Expanding Design Creativity with the PHR2 Model: Predicting Hedonic Responses in Architecture

Victor Sardenberg[1,*], Rafael Perrone[1]

[1] *Universidade Presbiteriana Mackenzie, Rua Itambé, 185 01239-001 São Paulo, Brazil.*

Abstract

This study advances computational aesthetics in architecture by refining the Predicted Hedonic Response (PHR) model to analyze and predict aesthetic preferences across diverse architectural typologies. Utilizing a dataset of 12,025 architectural images from the Aesthetic Visual Analysis (AVA) dataset, this research integrates fractal dimension, visual complexity, depth, and brightness as aesthetic criteria. The PHR2 model, powered by computer vision and artificial neural networks, captures elaborate relationships between these quantitative attributes and perceived aesthetic appeal. The study also explores the interplay between order, complexity, and perception, proposing a framework that enables aesthetic exploration. The findings provide insights into how computational methods can navigate uncharted territories of architectural form and perception. This research contributes to expanding architectural aesthetics through data-driven exploration of spatial and visual complexity in design.

Keywords

Computational aesthetics, Hedonic response, Aesthetic visual analysis, artificial neural network [1]

1. Introduction

In recent decades, architects utilizing computational tools like parametric modeling and artificial models as generative adversarial networks and diffusion models have been able to produce hundreds of thousands of design variations. Usually, architects rely on quantitative analyses such as structural and environmental behaviors to rank designs and assist them in decision-making. However, most criteria rely on structure and use to define the best solutions. However, aesthetics must also be included to complement other criteria. This paper describes the training for developing a computational aesthetics framework to predict the aesthetic preferences of subjects towards architectural images. It upgrades a previous framework [1] and complements it with more image quantitative aesthetics criteria such as depth, complexity, brightness, and fractal dimension. Moreover, the training utilizes a larger image dataset of 12.025 images related to architecture from the popular AVA dataset [2]. Its goal is to generalize the model to more types of architectural images and increase its accuracy.

A hedonic response is a reaction by a subject liking or disliking an object. The predicted hedonic response (PHR) model was introduced in 2022 and is an artificial neural network trained to predict how a specific group of subjects prefers images of architectural pavilions [3]. In its first version (nicknamed here PHR1.0) (Figure 1), it inputs parts and their relations that are recognized by the computer vision algorithm MSER (Maximally Stable Extremal Regions)[4]. In a second version from 2023 (Nicknamed PHR1.1) (Figure 2)[5], the neural network SAM (Segment Anything Model) [6] was also incorporated to recognize parts, and the network size was increased. The original version of the PHR focused on a specific pool of users, and, therefore, the aesthetic measure was calibrated toward their preferences. The PHR model inputs the number of parts, their relations, the aesthetic measure, and the calibrated aesthetic measure to output a predicted hedonic response from 0 to 10.

---

[*]Corresponding author.

✉ victor.sardenberg@mackenzista.com.br (V. Sardenberg); rafaelantonio.perrone@mackenzie.br (R. Perrone);

🆔 0000-0001-6089-0189 (V. Sardenberg); 0000-0003-0651-913X (R. Perrone);

