From Prompt to Palette: Enabling Brand-Specific Design with Datasets

Anonymous submission

Paper ID

1. Introduction

Multi-modal models currently exist like CLIP [12] and VQ-VAE [8], which allow for text-to-image generation, but their outputs often do not fully fulfill branding requirements and aesthetics (eg. "minimalist yet vibrant"). Brands cannot automate the process of "design," which can lead to overhead when looking to build or improve existing branding. Branding can be especially costly and has a magnitude of importance for entrepreneurs promoting their brand. Multi-modal models have revolutionized efficiency and clarity in many different domains, such as writing and coding, but this advancement remains restrained in terms of design.

1.1. Background and Motivation

Non-design-oriented founders and students often lack accessible AI-driven tools to create distinctive branding and marketing assets that align with their budget constraints, aesthetic preferences, and business objectives. In early-stage startups with limited resources and financial runway, effective marketing can significantly influence success, yet hiring professional designers may be cost-prohibitive for many founders. The creative outputs of current models are often too general to tell a unique story to users through an established design and aesthetic. AI-automated creative outputs still lack uniqueness and alignment with specific graphic design themes and overall branding, which does not allow founders and students to tell a compelling story about their product. The ineffectiveness of tools today results in human designers needing to iterate and tweak designs.

1.2. State of Current Work

Current multi-modal AI models like CLIP and VQ-VAE have advanced text-to-image generation by bridging semantic alignment between prompts and visuals, but struggle with brand-specific aesthetics such as precise color palettes or typography. Recent work, such as ACAI (AI Co-Creation for Advertising and Inspiration) [5], demonstrates progress and research in leveraging multi-modal large language models to generate personalized marketing materials. Mainstream tools like Adobe GenStudio for Performance Marketing show promise through AI-assisted marketing generation, but can be constrained by template-based workflows. These solutions are also closed-source, making them inaccessible to users with limited financial means.

Additionally, when looking at open-source or free tools available, popular LLM and text-to-image models fail to create coherent Instagram and online ads. Datasets for advertisements that relate to specific design attributes, such as color, are also hard to find.

1.3. Plan and Expected Challenges

We begin by searching and collecting online data from pre-existing datasets such as the MAdVerse Dataset. We are primarily focused on ads with text that seem to convey an image. These assets will be categorized into aesthetic styles (e.g., minimal, retro, bold) and with specific design attributes such as color and shape. To ensure broad stylistic coverage, we will augment the dataset through synthetic data generation. All assets will be resized and converted into tensor-compatible formats for model training.

We anticipate several challenges in data collection and classification. Since we are working with a dataset, there is always risk of stylistic bias due to prevailing design trends. This could potentially lead to over-representation of certain aesthetics and could skew a model's outputs and limit stylistic diversity. Translating natural language to design representations may also prove difficult as mapping generic user prompts—such as "use blue and red in this marketing asset" to precise Pantone values when the user prompt is unspecific could lead to undesired outputs. Finally, defining and curating "good design" practices for supervised learning remains subjective and may introduce further complexity during dataset construction.

1.4. Expected Outcomes

This paper aims to improve datasets for ad generation by adding more annotations that relate to aesthetic and branding requirements to enable more image-to-text semantic improvement for users to be able to fine-tune models that generate ads according to their style requirements to tell compelling stories and messages using the power of LLM models.

2. Related Works

Recent research has made significant strides in leveraging artificial intelligence for the creation and evaluation of digital advertisements, particularly on platforms such as Instagram. Several studies have examined how consumers perceive AI-generated content, the effectiveness of such ads, and the broader implications for digital marketing. For example, Exner et al. (2025) [3] conducted a large-scale quasi-experimental study and found that AI-generated ad images can outperform human-made images in terms of click-through rates, but only when the ads do not appear artificial. Their analysis identified specific visual cues, such as intense color saturation, that signal AI generation to consumers and can negatively impact ad performance if and when detected. Similarly, other research has shown that disclosing the use of AI in ad creation can lead to less favorable attitudes toward both the advertisement and the brand, especially among consumers with a high aversion to AI. [1]

Further studies have investigated the nuances of consumer attitudes toward AI-generated ads. [4] For instance, ads that emphasize action or empowerment are viewed in a more positive light by their audience, directly related to the consumer's sense of task self-efficacy. Additional research highlights that while AI-powered tools have transformed content creation and marketing strategies on Instagram, enabling more sophisticated analytics, captioning, and audience segmentation, these advances have not fully bridged the gap between automated content generation and the specific needs of brand-specific marketing campaigns. Most existing datasets and models focus on general ad generation or evaluation, lacking the granularity to capture specific aesthetic requirements or brand identities.

