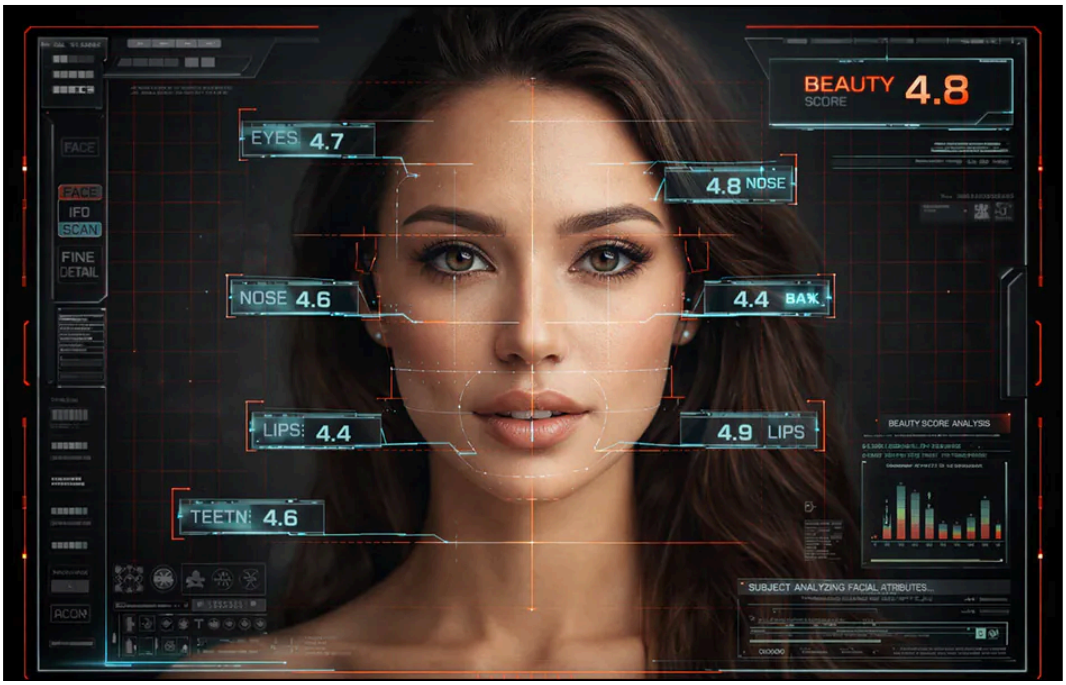
AI's Pursuit of Beauty

Originally published at https://www.unite.ai/ais-pursuit-of-beauty/



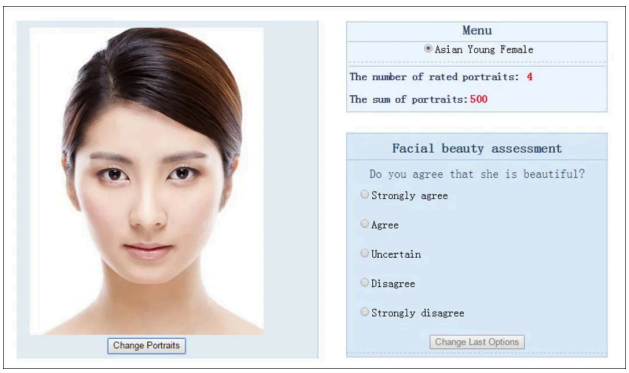
A new AI-driven beauty evaluation system rates how attractive faces appear, while training faster than typical deep learning models, potentially making large-scale automated beauty-scoring more practical.

Facial Beauty Prediction (FBP) is big business, and a fairly strong strand in the research literature. Even though it breaks practically every tenet behind combating bias in AI and machine learning practices, and even though in many ways it supports objectification and reductionism in algorithmic perceptions of women, it nonetheless attracts the interest of several multi-billion dollar industries, most of which are aimed directly at women, such as cosmetics, cosmetic facial surgery, livestreaming, and fashion, among others:

Women rated from 1-5, from the paper 'Asian Female Facial Beauty Prediction Using Deep Neural Networks via Transfer Learning and Multi-Channel Feature Fusion'. Source

Beyond these obvious female-centric business enclaves, advertising and multiple other industries, including entertainment and publishing, have notable stakes in understanding what both men and women find 'attractive', necessarily on a per-culture basis.

The fact that aggregate perceptions of beauty vary across regions means that no definitive globally-applicable datasets can be obtained, and that new research must either stay parochial, or else concentrate on 'high-level' methods that can be applied across diverse swathes of cultural data.