This paper presents the development of this model to (1) increase its accuracy - by using more aesthetic criteria such as depth and compression complexity and by introducing a new larger network with a better-fit architecture – and (2) its generalization – by training on a larger and more eclectic dataset of images.
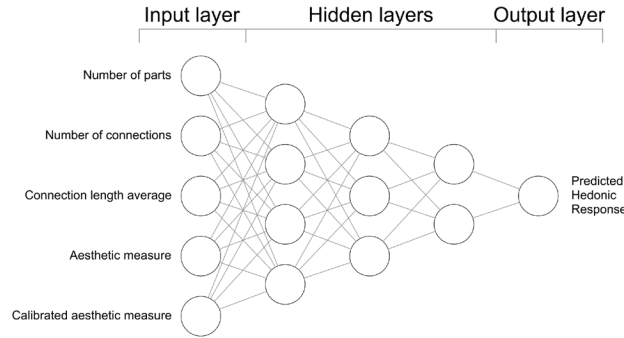
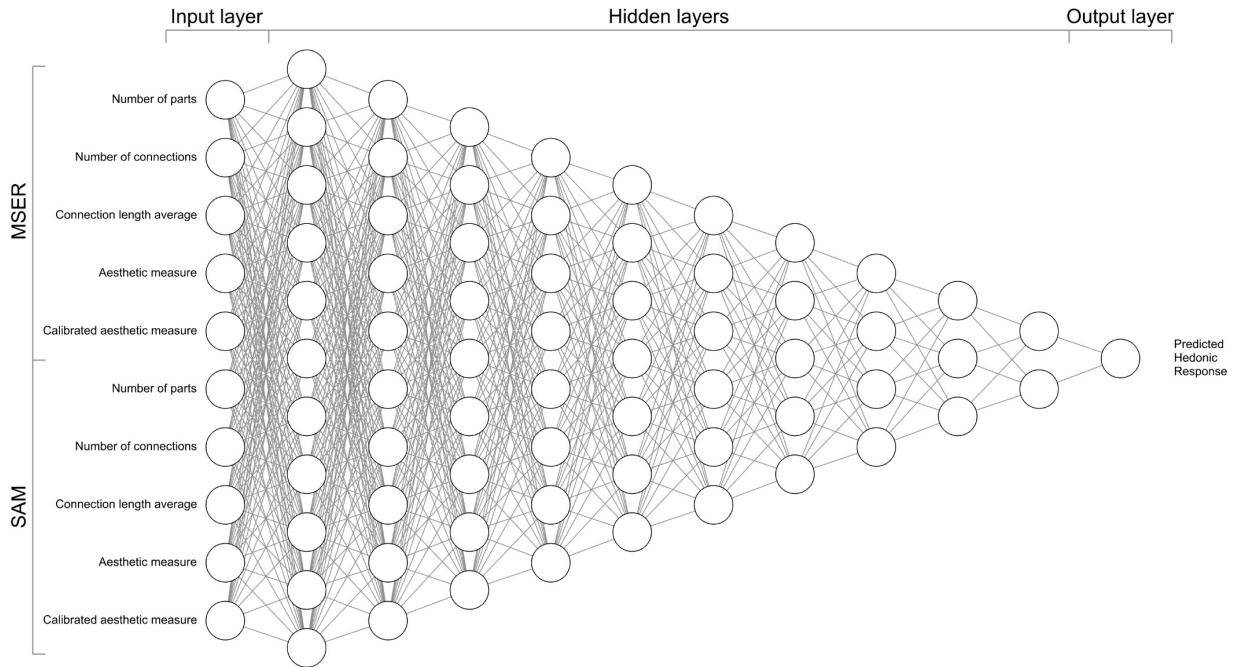**Figure 1:** Neural network architecture of the PHR1.0 model.

Figure 2: Neural network architecture of the PHR1.1 model.

2. Methods

This paper presents methods to (1) use a larger dataset, (2) implement more aesthetic criteria, and (3) train the network.

2.1. Image dataset

The PHR1.0 and PHR1.1 were trained using an image dataset of 141 perspective images of 87 pavilion designs. These were proposals for the MoMA PS1 Young Architecture Program competition for the same site in Queens, NY, USA. The models are only accurate for pavilion images because they were trained on this specific and narrow dataset.

The PHR2 was trained in a much larger image dataset of 12,025 pictures. These pictures are part of AVA, a large-scale database created to train and test aesthetic visual analysis. AVA images were

collected from the website www.dpchallenge.com, an online photo community. The dataset contains approximately 255,000 images from 1,447 challenges. A challenge was a specific topic, like "Cityscape," where participants would upload images relating to it and rank them. The aesthetic ratings range from 78 to 549 per image, averaging 210 [2]. Beyond the challenge name, semantic annotations belong to 66 textual tags. This research used all images with the tag architecture, resulting in a dataset of 12,025 images. This architectural images dataset is much more general than the previous one, which contained only images of pavilions.

2.2. Aesthetic criteria

There is a body of work applying quantitative methods to evaluate architectural aesthetics. The first one is Kiemle´s dissertation, advised by Max Bense, which applies information processing theory to evaluate if a subject is overwhelmed or bored by a building. [7]. Currently, there is considerable interest in the field of Computational Aesthetics, which may be defined as "research of computational methods that can make applicable aesthetic decisions in a similar fashion as humans can" [8]. PHR1.0 and 1.1 used computer vision to recognize parts and analyze their relationship to calculate an aesthetic measure and input these parameters into the model. PHR2 utilizes complementary criteria beyond those. These are:

Aesthetic measure (AM) is a quantitative formula originally introduced by G. D. Birkhoff [9] to evaluate the balance between order and complexity in an object's aesthetic appeal. In this research's computational framework, the aesthetic measure was explicitly adapted for architectural images, defining order as the number of visual connections between parts and their average length to capture the spatial organization of parts (e.g., dispersed versus compact layouts) within a composition and complexity as the number of distinct parts. Finally, the values are normalized by the square root of the number of pixels to make comparisons consistent among different image sizes and resolutions. Applying this principle makes it possible to systematically assess and compare the visual effectiveness of different architectural designs.

