

Rethinking Personalized Aesthetics Assessment: Employing Physique Aesthetics Assessment as An Exemplification

Supplementary Material

Appendix

A. More Details of Proposed Dataset

A.1. Data Collection

How to avoid selection bias by building a comprehensive PhysiqueAA dataset. Selection bias occurs when certain types of physiques or scenes are insufficiently collected in the dataset, which can negatively affect the model’s generalizability. To address this issue, we collected various human physique images from five open-source datasets [7, 48–51], to ensure the richness of physiques in our dataset. Additionally, our dataset includes three major categories (performance, sports, environment) and 36 subcategories (dance, yoga, street scene, etc.), ensuring a wide variety of physical activities.

How to Avoid Privacy Violations in Human Images. When annotating data, annotators may be particularly concerned with images that involve human faces. To safeguard privacy, we employed face-swapping technology, Roop [52]. This technology enables us to swap human faces while preserving essential facial features, including expressions and contours. The process consisted of three steps:

- 1) We created a facial dataset that includes a diverse range of faces, such as open-source, AI-generated, and digital human images, annotated with relevant labels such as age and gender.

- 2) For each human image, we selected the top five faces from this dataset based on cosine similarity of features, while also matching faces based on similar age and gender.

- 3) The images generated through Roop were then manually reviewed to identify the most suitable and natural-looking face.

How to mitigate long-tailed distributions with a preassessment approach. Aesthetic datasets [53, 54] often exhibit long-tailed distributions, which leads to a model bias toward the majority of samples. Existing IAA models can assess the overall aesthetics of physique-related images, making them a useful reference for preliminary screening. To mitigate long-tailed distributions, we performed an initial assessment using EAT [22]. We categorized the aesthetic quality of physique images into three groups: good, fair, and poor, aiming to achieve an even distribution across these categories. Finally, we selected a balanced set of samples from each group.

A.2. Data Annotation

Annotators should provide assessments across the following dimensions:

- 1) **The appearance score.** The score evaluates the individual’s ability to convey emotions and engage an audience through their physical presence, highlighting how their expressive qualities enhance the overall aesthetic impact of the images.

- 2) **The health score.** The score assesses the visual indicators of health and vitality as depicted in the images, focusing on attributes such as muscle definition, fitness, and balance that contribute to a striking aesthetic appeal.

- 3) **The posture score.** The score analyzes the individual’s posture in the images, considering its aesthetic qualities, such as stability, creativity, and expressiveness. It measures how effectively the posture enhances visual elegance and artistic expression, showcasing the interplay between technical accuracy and emotional resonance.

The initial score range for the three dimensions is from 0 to 10. However, given the decisive role of the healthy score in the PhysiqueAA and the difficulty of assessing health solely from a visual perspective, we refined the healthy score S_h based on the objective BMI metric [55], which defines a BMI between 18.5 to 25 as healthy. Thus, we established the following formula:

$$S'_h = S_h \cdot (1 - \sigma \cdot \mathbb{I}(\text{BMI} \notin [18.5, 25])), \quad (4)$$

where $\mathbb{I}(\cdot)$ denotes the indicator function, taking a value of 1 when the condition is true ($\text{BMI} \notin [18.5, 25]$) and 0 otherwise, and σ is a scaling constant set to 0.7. Finally, The health score ranges from 0 to 10.

A.3. Data Partition

Our PhysiqueAA50K dataset consists of three subsets: PAA-16-personality (for stage 1 validation), PAA-3-User (for stage 2 validation), and 400 images (for stage 3 validation), as outlined below:

- 1) PAA-16-personality contains 40,000 images annotated with PhysiqueAA scores and surveys from three experts, representing ISFJ, ESFJ, and ISTJ personalities. Each expert’s set of 40,000 images is split into 32,000 training images and 8,000 testing images.

- 2) PAA-3-User includes 10,000 images annotated with PhysiqueAA scores and surveys from three users (User-ISFJ, User-ESFJ, and User-ISTJ). Each user’s set of 10,000 images is divided into 8,000 for training and 2,000 for testing;

- 3) In stage 3, user feedback is immediate and dynamic during each update epoch, making it impossible to establish a fixed testing dataset. Therefore, the 400 images for stage 3 validation cannot be divided into training and testing sets.

