Aesthetic Enhancement: Automated Image Composition Suggestions

Haleigh Oeser
Brigham Young University
Provo, UT 84602 USA
oeserh@byu.edu

Anson Savage
Brigham Young University
Provo, UT 84602 USA
ansonsav@byu.edu

Riley Sinema
Brigham Young University
Provo, UT 84602 USA
rsinema@byu.edu

Abstract

Have you ever taken a picture and thought *It could be better... but how?* In photography, there are many guidelines for producing images with higher aesthetic quality by tuning various settings. However, not everyone has the time or interest to commit to learning and mastering the ability to create such an image. We propose a system that analyzes an image and suggests possible transformations to increase the aesthetic quality. Using an LLM to interact with the user and a ViT to assess the aesthetic quality of an image, our model can interpret and apply principles of photographic composition and, based on its initial analysis of the input image, will determine what changes will increase the aesthetic quality of the image. It will creatively navigate a space of possible transformations and select the best transformation by evaluating its aesthetic score, which is calculated based on ViT predictions for quality, composition, light, color, depth-of-field, etc. This enables users, even those with limited experience in photography or photo editing, to generate high-quality images that score well on aesthetic metrics and align with photography composition principles learned and selected by our model.

Introduction

Photography is a field that is easy for beginners to explore, but quickly becomes complex as one seeks to refine and enhance photo quality. As an art form, photography involves being able to look at the world around you and discover a snapshot that tells a story, but for the story to be compelling it is best to maximize the quality of the photo in several ways. One of the best ways to increase the impact of a photo is to increase its aesthetic quality, captivating viewers and inviting them to reflect on and uncover the story the photo conveys. While there is a significant amount of subjectivity involved in determining whether or not a photo is pleasing to an individual, there are certain guidelines set by photography professionals that help create images that are visually striking, aesthetically pleasing, and more effective at conveying a meaningful narrative.

These guidelines include photography composition rules, which offer principles to consider when capturing a photo and key elements to refine during post-processing. These principles encompass aspects such as where the subject is placed within the frame, the use of lines to guide the viewer's eye, and the control of lighting and focus to highlight key elements. Mastering these numerous rules takes time and effort, and often money as well from an aspiring photographer, and being able to internalize the rules to intentionally subvert them in a way that draws notice while remaining aesthetic requires an even higher level of dedication, time, and often money to get better equipment, travel to places where you might capture photos, and more. For professional photographers, this sacrifice is justified, as it directly supports their professional careers and source of income, but most amateurs do not have the time necessary to devote to mastering all of these principles.

Photography was once reserved for special occasions, but with the widespread availability of smartphones capable of high-quality images, it has become an everyday activity. As a result, basic photography skills have grown increasingly valuable, and many people attempt to create visually appealing images despite lacking formal training or deep knowledge of photographic principles. We designed a system that aims to help such people who want to produce more aesthetic images that align with photography composition principles without needing to dedicate as much time to doing so. Our system requires an input image that it will attempt to transform into a more aesthetic image, and produce suggestions to the user on how they might be able to reproduce the transformations to better their own photography skills. Our process involves three main steps: analyzing an image to identify potential aesthetic improvements, applying the proposed transformations, and evaluating the resulting image to determine whether it surpasses the original in aesthetic quality. It outputs the transformed image for the user to assess for themselves if it aligns with their personal thoughts on aesthetics, and gives the transformations in natural language aided by an LLM for the user to learn how to adjust their images to produce more aesthetics photos. Together, our system and its users can creatively achieve a more aesthetic photo by using both personal preferences and guidance from established photography composition principles.

Methods

Creative Inspiration

In designing a system that has the potential to achieve creativity, we took inspiration from Wiggins (2006) to define the search space to navigate for potential creative transformations. The space of all possible concepts $\mathscr{C}$ includes all