To compute the aesthetic measure, images are analyzed using computer vision algorithms such as Maximally Stable Extremal Regions (MSER) and the Segment Anything Model (SAM). These methods detect and segment architectural components, allowing the identification of distinct parts within an image. The Aesthetic Measure is then calculated using the formula:

$$Aesthetic\ Measure = \frac{Order}{Complexity} \tag{1}$$
$$= \frac{\text{Connection Length Average} * \text{Number of connections}}{Number\ of\ parts * \sqrt{Number\ of\ pixels}},$$

Fractal dimension measures how intricate and self-similar a pattern is across different scales [10]. It is used to analyze and quantify the visual complexity of architectural images by assessing how much detail is present at various magnification levels [11], [12]. The core idea is that specific structures, such as natural patterns or architectural designs, exhibit intricated geometric characteristics at multiple scales, which can be captured mathematically using fractal dimension analysis.

The box-counting method is applied to measure fractal complexity, which overlays a grid on an image and counts the number of occupied boxes at different scales. As the grid size decreases, the method observes how the complexity of the image changes. The fractal dimension quantifies how well a figure fills space, with values ranging between 1 (a simple line) and 2 (a completely filled plane). Fractal complexity is integrated to evaluate architectural images' visual richness and structural intricacy.

Compression complexity measures visual complexity based on how efficiently an image can be compressed using data compression techniques [13]. The fundamental idea is that the more structured and predictable an image is, the more it can be compressed, while highly intricate or chaotic images require more storage space due to their lack of repetitive patterns.

This method applies a lossless data compression algorithm, specifically PNG compression, to an image file. The compression ratio - the size of the compressed file relative to the original - serves as an indicator of complexity. A lower compression ratio suggests that the image contains highly repetitive or uniform patterns, making it easier to encode efficiently. Conversely, a higher compression ratio indicates greater visual complexity, with more details requiring additional storage. This measure is particularly useful in distinguishing between minimalist, highly ordered designs and more intricate, texturally rich compositions.

Brightness is a fundamental visual property that affects how architectural images are perceived. It refers to the overall luminance of an image, which influences clarity, contrast, and the visibility of architectural elements. Measuring brightness is done by converting the image to grayscale and computing the average luminance across all pixels. Higher average luminance indicates a brighter image, while lower values suggest a darker one. This method ensures that an objective measure of brightness is captured without being affected by color variations. It is instrumental in understanding how light conditions contribute to a design's perception of space, depth, and visual harmony.

Depth in architectural images refers to the perception of three-dimensional space within a two-dimensional representation. It is crucial in how viewers interpret spatial relationships, perspective, and composition. Images are analyzed using ZoeDepth [14] to measure depth, which determines the relative distance of objects within a scene. From this depth gradient, values are rounded to near, mid, and far distances, and their percentage is further used for analysis. Depth analysis helps distinguish between different architectural styles, from expansive, open environments to enclosed, intimate spaces.

2.3. Artificial neural network architecture

PHR1.0 (Figure 1) and PHR1.1 (Figure 2) are small neural networks, containing, respectively, 30 neurons and 40 connections and 76 neurons and 550 connections. Their number of inputs defined the size of these networks: PHR1.0 contains, as input neurons, the number of parts, connections, connection length, aesthetic measures, and calibrated aesthetic measures extracted from MSER, resulting in 5 inputs. PHR1.1 has double the neurons in the input layer because it calculates the same inputs from MSER and SAM. Each layer reduces one neuron until the output layer, which consists only of one neuron, the predicted hedonic response.

The PHR2 network architecture is significantly different. It has in its input layer 14 aesthetic characteristics:

- Brightness;
- Near depth percentage;
- Mid-depth percentage;
- Far depth percentage;
- Fractal dimension;
- Number of parts from SAM;
- Number of connections from SAM;
- Connection length average from SAM;
- Aesthetic measure from SAM;
- Number of parts from MSER;
- Number of connections from MSER;
- Connection length average from MSER;
- Aesthetic measure from MSER;

