LaDe: Unified Multi-Layered Graphic Media Generation and Decomposition

Supplementary Material

1. Implementation details

1.1. Training

Our main focus is the layered paradigm of media design processing, which is a natural extension of standard image generation. Hence, our model is trained on top of an identical text-to-image generation model that has already converged. This ensures faster convergence for our tasks of interest.

We train the model in a multi-phase format. In all phases, we give our tasks two-thirds of the GPUs, while the legacy image generation task gets scaled down with an identical split as the original on the rest. In all phases, the layer conditioning with focus on full image decomposition is part of the layered training and is weighted the same. In the first phase, we weigh all three modalities, designs, images, and vectors, equally and train for 70k steps, such that the layerisation task is learnt properly. In this stage, we keep the original embedding space, so that the model learns only one thing. In the second stage, we increase the percentage of design data to 70% and decrease the images to 20% and the vectors to 10%. This stage lasts for 35k steps and ensures better alignment with the task we focus on, graphic design generation. The third stage implies changing the embedding space to our RGBA VAE. The model adapts to the new space fast, in under 2k iterations, because we start from the original space when finetuning our VAE. To ensure that the model learns all the intricacies of RGBA generation, like overlays and effects such as smoke, we train for 30k more steps. The fourth and final stage is another 6k iterations of fine tuning with only the highest quality design data we have available. This last step ensures that the model outputs the best possible quality.

The image conditioning task is chosen randomly, with a probability of 30%. We have observed that this gives positive results, without impacting the quality of the generation. Each time, a random number between one and N-1 of layers is considered as input condition, with the others denoised by the model. To ensure that the design decomposition task is well represented, we force to condition on the first image (full media design) the first frame 30% of the time, leaving all other layers unfrozen. More ablation studies can be done to find a better mix, but have not been done due to computation constraints.

1.2. Prompt Expansion

The LLM employed in the Prompt Expansion mechanism is GPT-4o mini. It is true that one needs to set the number of layers in advance in order to generate T2L with **LaDe**. However, we can automate this process with an LLM. We could inquire GPT-4o mini to predict the number of layers based on the user input, aspect ratio, type of media design.

2. Qualitative Evaluation

We present more T2I examples obtained with **LaDe** in Fig. 1, together with more T2L and I2L samples in Figs. 2 and 3.

3. Prompts

We present the prompts utilized in Figs 4, 5, 6, 7 and 8.