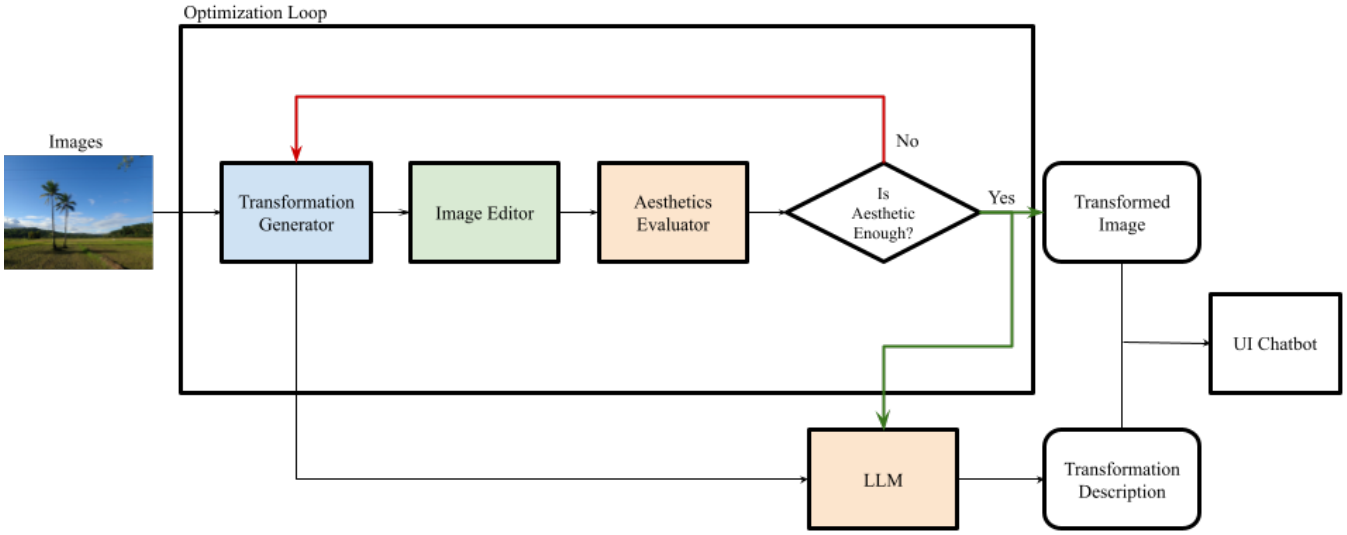

Figure 1: Architecture diagram.

possible photographic images. We define our rule set $\mathscr{R}$ as the set of defined composition rules for photography guiding the design of a photograph. Our goal is to navigate the space of possible transformation combinations, $\mathscr{T}$, to generate a new image that achieves a higher level of aesthetic quality than the original image. Finally, to evaluate how well the transformation is able to create a more aesthetic image, our evaluation rule set $\mathscr{E}$ computes an aesthetic score for evaluation criteria on whether the transformation improves the image. The transformation and evaluation of our system aim to promote creative results that suggest edits that generally follow accepted rules of photography composition while specifically targeting transformations that correlate to what humans perceive to be more aesthetic.

Architecture Overview

Our system consists of several components that work together to optimize an image and return both the resulting image and the suggested transformations to the user. To optimize the image, we have an algorithm designed to suggest and perform transformations on the input image, which is then evaluated by a model fine-tuned to calculate aesthetic scores on both the original and edited image. If the new image improves the aesthetic score of the original image, it is returned to the user along with the transformations performed to improve the image, which are passed through a standard LLM model to produce a description of the transformations in natural language to assist the user in making decisions to improve their photography.

Transformation Exploration

Our goal is to find a set of edits, or "transformations," to the image such that it maximizes s using the aesthetic evaluator A ($A(I) = s$, as described in the Aesthetics Evaluator section).

We treat a transformation, $\mathbf{t}$, as a tuple of operations that can be performed on an image: where

$$\mathbf{t} = (t_0, t_1, \ldots, t_n) \in T$$

where each $t_i \in \mathbb{R}$. For example, if the $i = 0$ dimension of $\mathbf{t}$ represents a *LeftCrop* operation, then t_0 represents the fraction of the image that is cropped from the left. Transformations were defined for each of the following:

- **LeftCrop**: Crops a fraction of the image width from the left side. The crop fraction ranges from $[0.0, 0.49]$, with a default of 0.0 (no crop).
- **RightCrop**: Crops a fraction of the image width from the right side. The crop fraction ranges from $[0.0, 0.49]$, with a default of 0.0 (no crop).
- **TopCrop**: Crops a fraction of the image height from the top. The crop fraction ranges from $[0.0, 0.49]$, with a default of 0.0 (no crop).
- **BottomCrop**: Crops a fraction of the image height from the bottom. The crop fraction ranges from $[0.0, 0.49]$, with a default of 0.0 (no crop).
- **Brightness**: Adjusts the image brightness via a scaling factor. The factor ranges from $[0.0, 2.0]$, with a default of 1.0 (no change).
- **Contrast**: Modifies the image contrast using a scalar multiplier. The multiplier ranges from $[0.0, 2.0]$, with a default of 1.0 (no change).
- **Color**: Changes the image's color saturation. The saturation factor ranges from $[0.0, 2.0]$, with a default of 1.0 (no change).
- **Rotate**: Rotates the image by a number of degrees (clockwise). The angle ranges from $[-30.0°, 30.0°]$, with a default of 0.0° (no rotation).
- **DepthOfField**: Applies a depth-aware background blur using a U^2-Net model (Qin et al. 2020) to segment the


Figure 2: Examples of one of our test images showing a suggested transformation and the associated improved scores. Left image is the original image for the "Dog" set, which has a CLIP score of 3.491 and right image shows the best human-perceived image, which has a CLIP score of 3.688.

foreground. The blur intensity parameter ranges from $[0.0, 2.0]$, with a default of 0.0 (no blur).

- **Vignette**: Applies a radial vignette effect that darkens the corners of the image. The vignette strength ranges from $[0.0, 1.0]$, with a default of 0.0 (no vignette).