Data is passed from the initial 14 neurons to the next layer, which contains 128 neurons to capture more intricate relations. It progressively reduces the number of neurons of the subsequent layers by half, allowing a hierarchical feature extraction process, where raw inputs are gradually transformed into abstract representations before reaching the final output. This architecture is well-suited for

complex data patterns but requires sufficient training data to perform effectively, which is possible because of the significantly larger dataset of 12,025 images.
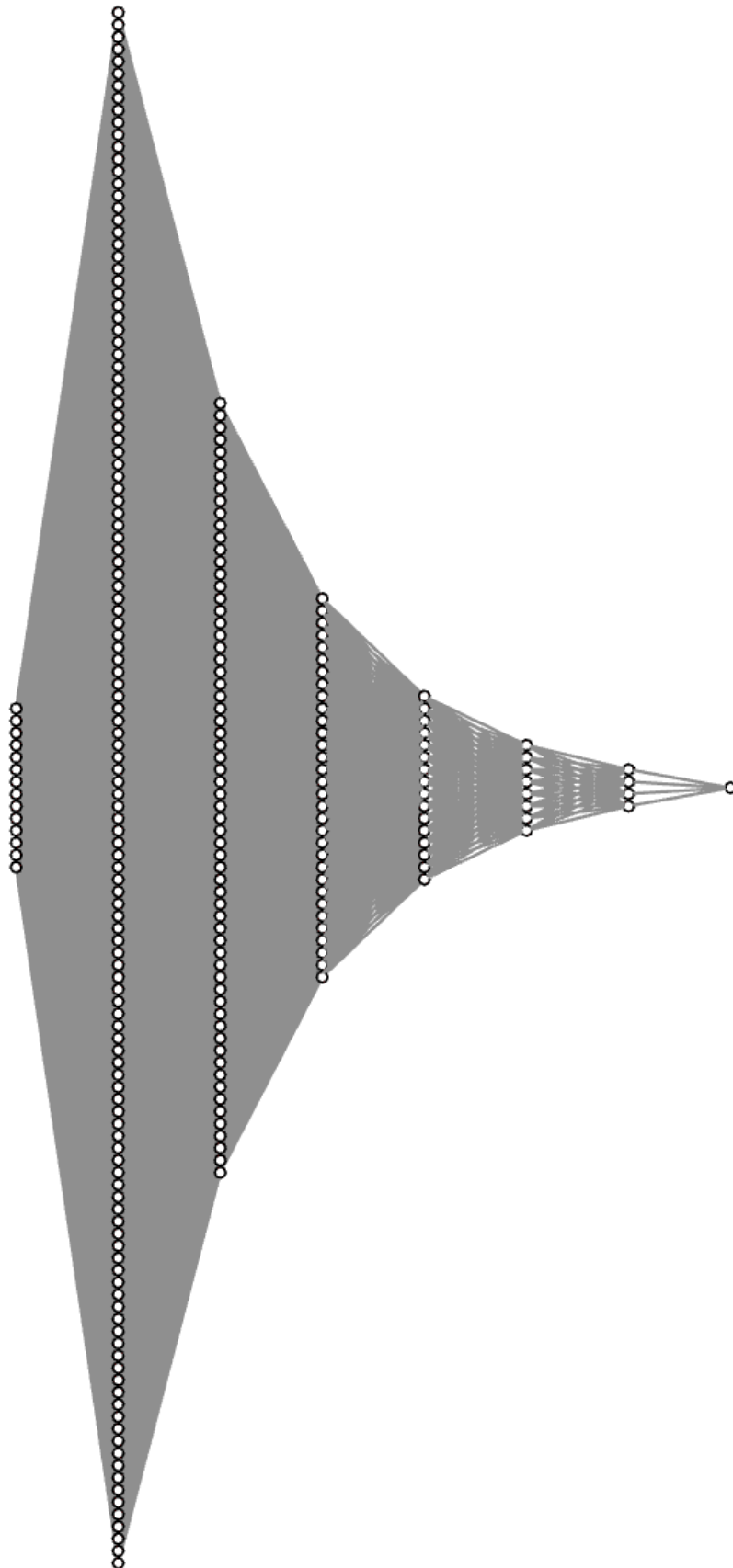

Figure 3: Neural network architecture of the PHR2 model.

Table 1
Comparison of network architecture.

Model	Neurons	Connections	Dataset size
PHR1.0	30	40	141
PHR1.1	76	550	141
PHR2	253	10,916	12,025

The scatter plot in Figure 4 visualizes the performance of the PHR2 model by comparing real outputs with predicted values. An ideal prediction model would yield points tightly clustered along the diagonal $y = x$, indicating perfect agreement between predicted and actual values. All datasets are split into 75% for training and 25% for testing.