B. 16 MBTI Personality Types

The Myers-Briggs Type Indicator (MBTI) is a psychological tool developed based on Carl Jung’s theory of personality types [28]. It identifies how individuals naturally perceive and make decisions, emphasizing inherent psychological preferences in these processes. The MBTI categorizes people into 16 distinct personality types, derived from four pairs of opposite preferences:

1) **I**ntroversion (**I**) vs. **E**xtraversion (**E**): Focuses on whether you gain energy from the outside world or from within yourself;

2) **S**ensing (**S**) vs. **I**ntuition (**N**): Describes whether you prefer concrete facts and details or look for patterns and possibilities;

3) **T**hinking (**T**) vs. **F**eeling (**F**): Refers to whether you make decisions based on logic and objectivity or personal values and emotions;

4) **J**udging (**J**) vs. **P**erceiving (**P**): Describes whether you prefer a structured, orderly lifestyle with careful planning to control your surroundings, or a flexible, spontaneous lifestyle focused on exploring and experiencing various ways of living without rigid control.

These preferences combine in different ways to form 16 unique types, each reflecting a stable pattern of cognitive and decision-making tendencies.

The user’s personality can reflect their stable aesthetic subjectivity [5, 6, 56, 57]. We analyze three common MBTI personality types: ISFJ, ESFJ, and ISTJ. The following is a brief analysis of their characteristics:

- **ISFJ** (Introversion, Sensing, Feeling, Judging): ISFJs tend to value classic and timeless elements of physical aesthetics. They appreciate appearances that convey warmth, harmony, and subtlety, favoring styles that emphasize comfort and approachability. ISFJs are likely drawn to soft, natural designs that prioritize balance and evoke a sense of care and security.
- **ESFJ** (Extraversion, Sensing, Feeling, Judging): ESFJs evaluate physical aesthetics with an emphasis on social relevance and interpersonal harmony. Their preferences often align with popular trends and socially accepted norms. They tend to appreciate fashionable, detail-oriented styles that enhance social interactions and positively influence others’ perceptions.
- **ISTJ** (Introversion, Sensing, Thinking, Judging): ISTJs prioritize practicality and precision when evaluating physical aesthetics. They tend to favor clean-cut, structured appearances that project order, reliability, and professionalism. Their preference leans towards classic and functional styles, prioritizing clarity and discipline over fleeting trends or extravagant designs.

There are two main reasons behind the rationale for linking MBTI types to physique preferences: 1) Previous stud-

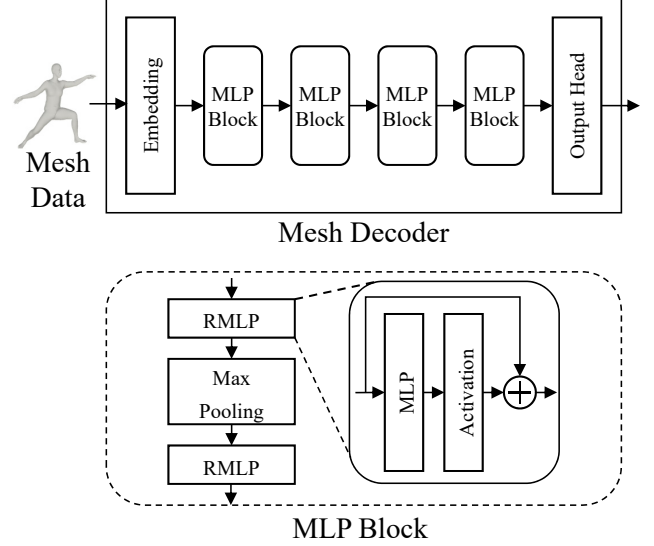


Figure 10. Comprehensive structure of the MPM.

ies, such as those in PAA [5, 6], have shown that personality traits reliably indicate stable aesthetic preferences. The widely recognized MBTI framework has also demonstrated a strong correlation with aesthetic preferences [57]. Therefore, physique preferences, as a type of aesthetic preference, are strongly correlated with MBTI types. 2) During data collection, we found that users with the same MBTI type tend to share similar physique preferences. For instance, around 80% of Extraverted (E) users preferred noticeable physiques, while only about 10% of Introverted (I) users did.