With $n = 11$, then T is the subset of $\mathbb{R}^{11}$ where each dimension is constrained as described above. So

$$\mathbf{t} \in T \subset \mathbb{R}^{11},$$

and we can now treat this as a standard optimization problem. Given an image editor $E(I, \mathbf{t})$ that applies transformation $\mathbf{t} \in T$ to an image I to produce a new image I^*, and an aesthetic evaluator $A(I^*)$ that returns a scalar score s, our goal becomes to find the transformation $\mathbf{t}^*$ that produces the most aesthetically pleasing result. Formally, this is an optimization problem seeking to find $\mathbf{t}^*$:

$$\mathbf{t}^* = \arg\max_{\mathbf{t} \in T} A\big(E(I, \mathbf{t})\big).$$

That is, we are searching for the transformation $\mathbf{t}^*$ within the feasible set T that maximizes the aesthetic score of the resulting image. In the following sections, we describe two optimization strategies we used to approximate this argmax.

Genetic Algorithm To explore the transformation space, we implemented a genetic algorithm that treats image transformations as a "genome." By mutating these sequences of transformations, we could try and improve image aesthetics. This is a gradient-free optimization method, so it has the advantage that A does not need to be differentiable with respect to the transformation operations.

Each "genome" in our genetic algorithm consists of a set of nucleotides, where each nucleotide corresponds to one of the image transformation operations described above. The genetic algorithm operates as follows:

1. **Population Initialization**: We start with a small population of randomly initialized transformation sets, each containing default transformation values. (line 1)
2. **Fitness Evaluation**: Each member of the population is evaluated using the aesthetic scorer A, which assigns a fitness score to each transformed image. (line 2)
3. **Selection**: Members are selected for reproduction with probability proportional to their fitness scores (or some user-definable exponent b of the fitness scores; in our implementation we use $b = 2$), implementing a "survival of the most aesthetic" mechanism. (lines 5–6)
4. **Crossover**: Selected members exchange each half of their transformation parameters with probability equal to the crossover rate. (lines 7–12)
5. **Mutation**: Each transformation parameter has a chance (the mutation rate) of being perturbed by a Gaussian random draw; out-of-bounds values are CLIPped. (lines 14–21)
6. **Population Management**: After crossover and mutation, we keep only the top-P fittest members. (line 23)

The algorithm terminates when the patience p iterations have elapsed without improvement. See Algorithm 1 for the full pseudocode.

Gradient-approximation based Optimization While genetic algorithms provide robust exploration, gradient-approximation based methods have the potential to offer a more directed optimization approach. Since the aesthetic scoring function A is non-differentiable with respect to the variety of image transformations we applied, we implemented gradient approximation techniques:

1. **Parameter Space**: We define a continuous parameter vector $\mathbf{t} \in \mathbb{R}^n$ corresponding to our transformation parameters.
2. **Gradient Approximation**: We implemented two primary methods:
   (a) **Finite Differences** approximates partial derivatives by perturbing each parameter slightly:

$$\frac{\partial A}{\partial t_i} \approx \frac{A(t_i + \varepsilon_i) - A(t_i - \varepsilon_i)}{2\varepsilon}$$

where ε_i is a small perturbation in the direction of the i-th parameter. Note that each gradient evaluation has time complexity $O(n)$ (where n is the dimensionality of t). However, we can take advantage of the fact that a separate loss is computed along each dimension to compute a dimension specific gradient if desired (see Figure 6).

(b) **Simultaneous Perturbation Stochastic Approximation (SPSA)** perturbs all parameters simultaneously:

$$\frac{\partial A}{\partial t_i} \approx \frac{A(\mathbf{t}+\boldsymbol{\varepsilon}) - A(\mathbf{t}-\boldsymbol{\varepsilon})}{2\varepsilon_i}$$

where $\boldsymbol{\varepsilon}$ is a random perturbation vector. SPSA offers computational efficiency as it requires only two function evaluations regardless of parameter dimensionality. However, the gradients that it produces are noisy approximations, and in practice we found that although it is $\frac{n}{2} = 5.5$ times faster at computing gradients at each step than Finite Differences, we generally observed no significant improvement in aesthetic quality by using this method to traverse the loss space.

3. **Optimization Algorithms**: For parameter updates we used the Adam optimizer with $\beta_1 = 0.9$, $\beta_2 = 0.999$, and a learning rate typically between $[0.01, 0.1]$.
4. **Constraint Handling**: Each transformation parameter has constraints (e.g., crop values between 0 and 0.49). We implemented constraint projection by clipping parameters to their valid ranges after each optimization step.

This was repeated in a typical optimization loop (pseudocode in Algorithm 2).

Aesthetics Evaluator

We employed a fine-tuned version of the vision encoder of CLIP (Radford et al. 2021) to evaluate the aesthetics of an image. Specifically, we utilized the Vision Transformer (ViT) variant of CLIP due to its strong performance on visual reasoning tasks.

