=0.2, \rho=1.92)$	$H(\Delta=0.2, \rho=5.96)$
	Selection consensus					
Individuals from FaceScrub	Most representative					
	PCAM	E,B	H,E,B,M	H,E,B,M	H,E,B	H,N,M
	Joint	$E(\Delta=0.2, \rho=4.50)$	$H,M(\Delta=0.4, \rho=1.67)$	$H(\Delta=0.6, \rho=0.35)$	$H,E(\Delta=0.2, \rho=0.18)$	$M(\Delta=0.4, \rho=1.11)$
	Independent	$E(\Delta=0.2, \rho=4.76)$	$H,M(\Delta=0.4, \rho=1.09)$	$H,M(\Delta=0.4, \rho=0.85)$	$H,E(\Delta=0.2, \rho=0.29)$	$N,M(\Delta=0.2, \rho=1.44)$
	Selection consensus					
Individuals from WebCaricature	Most representative					
	PCAM	N,M	E	N,M	N,M	E
	Joint	$N,M(\Delta=0, \rho=8.70)$	$E(\Delta=0, \rho=6.24)$	$M(\Delta=0.2, \rho=4.45)$	$M(\Delta=0.2, \rho=2.02)$	$E(\Delta=0, \rho=11.40)$
	Independent	$N,M(\Delta=0, \rho=10.21)$	$E(\Delta=0, \rho=4.33)$	$M(\Delta=0.2, \rho=4.29)$	$M(\Delta=0.2, \rho=2.44)$	$E(\Delta=0, \rho=8.42)$
	Selection consensus					
Individuals from FaceScrub	Most representative					
	PCAM	H,E,B	E,N	N	H,E,M	H,E,M
	Joint	$H(\Delta=0.4, \rho=0.72)$	$N(\Delta=0.2, \rho=3.14)$	$M(\Delta=0.4, \rho=3.16)$	$H(\Delta=0.4, \rho=2.07)$	$E(\Delta=0.4, \rho=3.36)$
	Independent	$H(\Delta=0.4, \rho=0.43)$	$N(\Delta=0.2, \rho=3.43)$	$M(\Delta=0.4, \rho=2.16)$	$H(\Delta=0.4, \rho=2.75)$	$E(\Delta=0.4, \rho=4.35)$
	Selection consensus					

Table 1: For each of the 20 individuals in the user study, we show the most representative portrait that was used by our method for selecting characteristic face parts (indicated under each portrait), as well as the joint and independent user selections. In parentheses we report the Hamming distances Δ from our method's selection, as well as the likelihood ratio ρ . Next we show the consensus of selections across users per face part, with selected parts indicated in magenta.

Table 2: The averages of the joint and independent scores of our user study.	Averages		
	WebCaricature	FaceScrub	All
Joint	$\Delta=0.26$ $\rho=3.26$	$\Delta=0.22$ $\rho=4.53$	$\Delta=0.24$ $\rho=3.89$
Independent	$\Delta=0.22$ $\rho=3.39$	$\Delta=0.22$ $\rho=4.28$	$\Delta=0.22$ $\rho=3.84$

6. Applications

6.1. Portrait ranking

Having trained the metrics M and M^H , we leverage them for determining the most and least representative portraits of an individual, by applying Eq. (9). Recall that Eq. (9) minimizes the norm of the difference between the features, before and after the transformation with the metric. Since the metric alters characteristic features, min-

imizing Eq. (9) amounts to searching for the portrait whose feature map is altered the least by the learnt metric, which we consider to be the most representative. Similarly, maximizing Eq. (9) amounts to searching for the least representative portrait.

In Figure 7, we show such portrait pairs for nine individuals, and suggest our explanations for the reason that the least representative portrait was selected as such. We identify six possible reasons, as mentioned in the caption.