Despite these advances, many student leaders and entrepreneurs follow specific design themes and specifications when building out a proportional campaign, and currently, there is a critical gap that remains: no publicly available dataset or model supports the generation of advertisements (such as Instagram ads or posters) tailored to specific campaign aesthetics or brand guidelines. Prior studies have primarily evaluated consumer perception, ad effectiveness, or the impact of AI disclosures, as well as novel ad generation through AI and editing existing designs, but have not addressed the challenge of producing ads that align with a brand's unique visual identity or marketing objectives. Visual identities can be thought of as aesthetics, such as having a retro or old-school aesthetic. This limitation highlights the need for a new dataset and modeling approach that enables the generation and evaluation of ads based on detailed aesthetic and branding criteria, a gap this project aims to fill by developing a dataset and tools designed specifically for this purpose.

3. Methodology

Acknowledging the gap in datasets, there is a need to build a dataset that categorizes aesthetic styles of design in order to build upon the needs of students and founders for an accessible way to create compelling marketing and advertising assets without the constraint of a design team or experts. We leverage data from the pre-existing MAdVerse ads dataset.

3.1. Data Collection and Curation

To build a dataset that reflects a diverse range of marketing and advertising graphics, it is essential that the collected data is annotated according to established design principles. We annotate each graphic on critical design elements, including color usage, font selection, text content, and shape composition, and overall graphic design theme.

We chose to annotate advertisements based on color, font, shape, and design style because these elements are fundamental to how visual messages are perceived and understood [6]. By focusing on these key aspects, we ensure that our dataset captures the essential visual characteristics that influence audience engagement, brand recognition, and the effectiveness of marketing materials. This approach allows us to analyze a wide variety of design styles, such as futuristic, retro, and minimalist, and to better understand how different design choices impact communication and appeal in advertising.

Since advertisements can incorporate an unlimited variety of colors, we chose to focus on identifying up to the seven most dominant colors in each design. This approach ensures a balanced representation of each color's prominence. For consistency and universal understanding, we represent these colors using their corresponding hex codes.

Similarly, since an advertisement can feature an unlimited variety of fonts, we referenced the extensive list of available Google Fonts and selected up to three of the most prominent fonts present in each advertisement for annotation. This approach ensures a manageable and consistent representation of typographic choices.

Shapes present in an advertisement was also an atrribte of the design that we focused on as different edges and shapes are characteristics of different design patterns and themes. We focused on general shapes of polygons, circles, squares, and triangles and included how many times those defined shapes occurred in the given advertisement.

Finally, we annotated images based on the overall graphic design theme that they represented. We focused on 8 overall themes with the following characteristics:

1. **Minimalist** [11]
   1. Simple layouts with few distractions
   2. Neutral or pastel color palettes
   3. Generous use of negative space (the empty or blank

areas in a design)
4. Light sans serif fonts for optimal readability

2. **Flat** [2]
   1. Two dimensional icons and graphics
   2. Minimal or no gradients
   3. Clean, straightforward layouts

3. **Retro** [14]
   1. Heavily nostalgic color palettes
   2. Vintage typography and graphic elements
   3. Visual Motifs from past cultural or technological ages

4. **Neobrutalism** [10]
   1. Strong color contrasts and stark backgrounds
   2. Visible grids and harsh drop shadows
   3. Brutalist, industrial-feeling typography

5. **Geometric** [7]
   1. Use of basic shapes: circles, squares, triangles
   2. Bright, often primary-based color schemes
   3. Strong grid-based layouts

6. **Interactive Media** [9]
   1. Combines visual design with interactive technology
   2. Clickable graphics that change state or content
   3. Interfaces with feedback loops through buttons, sliders, gesture controls

7. **Media Blending** [13]
   1. Combines photographs, illustrations, or textures
   2. Layered compositions with intentional overlap
   3. Playful and experimental visual themes

8. **Maximalism** [14]
   1. Bold, contrasting color choices
   2. Exaggerated design elements and typefaces
   3. Frequent use of abstract visuals

For all categories, characteristics were determined through extensive research into historical context of where these design themes emerged, as well as relevant research of these topics in the graphic design fields today.