The distribution of points in the plot (Figure 4) suggests that PHR2 demonstrates a strong correlation between real and predicted values. Most points align along the diagonal, suggesting that the model effectively captures the general trend in the data. However, the spread of points indicates some error. This accuracy is acceptable because this is a measure that is not entirely objective and because the PHR should be used to compare results instead of offering a definitive prediction of preference for an image. Partial code is available at: https://github.com/vsardenberg/PHR2

All trained models and Rhino/Grasshopper required to reproduce the experiment are publicly available at:

http://www.victorsardenberg.com/Aesthetics_Framework/PHR2_Analysis_and_Mapping.zip
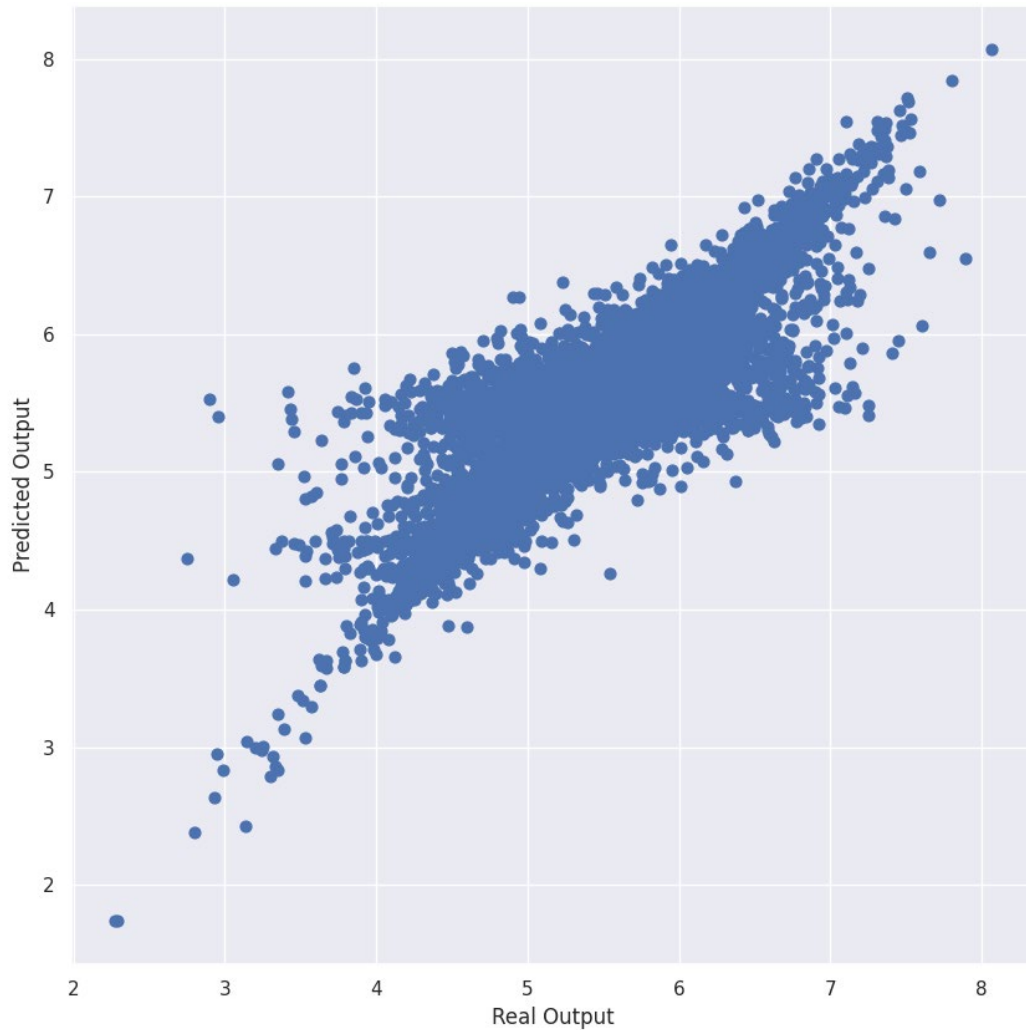
Figure 4: Correlation between Predicted and Real outputs of the PHR2 model.

3. Results

The performance of three predictive models, PHR1.0, PHR1.1, and PHR2, was evaluated using three key metrics: Root Mean Squared Error (RMSE), R² Score (R²), and Accuracy (%). Each model was tested on its respective dataset, with PHR1.0 and PHR1.1 trained and evaluated on MoMA PS1 pavilion images, while PHR2 was tested on AVA architectural images (Table 2). The results indicate significant improvements between iterations of the models, highlighting the effect of dataset characteristics and training steps on predictive performance. The number of training steps varied between models and was determined empirically. The training was stopped when performance metrics such as RMSE and R² no longer showed meaningful improvement on the validation set.

Table 2
Performance of the three models in their dataset.