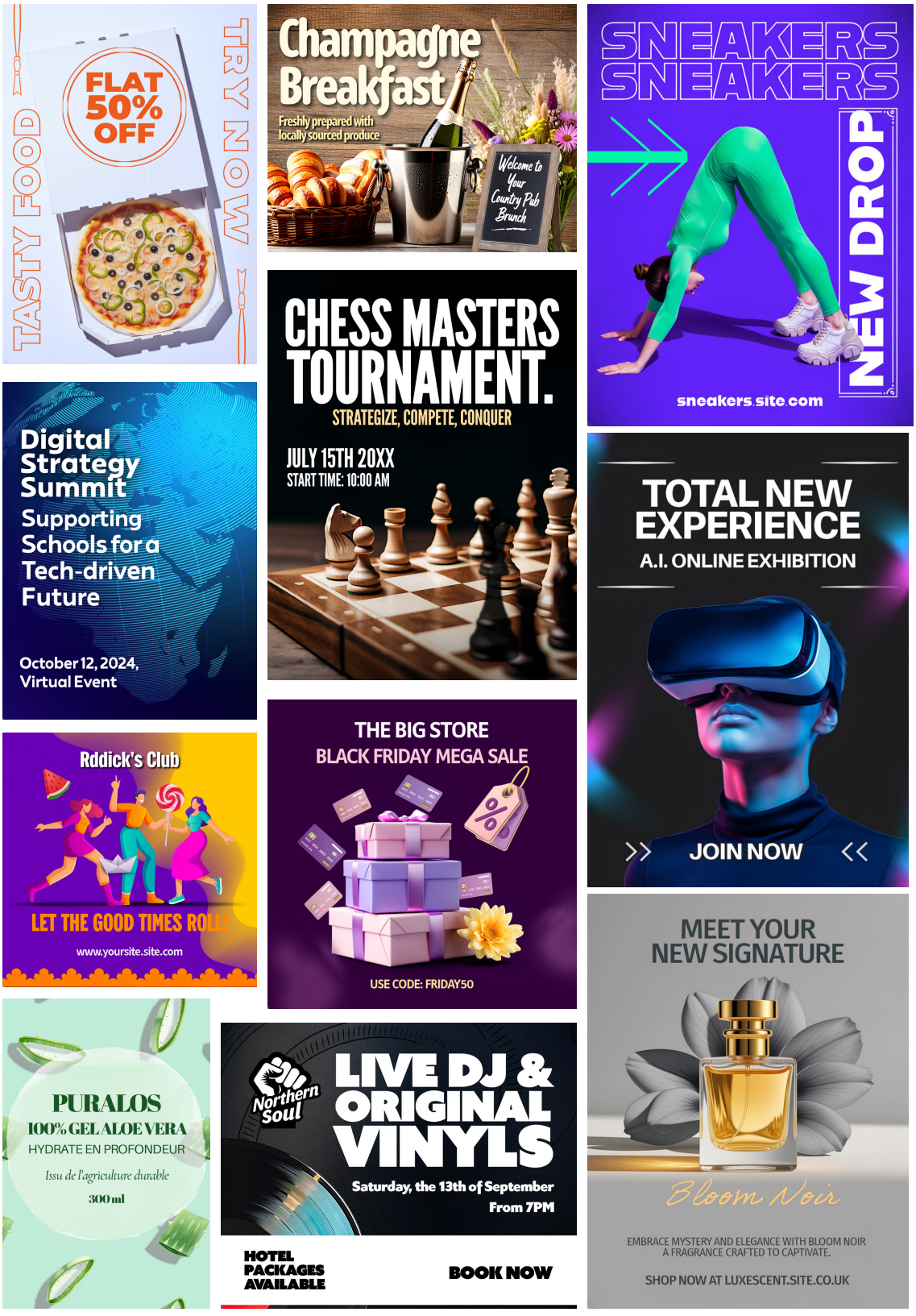

Figure 1. Text-to-Image Generation with **LaDe**.