Dataset The dataset used for fine-tuning was the Personalized image Aesthetics database with Rich Attributes (PARA) (Yang et al. 2022). PARA contains a diverse collection of images with human-annotated scores across multiple dimensions of aesthetic evaluation.

Each image in the dataset is accompanied by detailed human judgments on seven specific aesthetic attributes:

- Overall aesthetic score
- Quality score
- Composition score
- Color score
- Depth of field score
- Lighting score
- Content score

While PARA also includes subjective attribute information from judges such as emotional responses to images, the scope of our image aesthetics assessment (IAA) task focused on the objective aesthetic attributes listed above. This approach allows our model to learn generalizable aesthetic principles rather than highly personalized preferences.

The dataset's comprehensive attribute-specific annotations provide a rich training signal for our multi-headed architecture, enabling more targeted and explainable aesthetic assessments compared to models trained only on overall aesthetic scores.

Base Model Selection Hentschel, Kobs, and Hotho found in their study that the CLIP image encoder often outperformed other model architectures when fine-tuned for IAA tasks. They concluded that:

> CLIP extracts features from images that are related to human's image aesthetic perception due to its training on images and their corresponding human-generated descriptions. (Hentschel, Kobs, and Hotho 2022)

This made CLIP a strong candidate for the base model in our IAA task.

Architecture Modifications Our model uses a standard CLIP ViT-B/32 model from HuggingFace as its backbone, with seven fully connected layers as heads that independently predict scores for aesthetic attributes: composition, color, lighting, content, quality, depth of field, and overall aesthetic appeal.

Training Process The model was fine-tuned using the sum of the MSE loss on the seven aesthetic attributes from approximately 22,000 images in the PARA (Yang et al. 2022) dataset. We employed a learning rate of 1e-5 with AdamW optimizer and trained for 10 epochs.

User Interface

We developed a web-based interface using Streamlit (Snowflake Inc. 2025), structured as an intuitive chatbot for image aesthetic optimization. The interface enables users to upload images for evaluation, select preferred optimization algorithms (such as gradient approximation or genetic algorithms), and adjust relevant hyperparameters. Our conversational approach makes aesthetic optimization accessible to non-technical users, allowing them to visualize improvements in real-time and compare original images against optimized versions with their corresponding attribute scores.

Results

Aesthetic Scoring

One of the key factors we set out to find was whether or not an image could be quantitatively scored on its aesthetic quality. This was important as it provided a field we could use to optimize evaluation of aesthetic quality improvement. Using our evaluator, we determined that we could produce a measure that reflected the aesthetic quality of an image. To test the ability of our model to reflect aesthetic quality, we scored an image we determined to be relatively aesthetic as well as a blank white image to compare the scores and show that the unaesthetic blank image scored consistently lower in all categories compared to the aesthetic image. This shows that our model is able to produce a score that relates

image	Aesthetic	Quality	Composition	Light	Color	Depth-of-Field	Content
Aesthetic	3.8771	3.9893	3.9620	3.998	3.7941	3.8808	3.6442
Non-aesthetic	2.2450	2.3347	2.5573	2.2898	2.10888	2.3782	2.3136

Figure 3: Comparison between aesthetic and unaesthetic image scores. Image shown is the aesthetic image; non-aesthetic image is blank white square.

to the aesthetic quality of the image. Results can be seen in Figure 3.

Quantitatively producing an aesthetic score was a large step towards being able to give suggestions on more aesthetic images, but we determined that even among our group, there were sometimes discrepancies on the best image when performing and scoring our transformations. To look further into how human perception affects the aesthetic of an image, we chose to conduct a study and compared the human perception to our system's aesthetic evaluation of various transformed images.

Survey

To test how well our transformed images achieved an increase in human-perceived aesthetic quality, we performed a user study where participants scored five images on a scale of 1-5, where 1 is the lowest score and 5 is the highest score. The five images consist of the original image and four transformations, all of which have different kinds of transformations that result in computed aesthetic scores higher than that of the original image. The study included 30 anonymous participants scoring 6 different sets of 5 images. The results from the survey are shown in Table 1.

The image sets are defined by the subject of the image and consist of dog, car, boat, flower, landscape, and monopoly, while each image in the set is determined by the algorithm parameters used to produce the transformations. The possible algorithms include original, which is the untransformed image; LAION genetic, utilizing a LAION model and our genetic algorithm; CLIP Gradient and CLIP Genetic, which utilize a CLIP model and either the gradient or genetic algorithm described above; and CLIP+, which again uses either the gradient or genetic algorithm described above but includes some of our more advanced and targeted transformations including depth of field and vignette.

Image	Top Scoring	Top Choice
Dog	CLIP Genetic+	Original
Car	CLIP Gradient	Both CLIPs (tie)
Boat	CLIP Gradient+	Original
Flower	CLIP Gradient	CLIP Gradient
Landscape	CLIP Genetic+	CLIP Genetic+
Monopoly	CLIP Genetic+	CLIP Genetic+

Table 2: Comparison of image scored the highest with image selected as the best image by our study participants.

