Supplementary Material: A Paradigm for Building Generalized Models of Human Image Perception through Data Fusion

Shaojing Fan[1], Tian-Tsong Ng[2], Bryan L. Koenig[3], Ming Jiang[1], and Qi Zhao[1]

[1]National University of Singapore
[2]Institute for Infocomm Research
[3]Washington University in St. Louis

1. Human annotations of perception datasets

The full set of human annotated attributes of Visual Realism Dataset [3] and Memorability Dataset [5] are listed in Tables S1 and S2. Those used in our perception modeling are marked with *.

2. Empirical modeling on separate datasets

2.1. Exploratory and confirmatory factor analysis results

The results of EFA for Visual Realism and Memorability Datasets are reported in Tables S3 and S4, respectively. The measurement models of CFA are shown in Fig. S1. The measurement models have same latent factors and loadings as the final models resulting from path analysis (Fig. 2 in main text), except they are unlayered. Thus in subsequent modelings, we only report the results of path analysis.

2.2. Path analysis results

The estimates of separate models are shown in Tables S5—S8.

3. Data fusion

3.1. Empirical modeling on fused datasets

The EFA results of fused datasets are reported in Table S9. The results of path analysis are reported in Tables S10, S11.

3.2. MI evaluation

To validate MI in our data fusion, we computed the frequency distribution of observed attributes and imputed attributes. The results are shown in Fig. S2. We further tested the effects of imputation, including those on model coefficients, model completeness, and changes in model fit. We built four restricted models, two for Visual Realism Dataset and two for Memorability Dataset. For each, we built two models, one that excluded the imputed data (hereafter referred to as *incomplete model*), and another that included the imputed data (hereafter referred to as *complete model*). In all cases we attempt as much as possible given the restriction to recreate our final model. The two models with the imputed data would be complete models (in the sense that they have all variables), but would be based on only one dataset (with some variables imputed). The two models without imputed data would not be complete models in the sense that they would actually be missing some variables. We then compare the four restricted models with the complete model regarding coefficients, model completeness, and model fit. All models have acceptable fit to the data, $CFIs \geq .94$, $RMSEA \leq .093$. The estimates are shown in Tables S12 and S13. The consistency between the complete model and the model based on the fused dataset supports the validity of our data imputation.

4. Predicting visual sentiment, visual realism, and interestingness

4.1. Data fusion

The results of EFA in three fused datasets (Visual Realism Dataset, Memorability Dataset, Sentiment Dataset) are reported in Table S14. The results of path analysis are reported in Tables S15 and S16.

4.2. Features design for computational perception

Here is a detailed description on how we designed our computational features for layer-1 factors in fused model 2 (Fig. 5 in main text).

***Natural* & *familiar*:** As shown in our fused model 1 (Fig. 3 in main text), *natural* strongly correlates with *familiar*, so we model them jointly. In [9], several statistical models were introduced to represent the regularities inherent in natural images. High contrast local image patches which mainly correspond to the edge structures were studied and shown to display some regular patterns. This motivated us to use gradient information in modeling image naturalness. Let $I(x, y)$ denote the image intensity, we computed the surface gradient of the image intensity with a scaled constant α as Equation 1:

$$|grad(\alpha I)| = \sqrt{\frac{|\nabla I|^2}{\alpha_{-2} + |\nabla I|^2}} \quad where\ |\nabla I| = \sqrt{I_x^2 + I_y^2} \tag{1}$$

The constant α was to control the weight of emphasis on the low gradient region versus the high gradient region. We computed the gradient on R, G, B channels ($|grad(\alpha I)|_R$, $|grad(\alpha I)|_G$, $|grad(\alpha I)|_B$) at every pixel of an image with $\alpha = 0.25$. We used spatial pooling to reduce the dimensions to 72 in the final algorithm.

***Artistic*:** To capture the *artistic* factor, we applied Ke's method [6] for extracting five aesthetic features, which are HSV statistics, contrast, edge distribution, blur and color distribution.

1. Saturation, hue, and illumination: We computed features defined in the HSV space. Saturation indicates chromatic purity. Pure colors in a photo tend to be more appealing than dull or impure ones [2]. We computed the average saturation $f_s = \frac{1}{XY}\Sigma_{x=0}^{X-1}\Sigma_{y=0}^{Y-1} I_S(x, y)$ as the saturation indicator. Hue and illumination were similarly computed by averaging over I_H and I_V separately. Although the interpretation of such features is not as clear as saturation, they were found to be predictive of image aesthetics [2, 6].

2. Contrast: We used a contrast quality measure similar to [6], except that we computed the gray-scale level histogram of each image on R, G, B channels separately, and measured the width of the middle 98% gray level mass on each channel.