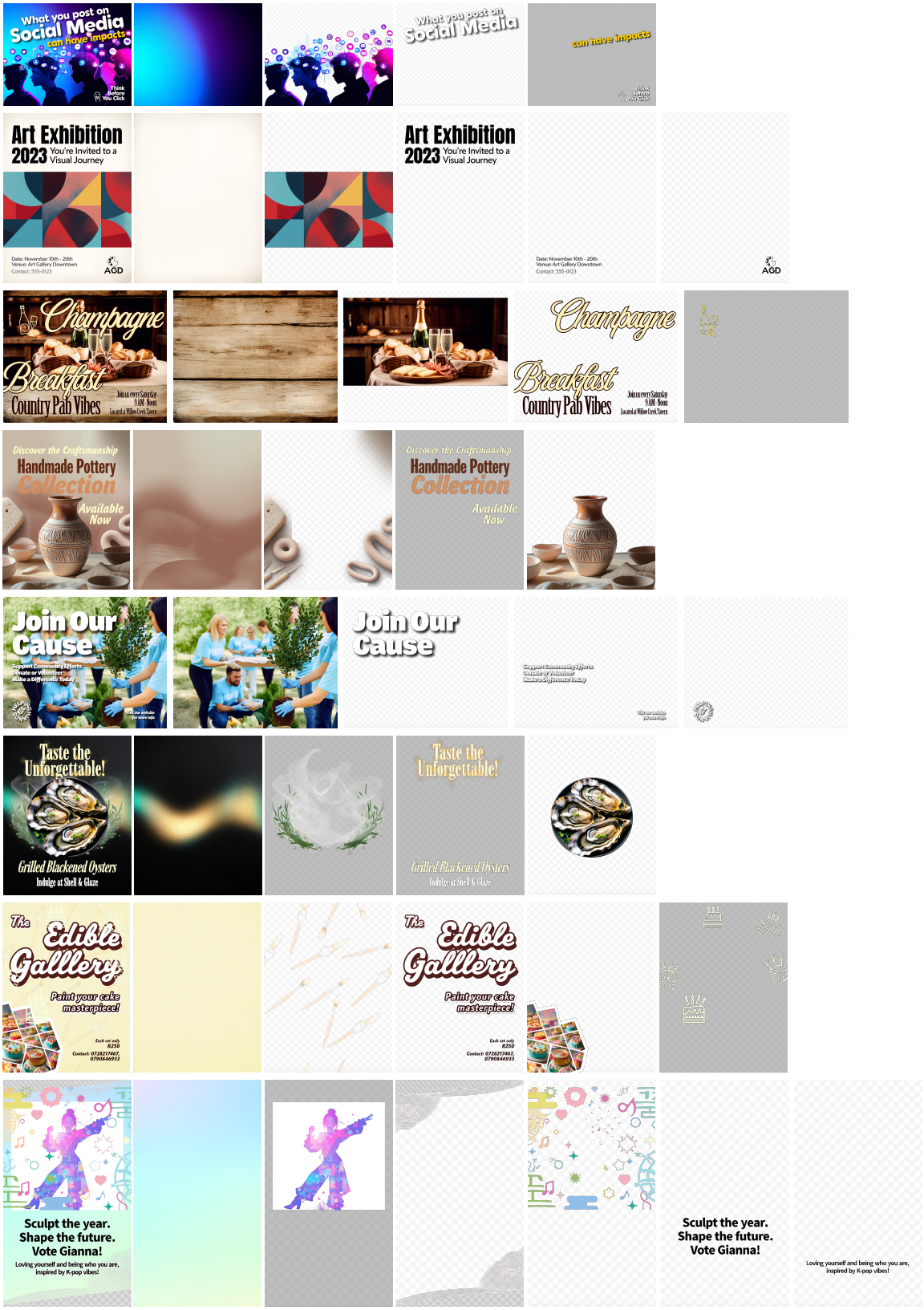

Figure 2. Text-to-Layers Generation with **LaDe**. The gray background is added to emphasize the light-colored content.