Additionally, we asked users to choose which image they believed to be the most aesthetic. In doing so, we wanted to see whether there was consistency with the image that was highest scored overall with the image that was most selected as being the most aesthetic.

Evaluation

Survey Implications

Correlation between human preference and score element One question we wanted to answer was, given our 6 selected images with four transforms each (30 images total), is there correlation between user preference in the survey and each of the aesthetic scores, and if so, which score is most correlated with human preference?

We found that the correlation was generally low. For a survey size of $n = 30$, R values greater than 0.361 are statistically significant ($p < 0.05$). These values are bolded in Table 3.

Metric	R
Composition	0.2446
Content	0.2964
Depth of Field	0.3179
Overall	**0.3774**
Quality	**0.3970**
Color	**0.4064**
Lighting	**0.4326**

Table 3: Pearson correlation coefficients (R) for mean response value vs. each metric.

We observe that the *Lighting* score predicted by our fine-tuned CLIP scorer is most correlated, and plot the correlation plot in Figure 4.

Correlation between human preference rankings and computationally evaluated aesthetics Two research questions that we wanted to answer were do the images rank in the same order as their evaluated aesthetic scores, and do the edited images rank higher than the original images?

Our findings suggest a nuanced relationship between computational aesthetic assessment and human judgment. In some cases, survey participants preferred the original images over those transformed by our system. The optimized enhancements did not always match human preference for aesthetics in images. Figure 5 shows the difference in how edited images were evaluated by humans.

Aesthetic Evaluation Implications

The discrepancy between our system's rankings and human preferences may be attributed to several factors. Aesthetic

Image	Original	LAION Genetic	CLIP Gradient	CLIP Genetic	CLIP+
Dog	3.167	2.967	3.133	2.433	**3.433**
Car	2.967	2.867	3.000	**3.100**	2.967
Boat	3.733	2.800	3.233	2.867	**3.867**
Flower	3.067	4.133	**4.200**	2.433	3.133
Landscape	2.500	2.433	2.067	2.467	**3.267**
Monopoly	3.167	2.867	3.367	—	**3.733**

Table 1: Human-Perception Aesthetic Scoring Results. Note: for the monopoly image, the CLIP Genetic image was replaced with another CLIP+, which did not affect any results positively or negatively and so is excluded from this table for clarity.

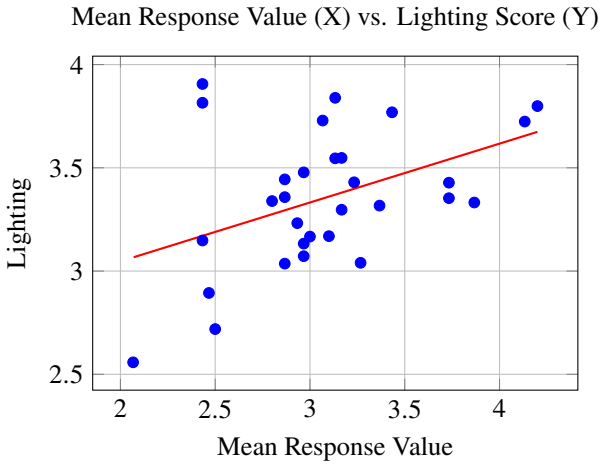

Figure 4: Scatter plot with trend line, $R = 0.4326$

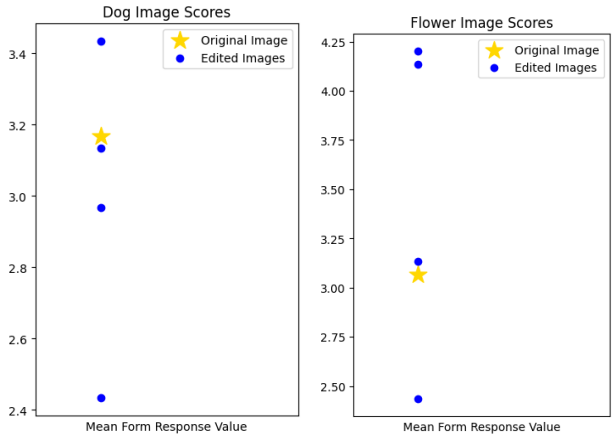

Figure 5: Human preference scores of the original image and the 4 edited images from the survey for the dog and flower images. Note that the flower image had 3 edited images ranked higher than the original, while the dog image only had 1 edited image rank higher than the original.

judgment is inherently subjective and influenced by personal taste, cultural background, and individual experiences. This subjectivity presents a fundamental challenge for IAA systems and explains why our survey participants sometimes disagreed with our system's assessments, because our aesthetics evaluator was trained to approximate a mean of preferences rather than a specific person's preferences.

Despite this challenge, our approach demonstrates that computational enhancement can consistently produce improvements in aesthetic quality. For each original image in our study, at least one algorithmically enhanced version received higher ratings from human evaluators, suggesting that our system successfully explores a diverse solution space that accommodates varying aesthetic preferences. Figure 5 shows 2 examples of images from our survey, and for each the highest rated image was an edited image.

