

Supplementary Material for PosterIQ

A Design Perspective Benchmark for Poster Understanding and Generation

We first present the statistical details of PosterIQ, followed by a description of how we obtain the evaluation results for each task. To validate the automatic evaluation, we also conduct a human evaluation and provide the annotator guideline used to construct the benchmark. Finally, we provide visual examples for each task to aid understanding.

1. Benchmark Statistics

Figure 1 summarizes the data distribution of our benchmark. For the **understanding** part (top), the dataset contains 7,765 items in total, with font-related tasks taking the largest share: *Font Attributes* (1,813, 23.3%) and *Font Size OCR* (1,400, 18.0%) together account for over 40% of all instances. OCR and layout-related tasks, including *Logo OCR*, *Poster OCR*, *Simple/Hard OCR*, *Text Localization*, *Layout Comparison*, *Empty Space*, and *Layout Generation*, form the bulk of the remaining samples, while *Style Understanding*, *Composition Understanding*, *Intention Understanding*, and *Overall Rating* provide higher-level assessments of visual design and semantics.

For the **generation** part (bottom), the 822 instances are evenly distributed: *Style Generation* (256, 31.1%) and *Intention Generation* (200, 24.3%) dominate the set, whereas *Font Generation* (135), *Composition Generation* (117), and *Dense Generation* (114) each contribute roughly 14–16% of the total, ensuring balanced coverage across different aspects of poster synthesis.

2. Task Evaluation

OCR Accuracy (Text Instance Level): For **logo OCR** and **poster OCR**, we evaluate accuracy at the text-instance level. Invisible characters (e.g., spaces and line breaks) are stripped from both prediction and ground truth, and an instance is counted as correct only under exact match. The overall accuracy (AC) is then given by the average proportion of correctly recognized text instances:

$$AC = \frac{1}{N} \sum_{i=1}^N \mathbb{1}(\hat{T}_i = T_i), \quad (1)$$

where N is the total number of text instances, \hat{T}_i is the predicted text for instance i , and T_i is the corresponding

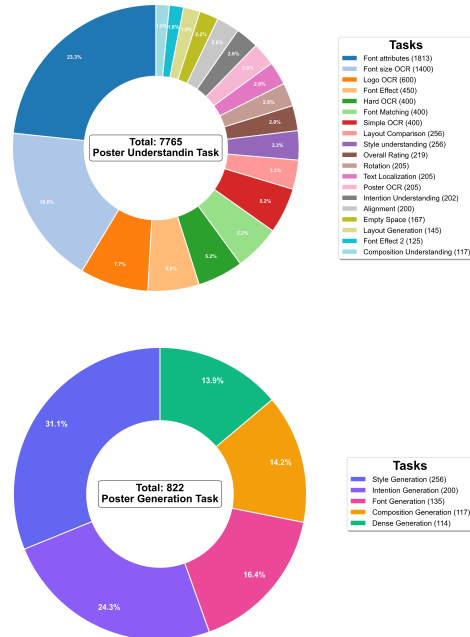


Figure 1. Benchmark statistics for understanding tasks (top) and generation tasks (bottom).

ground-truth text. The function $\mathbb{1}(\cdot)$ is an indicator that returns 1 if the condition holds and 0 otherwise. Each logo is counted as a single text instance, while a poster can contain multiple text instances.

Word-level Recall for Robust OCR: This metric is used for the synthetic OCR tasks: **Simple OCR**, **Hard OCR**, and **Font-Size OCR**. Given a ground-truth text string, we split it into consecutive units g , each consisting of five characters. For each unit g , we check whether it appears as a substring in the model-predicted long text *output*. This normalization makes scores comparable across texts of different lengths. The *word-level recall* (WR) is then defined as the proportion of ground-truth units that are recovered in the prediction:

$$WR = \frac{\sum_{g \in G} \mathbb{1}(g \subseteq \text{output})}{|G|}, \quad (2)$$

where G denotes the set of all five-character units extracted from the ground-truth text, and $\mathbb{1}(\cdot)$ is an indicator function that returns 1 when the condition is satisfied and 0 otherwise. Before segmentation, spaces and escape characters are stripped, and each word is guaranteed to appear at most once in a given image.