Model	Steps	RMSE	R2	Accuracy
PHR1.0	20000	1.10	0.05	85%
PHR1.1	30000	0.32	0.93	95%
PHR2	33000	0.35	0.71	93%

A comparative analysis of PHR1.0 and PHR1.1 reveals a substantial performance enhancement in the latter. PHR1.0, trained with 20,000 steps, achieved an RMSE of 1.10, a low R² score of 0.05, and an accuracy of 85%. These values indicate that the model struggled to establish strong predictive relationships within the dataset. In contrast, PHR1.1, trained with an increased 30,000 steps, exhibited a dramatic reduction in RMSE to 0.32 and an increase in R² to 0.93, demonstrating a much stronger correlation between predictions and ground truth values. Furthermore, accuracy improved to 95%, marking a substantial refinement in predictive precision. This improvement is also the product of having a more extensive network and using MSER and SAM as inputs.

When analyzing PHR2, which was trained and tested on AVA architectural images, the model demonstrated competitive but slightly lower performance than PHR1.1. With an RMSE of 0.35, an R² score of 0.71, and an accuracy of 93%, PHR2 outperformed PHR1.0 significantly but did not achieve the predictive strength of PHR1.1. Interestingly, despite being trained for 33,000 steps, a more significant number than PHR1.0 and PHR1.1, PHR2 did not surpass the latter's performance. This discrepancy may be attributed to differences in dataset richness, indicating that the AVA dataset contains more significant visual variability.

3.1. Network Architecture and Its Influence on Performance

The network architecture plays a crucial role in explaining the observed performance differences. PHR1.0 and PHR1.1 are relatively small networks. The inputs of these models were limited to five and ten features, respectively, extracted from MSER and SAM, such as the number of parts, connections, connection length, aesthetic measure, and calibrated aesthetic measure. Each layer progressively reduces the number of neurons until a single output neuron predicts the hedonic response.

On the other hand, PHR2 employs a significantly different and more complex architecture. Its input layer consists of 14 aesthetic characteristics, including brightness, depth percentages, fractal dimension, and multiple extracted features from MSER and SAM. The subsequent layer expands to 128 neurons to capture intricate relationships, followed by a progressive reduction in neurons through deeper layers. This hierarchical structure facilitates more feature extraction, enabling the model to better represent complex aesthetic patterns. Additionally, PHR2 was trained on a much larger dataset of 12,025 images compared to the 141 images used for PHR1.0 and PHR1.1. This substantial dataset size likely contributed to the model's ability to generalize better. The comparison

of PHR1.0, PHR1.1, and PHR2 highlights the importance of dataset size and network size in aesthetic analysis.

4. Discussion

The comparative evaluation of the PHR models highlights critical insights regarding architectural aesthetic predictions using artificial neural networks. The progressive evolution from PHR1.0 to PHR2 demonstrates the impact of neural network size, dataset size, and feature selection on model performance. While PHR1.1 significantly improved predictive accuracy over PHR1.0 within the MoMA PS1 dataset, the generalization capabilities of PHR2 reveal the potential benefits and challenges of applying a broader dataset to train neural networks for aesthetic evaluation.

The scatter plot visualization further illustrates PHR2's predictive performance, showing a strong correlation between predicted and real values, with a clear alignment along the diagonal. However, deviations suggest that the model has room for improvement. These discrepancies may stem from underlying biases in the dataset distribution, where certain architectural styles are underrepresented, affecting the network's ability to generalize across the full aesthetic spectrum. Future iterations of the model may benefit from additional data augmentation techniques and improved training strategies.

Overall, these findings underscore the trade-off between specialization and generalization in computational aesthetics models. While smaller, domain-specific networks such as PHR1.1 can achieve high accuracy within a constrained dataset, their applicability beyond that domain remains limited. In contrast, broader models such as PHR2 exhibit greater versatility but may require further refinement to enhance their predictive precision across diverse architectural contexts. The application of computational aesthetics in architecture continues to evolve, offering promising avenues for assessing and quantifying aesthetic perception through machine learning.