Creativity

In evaluating the creative capabilities of our system, we used suggestions by Jordanous (2019) to determine how well we were able to achieve creativity in our system. The main creative components we determined were essential to our system for it to be considered creative are domain competency, intentionality, and value. On the road to achieving creativity, we want to show the knowledge of photography composition principles in play in our system, intentionally use these principles to transform our image, and create a transformed image that values at a higher aesthetic quality level than the original image.

It's difficult to determine exactly how well we are able to achieve complete domain knowledge in our system, so we chose our transformations to target some of the photography principles: in addition to rotations, the various crops allow us to alter subject placement; brightness, contrast, and color alterations assist with balancing lighting; we employ depth of field, which directly relates to depth of field qualities in images; and the vignette transform works to centralize focus. Utilizing these transforms via the genetic and gradient algorithms allows us to traverse the space of possible transformations with a relation to the photography domain knowledge.

By optimizing the transformations based on aesthetic score, we introduce intentionality since the scores follow along several of the composition guidelines, including composition, color, depth of field, and lighting–directly related to the composition guidelines our transformations target. When we are able to find an image that increases the score, the system displays a degree of intentionality. Additionally, the aesthetic score optimization allows us to determine that the new image achieves a higher value than the original, though the users themselves are the final determiner on how high the

increase in value may be.

With this in mind, we evaluated on the scale proposed by Ventura (2012) we fall somewhere in the range between filtration and inception. Using the aesthetic scores as an optimizer, we are able to clearly show we achieve the fitness function requirement of filtration, but the ambiguity on how aware the system is of the domain knowledge prevents us from confidently stating we achieve the level of inception.

Conclusion

Contributions

We provide a proof-of-concept chatbot that is shown to, in many cases, suggest images that are rated more aesthetically pleasing by human viewers than the original images.

More broadly, our system contributes to the growing space of human-AI collaboration tools by acting as a creativity augmenter—a system designed not to replace human artistic judgment, but to enhance and support it.

While aesthetics are subjective, our user study shows that our system is capable of suggesting improvements that resonate with human preferences in a nontrivial number of cases. Notably, our "CLIP+" algorithm frequently achieved the highest user scores, reinforcing that automated composition enhancements can align with human judgments of beauty.

Even in cases where transformed images did not outperform the original, the diversity of improvements offered by the system presents users with new perspectives on how to approach composition and edit their images. By suggesting edits and providing natural language explanations, our tool supports the creative process, giving users the power to choose the transformation that best aligns with their vision or learn new compositional techniques.

This work shows that human-in-the-loop aesthetic optimization can meaningfully contribute to making photography more accessible and empowering for a broader audience.

Future Work

While our current system demonstrates promising results in automating aesthetic enhancement of images, several avenues for improvement remain. Our optimization approach could be enhanced through more sophisticated techniques such as Bayesian optimization or reinforcement learning. Aesthetic scoring could be improved by using a model that provides uncertainty quantification with its predictions. Additionally, leveraging recent advances in multimodal foundation models could significantly improve aesthetic evaluation by incorporating both visual features and broader understanding of aesthetic principles, potentially addressing the subjective nature of aesthetics highlighted in our survey results.

The transformation capabilities of our system could be expanded through more sophisticated photography-inspired operations, such as guided content-aware fill for distracting elements, perspective correction, or golden ratio-based composition adjustments. More extensive surveys with larger and more diverse participant pools, including both novice and expert photographers, would provide more statistically robust insights into the relationship between computational aesthetic metrics and human judgments.

Finally, the principles underlying our system could be adapted to other visual art forms, such as drawings or illustrations, and work as a creativity enhancing tool in those domains. By pursuing these directions, future research can address the limitations identified in our current work and move closer to the goal of accessible, high-quality creativity enhancement for all users, regardless of their aesthetic expertise.

Final Thoughts

At the heart of our work is the belief that creativity should be accessible. While professional image editing tools and artistic expertise remain invaluable, not everyone has the time or resources to master them. Our system demonstrates that it is possible to empower users—regardless of their technical or artistic background—to engage meaningfully in the creative process. By blending optimization techniques with learned aesthetic principles and offering suggestions in natural language, we reduce the barrier between intention and expression.

More importantly, this system is not meant to dictate artistic decisions, but to collaborate with users as a creativity support tool. It encourages exploration, supports learning, and fosters confidence in creative choices. As technology continues to evolve, we believe AI tools like this will not only enhance aesthetic quality, but also inspire new forms of creative expression across domains. Ultimately, this project is a step toward that vision: a world where computational creativity supports and amplifies human creativity, rather than replaces it.

References

[2022] Hentschel, S.; Kobs, K.; and Hotho, A. 2022. Clip knows image aesthetics. *Frontiers in Artificial Intelligence* 5:976235.

[2019] Jordanous, A. 2019. Evaluating evaluation: Assessing progress and practices in computational creativity research. In Veale, T., and Cardoso, A. F., eds., *Computational Creativity: The Philosophy and Engineering of Autonomously Creative Systems*. Cham, Switzerland: Springer. 211–236.