3. Edge distribution: The spatial distribution of the high frequency edges of an image was computed in order to capture its *simplicity*. A uniform distribution of edges might indicate snapshots having cluttered backgrounds, while the opposite may indicate aesthetic photos that have well defined subjects and objects in focus (*i.e.* higher simplicity) [6]. Similar to [6], we applied a 3×3 Laplacian filter with $\alpha = 0.2$ to the R, G, B channels of an image separately and took the mean across the channels. We then normalized the Laplacian image sum to 1. We calculated the area of the bounding box that encloses the top 96.04% of the edge energy of the Laplacian image L by projecting it to the x and y axes independently, so that the area of the bounding box is denoted by $1 - w_x w_y$, with w_x and w_y being the box's normalized width and height.

4. Blur: The degree of blur of an image is a strong indication for its quality and aesthetics. A blurry photo of a scene is almost always worse than a sharp photo of the same scene [6]. For blur prediction, we estimated the maximum frequency of the image I_b by taking its two dimensional Fourier transform and counting the number of frequencies whose power was greater than some threshold θ. We then normalized it by the size of the image [6]. We set $\theta = 5$ in our algorithm.

5. Color distribution: We measured the color distribution as Earth Mover distance (in the LUV color space) of the color histogram of an image H_I to a uniform color histogram H_{uni}. The smaller the distance the more distributed in color.

***Space*:** GIST descriptors [8] are designed to capture spatial layout properties of the scene by estimating the mean of global image features. We modeled *Space* using 4×4 image block by GIST and obtained a 512 dimension feature.

***Weird*:** In anomaly detection, the Local Outlier Factor (LOF) algorithm [1] is an algorithm proposed for finding anomalous data points by measuring the local deviation of a given data point with respect to its neighbours. We modeled *weird* by applying LOF to global image descriptors. We use a 10-distance neighborhood and use GIST [8] and SIFT [7] as global features.

4.3. Predicting visual realism and interestingness

To guide visual realism and interestingness prediction, we modified the layer-2 factor in fused model 2 to the corresponding perception factors. The modified models are shown in Fig. S3. The models had acceptable fit to the data, $CFIs \geq .92$, $RMSEA \leq .091$. The models guided the prediction of visual realism and interestingness by providing a more comprehensive set of attributes. The computational results are reported in Table. 3 and Fig. 7 in main text.

4.4. Predicting visual sentiment using MI and the perception model

In this section, we use MI and our perception model to predict image sentiment directly, and compare their results with those from SVM. In MI, sentiment score was computed by averaging the *exciting* and *interesting* attributes. In the perception model, we calculated sentiment by computing the separate layer-1 factors and then used their weighted sum. The weights were set to their respective links towards *liking* in fused model 2. For example:

$$\begin{aligned} Sentiment = .37 \times Familiar + .20 \times Artistic \\ +.31 \times Dynamic + .37 \times Weird \\ -.14 \times Natural + .17 \times Space. \end{aligned} \tag{2}$$

The actual values for layer-1 factors were created by computing unweighted means of the attributes that loaded on them[1], as illustrated below:

$$Artistic = (Image\_Sharp + Image\_Quality)/2. \quad (3)$$

As shown in Table S17, the results are inferior to those based on SVM (AUCs $<= .59$ *v.s.* AUCs $>= .64$). This might be because MI and the perception model are based on linear regression, whereas our SVM used non-linear kernel. The future work will include non-linear models such as Isomap [10].

References

[1] Markus M Breunig, Hans-Peter Kriegel, Raymond T Ng, and Jörg Sander. Lof: identifying density-based local outliers. In *ACM sigmod record*, volume 29, pages 93–104. ACM, 2000.

[2] Ritendra Datta, Dhiraj Joshi, Jia Li, and James Z Wang. Studying aesthetics in photographic images using a computational approach. In *ECCV*, pages 288–301. Springer, 2006.

[3] Shaojing Fan, Tian-Tsong Ng, Jonathan S Herberg, Bryan L Koenig, Cheston Y-C Tan, and Rangding Wang. An automated estimator of image visual realism based on human cognition. In *CVPR*, pages 4201–4208. IEEE, 2014.

[4] Matthew W Gallagher and Timothy A Brown. Introduction to confirmatory factor analysis and structural equation modeling. In *Handbook of Quantitative Methods for Educational Research*, pages 289–314. Springer, 2013.

[5] Phillip Isola, Devi Parikh, Antonio Torralba, and Aude Oliva. Understanding the intrinsic memorability of images. In *NIPS*, pages 2429–2437, 2011.

[6] Yan Ke, Xiaoou Tang, and Feng Jing. The design of high-level features for photo quality assessment. In *CVPR*, volume 1, pages 419–426. IEEE, 2006.