C. More Details of PhysiqueFrame

C.1. Details of the Mesh Decoder

After processing the mesh data, we employ a mesh decoder (Fig. 10) for further analysis. This structure consists of an Embedding layer and four stages. Each stage is composed of two Residual Multilayer Perceptron (RMLP) blocks and a max-pooling layer alternately connected, which can be defined as:

$$f_{mpm} = \{R_{2nd}(M(R_{1st}(v_{i,j}))) \mid i \in \mathbb{R}^n, j \in \mathbb{R}^k, i \neq j\}, \quad (5)$$

The first residual block $R_{1st}(\cdot)$ learns shared weights from different local regions of the torso, enabling effective modeling of localized physique aesthetics. The max-pooling layer $M(\cdot)$ is used for feature aggregation. Then, the second residual block $R_{2nd}(\cdot)$ extracts deeply aggregated features f_{mpm} , which capture both local and global aspects of the physique. The alternating use of two RMLP blocks effectively balances the extraction of local features (such as limb details) and global features (such as overall posture). This

enables the model to capture the full characteristics of the physique while retaining crucial details.

C.2. Details of Five Specific Physique Factors

We analyzed literature on physique aesthetics [12, 13], so-maesthetics [14, 58], and classical photography [59, 60]. Additionally, we referenced standards from real-world competitions, such as modeling [61], bodybuilding [62], and dancing [63]. These standards emphasize core elements such as posture, symmetry, and muscle definition, which are pivotal in evaluating physique aesthetics across different contexts. By synthesizing insights from these domains, we identified five factors highly relevant to PhysiqueAA: physique expressiveness, physique style, physique shape, physique posture, and personality traits.

The five specific physique factors differ from PhysiqueAA’s dimensions (appearance, health, posture) by offering a completely objective framework to describe physique aesthetics. These factors facilitate the integration of user preference data, gathered from multimodal inputs such as surveys and feedback, into a structured and interpretable format. This approach establishes a coherent connection between subjective user preferences and the assessment of physique aesthetics.

C.3. Details of the feature aggregation from PENet and PANet

The features of two modules are concatenated and then passed through MLPs for non-linear mapping.

Metric		CNN-based models				Transformer-based models				Ours
		NIMA	BIAA	HGCN	TANet	MaxViT	TCFormer	EAT	Q-Align	
Appearance	$\mathcal{S} \uparrow$.560	.503	.401	.548	.539	.405	.569	.549	.649
	$\mathcal{L} \uparrow$.588	.531	.418	.564	.555	.428	.591	.571	.678
	$\mathcal{A} \uparrow$.702	.684	.637	.714	.728	.643	.737	.716	.744
Health	$\mathcal{S} \uparrow$.496	.425	.333	.454	.447	.345	.506	.482	.540
	$\mathcal{L} \uparrow$.527	.457	.358	.492	.474	.382	.523	.521	.567
	$\mathcal{A} \uparrow$.718	.674	.591	.700	.708	.598	.726	.716	.740
Posture	$\mathcal{S} \uparrow$.584	.542	.467	.566	.577	.501	.604	.588	.657
	$\mathcal{L} \uparrow$.631	.571	.503	.610	.625	.519	.624	.616	.695
	$\mathcal{A} \uparrow$.707	.690	.603	.712	.703	.628	.744	.725	.761

Table 6. Comparison of the night models conducted on the User-ESFJ dataset for stage 2. We **retrained the models to maximize performance**, using the recommended parameters and settings.

Metric		CNN-based models				Transformer-based models				Ours
		NIMA	BIAA	HGCN	TANet	MaxViT	TCFormer	EAT	Q-Align	
Appearance	$\mathcal{S} \uparrow$.527	.486	.409	.503	.502	.416	.532	.529	.617
	$\mathcal{L} \uparrow$.548	.503	.425	.528	.524	.442	.560	.551	.650
	$\mathcal{A} \uparrow$.708	.641	.630	.698	.659	.621	.692	.686	.716
Health	$\mathcal{S} \uparrow$.499	.405	.318	.458	.451	.327	.508	.487	.546
	$\mathcal{L} \uparrow$.525	.438	.334	.484	.477	.343	.529	.503	.570
	$\mathcal{A} \uparrow$.704	.654	.555	.679	.671	.575	.705	.697	.726
Posture	$\mathcal{S} \uparrow$.546	.487	.418	.488	.501	.426	.540	.518	.597
	$\mathcal{L} \uparrow$.588	.523	.456	.532	.541	.462	.583	.566	.641
	$\mathcal{A} \uparrow$.690	.610	.583	.669	.652	.589	.693	.678	.701

Table 7. Comparison of the night models conducted on the User-ISTJ dataset for stage 2. We **retrained the models to maximize performance**, using the recommended parameters and settings.