While it is difficult to assess to what degree our approach indeed selects the most representative portrait in a set, it is instructive to examine the reasons for selecting the least representative portrait. In many cases, our method determines a portrait as least representative when it is partially occluded, was taken in an extreme pose or illumination, or even cases where the portrait is not a natural one (e.g., a pencil drawing). In several cases where the set of portraits

contained by mistake a portrait of another individual, our method has correctly identified the outlier as the least representative portrait. Thus, our approach may be helpful for distilling datasets. We present more such results in the supplementary material.



Figure 7: Most and least representative portraits for nine different individuals, selected from nine corresponding portrait sets using Eq. (9). In each pair, the most representative portrait appears on the left. We identify six recurring reasons for the least representative portrait selections as follows:

- (i) The portrait depicts the individual at a younger or older age
- (ii) The portrait depicts the individual in an extreme pose, illumination or colors
- (iii) The portrait is of another individual (outlier in the set)
- (iv) The face is partially occluded (beard, sunglasses, cigar)
- (v) Unexplained failures of our ranking algorithm (cases where we found the least representative portrait to resemble many of those in \mathcal{P}_{ind})
- (vi) Other

6.2. Synthesizing facial hybrids

Overview. Our face characterization method may be used for automatic creation of facial hybrids, such as those produced by the two artists *GesichterMix* and *ThatNordicGuy*. These hybrids are created by combining together recognizable facial features of two individuals. To ensure the recognizability of facial parts, they must be composed together, similarly to the process of creating facial composites by the police, rather than blended, as commonly done in image morphing [GDCV98].

Our method can be used to determine which facial parts of each of the two individuals should be selected for inclusion in the hybrid. In order to increase recognizability, the parts selected from each individual should be as characteristic as possible, with respect to the two individuals. Thus, rather than performing our face characterization on each of the two individuals separately (using the sets \mathcal{P}_{ind}

and \mathcal{P}_{arb}), we characterize the two individuals with respect to each other, using their two sets of portraits \mathcal{P}_{ind1} and \mathcal{P}_{ind2} .

Having performed the pairwise face characterization analysis, we automatically select a pair of aligned portraits $P_1^A \in \mathcal{P}_{ind1}^A$ and $P_2^A \in \mathcal{P}_{ind2}^A$, which are segmented as described in Section 4.5.

Next, we compute PCAMs for each of the two aligned portraits. Using the PCAMs, we determine for each portrait which of its segments should be included in the hybrid, and proceed to fuse them together using nonlinear spectral fusion [BMN*17]. Below we elaborate on the main steps.

Portrait selection. To create a visually plausible facial hybrid, the two source portraits should be compatible in terms of their global properties, such as pose and illumination. Rather than performing precise recovery and matching of these properties, we *automatically* select a pair of portraits where both the pose and the illumination are as frontal as possible. Needless to say, the automatic selection of portraits is optional to our approach; the user can choose to exercise artistic control and indicate which two portraits to use.

We compute the CNN features for three versions of each portrait in the two sets: the original (aligned) image P_i^A and two horizontally reflected versions of it. The two versions are created by cropping the left or the right half of the portrait and reflecting the remaining half horizontally, yielding two symmetric images denoted by P_i^L and P_i^R . We extract the deep features of these three images, which we denote by f^i , $f^{i,L}$ and $f^{i,R}$. We choose one portrait from each set, for which the maximal L^2 distance between the features of the original portrait and those of its two reflected versions is minimal. More formally, we sort the images within each set according to

$$\max_{r \in \{L,R\}} \|f^i - f^{i,r}\|^2, \quad (10)$$

and select the image with the smallest score. The benefit of selecting this most symmetric image by examining deep features, rather than using the image directly is that we disregard the hair and wrinkles. We also found that this method effectively filters out portraits with occluded facial regions. We demonstrate the effectiveness of this simple approach in the supplementary material.

Segmentation. Each of the two selected portraits is segmented into three types of components, (i) facial parts, (ii) face base and (iii) the background, as illustrated in the inset previously. The facial parts are the regions containing the eyes, eyebrows, nose, mouth and hair (shown in yellow). The corresponding segments are obtained from the detected facial landmarks and using our hair segmentation FCN. The face base, shown in purple, contains the rest of the face, and is segmented using the method of Nieuwenhuis and Cremers [NC13]. The background is defined as the remaining image regions (shown in green), belonging to neither the face parts, nor to the base.