An interface for a facial beauty assessment system for the 2015 SCUT-FBP project. Source

Often, geographic location is not the only restriction, since attractiveness-focused datasets may struggle to provide equal efficacy across genders, or may have been curated with a particular application in mind – and this may restrict the collection's use in other domains.

For instance, in 2025 I reported on the development of a relatively high-scale (100,000+ identities) dataset to assess attractiveness in live-streaming, whose close-cropped standards might need notable adaptation to broader projects, despite the enormous effort behind the initiative.

Facial Rendition

As may be evident from the links and images above, Asian research bodies are often not operating under the same cultural restrictions as their western counterparts, who would be hard-pressed to dare publish a scientific illustration rating five western women from least to most attractive, as we see in the above-illustrated study.

It could be argued that where Asian-originated systems of this kind are proven effective in public, without fear of local censure, western interests can use or adapt such research into proprietary, private implementations. The task of 'rating women', in that scenario, is renditioned to a locale where it can be pursued without criticism.

Whether this is common or whether less-publicized western equivalent systems tend to be developed away from open source collaboration and from public scrutiny, it's reasonable to assume that the target goal is of global interest, due to the large number of professional sectors that can or could benefit from accurate evaluations of attractiveness.

Survival of the Fittest

It may seem that massive web-scrapable corpora such as Tik Tok, Instagram and YouTube would prove excellent arbiters of beauty, by correlating followers, likes and traffic to attractiveness, since this is a common and reasonable association (albeit with some exceptions).

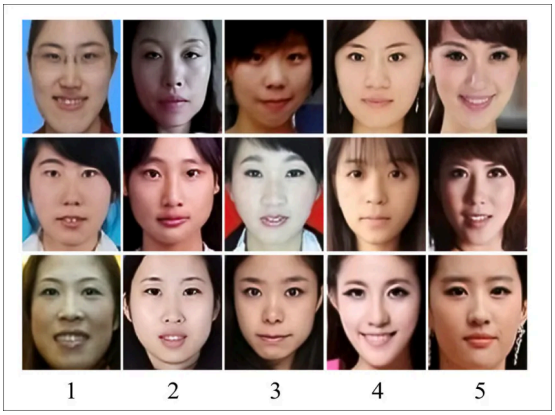
Likewise, existing collections – such as ImageNet and LAION – featuring actors and models who have 'risen to the top'– will typically feature attractive individuals (though often with too many data points of too few people), allowing wider cultural mechanisms to act as a proxy for attractiveness.

However, this does not account for shifting tastes in what people find attractive over time (never mind geographically). Therefore, again, high-level and data-agnostic systems are needed, not individual and specious collections or curations which will fail to reflect changing tastes.

Combination Skin

The latest academic entry to tackle these challenges comes from China, where transfer learning and Broad Learning System (BLS) are combined to address the long-standing trade-off between accuracy and computational cost.

Conventional neural networks tend to achieve strong results only with heavy training, while lighter systems such as BLS train quickly, but struggle to capture enough detail. The new work bridges this gap by using a pre-trained visual model to extract facial features, which are then passed to a fast BLS-based system for scoring, allowing features to be reused instead of learned from scratch, while keeping training efficient:



Sample images from the LSAFBD dataset, showing female faces grouped by human-assigned beauty scores from 1 to 5. Ratings were obtained from multiple annotators, and used as supervised labels for training and evaluating facial beauty prediction models across variations in pose, lighting, and appearance. Source

The first of two variations introduced in the work (E-BLS, see below), feeds extracted features directly into the lightweight system, while the second, ER-BLS (also see below), adds a simple intermediate step that standardizes and refines those features before evaluation, helping improve consistency without slowing the process.

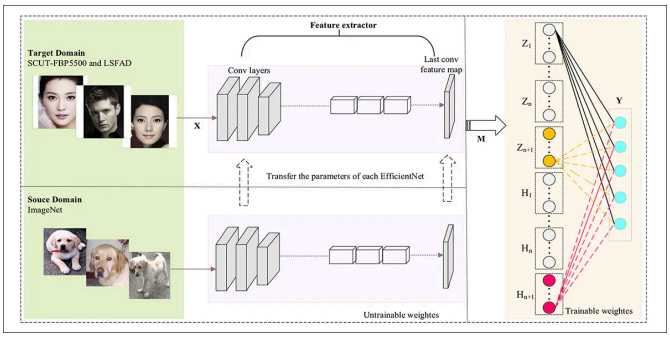
Tests conducted by the authors prove, they claim, that their approach is superior to either method by itself, and to other competing methods.

The new paper is titled *Facial beauty prediction fusing transfer learning and broad learning system*, and comes from six researchers at Wuyi University, Jiangmen.