We define Δ as the difference between the WR scores on **Simple OCR** and **Hard OCR**. This gap reflects the robustness of the model to noise: a larger Δ indicates higher sensitivity to background clutter, missing context, rotation, and other perturbations.

For the **Font-Size OCR** task, we additionally report the standard deviation Std of the WR scores across 14 different font sizes. This metric captures the robustness of the model’s OCR performance with respect to changes in font size: lower Std indicates more stable recognition across scales.

K-Option Scoring for Multiple-Choice Tasks: For tasks where the model predicts a label from a finite set of options, we use multiple-choice style metrics. This applies to the font-related tasks (**Font Matching**, **Font Attributes**, and **Font Effects**), where we report $Score$, $Effect Score$, and $Color Score$; to the **Style Understanding** task, where we report a style classification $Score$; to the **Text Position** task, where $Alignment$ and $Rotation$ are cast as discrete choices; and to **Layout Comparison**, where we also use a multiple-choice $Score$. For multiple-choice tasks with k answer options, let the model’s accuracy be a , where $a \in [0, 1]$. The scoring formula normalizes the score such that random guessing results in a score of zero, and perfect accuracy results in a score of one:

$$Score = \max\left(0, \frac{k \cdot a - 1}{k - 1}\right). \quad (3)$$

Under random guessing, the expected accuracy is $\frac{1}{k}$, which maps to a score of zero in our formulation. When the model attains perfect accuracy $a = 1$, the score reaches one. This scoring scheme is used for the font-related tasks, the text positioning task, and the layout comparison task.

Bbox-Related Metrics: For the **Text Localization** task, ground-truth bounding boxes are first sorted in descending order by area. We then evaluate the average Intersection over Union (IoU) over the top- n predicted boxes:

$$IoU = \frac{1}{n} \sum_{j=1}^n \frac{|B_j^{\text{pred}} \cap B_j^{\text{gt}}|}{|B_j^{\text{pred}} \cup B_j^{\text{gt}}|}, \quad (4)$$

where B_j^{pred} and B_j^{gt} denote the predicted and ground-truth boxes for the j -th instance.

To further assess prompt-following behavior, we examine how well the number of predicted boxes matches the number of queried text instances. For each sample, if the model predicts fewer boxes than requested, we compute the recall as the ratio between the number of predicted boxes and the number of queried objects. If it predicts more boxes than requested, we assign a recall of 1. The final recall score is the average over all samples:

$$\text{Recall Rate} = \frac{1}{N} \sum_{i=1}^N \begin{cases} \frac{n_i^{\text{pred}}}{n_i^{\text{query}}}, & \text{if } n_i^{\text{pred}} \leq n_i^{\text{query}}, \\ 1, & \text{otherwise,} \end{cases} \quad (5)$$

where N is the total number of evaluation samples, n_i^{pred} is the number of predicted boxes for sample i , and n_i^{query} is the number of query objects specified in the prompt.

In the **Layout Generation** task, there is no uniquely correct placement of layout boxes, so standard IoU-based matching is not directly applicable. Instead, we evaluate the predicted layout by comparing the relative positions and areas of predicted boxes with those of the ground truth. Higher-quality layouts exhibit smaller $Center Bias$ and an $Area Ratio$ closer to 1.

$Center Bias$ quantifies the normalized Euclidean distance between the centers of the predicted and ground-truth boxes:

$$\text{Center Bias} = \frac{1}{N} \sum_{i=1}^N \|C_i^{\text{pred}} - C_i^{\text{gt}}\|_2, \quad (6)$$

where C_i^{pred} and C_i^{gt} are the normalized center coordinates of the predicted and ground-truth box for the i -th element.