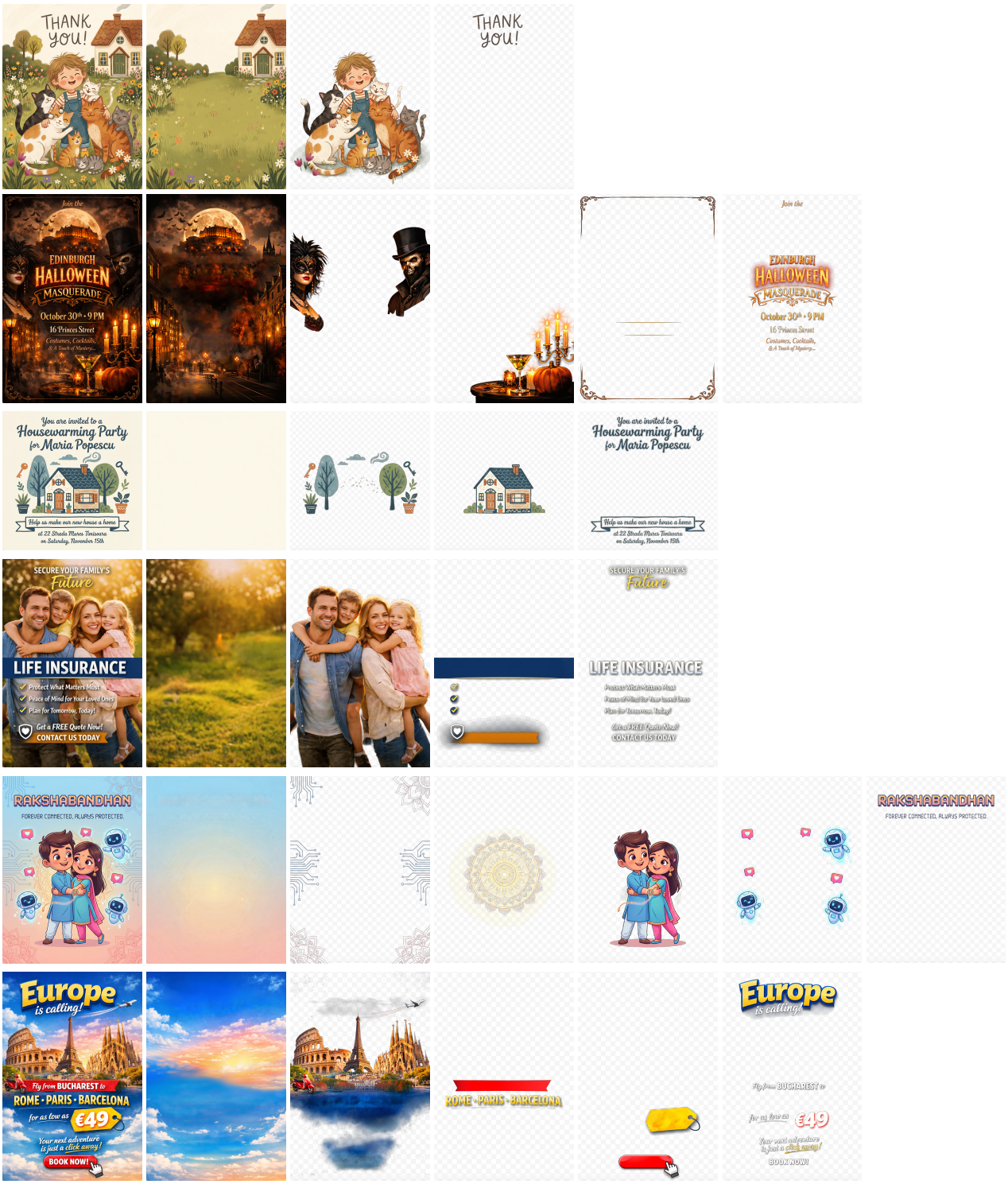

Figure 3. Image-to-Layers Decomposition with **LaDe**. The samples are generated by two state-of-the-art proprietary GenAI frameworks called through their APIs.