Segment selection. For each face part, we sum the PCAM activation across the relevant segment in both portraits, P_1^A and P_2^A , and choose the one for which the total activation is higher. The base is chosen in the same manner, while the background is selected from the same portrait as the hair.

Figure 8 shows a few of our results. Note that both of the combined individuals (shown in the two bottom corners of each image) may



Figure 8: Results of our facial hybrid synthesis algorithm. The two small images within each sub-figure are the inputs. The large image within each sub-figure is our resulting facial hybrid.

be recognized in the resulting facial hybrid. Additional hybrids are included in the supplementary material.

6.3. Doppelgänger search

Face recognition approaches work best on aligned portraits, captured under controlled settings. Recognizing faces in-the-wild, which may exhibit extreme pose, illumination, or expression variations, is much more challenging. Thus, searching for an individual's look-alikes, or Doppelgängers, in a set of portraits captured in-the-wild, is a challenging task. Here, we demonstrate that by measuring face similarity using our learned metrics the ability to find Doppelgängers is consistently and significantly improved.

Schroff et al. [STKB11] develop a face-similarity measure that is largely invariant to differences in pose, illumination or expression. We do not explicitly attempt to tackle such portrait variations, but merely show that our metric improves robustness to such conditions implicitly. We note that Schroff's method relied on a relatively large dataset, the CMU Multi-PIE database, which includes images of 337 individuals in more than 2000 pose, lighting and illumination combinations, in order to achieve the desired robustness. Our method, on the other hand, learns a metric from a sets of only ~100 images, and does not explicitly target pose or illumination invariance.

To demonstrate the effectiveness of our approach for the task of finding Doppelgängers, we searched for Doppelgänger images for several celebrities on Google (using the search term “\$individual\$ DoppelGanger”, where the key \$individual\$ was replaced with the individual's name). For each target individual, we add the found Doppelgänger portraits (between 1 to 5) to a set of randomly selected 100 portraits of other individuals. We next sort the portraits in the resulting set according to their similarity to the representative portrait of the target individual. Ideally, such sorting should result in the “true” Doppelgänger images ranked at the top. Similarity was

measured using our learned metric:

$$\left\| \mathbf{M}_{ind}(f_{ind}^{k_{rep}} - f_{ind}) \right\| + \nu \left\| \mathbf{M}_{ind}^H(f_{ind}^{H,k_{rep}} - f_{ind}^H) \right\|, \quad (11)$$

where $f^{k_{rep}}$ and $f^{H,k_{rep}}$ are the features of the representative portrait of the individual for which we are searching for a Doppelgänger. As a performance baseline, we also sort the same set using a naive metric based on VGG-Face features:

$$\left\| f_{ind}^{k_{rep}} - f_{ind}^k \right\|. \quad (12)$$

Table 2 shows the results of this experiment on 12 celebrities. It may be seen that the ranking produced using our metric is consistently (and often significantly) higher for the “planted” Doppelgänger images, compared to the baseline metric. In fact, in 11 out of the 12 cases, one of the Doppelgänger images was ranked first using our metric, compared to only 3 out of 12 using the baseline.

A useful property of our metric is that the sorted distances exhibit a tendency to increase faster. We attribute this to our metric's explicit sensitivity to the characteristic face parts. In Figure 9 we plot the distances sorted by our metric (dashed) and the baseline metric (solid), for four individuals. The distances are normalized by the smallest distance in the set (of the portrait ranked first) in order to plot both graphs on the same scale. It is noticeable that when distances are computed using our metric, they increase faster.

7. Conclusions and limitations

We have presented a technique for characterizing an individual's face based on a small set of his/her portraits. The key component of our approach is a weakly supervised metric learning scheme that analyzes two sets of portraits. Our results reconfirm the connection between the deep features a network uses for face recognition and those used in human perception.