Along with deciding the guiding principles from above we also took careful consideration to make sure the dataset is comprehensive of many different projects and company types. The MAdVerse dataset does a great job at this as it covers eight different product domains with over 50,000 instances of advertisements.

3.2. Metadata Annotation Framework

The new annotation taxonomy that we are adding to the existing MAdVerse data set is ultimately split into the following:

Style category – minimalist, retro, bold, etc.
Font type – Arial, Helvetica, etc.
Color encoding – HEX values
Shape semantics – curves, squares, circles, etc.

Font recognition can be collected via Tesseract OCR tools in order to identify text regions using bounding box detection in order to identify specific fonts from images. The individual fonts are then cropped from the image in the dataset and put through a ONNX font detection model that identifies the closest Google Font. Each distinguishable color on each image was converted into RGB values and converted into HEX values through tools such as digital Color Meter. Shapes are also identified throguh edge detection using OpenCV.

3.3. Future Alignment Validation

CLIP-based scoring can be used in order to quantify the semantic similarity between a text and image, and thus prompt and generated output. We will be able to use this score as a way to understand how adding these annotations to the data set will elevate, decrease, or maintain semantic similarity between image and text (”build me a retro ad for my brand” with the new annotations v.s. the previous ones). Its ability to also compare color harmony, typographic coherence, and layouts will also ensure that in the future assessment of design will be able to be less dependent on human review by introducing these benchmarks. [15]

4. Experiments

The project takes on an automated annotation generation pipeline in order to take in images from the MAdVerse Ads Dataset in order to produce a new set of more informed design information, and then combine this JSON file with the existing MAdVerse Dataset Annotations.

4.1. Annotation Extraction Process

There are three sets of extraction processes that occur for all images in the dataset that extract text/font, colors, shapes, and design style.

For text/ font extraction, within each directory of advertisement category, the pipeline goes through each folder and utilizes Tesseract OCR technology in order to crop the fonts included in each image and run them through an ONNX font detection model that that already has been trained on a set of Google fonts. This way inference can just be run and identify the closest identifying Google fonts for each cropped image from segmentation, and then save the top 3 most occuring fonts in the image in a JSON file.

For colors and shapes, the extraction process involves edge detection using k-means as well as the OpenCV library in order to identify which are the top three most occurring shapes and HEX colors and put this information into their respective JSON files.

For design style, as mentioned earlier we categorized characteristics of 8 different design styles and utilized the OpenAI API, Tesseract OCR, and OpenCV in order to prompt this LLM for the most relevant design style within a given image for each image in the dataset.


```
"GoAir_26.jpg": {
    "full_path":
"travel/airlines/GoAir/GoAir_26.jpg",
    "results": [
      [
        "GoAir_26.jpg_14.png",
        [
          "Montserrat Subrayada",
          "bold"
        ]
      ],
   ],
    "dominant_colors": [
      "#0439c3",
      "#6088c7",
      "#4b3233",
      "#cc925b",
      "#f5f7f9"
    ],
    "dominant_shapes": {
      "Rectangle": 7,
      "Circle": 47,
      "Polygon": 4
    }
```


Figure 1. Example Consolidated JSON

4.2. Merging Annotations

Once all of the annotations for each image in the dataset occurs, merging all JSON files into one is necessary in order to consolidate this information and be able to update the MAdVerse Dataset with these more robust annotations for enhanced prompt adherence.