[2020] Qin, X.; Zhang, Z.; Huang, C.; Dehghan, M.; Zaiane, O. R.; and Jagersand, M. 2020. U^2-net: Going deeper with nested u-structure for salient object detection. *Pattern Recognition* 106:107404.

[2021] Radford, A.; Kim, J. W.; Hallacy, C.; Ramesh, A.; Goh, G.; Agarwal, S.; Sastry, G.; Askell, A.; Mishkin, P.; Clark, J.; Krueger, G.; and Sutskever, I. 2021. Learning transferable visual models from natural language supervision.

[2025] Snowflake Inc. 2025. Streamlit: An open-source python framework for data apps.

[2012] Ventura, D. 2012. A redefinition of creativity—a mereological perspective. In *Proceedings of the Third International Conference on Computational Creativity*, 117–122. Dublin, Ireland: University College Dublin.

[2006] Wiggins, G. A. 2006. A preliminary framework for description, analysis and comparison of creative systems. *Knowledge-Based Systems* 19(7):449–458.

[2022] Yang, Y.; Xu, L.; Li, L.; Qie, N.; Li, Y.; Zhang, P.; and Guo, Y. 2022. Personalized image aesthetics assessment with rich attributes.

Appendix

```
parameters_to_loss_mapping = {
    0: 2,  # LeftCropTransform -> Composition score
    1: 2,  # RightCropTransform -> Composition score
    2: 2,  # TopCropTransform -> Composition score
    3: 2,  # BottomCropTransform -> Composition score
    4: 3,  # BrightnessTransform -> Lighting score
    5: 3,  # ContrastTransform -> Lighting score
    6: 2,  # RotateTransform -> Composition score
    7: 4,  # ColorTransform -> Color score
    8: 5,  # DepthOfFieldTransform -> DoF score
    9: (0,1,2,3),  # Vignette -> (Overall, Quality, Composition, Lighting)
}
```

Figure 6: Mapping from transformation dimensions to loss channels.

Pseudocode

Algorithm 1 GeneticAlgorithmOptimizeImage

Require: Image I, population size P, mutation rate m, crossover rate c, max generations G, patience p

$\mathcal{P} \leftarrow$ InitializePopulation(P)
$best \leftarrow$ ArgMax($\mathcal{P}, A$)
for $g = 1$ **to** G **do**
    // Selection
    $\mathcal{M} \leftarrow$ RouletteWheelSelect($\mathcal{P}, A$)
    // Crossover
    for all $(p_1, p_2) \in \mathcal{M}$ **do**
        if Rand() $< c$ **then**
            $(o_1, o_2) \leftarrow$ Crossover(p_1, p_2)
            insert offspring into $\mathcal{P}$
        end if
    end for
    // Mutation
    for all $p \in \mathcal{P}$ **do**
        for all gene t_i in p **do**
            if Rand() $< m$ **then**
                $t_i \leftarrow t_i + \mathcal{N}(0, \sigma^2)$
                $t_i \leftarrow$ CLIPToBounds(t_i)
            end if
        end for
    end for
    // Survivor selection
    $\mathcal{P} \leftarrow$ TopK($\mathcal{P}, P, A$)
    // Early-stop check
    if $A(\mathcal{P}[0]) > A(best)$ **then**
        $best \leftarrow \mathcal{P}[0]$ $stall = 0$
    else
        $stall \leftarrow stall + 1$
        if $stall \geq p$ **then break**
        end if
    end if
end for
return $best$

Algorithm 2 GradientBasedOptimizeImage

Require: Image I, initial parameters $\mathbf{t}_{start}$, step size η, max iterations T, method FD or SPSA

$\mathbf{t} \leftarrow \mathbf{t}_{start}$; $best \leftarrow \mathbf{t}_{start}$
$m \leftarrow 0$; $v \leftarrow 0$
for $iter = 1$ **to** T **do**
  // Approximate gradient
  if method = FD **then**
    $\nabla A \leftarrow \text{FiniteDifferences}(A, \mathbf{t})$
  else
    $\nabla A \leftarrow \text{SPSA}(A, \mathbf{t})$
  end if
  // Adam moment update
  $m \leftarrow \beta_1 \cdot m + (1 - \beta_1) \cdot \nabla A$
  $v \leftarrow \beta_2 \cdot v + (1 - \beta_2) \cdot (\nabla A)^2$
  $\hat{m} \leftarrow m / (1 - \beta_1^{iter})$
  $\hat{v} \leftarrow v / (1 - \beta_2^{iter})$
  // Parameter update
  $\mathbf{t} \leftarrow \mathbf{t} - \eta \cdot \hat{m} / (\sqrt{\hat{v}} + \varepsilon)$
  $\mathbf{t} \leftarrow \text{CLIPToBounds}(\mathbf{t})$
  if $A(E(I, \mathbf{t})) > A(E(I, best))$ **then**
    $best \leftarrow \mathbf{t}$
  end if
end for
return $best$