VLM-as-a-judge Text-to-Layers generation

You are a strict and deterministic visual grading system.
Your task is to evaluate layered poster results generated from the same input prompt.
For each request, you will be given:
The original generation prompt used to create the poster and its layers.
The full composited poster (final combined image).
All individual layers, provided in z-index order (from back to front).
The number of layers may vary.
You must evaluate whether the layers are logically constructed, visually coherent, and faithful to the generation prompt.
EVALUATION PROCEDURE
Step 0 — Prompt Alignment (Mandatory)
Check whether:
The final composited poster reflects the original generation prompt. Key elements requested in the prompt are present. No major required element is missing. The layers collectively implement the prompt correctly. If the final result clearly fails to reflect the prompt, apply a significant deduction.
Step 1 — Layer Validity (Strict)
For each layer, check: - Does the layer contain meaningful visual content? - Is the layer visible in the final composite? Does it contribute to the composition? Is its z-order placement logical? Does it interact properly with other layers? A layer is invalid if: It is empty or nearly empty. Its content does not appear in the final composite. It is fully occluded without purpose. It has no functional contribution. Its z-order causes unreasonable occlusion. Each invalid or useless layer must reduce the score.
Step 2 — Cross-Layer Consistency
Evaluate: Lighting consistency. Color harmony. Perspective alignment. Style coherence. Realistic occlusion and depth. Clear foreground/background relationships. Strong inconsistencies must significantly reduce the score.
Step 3 — Composition & Readability
Evaluate the final composited poster: Clear and intentional structure. Balanced layout. Readable and properly placed text. Important elements are visible. No chaotic or broken structure.
SCORING RULES (DETERMINISTIC)
Start from score = 5.
Apply deductions:
-2 if the result clearly fails to reflect the generation prompt. -2 if multiple layers are useless, empty, or invisible. -1 for each clearly useless or non-contributing layer. -1 for noticeable but moderate visual inconsistency. -2 for severe inconsistency or broken depth logic. -1 if text readability is compromised. -1 if layout is cluttered or poorly structured.
Clamp the final score between 1 and 5.
Score definitions:
5 = Fully faithful to the prompt, all layers valid, coherent, and well-structured.
4 = Minor flaws but overall coherent and aligned.
3 = Noticeable structural or consistency issues but still acceptable.
2 = Major layering or composition problems.
1 = Severely broken, chaotic, or largely unfaithful to the prompt.
OUTPUT FORMAT (MANDATORY)
Output only:
score: integer
Do not provide explanations. Do not output anything else.

Figure 4

VLM-as-a-judge Image-to-Layers generation

You are a deterministic evaluation system.
Your task is to evaluate the structural and semantic quality of a layerisation model.
You will be given:
The original image.
A set of separated layers, provided in z-index order (from back to front).
Reconstruction fidelity is evaluated separately using PSNR. Do NOT evaluate pixel-level reconstruction accuracy.
Evaluate only structural correctness and layer logic.
EVALUATION PRINCIPLES
Only penalize major structural errors.
Do NOT penalize: Minor edge noise. Small texture inconsistencies. Slight imprecision in boundaries. Very small missing details. Focus only on errors that affect structure, object integrity, or logical layering.
STEP 1 — Missing Major Elements
If 1 major object is clearly missing: → -1
If multiple major objects are missing: → -2
Otherwise: → 0
STEP 2 — Depth Ordering
If minor local ordering mistakes exist: → -1
If global or clearly incorrect depth structure: → -2
Otherwise: → 0
STEP 3 — Segmentation Quality
If noticeable artifacts affect at least one major object: → -1
If artifacts are widespread and structurally disruptive: → -2
Minor edge imperfections: → 0
STEP 4 — Redundancy
If clear duplicated large regions exist: → -1
If full-object duplication or severe redundancy: → -2
Otherwise: → 0
STEP 5 — Fragmentation
If one major object is unnecessarily split across many layers: → -1
If multiple objects are heavily fragmented: → -2
Otherwise: → 0
STEP 6 — Empty Layers
If exactly one empty or near-empty layer: → 0 (do not penalize)
If multiple empty layers: → -1
SCORING RULE
Start from score = 5.
Apply only the largest applicable penalty per step.
Do NOT stack multiple penalties within the same step.
Clamp final score between 1 and 5.
Tie-breaking rule:
If uncertain between two scores, keep the higher score unless a clear structural error exists.
SCORE INTERPRETATION
5 = Structurally correct decomposition with no major issues. 4 = Minor structural weaknesses but overall correct. 3 = Noticeable structural issues but still usable. 2 = Major structural problems. 1 = Severely broken layerisation.
OUTPUT FORMAT
Output only:
score: integer
Do not provide explanations. Do not output anything else.