A limitation of our method is that it lacks the ability to report face



































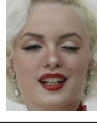
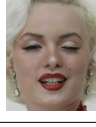
Baseline ↓ Metric	Rep. portrait	Our metric	Baseline metric	Baseline ↓ Metric	Rep. portrait	Our metric	Baseline metric	Baseline ↓ Metric	Rep. portrait	Our metric	Baseline metric
(36) ↓ (3)				(18,3,2) ↓ (5,3,1)				(50) ↓ (1)			
(8) ↓ (1)				(7,4) ↓ (3,1)				(79,34,17,2) ↓ (4,3,1,2)			
(25,15,14,2,1) ↓ (8,4,1,2,3)				(28,25,22) ↓ (8,4,1)				(72,47,7) ↓ (7,2,1)			
(37,19) ↓ (3,1)				(20,9,1) ↓ (2,4,1)				(5,3,2,1) ↓ (4,3,2,1)			

Table 2: Our metric in Eq. (11) results in higher rankings for the Doppelgängers present in a portrait set, compared to the baseline metric in Eq. (12). Columns 1, 5, and 9 show the change in the ranking order of the Doppelgänger portraits when switching from the baseline metric to ours. It may be seen that in nearly all cases, the ranking is improved and the Doppelgänger images are ranked higher (smaller indices) when the set is sorted according to our metric. In each block of images, the left column shows the representative portrait of the individual for whom Doppelgängers are sought, while the middle and right columns show the top ranked portrait in the set according to our metric and the baseline metric, respectively.

characteristics which are geometrical, such as the shape of one's face (e.g., oblong vs. square shaped faces). In the future, we find applying our proposed metric learning setting on 3D face representations as a promising direction for gaining sensitivity to facial geometry. In addition, since our method relies on the presence of activations, it lacks the ability to indicate that the absence of a face part, such as baldness, is characteristic.

Another drawback of our method is that it relies on face segmentation, which we apply to enable comparing the activity of common face parts. In future work, it would be interesting to bypass the need for such segmentation and attempt to select characterizing regions at a finer granularity. This would enable defining tiny face parts such as moles (e.g., Marilyn Monroe's mole) as characteristic.

In Figure 10 we show some examples where our method failed to select face parts for creating an effective facial hybrid. These failure cases are caused by differences in face pose (as shown in the leftmost hybrid), occlusions (as shown in the middle hybrid), and face expressions (as shown in the rightmost result). Although our method does not explicitly attempt to cope with such differences, we believe that a more careful facial analysis should be able to recover from such cases quite easily.

Observing numerous learned metrics, we notice that they seem to capture rather similar structures (turning on or off certain face parts, depending on the specific individual at hand). We feel that the differences between metrics are mainly either small refinements for

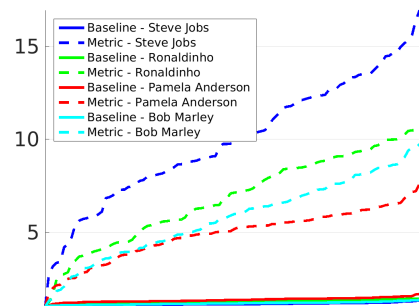


Figure 9: Plots of sorted distances between a target portrait and the portraits in the set containing the planted Doppelgängers. When distances are computed using our metric they increase faster (dashed line) compared to the distances computed using the baseline metric (solid line).

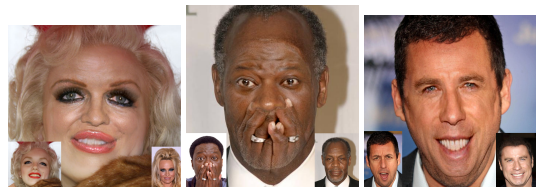


Figure 10: Failures of our method, attributed to the inability to cope with differences in face pose, occlusions or face expressions.

better alignment of features, or affecting the decision to turn face parts on or off. Assuming these differences are indeed typically small, a promising research direction would be to try and apply transfer learning for training these metrics. This will enable faster analysis using much smaller sets of portraits, perhaps even from a single portrait per individual. This, in turn, may support new applications, such as creating hybrids of two parents, as a means of hallucinating a possible portrait of their potential child, or synthesizing portraits of new siblings by analyzing the distinctive features of two already born siblings.