[7] Svetlana Lazebnik, Cordelia Schmid, and Jean Ponce. Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories. In *CVPR*, volume 2, pages 2169–2178. IEEE, 2006.

[8] Aude Oliva and Antonio Torralba. Modeling the shape of the scene: A holistic representation of the spatial envelope. *International Journal of Computer Vision*, 42(3):145–175, 2001.

[9] Anuj Srivastava, Ann B Lee, Eero P Simoncelli, and Song-Chun Zhu. On advances in statistical modeling of natural images. *Journal of Mathematical Imaging and Vision*, 18(1):17–33, 2003.

[10] Joshua B Tenenbaum, Vin De Silva, and John C Langford. A global geometric framework for nonlinear dimensionality reduction. *Science*, 290(5500):2319–2323, 2000.

[1] Although it is tempting to use the factor loadings as weights, those weights are unstable across samples and result in overfitting [4]

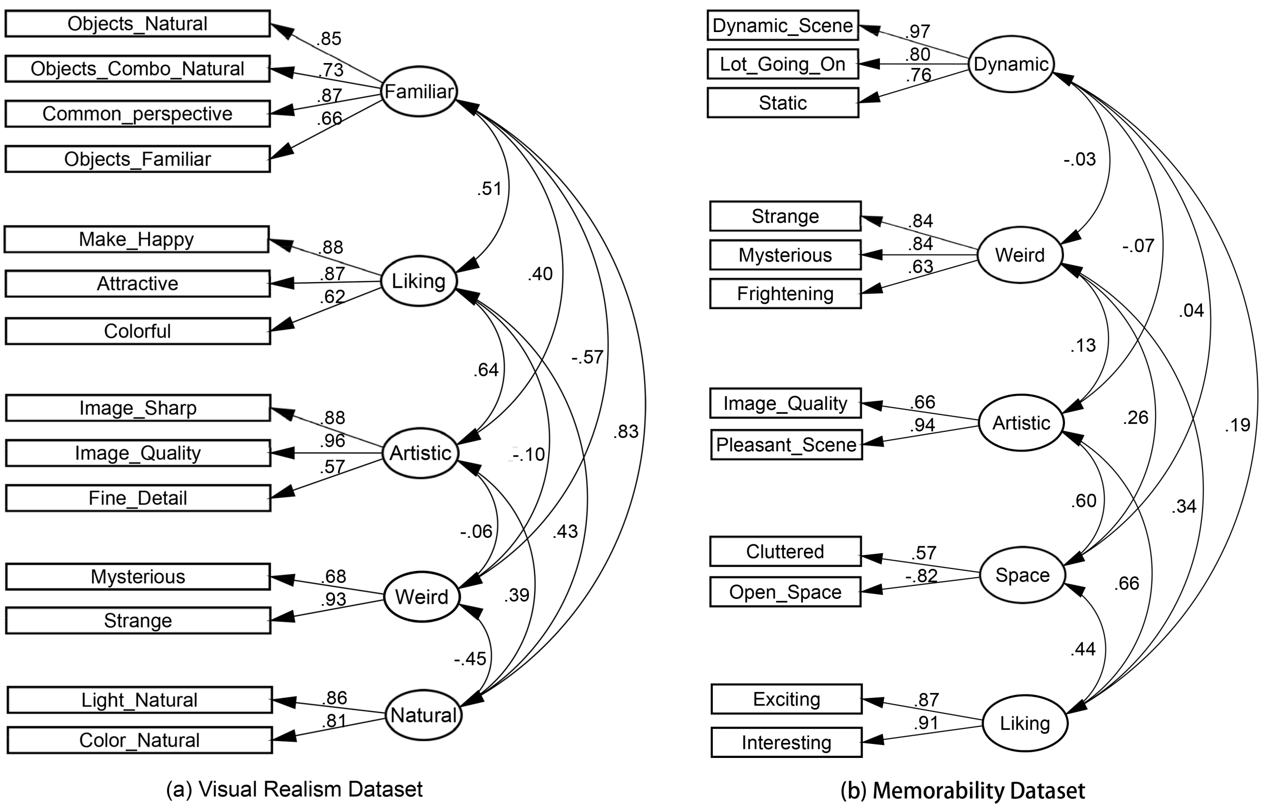


Figure S1: Measurement models of two separate datasets.