Metric		CNN-based models				Transformer-based models				Ours
		NIMA	BIAA	HGCN	TANet	MaxViT	TCFormer	EAT	Q-Align	
Appearance	$\mathcal{S} \uparrow$.485	.489	.409	.509	.506	.422	.516	.509	.580
	$\mathcal{L} \uparrow$.491	.501	.416	.512	.527	.447	.536	.514	.603
	$\mathcal{A} \uparrow$.643	.654	.607	.661	.655	.610	.678	.674	.706
Health	$\mathcal{S} \uparrow$.413	.415	.332	.428	.431	.352	.447	.455	.539
	$\mathcal{L} \uparrow$.451	.457	.402	.463	.476	.405	.482	.494	.599
	$\mathcal{A} \uparrow$.612	.622	.507	.608	.613	.539	.646	.635	.687
Posture	$\mathcal{S} \uparrow$.493	.499	.377	.513	.518	.435	.524	.522	.588
	$\mathcal{L} \uparrow$.518	.523	.396	.544	.557	.443	.561	.545	.637
	$\mathcal{A} \uparrow$.729	.732	.680	.742	.743	.593	.752	.746	.777

Table 8. Comparisons of night models conducted on the PAA-16-personality (ISFJ) dataset for stage 1. We **retrained the models for the best performance** with the recommended parameters and settings.

Metric		CNN-based models				Transformer-based models				Ours
		NIMA	BIAA	HGCN	TANet	MaxViT	TCFormer	EAT	Q-Align	
Appearance	$\mathcal{S} \uparrow$.447	.441	.401	.458	.469	.414	.473	.465	.502
	$\mathcal{L} \uparrow$.456	.449	.412	.472	.481	.436	.486	.474	.502
	$\mathcal{A} \uparrow$.633	.625	.605	.656	.652	.612	.671	.662	.689
Health	$\mathcal{S} \uparrow$.385	.381	.322	.395	.400	.345	.420	.402	.465
	$\mathcal{L} \uparrow$.423	.426	.375	.439	.451	.391	.478	.483	.559
	$\mathcal{A} \uparrow$.577	.570	.516	.574	.592	.541	.603	.581	.621
Posture	$\mathcal{S} \uparrow$.471	.479	.403	.483	.486	.425	.505	.497	.545
	$\mathcal{L} \uparrow$.476	.485	.419	.497	.505	.436	.521	.505	.551
	$\mathcal{A} \uparrow$.644	.655	.607	.660	.661	.615	.674	.663	.698

Table 9. Comparisons of night models on the PAA-16-personality (ESFJ) dataset for stage 1. We **retrained the models for the best performance** with the recommended parameters and settings.

Metric		CNN-based models				Transformer-based models				Ours
		NIMA	BIAA	HGCN	TANet	MaxViT	TCFormer	EAT	Q-Align	
Appearance	$\mathcal{S} \uparrow$.517	.515	.449	.541	.514	.463	.555	.549	.616
	$\mathcal{L} \uparrow$.531	.528	.466	.556	.531	.485	.578	.564	.628
	$\mathcal{A} \uparrow$.684	.675	.635	.686	.683	.654	.701	.698	.730
Health	$\mathcal{S} \uparrow$.501	.494	.389	.511	.503	.411	.536	.524	.593
	$\mathcal{L} \uparrow$.536	.523	.408	.538	.530	.437	.561	.547	.685
	$\mathcal{A} \uparrow$.615	.607	.551	.644	.635	.563	.672	.665	.700
Posture	$\mathcal{S} \uparrow$.552	.563	.513	.590	.575	.535	.603	.605	.660
	$\mathcal{L} \uparrow$.576	.585	.536	.623	.596	.561	.637	.640	.683
	$\mathcal{A} \uparrow$.677	.682	.634	.695	.689	.656	.701	.704	.732

Table 10. Comparisons of night models on the PAA-16-personality (ISTJ) dataset for stage 1. We **retrained the models for the best performance** with the recommended parameters and settings.

D. More Details of Experiment

D.1. More Performance Comparison with PANet Alternatives

As shown in Table 6 and Table 7, our model achieves the best performance across all metrics when compared with alternative models on the User-ESFJ and User-ISTJ dataset. These results demonstrate our model’s effectiveness in capturing physique aesthetics features.