Acknowledgments

We thank the anonymous reviewers for their valuable comments. This work was supported in part by the Israel Science Foundation (2366/16) and the ISF-NSFC Joint Research Program (2472/17).

References

- [AECOKC17] AVERBUCH-ELOR H., COHEN-OR D., KOPF J., COHEN M. F.: Bringing portraits to life. *ACM Trans. Graph. (Proc. SIGGRAPH Asia 2017)* 36, 4 (2017). 3
- [AY16] ABUDARHAM N., YOVEL G.: Reverse engineering the face space: Discovering the critical features for face identification. *Journal of Vision* 16, 3 (2016), 40–40. 1
- [AY17] ABUDARHAM N., YOVEL G.: Critical features for face recognition in humans and machines. In *Proc. European Conference on Visual Perception (ECVP)* (2017). 2, 3
- [BHK97] BELHUMEUR P. N., HESPANHA J. P., KRIEGMAN D. J.: Eigenfaces vs. fisherfaces: Recognition using class specific linear projection. *IEEE Transactions on pattern analysis and machine intelligence* 19, 7 (1997), 711–720. 3
- [BKD*08] BITOUK D., KUMAR N., DHILLON S., BELHUMEUR P., NAYAR S. K.: Face swapping: Automatically replacing faces in photographs. *ACM Trans. Graph.* 27, 3 (Aug. 2008), 39:1–39:8. doi:10.1145/1360612.1360638. 3
- [BMN*17] BENNING M., MÖLLER M., NOSSEK R. Z., BURGER M., CREMERS D., GILBOA G., SCHÖNLIEB C.-B.: Nonlinear spectral image fusion. In *Proc. International Conference on Scale Space and Variational Methods in Computer Vision* (2017), Springer, pp. 41–53. 9
- [CWS95] CHELLAPPA R., WILSON C. L., SIROHEY S.: Human and machine recognition of faces: a survey. *Proceedings of the IEEE* 83, 5 (May 1995), 705–741. doi:10.1109/5.381842. 2
- [DTM14] DU S., TAO Y., MARTINEZ A. M.: Compound facial expressions of emotion. *Proceedings of the National Academy of Sciences* 111, 15 (2014), E1454–E1462. 3
- [GDCV98] GOMES J., DARSA L., COSTA B., VELHO L.: *Warping and Morphing of Graphical Objects*. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 1998. 9
- [Ges17] GESICHTERMIX: GesichterMix, 2017. URL: <https://www.instagram.com/gesichtermix/>. 2
- [HL01] HJELMÅS E., LOW B. K.: Face detection: A survey. *Computer Vision and Image Understanding* 83, 3 (2001), 236–274. 2
- [HLS*17] HUO J., LI W., SHI Y., GAO Y., YIN H.: Webcaricature: a benchmark for caricature face recognition. *CoRR abs/1703.03230* (2017). URL: <http://arxiv.org/abs/1703.03230>, arXiv:1703.03230. 7
- [JE09] JOHNSTON R. A., EDMONDS A. J.: Familiar and unfamiliar face recognition: A review. *Memory* 17, 5 (2009), 577–596. 1
- [KSS11] KEMELMACHER-SHLIZERMAN I., SEITZ S. M.: Face reconstruction in the wild. In *Proc. IEEE ICCV* (2011), IEEE, pp. 1746–1753. 3
- [KSS12] KEMELMACHER-SHLIZERMAN I., SEITZ S. M.: Collection flow. In *Proc. IEEE CVPR* (2012), IEEE, pp. 1792–1799. 3
- [LSD15] LONG J., SELHAMER E., DARRELL T.: Fully convolutional networks for semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (2015), pp. 3431–3440. 5
- [MLGM02] MAURER D., LE GRAND R., MONDLOCH C. J.: The many faces of configural processing. *Trends in Cognitive Sciences* 6, 6 (2002), 255–260. 3