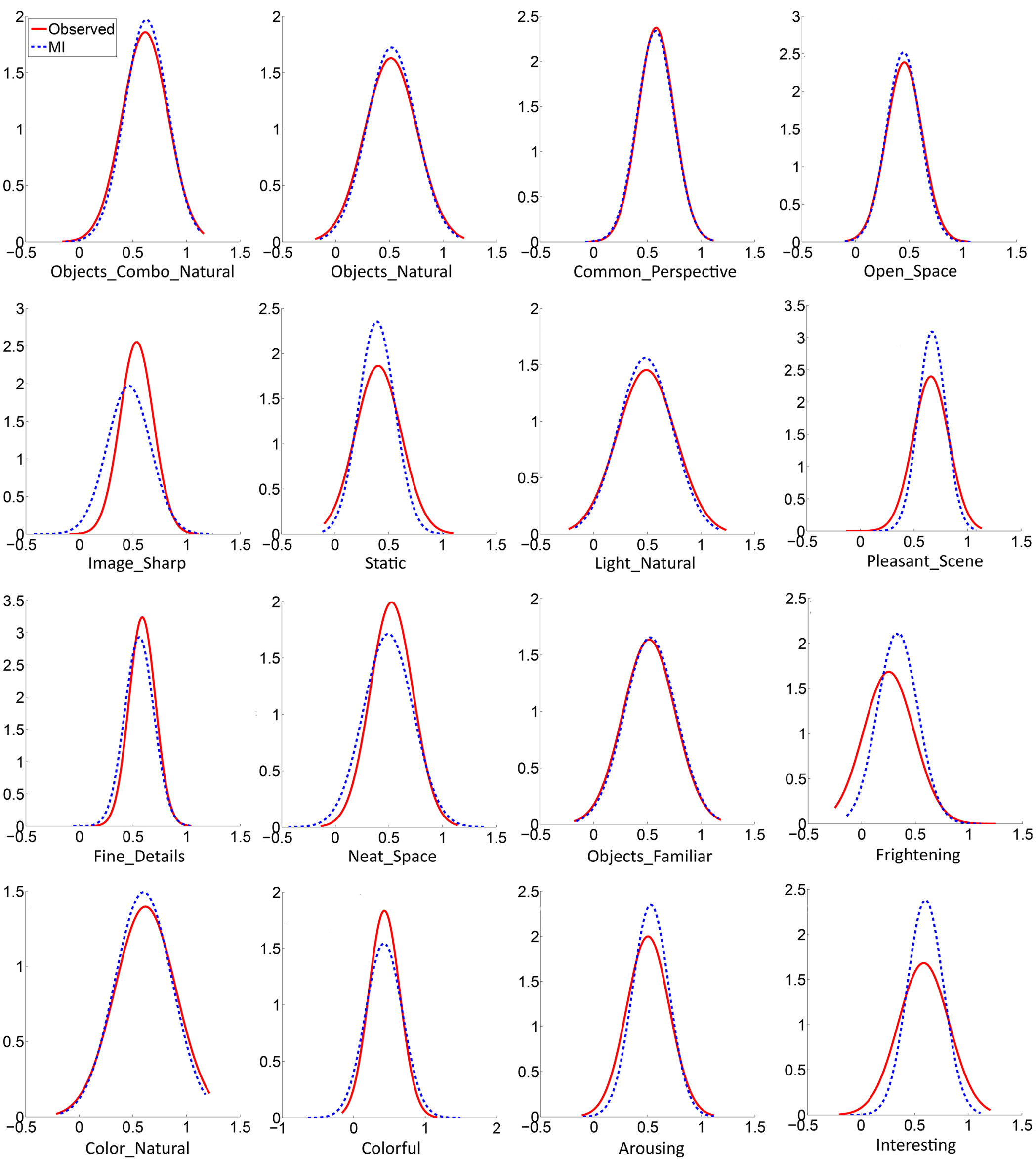
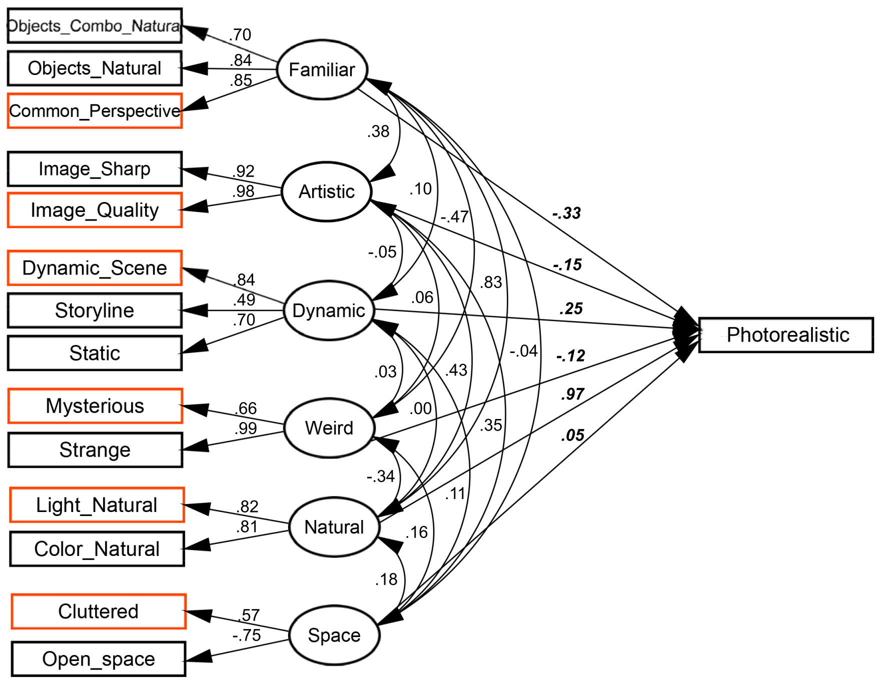
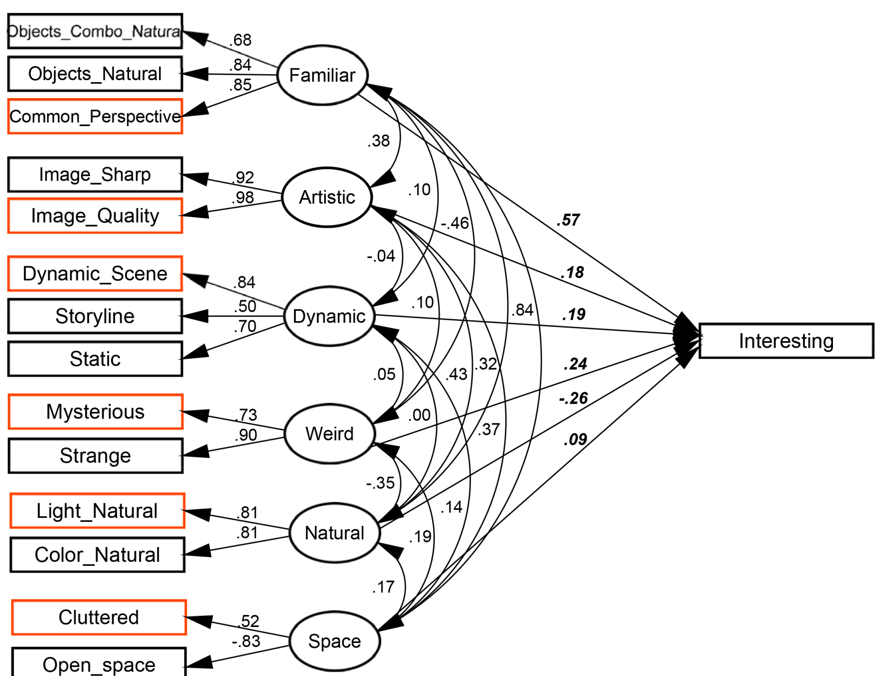