The goal of computational aesthetics applied to architecture should not be to optimize design towards the most popular one because it will not necessarily be the most exciting or interesting design. Aesthetics as a criterion is more complicated than simply minimizing stress in building components. In the current stage of this research, the position is to use computational aesthetics to navigate the myriad of design variations and present solutions to the architect that are counter-intuitive or not imagined at the early design stages. Computational aesthetics should be a tool to expand the creativity of architects, introducing designs that are simultaneously aesthetically pleasing and not cliché. To achieve this goal, computational aesthetics can be applied to produce a map of generative and parametric models that rank designs and present them to architects (Figure 5).
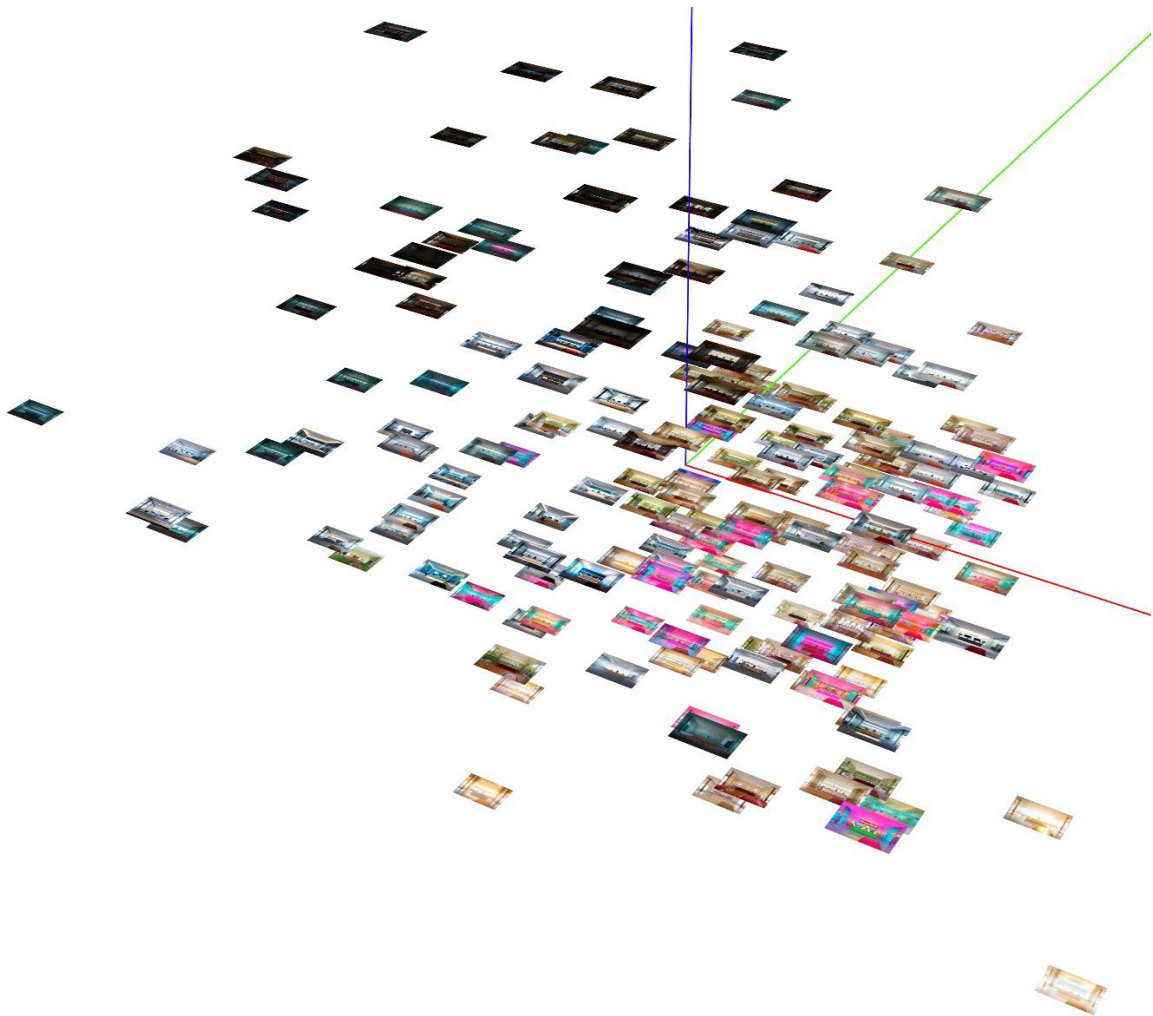
Figure 5: Map of the solution space of images generated by Stable Diffusion.

To produce such a map, Principal Component Analysis is used to flatten the multi-dimensional analysis that utilizes the 14 aesthetic characteristics (i.e., brightness, fractal dimension, depth) into two dimensions, clustering together similar designs and moving further those that are dissimilar. The PHR is introduced as a 3rd axis that puts the highest-ranked designs at the top. This strategy allows architects to visualize first the most appealing designs and dispatch those that rank too low. In this application, it is still possible to visualize very discrepant designs, allowing architects to be presented with new, uneasy, and dormant design solutions. The application of the PHR2 as a computational aesthetics tool allows human creativity to be boosted by artificial creativity, expanding the use of artificial intelligence beyond optimization and toward a hybrid imagination – half human and half machine.