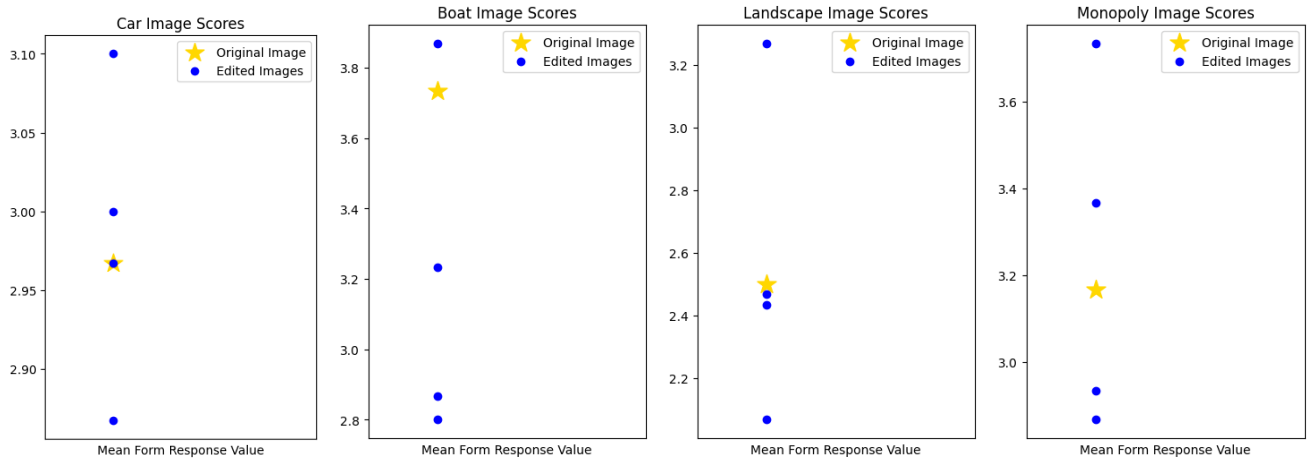

Figure 7: Human preference scores of the original image and the 4 edited images from the survey for the car, boat, landscape, and Monopoly images. Note that in for each image a edited version scored the highest.

Survey Images

All images used in our survey. They are shown in the following order: original, LAION genetic, CLIP gradient, CLIP genetic, CLIP+. The monopoly section shows the extra CLIP+ image, which is in place of CLIP genetic.

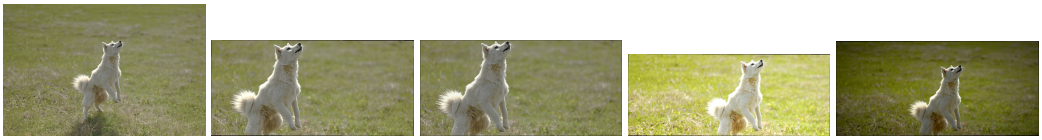

Figure 8: All dog images used in survey. Shown in same order as results in table 1.

Figure 9: All car images used in survey. Shown in same order as results in table 1.

Figure 10: All boat images used in survey. Shown in same order as results in table 1.

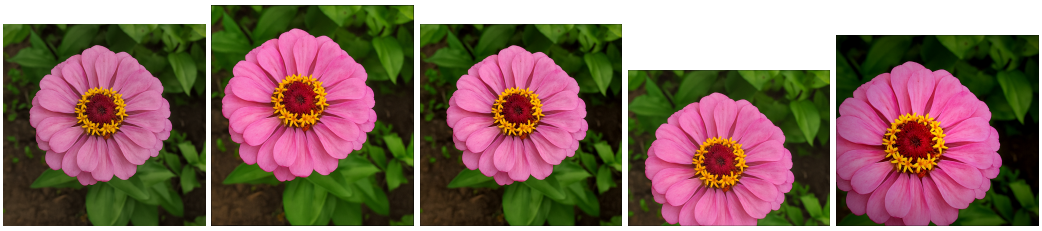

Figure 11: All flower images used in survey. Shown in same order as results in table 1.

Figure 12: All landscape images used in survey. Shown in same order as results in table 1.

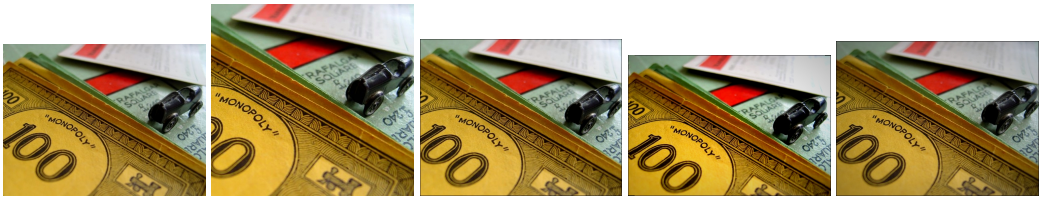

Figure 13: All monopoly images used in survey. Shown in same order as results in table 1.