Method

The aforementioned *Broad Learning System* is a lightweight alternative to deep neural networks, that skips stacking multiple layers, and instead spreads learning across a wide set of simpler connections, allowing models to train quickly – but usually at the cost of missing finer visual detail.

The first of the two variants, **E-BLS**, combines EfficientNet-based transfer learning with BLS, extracting detailed visual features from a face, and then passing them to BLS, entailing a final prediction that avoids the need to train a full deep neural network from scratch:

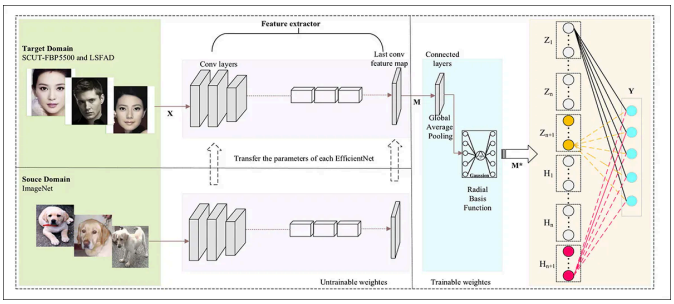


Architecture schema for the E-BLS model.

EfficientNet, pre-trained on ImageNet-1k, and largely kept unchanged, converts each input image into a compact set of feature values that describe the face in a structured way, while BLS takes those values and processes them through a network of simple, randomly connected nodes that transform and combine the information, before producing the final attractiveness score.

Because BLS does not rely on deep layered structures, E-BLS can be updated by adding more nodes instead of retraining the entire system,. This keeps training fast, and makes it easier to improve the model as new data is introduced.

The second of the two variants, ***ER*-BLS**, builds on E-BLS by inserting an additional processing stage between the EfficientNet feature extractor and BLS, with the goal of improving how those extracted features are prepared before being used for prediction:



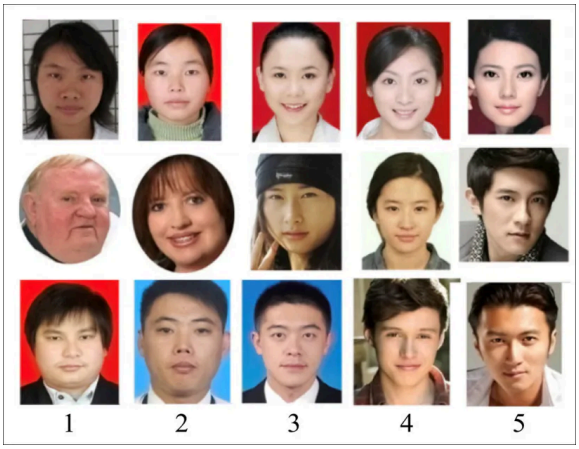
Architecture of the ER-BLS model.

Instead of sending the raw EfficientNet features directly into BLS, ER-BLS first passes them through a refinement layer that standardizes and reshapes the data, helping to reduce noise, and making the features more consistent across different images. This step is designed to improve how well the system generalizes, especially when faces vary in lighting, pose, or other visual conditions that can otherwise introduce instability into the predictions.

The refined features are then fed into the same BLS structure used in E-BLS, where feature nodes and enhancement nodes transform and combine the information to produce the final attractiveness score.

Data and Tests

To test their approach, the authors leveraged the SCUT-FBP5500 dataset, a facial beauty prediction collection from South China University, containing 5,500 frontal face images at 350x350px, featuring diverse races, genders and ages:



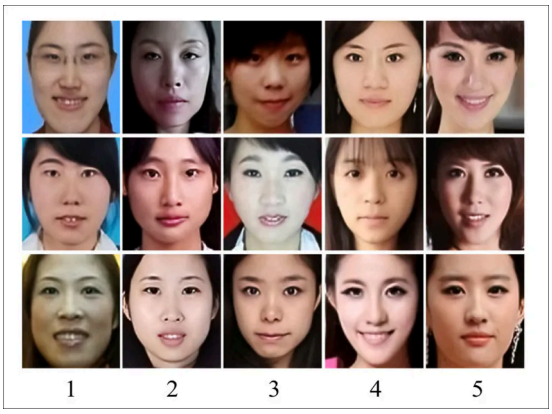
Sample SCUT-FBP5500 dataset facial images rated from least (1) to most (5) attractive.

Each image was rated with a beauty score by 60 volunteers, on a 1-5 scale, ranging from *extremely unattractive* (1) to *extremely attractive* (5):

Beauty Level	Rating	Number of Images
1	Extremely unattractive	76
2	Unattractive	821
3	Average	3278
4	Attractive	1226
5	Extremely attractive	99

The division of proportions of images by beauty rating.