Future work may include:

- Applying segmentation algorithms to preprocess images in training and inference to extract buildings from their context and exclude skies that may influence the complexity of the images.
- Incorporate exposure normalization techniques or separate indoor/outdoor classifiers to refine the metric further.
- Testing the application of the PHR2 with students and practitioners will give feedback on its potential and limitations.

Acknowledgments

Declaration on Generative AI

While preparing this work, the authors used ChatGPT 4o to produce a draft of the text entirely rewritten by the authors and Grammarly to check grammar and spelling. After using these tools, the authors reviewed and edited the content as needed and take full responsibility for the publication's content.

References

[1] V. Sardenberg, I. Guatelli, and M. Becker, "A computational framework for aesthetic preferences in architecture using computer vision and artificial neural networks," *International Journal of Architectural Computing*, p. 14780771241279350, Sep. 2024, doi: 10.1177/14780771241279350.

[2] N. Murray, L. Marchesotti, and F. Perronnin, "AVA: A large-scale database for aesthetic visual analysis," in *2012 IEEE Conference on Computer Vision and Pattern Recognition*, Jun. 2012, pp. 2408–2415. doi: 10.1109/CVPR.2012.6247954.

[3] V. Sardenberg and M. Becker, "Computational Quantitative Aesthetics Evaluation - Evaluating architectural images using computer vision, machine learning and social media," in *Pak, B, Wurzer, G and Stouffs, R (eds.), Co-creating the Future: Inclusion in and through Design - Proceedings of the 40th Conference on Education and Research in Computer Aided Architectural Design in Europe (eCAADe 2022) - Volume 2, Ghent, 13-16 September 2022, pp. 567–574*, CUMINCAD, 2022. Accessed: Dec. 14, 2022. [Online]. Available: http://papers.cumincad.org/cgi-bin/works/paper/ecaade2022_75

[4] J. Matas, O. Chum, M. Urban, and T. Pajdla, "Robust wide-baseline stereo from maximally stable extremal regions," *Image and Vision Computing*, vol. 22, no. 10, pp. 761–767, Sep. 2004, doi: 10.1016/j.imavis.2004.02.006.

[5] V. Sardenberg and M. Becker, "Aesthetics as a Criterion: Navigating Solution Spaces Utilizing Computer Vision, the Aesthetic Measure, and Artificial Neural Networks," in *2023 Annual Modeling and Simulation Conference (ANNSIM)*, May 2023, pp. 496–507. [Online]. Available: https://ieeexplore.ieee.org/document/10155349

[6] A. Kirillov *et al.*, "Segment Anything," Apr. 05, 2023, *arXiv*: arXiv:2304.02643. doi: 10.48550/arXiv.2304.02643.

[7] M. Kiemle, *Ästhetische Probleme der Architektur unter dem Aspekt der Informationsästhetik*. Schnelle, 1967.

[8] F. Hoenig, *Defining Computational Aesthetics*. The Eurographics Association, 2005. doi: 10.2312/COMPAESTH/COMPAESTH05/013-018.

[9] G. D. Birkhoff, *Aesthetic Measure*. Cambridge, MA: Harvard University Press, 1933.

[10] B. Mandelbrot, "How Long Is the Coast of Britain? Statistical Self-Similarity and Fractional Dimension," *Science*, vol. 156, no. 3775, pp. 636–638, May 1967, doi: 10.1126/science.156.3775.636.

[11] M. Kulcke and W. Lorenz, "Spherical Box-Counting: Combining 360° Panoramas with Fractal Analysis," *Fractal and Fractional*, vol. 7, pp. 1–20, Apr. 2023, doi: 10.3390/fractalfract7040327.

[12] M. J. Ostwald and J. Vaughan, *The Fractal Dimension of Architecture*. Cham: Springer International Publishing, 2016. doi: 10.1007/978-3-319-32426-5.

[13] G. Birkin, "Aesthetic complexity: practice and perception in art & design," doctoral, Nottingham Trent University, 2010. Accessed: Jan. 17, 2023. [Online]. Available: http://irep.ntu.ac.uk/id/eprint/91/

[14] S. F. Bhat, R. Birkl, D. Wofk, P. Wonka, and M. Müller, "ZoeDepth: Zero-shot Transfer by Combining Relative and Metric Depth," Feb. 23, 2023, *arXiv*: arXiv:2302.12288. doi: 10.48550/arXiv.2302.12288.