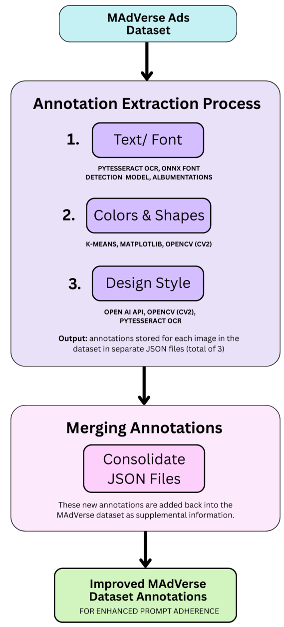

Figure 2. Outlines the project annotation pipeline

5. Results

Testing the use of improved annotations and their capabilities for founders, entrepreneurs, and students to be able to fine tune models with these datasets for improved ad generation according to aesthetic style, we see the following case.

5.1. Text to Image Model Testing

Let's say that a user would like to create a minimalist Instagram ad for their company ”boone” that sells vitamin D based lamps and is having a 30% off sale. In the following Figure 3, we can see results for the prompt into the text to image model are a lot more cluttered visually and semantic meaning is harder to grasp. In comparison to Figure 4 that utilizes the improved annotations, we can see the characteristics matching up more with what we previously defined as ”minimalist” which includes more empty space, neutral colors, and still gets semantic meaning across.

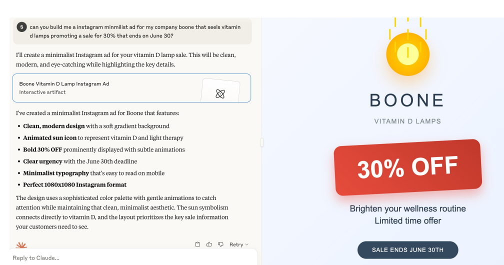

Figure 3. Ad generation without new annotations

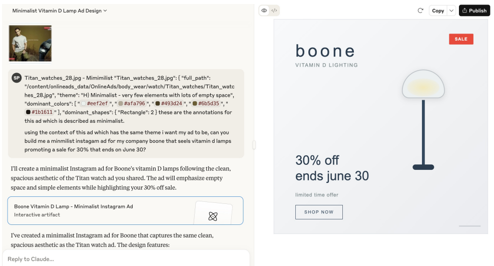

Figure 4. Ad generation with new annotations

5.2. Quantifying Results

For the previous example, using CLIP score [12] as a way to measure human judgment of aesthetic alignment to text, we take the vectorized image and the vectorized text or prompt in this case, and calculate the difference in cosine distance between them. This measures the angles between image and text and thus semantic similarity. Higher cosine scores indicate a much stronger semantic correlation between the image and text. As we can see in Figure 4, there is overall greater CLIP score in the different categories for the ad and an overall 36% improvement across categories showing that with enhanced annotations the MAdVerse dataset has capabilities to help achieve user aesthetic requirements.

CLIP Score Performance Comparison

Aesthetic Criteria	With Annotations	Without Annotations	Improvement
Minimalism	78.4	62.1	26.2%
Typography	82.7	58.9	40.4%
White Space	85.2	55.3	54.1%
Brand Alignment	76.8	61.2	25.5%
Average Score	80.8	59.4	36.0%

Figure 5. Ad generation with new annotations

6. Future Work

While this project demonstrates the feasibility of generating advertisements and social media posts using text to image models with an enhanced dataset, several promising avenues remain for future research and development.

One immediate direction is the expansion and refinement of the dataset. As our work highlights, there is a lack of publicly available datasets annotated for both visual and textual features specific to branded advertisements, such as Instagram ads or campaign posters, which is why this dataset is novel. Future efforts could focus on collecting and annotating a larger-scale dataset that includes not only representations of ads but also expands to campaign objectives and broader brand guidelines.

Another important area is the enhancement of text to image generative model ability to capture and reproduce specific aesthetics and bring the user's vision to life. While current models can generate ad layouts and text, they struggle to align outputs with given specific branding. Future work could explore the integration of brand style embeddings or reinforcement learning from human feedback to better align the output with the vision. Incorporating human evaluation and critique into the output, similar to approaches used in UIcrit for UI design, could improve both quality and relevance as well. Long-term, it is essential to study the real-world impact of AI-generated ads through A/B testing and user studies on social platforms, providing evidence of effectiveness for brands and individual users.

Future work in this department can help close the gap between automated ad generation and modern-day marketing campaigns, enabling more effective and personalized advertising solutions.