Additionally, as shown in Table 8, Table 9, and Table 10, we compared our model with alternative models on PAA-16-personality datasets (ISFJ, ESFJ, ISTJ) for stage 1. Our model achieves state-of-the-art performance across all metrics.

D.2. More Quantitative Results

- Fig. 11 provides more qualitative examples illustrating the impact of personalized surveys on model predictions.
- Fig. 12 provides additional results from PANet Alternatives.

tives. Our model predicts the results that are closest to the ground truth.

- Fig. 13 presents examples of images with the prediction results from PhysiqueFrame, illustrating the evaluation process from both a 3D (mesh) perspective and a posture perspective.

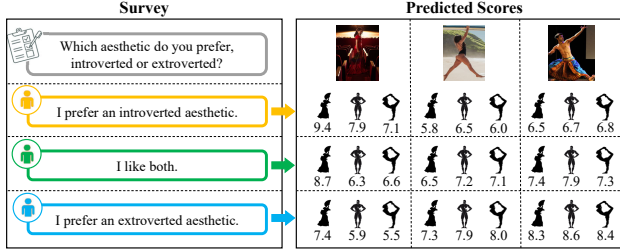


Figure 11. Impact of personalized surveys on PhysiqueAA predictions, with icons \mathbf{A} , \mathbf{H} and \mathbf{P} representing scores of appearance, health, and posture, respectively.

D.3. More ablation studies of different modules

As shown in Table 11, the scores from PENet alone perform worse than those from PANet + PENet. Besides, without GCN or Affine Geometry, the model’s performance decreases.




Method	Appearance		Health		Posture	
	$\mathcal{S} \uparrow$	$\mathcal{L} \uparrow$	$\mathcal{S} \uparrow$	$\mathcal{L} \uparrow$	$\mathcal{S} \uparrow$	$\mathcal{L} \uparrow$
w/o GCN in PAM	.675	.704	.604	.633	.643	.684
w/o Affine Geometry in MPM	.679	.711	.606	.630	.646	.688
only PENet	.687	.716	.611	.639	.657	.693
w/ preferences	.699	.724	.625	.657	.664	.703

Table 11. Ablation studies on the performance of different modules, conducted on the User-ISFJ dataset for stage2.

E. Safeguards and Licenses for Existing Assets

The original owners of assets (e.g., code, data, models, personalized surveys) used in the paper are properly credited, and the licenses and terms of use are explicitly mentioned and properly respected, ensuring that there are no copyright issues and no risk of misuse.

Input	NIMA	BIAA	HGCN	TANet	MaxViT	TCFormer	EAT	Q-Align	Ours	Ground Truth
	 6.9	 6.8	 7.3	 7.6	 7.8	 6.8	 7.5	 8.2	 9.1	 9.2
	 7.0	 6.5	 6.8	 7.0	 7.1	 6.3	 7.0	 7.3	 8.6	 8.4
	 7.3	 7.5	 6.8	 7.7	 7.5	 5.5	 7.6	 8.1	 9.5	 9.5
Input	NIMA	BIAA	HGCN	TANet	MaxViT	TCFormer	EAT	Q-Align	Ours	Ground Truth
	 6.0	 4.2	 4.6	 5.6	 5.5	 4.8	 6.3	 5.4	 7.5	 7.4
	 6.2	 4.5	 5.8	 8.1	 7.0	 5.6	 6.9	 7.8	 9.5	 9.7
	 4.7	 4.1	 3.9	 4.8	 5.0	 5.5	 5.2	 4.3	 6.1	 6.3
Input	NIMA	BIAA	HGCN	TANet	MaxViT	TCFormer	EAT	Q-Align	Ours	Ground Truth
	 6.5	 6.0	 6.6	 7.3	 6.8	 6.3	 7.3	 7.1	 8.0	 8.2
	 7.6	 5.5	 5.8	 7.0	 7.0	 6.4	 7.5	 7.2	 8.5	 8.5
	 5.2	 7.1	 6.9	 8.3	 7.3	 6.8	 8.2	 7.7	 9.4	 9.5
Input	NIMA	BIAA	HGCN	TANet	MaxViT	TCFormer	EAT	Q-Align	Ours	Ground Truth
	 6.0	 5.8	 3.9	 5.7	 6.0	 4.8	 5.4	 6.1	 7.6	 7.8
	 5.9	 6.1	 3.8	 6.1	 6.5	 5.2	 5.8	 6.5	 7.7	 7.5
	 4.8	 4.9	 4.8	 3.7	 5.4	 4.6	 5.1	 4.8	 6.3	 6.1

Figure 12. The prediction results from the 9 models are displayed, with icons ,  and  representing scores of appearance, health, and posture, respectively. We *retrained the models on the User-ISFJ dataset for the best performance*.