Figure S2: Frequency distribution (by normal fitting) of observed attributes and imputed attributes (MI) from fused dataset.



(a)



(b)

Figure S3: Human perception model for predicting image visual realism (a) and interestingness (b). Attributes with red border in Sentiment Dataset were collected from survey on AMT.

Table S1: Human annotated attributes of Visual Realism Dataset [3]

Familiar: Familiar with the scene*? Familiar with the objects*? Unusual or Strange*? Mysterious*?
Illumination and color: Lighting effect natural*? Shadows in the image? Sharp shadow? Color appearance natural*? Colors go well together*? Colorful*?
Aesthetics: High quality vs. Low quality*; Sharp vs. Blurry*; Expert photography*? Attractive to you*?
Spatial layout: Clean scene*?; Close-range vs. Distant-view*; Have objects of focus? Neat Space?*; Empty space vs. Full space*? Common perspective*?
Emotions: Makes you happy*? Makes you sad*? Exciting*?
Semantics: Contain fine details*? Dynamic or energetic scene*? Is there a storyline*? Contain living objects? Object appearance natural*? Naturally-occurring objects combinations*? Total number of objects? Unique number of objects?
Human semantics: Number of people in the image? Face visible? Is the person attractive? Making eye contact with viewer? Human activities? Expression genuine? Posing for the image?
Visual Realism: Appears to be a photograph?

* Attributes used in our perception modeling.

Table S2: Human annotated attributes of Memorability Dataset [5].