- [MSLB18] MUHAMMAD U. R., SVANERA M., LEONARDI R., BENINI S.: Hair detection, segmentation, and hairstyle classification in the wild. *Image and Vision Computing* (2018). 5
- [NC13] NIEUWENHUIS C., CREMERS D.: Spatially varying color distributions for interactive multilabel segmentation. *IEEE Trans. Pattern Analysis and Machine Intelligence* 35, 5 (2013), 1234–1247. 9
- [NMaTa*17] NIRKIN Y., MASI I., AN TRẦN A. T., HASSNER T., MEDIONI G.: On face segmentation, face swapping, and face perception. *CoRR abs/1704.06729* (April 2017). 3
- [NW14] NG H.-W., WINKLER S.: A data-driven approach to cleaning large face datasets. In *Image Processing (ICIP), 2014 IEEE International Conference on* (2014), IEEE, pp. 343–347. 7
- [PVZ15] PARKHI O. M., VEDALDI A., ZISSERMAN A.: Deep face recognition. In *Proc. British Machine Vision Conference* (2015). 3
- [SBOR05] SINHA P., BALAS B., OSTROVSKY Y., RUSSELL R.: Face recognition by humans: 20 results all computer vision researchers should know about. *Department of Brain and Cognitive Sciences Massachusetts Institute of Technology Cambridge, MA* (2005). 3
- [SPVZ13] SIMONYAN K., PARKHI O. M., VEDALDI A., ZISSERMAN A.: Fisher vector faces in the wild. In *BMVC* (2013), vol. 2, p. 4. 4
- [SSKS15] SUWAJANAKORN S., SEITZ S. M., KEMELMACHER-SHLIZERMAN I.: What makes Tom Hanks look like Tom Hanks. In *Proc. IEEE ICCV* (2015). 3
- [STKB11] SCHROFF F., TREIBITZ T., KRIEGMAN D., BELONGIE S.: Pose, illumination and expression invariant pairwise face-similarity measure via doppelgänger list comparison. In *Computer Vision (ICCV), 2011 IEEE International Conference on* (2011), IEEE, pp. 2494–2501. 10
- [SZ14] SIMONYAN K., ZISSERMAN A.: Very deep convolutional networks for large-scale image recognition. *CoRR abs/1409.1556* (2014). 5
- [TAAMA11] TAUBERT J., APTHORP D., AAGTEN-MURPHY D., ALAIS D.: The role of holistic processing in face perception: Evidence from the face inversion effect. *Vision Research* 51, 11 (2011), 1273–1278. 3
- [TF93] TANAKA J. W., FARAH M. J.: Parts and wholes in face recognition. *The Quarterly Journal of Experimental Psychology* 46, 2 (1993), 225–245. 1
- [Tha18] THATNORDICGUY: ThatNordicGuy, 2018. URL: <https://thatnordicguy.deviantart.com/>. 2
- [TKB12] TOSEEB U., KEEBLE D. R., BRYANT E. J.: The significance of hair for face recognition. *PloS one* 7, 3 (2012), e34144. 5
- [TYRW14] TAIGMAN Y., YANG M., RANZATO M., WOLF L.: Deep-face: Closing the gap to human-level performance in face verification. In *Proc. IEEE CVPR* (2014), pp. 1701–1708. 2, 3
- [ZKL*16] ZHOU B., KHOSLA A., LAPEDRIZA A., OLIVA A., TORRALBA A.: Learning deep features for discriminative localization. In *Proc. IEEE CVPR* (2016), pp. 2921–2929. 2, 3, 6
- [ZLO10] ZHAN C., LI W., OGUNBONA P.: Finding distinctive facial areas for face recognition. In *Control Automation Robotics & Vision (ICARCV), 2010 11th International Conference on* (2010), IEEE, pp. 1848–1853. 2
- [ZR12] ZHU X., RAMANAN D.: Face detection, pose estimation, and landmark localization in the wild. In *Proc. IEEE CVPR* (2012), IEEE, pp. 2879–2886. 3, 5