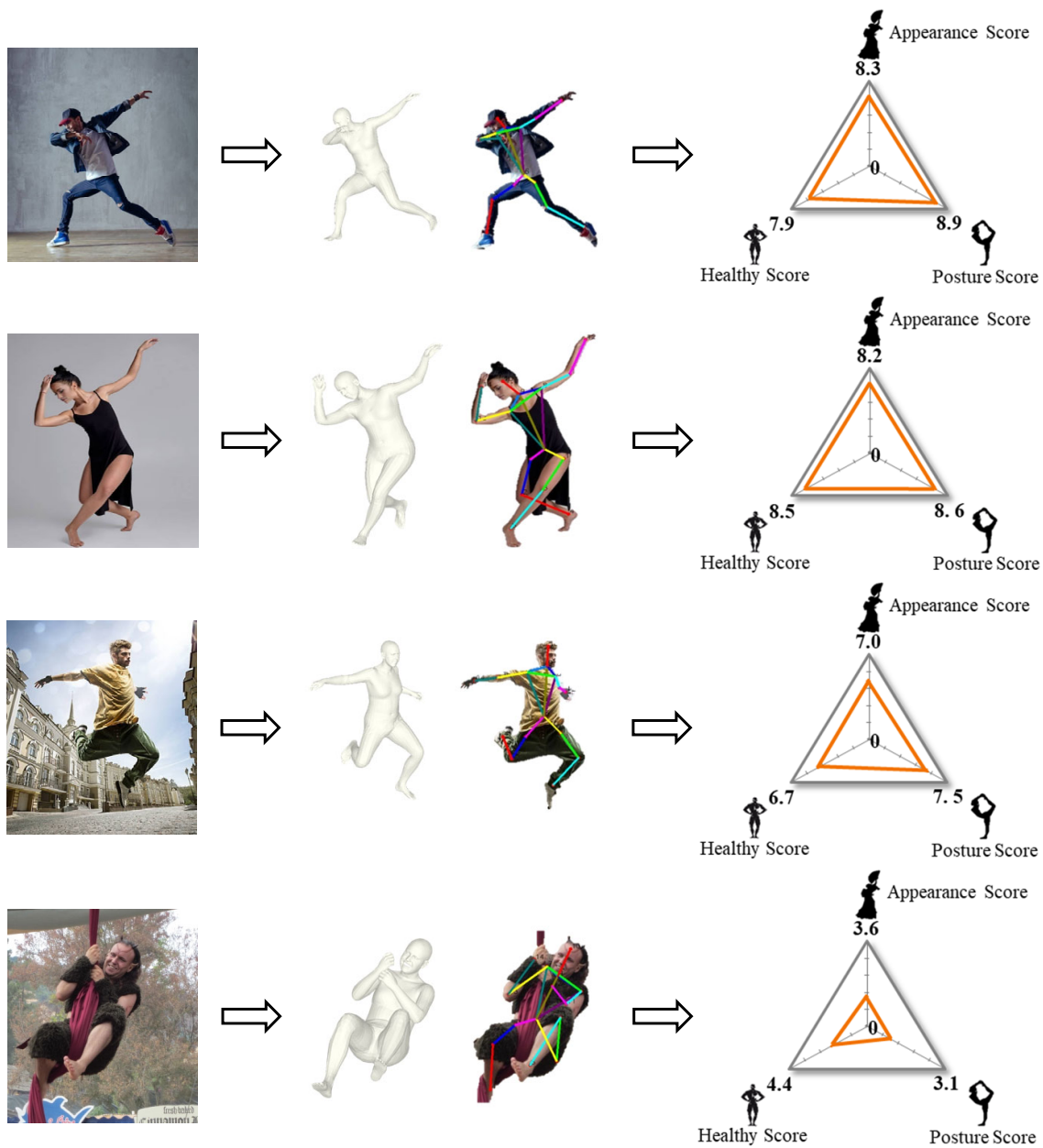


Figure 13. Examples of images are displayed with the prediction results from PhysiqueFrame. This section also illustrates the evaluation process from both a 3D (mesh) perspective and a posture perspective.