Spatial layout: Enclosed space vs. Open space*; Perspective view vs. Flat view*; Empty space vs. Cluttered space*; Mirror symmetry vs. No mirror symmetry*
Aesthetics: Post-card like*? Buy this painting? Hang-on wall? Is aesthetic*? Pleasant vs. Unpleasant*; Unusual or strange vs. Routine or mundane*; Boring vs. Striking colors*; High quality (expert photography) vs. Poor quality photo*; Attractive vs. Dull photo*; Memorable vs. Not memorable; Sky present? Clear vs. Cloudy sky; Blue vs. Sunset sky; Picture of mainly one object vs. Whole scene; Single focus vs. Many foci; Zoomed-in vs. Zoomed-out; Top down view vs. Side view
Emotions: Frightening*? Arousing*? Funny*? Engaging*? Peaceful*? Exciting*? Interesting*? Mysterious*? Strange*? Striking*? Makes you happy*? Makes you sad*?
Dynamic: Action going on*? Something moving in scene*? Picture tells a story*? About to happen*? Lot going on*? Dynamic scene*? Static scene*? Have a lot to say*; Length of description*
Location: Famous place? Recognize place? Like to be present in scene*? Many people go here?
Contains a person?[1]

* Attributes used in our perception modeling.
[1] For detailed attributes describing people in the image (which are not used in our work), please refer to [5].

Table S3: The loadings of attributes on the 5 major principal components (PC) for Visual Realism Dataset. Only strongest loadings (> .5) are shown. The “Cumulative variability” row shows how each PC cumulatively explains the variability of attributes in presented sequence.

PC / Attributes	1 (Liking)	2 (Familiar)	3 (Artistic)	4 (Weird)	5 (Natural)
Exciting	.89				
Make_happy	.78				
Dynamic_Scene	.77				
Attractive	.70				
Colorful	.61				
Objects_Natural		.87			
Objects_Combo_Natural		.87			
Common_Perspective		.56			
Objects_Familiar		.51			
Image_Sharp			.99		
Image_Quality			.91		
Fine_Details			.54		
Mysterious				.89	
Strange				.67	
Light_Natural					.80
Color_Natural					.70
Cumulative variability explained (%)	34.96	52.80	60.31	64.84	67.33

Table S4: The loadings of attributes on the 5 major principal components (PC) for Memorability Dataset. Only strongest loadings (> .5) are shown. The “Cumulative variability” row shows how each PC cumulatively explains the variability of attributes in presented sequence.

PC / Attributes	1 (Liking)	2 (Dynamic)	3 (Weird)	4 (Artistic)	5 (Space)
Exciting	.97				
Make_Happy	.86				
Interesting	.83				
Arousing	.63				
Dynamic_Scene		.98			
Lot_going_on		.78			
Static		.77			
Strange			.86		
Mysterious			.77		
Frightening			.70		
Image_Quality				.87	
Dull_Colors				.-72	
Pleasant_scene				.56	
Cluttered					.89
Open_Space					-.54
Cumulative variability explained (%)	16.58	42.86	55.40	61.50	64.98

Table S5: Standardized regression weights (γ) in the perception model from Visual Realism Dataset.

Dependent variable	Predictor	γ
Liking	Artistic	$.49^*$
Liking	Familiar	$.51^*$
Liking	Weird	$.17^*$
Liking	Natural	$-.11^*$
Image_Sharp	Artistic	$.88^*$
Image_Quality	Artistic	$.96^*$
Fine_Detail	Artistic	$.57^*$
Objects_Combo_Natural	Familiar	$.73^*$
Objects_Natural	Familiar	$.85^*$
Common_Perspective	Familiar	$.87^*$
Objects_Familiar	Familiar	$.66^*$
Make_Happy	Liking	$.88^*$
Attractive	Liking	$.87^*$
Colorful	Liking	$.62^*$
Mysterious	Weird	$.68^*$
Strange	Weird	$.93^*$
Light_Natural	Natural	$.86^*$
Color_Natural	Natural	$.81^*$

* p-value less than .05.

Table S6: Factor correlations (ϕ) in the perception model from Visual Realism Dataset.

Factor 1	Factor 2	ϕ
Artistic	Familiar	$.40^*$
Artistic	Weird	$-.06^*$
Artistic	Natural	$.39^*$
Familiar	Weird	$-.58^*$
Familiar	Natural	$.84^*$
Weird	Natural	$-.45^*$

* p-value less than .05.

Table S7: Standardized regression weights (γ) in the perception model from Memorability Dataset.

Dependent variable	Predictor	γ
Liking	Dynamic	$.24^*$
Liking	Artistic	$.66^*$
Liking	Weird	$.27^*$
Liking	Space	-.03
Exciting	Liking	$.87^*$
Interesting	Liking	$.91^*$
Dynamic_Scene	Dynamic	$.97^*$
Static	Dynamic	$.76^*$
Lot_Going_On	Dynamic	$.81^*$
Image_Quality	Artistic	$.66^*$
Strange	Weird	$.84^*$
Mysterious	Weird	$.84^*$
Frightening	Weird	$.63^*$
Pleasant_Scene	Artistic	$.94^*$
Open_Space	Space	$-.82^*$
Cluttered	Space	$.57^*$

* p-value less than .05.the perception model of Memorability Dataset.

