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

Sthiti Patnaik, Caitlyn Widjaja

Motivation

- Multi-modal models currently exist like CLIP and VQ-VAE, which allow for text-to-image generation to create advertisements, but their **outputs often do not fully fulfill branding requirements and aesthetics** (eg. “minimalist yet vibrant”).
- Datasets with advertisements and elements relating to theme such as color, font, text **do not exist currently**
- Thus, we updated the Madverse dataset to make it higher quality for models to understand different graphic design themes and product quality outputs

Dataset

- We built our dataset off the MAdVerse dataset an extensive multilingual dataset with over 50,000 ads
- These ads spanned across **10 different domains**, from shopping to baby products to ads for financial institutions, providing a wide variety of advertisements to look from

Goal

- Create a pipeline that, given any advertisement, can pull relevant information relating to the graphic design theme of the advertisement
- Our focus: **automating dataset generation** for online advertisements to finetune Claude 4.0 to enable entrepreneurs and founders to create high quality advertising campaigns without the need for human capital

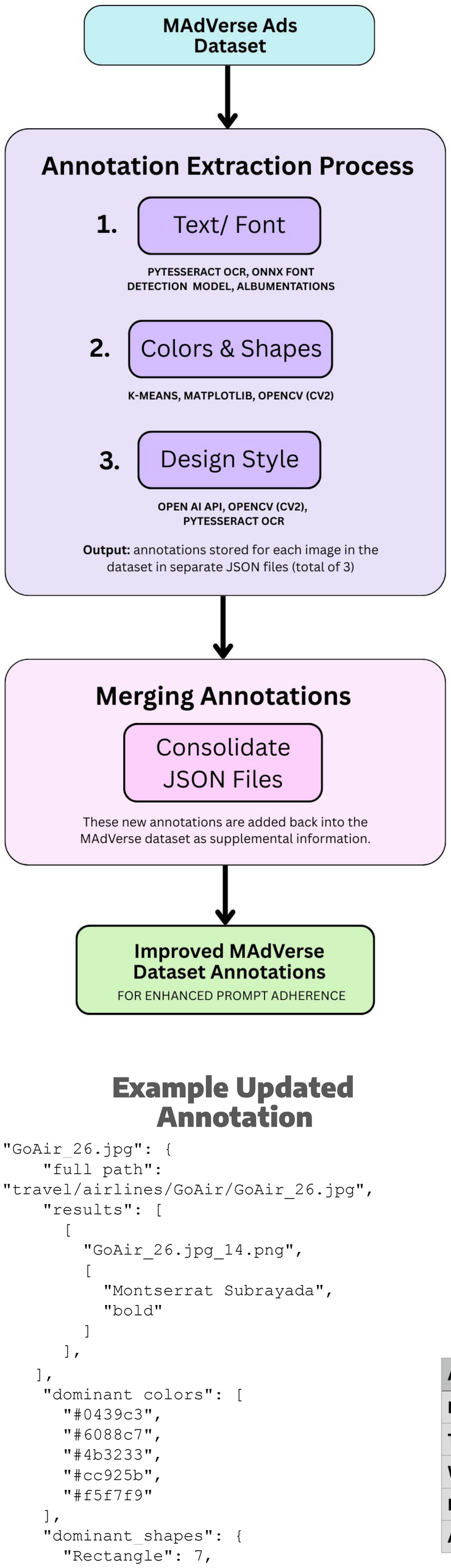
References

[1] 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. 4

[2] Mathew Pelowski, Yuna Hur, Katherine Cotter, Tomoki Ishizu, Anders Christensen, and Helmut Leder. Does beauty capture the eye, even if it’s not (overtly) adaptive? a comparative eye-tracking study of spontaneous attention and visual preference with vast abstract art. *Acta Psychologica*, 209:103133, 2020. Available at https://doi.org/10.1016/j.actpsy.2020.103133. 4

Architecture

Automated Annotation Generation Pipeline



Example Updated Annotation

```
"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
    }
```

Model

Multimodal Neural Networks

Annotation Reasoning

- Additional annotations to the MAdVerse dataset are added about **typography**, **colors**, **shapes**, and **design** style because these elements are recognized as the most important visual elements of graphic design [1]
- By improving model capabilities to output more aligned images, users can achieve aesthetic requirements which increases viewing time of the product being advertised
  - Positive correlation between time spent looking at preferred designs [2]

Results

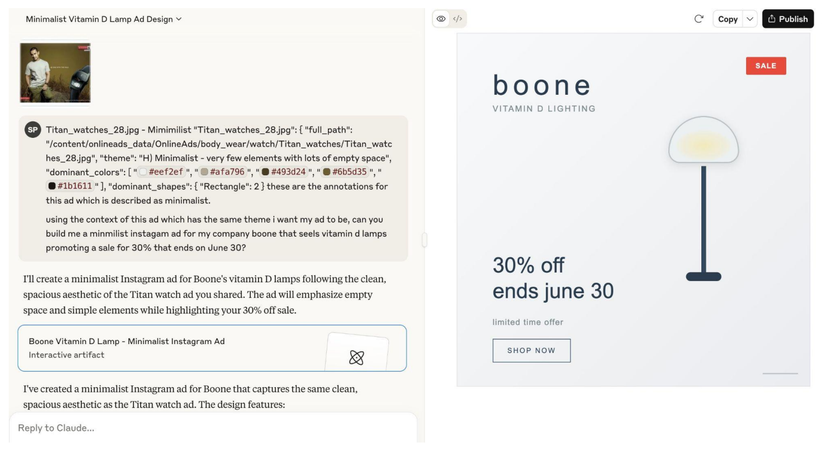


Figure 1: Model Performance with New Annotations

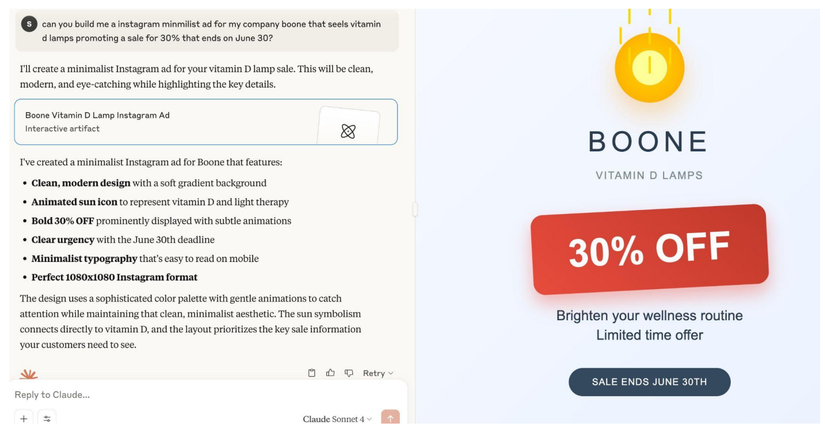


Figure 2: Model Performance without Annotations

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%

Shows 36% overall improvement, validating enhanced MAdVerse Dataset achieves user aesthetic requirements