The other database used was the Large-Scale Asian Female Beauty Dataset (LSAFBD) collection, a dataset curated by the authors themselves.



Sample LSAFBD dataset facial images rated from least (1) to most (5) attractive.

The collection consists of 80,000 unlabeled images at 144x144px resolution, with variations in pose and background, as well as age. These were rated by 75 volunteers for the same criteria as the prior dataset, this time on a 0-4 scale:

Beauty Level	Rating	Number of Images
0	Extremely unattractive	948
1	Unattractive	1148
2	Average	3846
3	Attractive	2718
4	Extremely attractive	1333

The divisions for the LSAFBD dataset.

Each dataset was split into training and testing segments at an 8/20 ratio, and cross-validation used to stabilize results across runs. The BLS component was configured through the number of feature windows; the number of nodes per window; and the number of enhancement nodes, with Hyperopt used to search for effective combinations.

To establish a baseline, a standard BLS model was trained under identical settings, after which a series of transfer learning models were introduced, including ResNet50, Inception-V3, DenseNet121, InceptionResNetV2, EfficientNetB7, MobileNetV2, NASNet, and Xception – all initialized with ImageNet-1k weights, and trained with their final layers unfrozen.

Training used a learning rate of 0.001 (reduced when progress stalled), and a batch size of 16, across 50 epochs, with regularization and rectified linear activation (ReLU) applied throughout.

Performance was evaluated using accuracy and Pearson correlation, alongside total training time, with results averaged across five runs.

The authors report the training setup as an Intel-i7 3.6 GHz CPU and 64GB RAM on a ‘desktop computer’:

Model	Training time (s)	Testing AC (%)	Training AC (%)
E-BLS (ours)	*1399.88*	*73.13*	*75.38*
BLS	640.5	65.85	68.83
NASNet	5493.06	68.11	73.92
MobileNetV2	4224.42	69.03	79.3
DensNet121	9862.52	70.77	75.97
ResNet50	8211.02	71.23	76.63
InceptionResNetV2	25,896.82	71.69	74.28
EfficientNetB7	17,589.68	72.70	76.68
Xception	6770.64	72.98	75.01
ER-BLS (ours)	*1291.83*	*74.69*	*76.76*
InceptionV3	11,630.7	73.16	78.02

Performance comparison on SCUT-FBP5500, where E-BLS and ER-BLS achieve competitive accuracy against deep CNN models including ResNet50, EfficientNetB7, InceptionV3, and Xception, while requiring substantially less training time – highlighting the efficiency gains of combining transfer learning with a Broad Learning System.

Results indicated that E-BLS improved accuracy from 65.85% to 73.13%, while ER-BLS reached 74.69%, exceeding all compared models. Training time remained notably lower than deep CNNs, at roughly 1,300 seconds, versus several thousand to over 25,000 seconds.

For the tests on LSAFBD, results showed that E-BLS improved accuracy over plain BLS, while ER-BLS achieved the highest accuracy among all compared methods:

Model	Training time (s)	Testing AC (%)	Training AC (%)
ER-BLS (ours)	*2286.54*	*62.13*	*72.34*
E-BLS (ours)	*2300.48*	*60.82*	*71.58*
BLS	1446.37	52.96	67.89
NASNet	12,043.95	53.12	60.72
MobileNetV2	7830.47	54.74	66.66
Xception	16,852.60	57.01	62.91
InceptionResNetV2	36,853.53	57.31	60.91
EfficientNetB7	21,165.61	57.91	61.32
ResNet50	20,014.05	58.37	65.10
InceptionV3	25,201.29	58.42	66.47
DensNet121	19,520.28	59.93	63.54

Performance on LSAFBD, where ER-BLS and E-BLS deliver higher accuracy than all baseline and transfer-learning models while requiring only a fraction of their training time, indicating a consistent advantage in efficiency without sacrificing predictive quality.

Both variants maintained substantially lower training time than deep CNN models, indicating a more efficient balance between performance and computational cost.

Conclusion

This is somewhat of a ‘throwback’ publication, as evidenced by its use of pre-boom favorites such as CNNs, and by the lowest-level training equipment I have encountered in a new paper in many years.

Nonetheless, it deals with a surprisingly resilient objective in computer vision; one which touches heavily on human experience and subjective interpretation, and which demands a schema that transcends the aesthetic trends of the moment, and can furnish a truly resilient pipeline for the task.

First published Thursday, March 19, 2026

Writer on machine learning, domain specialist in human image synthesis. Former head of research content at Metaphysic.ai, until its dissolution into DNEG's Brahma.ai.
Portfolio site: martinanderson.ai
Contact: martin@martinanderson.ai

Discover More