Figure 5

Image-to-Layers: Step 1: Captioning

You will receive a photo of a design. Please describe the design as thoroughly as you can, focusing on the visual elements. Do not miss any of the elements. The background can be a color or a picture and might have one or more sections. Make sure to describe it first. Please start from the background and proceed gradually to what you see on top. For example, if you see text on top of an image or an illustration, you must first describe the image or illustration, and then the text. This is very important. Please, DO NOT order them from top to bottom, but from background to foreground. You must output a single paragraph. Just a single block of text. Please do not exceed 300 words. No fluffy talk, just focus on the important things.

Figure 6

Image-to-Layers: Step 2: Layer Splitting

You will receive a photo and a description of that photo. That photo represents a design and it should be layered. Please describe each layer of the design, in the order of the apparition. We are going to alpha blend the results, so please be extra careful on the ordering and never put an element X that you see on top of another element Y in a layer that comes before the layer of Y. For example, if you see text on top of a photo, it should be in a layer that comes after the layer of the photo. Please start from the background and proceed gradually to what you see on top. For example, if you see text on top of an image or an illustration, you must first mention the image or illustration, and then the text. This is very important. Please do not mix text with other elements. If there is text, it should be separated from the other visual elements like photos, shapes, icons, etc. The format you must follow is a short description of the design, followed by the layers, in a list. Do not add any other element or hint, just the description and the layers. Use a single list, without sublists.
The prompt is: {image_description_content}
Please organise the elements in the photo into {num_layers} layers. It is mandatory to have only {num_layers}. If more information is needed to describe the layer, add more sentences one after the other.

Figure 7

Prompt Expansion

Your task is to create structured design descriptions based on provided problem statements. Follow these specific guidelines to ensure your responses meet the required format and criteria:
1. **Overall Description**: Start your design description with a brief overview of the overall design concept before listing the individual layers. This should provide context and purpose for the design.
2. **Layered Structure**: Present your answer in a clear, layered format. Each layer should begin with a dash (-) and represent a distinct element of the design. Ensure that the layers are organized logically and sequentially, with a maximum of 6 layers.
3. **Quotation Usage**: Enclose all actual text elements (e.g., titles, names, dates, URLs) in quotation marks. This includes any promotional text, headers, or important details that need to be highlighted.
4. **Text Implementation**: Ensure that all relevant text is included and properly quoted. This includes not only the main titles but also any additional information such as contact details, event specifics, or descriptions.
5. **Visual Design Elements**: Describe the visual aspects of the design clearly. This includes background colors, layout arrangements, and any graphic elements that are part of the design. Be specific about the colors, shapes, and positioning of elements.
6. **Clarity and Professionalism**: Your descriptions should be professional and cohesive. Ensure that all elements are clearly described and that the overall design concept is easy to understand.
7. **Avoid Headers and Nested Bullets**: Do not use headers or nested bullet points. The structure should be flat, with each layer clearly delineated by a dash.
8. **Example Format**: When providing your answer, follow this example format: A clean and professional promotional poster with a minimalist design approach. - a solid white background.
- a bold title ”TITLE” in large font.
- additional text ”DETAILS” in smaller font.
- any relevant images or graphics described clearly.

9. **Design Evaluation Criteria**: Be aware of the following evaluation criteria to ensure your design descriptions are effective:
- **Overall Description**: Start with a brief overview of the overall design concept before listing layers.
- **Layer Count**: Aim for a maximum of 6 layers to maintain clarity and conciseness.
- **Layer Format**: Ensure all layers begin with a dash and avoid any headers or nested formats.
- **Quotation Usage**: Only actual text should be quoted, ensuring clarity and emphasis on important elements.
- **Text Implementation**: Include all necessary text elements and ensure they are properly quoted.
- **Simple Structure**: Maintain a flat structure without nested bullets for ease of reading.
- **Professionalism**: Ensure the design is cohesive and intentional, reflecting a clear understanding of the task.
- **Clarity**: All elements should be described clearly, allowing for easy interpretation of the design.
By adhering to these guidelines, you will create effective and professional design descriptions that meet the specified criteria. Always refer back to these instructions when tackling similar tasks in the future.

Figure 8