Table S8: Factor correlations (ϕ) in the perception model from Memorability Dataset.

Factor 1	Factor 2	ϕ
Dynamic	Artistic	$-.07^*$
Artistic	Weird	$.13^*$
Space	Weird	$.26^*$
Space	Artistic	$.60^*$
Dynamic	Weird	-.03
Dynamic	Space	.04

* p-value less than .05.

Table S9: The loadings of attributes on the 7 major principal components (PC) based on fusion of two datasets. Only strongest loadings (> .5) are shown. The "Cumulative variability" row shows how each PC cumulatively explains the variability of attributes in presented sequence.

Attributes \ PC	1 (Artistic)	2 (Liking)	3 (Familiar)	4 (Dynamic)	5 (Weird)	6 (Natural)	7 (Space)
Image_Sharp	.97						
Image_Quality	.94						
Fine_Details	.69						
Interesting		.89					
Exciting		.86					
Make_Happy		.83					
Objects_Combo_Natural			.85				
Objects_Natural			.79				
Common_Perspective			.53				
Dynamic_Scene				.88			
Storyline				.75			
Static				.66			
Mysterious					.86		
Strange					.72		
Make_Sad					.65		
Frightening					.60		
Light_Natural						.86	
Color_Natural						.64	
Cluttered							.75
Open_Space							-.56
Cumulative variability explained (%)	22.71	38.03	50.39	57.97	61.54	64.46	67.32

Table S10: Standardized regression weights (γ) in the perception model based on the fusion of two datasets.

Dependent variable	Predictor	γ
Liking	Familiar	.46*
Liking	Artistic	.14*
Liking	Dynamic	.35*
Liking	Weird	.44*
Liking	Natural	-.10*
Liking	Space	.14*
Objects_Combo_Natural	Familiar	.67*
Objects_Natural	Familiar	.82*
Common_Perspective	Familiar	.89*
Exciting	Liking	.91*
Interesting	Liking	.82*
Image_Sharp	Artistic	.95*
Image_Quality	Artistic	.96*
Dynamic_Scene	Dynamic	.85*
Storyline	Dynamic	.48*
Static	Dynamic	.68*
Mysterious	Weird	.77*
Strange	Weird	.86*
Light_Natural	Natural	.83*
Color_Natural	Natural	.80*
Cluttered	Space	.44*
Open_space	Space	-.91*

* p-value less than .05.

Table S11: Factor correlations (ϕ) in the perception model based on the fusion of two datasets.

Factor 1	Factor 2	ϕ
Familiar	Artistic	.41*
Familiar	Dynamic	-.02
Familiar	Weird	-.44*
Familiar	Natural	.85*
Artistic	Dynamic	-.07*
Artistic	Weird	.12*
Artistic	Natural	.39*
Dynamic	Weird	.07*
Dynamic	Natural	.06
Weird	Natural	-.41*
Natural	Space	-.19*
Weird	Space	.19*
Dynamic	Space	.14*
Artistic	Space	.27*
Familiar	Space	.15*

* p-value less than .05.

Table S12: Standardized regression weights in complete and incomplete models of Visual Realism (VR) and Memorability (MEM) datasets. "n. a." means the corresponding relation does not show up in the model.

Dependent variable	Predictor	Complete model		Incomplete model	
		VR	MEM	VR	MEM
Liking	Familiar	.27*	.73*	.15*	n.a.
Liking	Artistic	.29*	.31*	.28*	.31*
Liking	Dynamic	.69*	.25*	.74*	.22*
Liking	Weird	.13*	.64*	.03	.27*
Liking	Natural	-.08	-.03	-.03	n.a.
Liking	Space	-.01	-.05	n.a.	-.23*
Objects_Combo_Natural	Familiar	.70*	.56*	.71*	n.a.
Objects_Natural	Familiar	.85*	.80*	.85*	n.a.
Common_Perspective	Familiar	.87*	.89*	.87*	n.a.
Exciting	Liking	.96*	.86*	.99*	.87*
Interesting	Liking	.65*	.93*	n.a.	.91*
Image_Sharp	Artistic	.84*	.95*	.84*	n.a.
Image_Quality	Artistic	.99*	.97*	.99*	.99*
Dynamic_Scene	Dynamic	.87*	.95*	.83*	.95*
Storyline	Dynamic	.57*	.77*	.59*	.78*
Static	Dynamic	.38*	.69*	n.a.	.69*
Mysterious	Weird	.72*	.81*	.73*	.99*
Strange	Weird	.87*	.82*	.86*	.62*
Light_Natural	Natural	.85*	.84*	.85*	n.a.
Color_Natural	Natural	.82*	.82*	.82*	n.a.
Cluttered	Space	-.17*	-.52*	n.a.	-.52*
Open_Space	Space	.99*	.91*	n.a.	.91*

* p-value less than .05.