References

[1] Anonymous. Consumer attitudes toward ai-generated ads: Appeal types, self-efficacy and ai's social role. *Journal of Business Research*, 185:114867, 2024. Available at https://www.sciencedirect.com/science/article/pii/S0148296324003710. 2

[2] Ivan Burmistrov, Tatiana Zlokazova, Anna Izmalkova, and Anna Leonova. Flat design vs traditional design: Comparative experimental study. In *Human-Computer Interaction – INTERACT 2015, Lecture Notes in Computer Science*, volume 9297, pages 106–114. Springer International Publishing, 2015. Available at https://doi.org/10.1007/978-3-319-22668-2$_1$0. 3

[3] Yannick Exner, Jochen Hartmann, Oded Netzer, and Shunyuan Zhang. Ai in disguise—how ai-generated ads' visual cues shape consumer perception and performance. *SSRN Working Paper Series*, (5096969), 2025. Available at https://ssrn.com/abstract=5096969. 2

[4] Yannick Exner, Jochen Hartmann, Oded Netzer, and Shunyuan Zhang. Ai in disguise—how ai-generated ads' visual cues shape consumer perception and performance. *Unpublished manuscript*, 2025. Manuscript dated May 23, 2025. 2

[5] N. Karnatak, A. Baranes, R. Marchant, T. Butler, and K. Olson. Expanding the generative ai design space through structured prompting and multimodal interfaces. *arXiv preprint*, arXiv:2504.14320v1, 2025. Available at https://arxiv.org/abs/2504.14320. 1

[6] Zhuoran Luo, Zhen Zhang, Yilun Xu, Ji Lin, and Song Han. Super-mario: Transformer as super memory for continual reinforcement learning. *arXiv preprint arXiv:2309.01371*, 2023. Available at https://arxiv.org/pdf/2309.01371. 2

[7] Ghassan Zainel Mahmood and Nsiyf Jassem Mohammed. The influence of the bauhaus in contemporary graphic design. *Unpublished manuscript*, 2023. Available at https://www.researchgate.net/publication/369374365$_The_influence_of_the_Bauhaus$

[8] K. Miyazawa and T. Nagai. Concept formation through multimodal integration using multimodal bert and vq-vae. *Advanced Robotics*, 37(4):281–296, 2022. 1

[9] Clement Mok. Teaching interactive media design. *International Journal of Technology and De-*

sign Education, 12:161–171, 2002. Available at https://www.researchgate.net/publication/225658479$_{T}eaching_{I}nteractive_{M}edia_{D}esign$.3

[10] Nielsen Norman Group. Neobrutalism: A bold (and beautifully brutal) ux/ui trend. Online article, 2022. Available at https://www.nngroup.com/articles/neobrutalism/. 3

[11] Nanang Ganda Prawira, Aditya Aditama Putri Hikmatyar, and Fathia Salma Ramadhani. Redesign and development of cirebon city msme digital promotion media as implementation of project-based learning mbkm consultation project. *Journal of Visual Communication Design (VCD)*, 9(2):179–196, 2024. Available at https://journal.uc.ac.id/index.php/VCD/article/download/4663/3097/17663. 2

[12] Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, Gretchen Krueger, and Ilya Sutskever. Learning transferable visual models from natural language supervision. *arXiv preprint arXiv:2103.00020*, 2021. 1, 5

[13] Zhida Sun, Zhenyao Zhang, Yue Zhang, Min Lu, Dani Lischinski, Daniel Cohen-Or, and Hui Huang. Creative blends of visual concepts. *arXiv preprint arXiv:2502.16062*, 2025. Available at https://arxiv.org/abs/2502.16062. 3

[14] Shaelyn J. Vinson. Exploring the contemporary use and understanding of precedent in architectural design via a comparative analysis of brunelleschi and le corbusier. *Architecture Undergraduate Honors Theses, ScholarWorks@UARK*, 2016. Retrieved from https://scholarworks.uark.edu/archuht/15. 3

[15] Zhuokai Zhao, Yiming Zhang, Zhaorun Chen, Zhili Feng, Zenghui Ding, and Yining Sun. Rankclip: Ranking-consistent language-image pretraining. *arXiv preprint*, arXiv:2404.09387v1, 2024. Available at https://arxiv.org/abs/2404.09387. 3