Table S13: Factor correlations in complete and incomplete models of Visual Realism (VR) and Memorability (MEM) datasets. "n. a." means the corresponding relation does not show up in the model.

Factor1	Factor2	Complete model		Incomplete model	
		VR	MEM	VR	MEM
Familiar	Artistic	.38*	.49*	.38*	n.a.
Familiar	Dynamic	.05	-.07	.05	n.a.
Familiar	Weird	-.56*	-.35*	-.56*	n.a.
Familiar	Natural	.86*	.83*	.86*	n.a.
Artistic	Dynamic	.19*	.15*	.18*	.13*
Artistic	Weird	-.08*	-.12*	-.07	-.17*
Artistic	Natural	.36*	.40*	.36*	n.a.
Dynamic	Weird	.29*	-.01	.35*	-.02
Dynamic	Natural	-.03	n.a.	-.10*	n.a.
Weird	Natural	-.53*	n.a.	-.53*	n.a.
Natural	Space	-.10*	-.25*	n.a.	n.a.
Weird	Space	.03	-.20*	n.a.	-.25
Dynamic	Space	-.11*	-.08*	n.a.	-.08
Artistic	Space	-.11*	-.36*	n.a.	-.36
Familiar	Space	-.16*	-.38*	n.a.	n.a.

* p-value less than .05.

Table S14: The loadings of attributes on the 7 major principal components (PC) based on the fusion of three datasets. Only strongest loadings ($> .4$) are shown. The "Cumulative variability" row shows how each PC cumulatively explains the variability of attributes in presented sequence.

Attributes \ PC	1 (Familiar)	2 (Artistic)	3 (Liking)	4 (Weird)	5 (Dynamic)	6 (Space)	7 (Natural)
Objects_Combo_Natural	.89						
Objects_Natural	.82						
Common_Perspective	.65						
Image_Sharp		.96					
Image_Quality		.94					
Fine_Details		.70					
Interesting			.88				
Exciting			.87				
Make_Happy			.74				
Mysterious				.73			
Strange				.68			
Make_Sad				.65			
Frightening				.65			
Dynamic_Scene					.88		
Storyline					.70		
Static					.69		
Cluttered						.76	
Open_Space						-.48	
Light_Natural							.77
Color_Natural							.56
Cumulative variability explained (%)	23.87	37.72	49.54	56.43	61.12	63.61	66.48

Table S15: Standardized regression weights (γ) in the perception model based on the fusion of three datasets.

Dependent variable	Predictor	γ
Liking	Familiar	.37*
Liking	Artistic	.20*
Liking	Dynamic	.31*
Liking	Weird	.37*
Liking	Natural	-.14*
Liking	Space	.17*
Objects_Combo_Natural	Familiar	.69*
Objects_Natural	Familiar	.84*
Common_Perspective	Familiar	.87*
Exciting	Liking	.91*
Interesting	Liking	.82*
Light_Natural	Natural	.81*
Color_Natural	Natural	.81*
Image_Quality	Artistic	.97*
Image_Sharp	Artistic	.93*
Cluttered	Space	.45*
Open_Space	Space	-.91*
Dynamic_Scene	Dynamic	.85*
Storyline	Dynamic	.49*
Static	Dynamic	.67*
Mysterious	Weird	.74*
Strange	Weird	.88*

* p-value less than .05.

Table S16: Factor correlations (ϕ) in the perception model based on the fusion of three datasets.

Factor 1	Factor 2	ϕ
Familiar	Artistic	.38*
Familiar	Dynamic	.08*
Familiar	Weird	-.44*
Familiar	Natural	.83*
Familiar	Space	.16*
Natural	Artistic	.43*
Natural	Dynamic	-.03
Natural	Weird	-.37*
Artistic	Dynamic	-.05*
Artistic	Weird	.11*
Space	Dynamic	.14*
Space	Artistic	.28*
Space	Natural	.11*
Space	Weird	.17*
Dynamic	Weird	.07*

* p-value less than .05.

Table S17: Area under ROC curve (AUC) results based on MI, perception model, and non-linear SVM [1] for different fusion from Visual Realism Dataset (VR), Memorability Dataset (Mem), and Sentiment Dataset (Senti).

Fused Datasets	MI	Model	SVM
VR+Mem+Senti	.55	.55	.70
VR+Senti	.55	.59	.64
Mem+Senti	.58	.56	.65

[1] The SVM results are from Sec.5.3 in main text.