$Area Ratio$ measures how similar the box areas are by taking the ratio between the smaller and larger area:

$$\text{Area Ratio} = \frac{1}{N} \sum_{i=1}^N \frac{\min(A_i^{\text{pred}}, A_i^{\text{gt}})}{\max(A_i^{\text{pred}}, A_i^{\text{gt}})}, \quad (7)$$

where A_i^{pred} and A_i^{gt} denote the areas of the predicted and ground-truth boxes, respectively.

Empty-Space Evaluation. In the **Empty-Space** task, both the ground truth and the model output are represented as sets of patch IDs. The ground truth set corresponds to the patches annotated as suitable empty regions, and the model is asked to predict a set of patch IDs for placing new content. We first measure the agreement between these sets using Intersection over Union (IoU):

$$\text{Patch IoU} = \frac{1}{N} \sum_{i=1}^N \frac{|\mathcal{P}_{\text{pred}} \cap \mathcal{P}_{\text{gt}}|}{|\mathcal{P}_{\text{pred}} \cup \mathcal{P}_{\text{gt}}|}, \quad (8)$$

where $\mathcal{P}_{\text{pred}}$ and \mathcal{P}_{gt} denote the predicted and ground-truth patch ID sets, respectively.

Match Accuracy. The prompt also specifies how many patch IDs should be returned. We therefore evaluate prompt-following behavior by checking whether the predicted set size matches the requested size. Match Accuracy is defined as the proportion of samples that satisfy this constraint:

$$\text{Match Accuracy} = \frac{1}{N} \sum_{i=1}^N \mathbb{1}(|\mathcal{P}_{\text{pred}}| = |\mathcal{P}_{\text{gt}}|), \quad (9)$$

where N is the number of evaluation samples, $\mathcal{P}_{\text{pred}}$ is the predicted patch set for sample i , $|\mathcal{P}_{\text{gt}}|$ is the number of patch IDs requested in the prompt, and $\mathbb{1}(\cdot)$ is an indicator function that returns 1 if the condition holds and 0 otherwise.

Point Score for Advanced Understanding Metrics: For **intention understanding**, each creative advertisement poster is paired with a set of manually annotated key elements that summarize the intended semantic or conceptual message of the design. To evaluate whether an MLLM can correctly capture and verbalize these elements, we use GPT-5 as an automatic judge. Given a model-generated caption, the judge checks for each key element whether it is correctly identified and explicitly mentioned. For a given poster, the prediction is labeled Yes if all annotated key points are covered, and No otherwise. The resulting *Point Score* is defined as the fraction of posters judged as Yes :

$$\text{Point Score} = \frac{N_{\text{Yes}}}{N_{\text{Total}}}, \quad (10)$$

where N_{Yes} is the number of posters whose model-generated captions successfully cover all key points, and N_{Total} is the total number of evaluated posters.

For **Composition Understanding**, we adopt the same Point Score metric.

Overall Rating Metric: In the **Overall Rating** task, both humans and MLLMs assign a quality score in the range 0–10 for each poster. We first normalize human and model scores to have zero mean, and then measure their agreement via cosine similarity. Formally, let $\mathbf{h} \in \mathbb{R}^N$ and $\mathbf{m} \in \mathbb{R}^N$ denote the human and model score vectors over N posters. We compute the zero-mean versions

$$\tilde{\mathbf{h}} = \mathbf{h} - \bar{h}\mathbf{1}, \quad \tilde{\mathbf{m}} = \mathbf{m} - \bar{m}\mathbf{1}, \quad (11)$$

where \bar{h} and \bar{m} are the mean human and model scores, and $\mathbf{1}$ is an all-ones vector. The final metric is the cosine similarity between the two normalized vectors:

$$\text{Overall Rating} = \frac{\tilde{\mathbf{h}}^\top \tilde{\mathbf{m}}}{\|\tilde{\mathbf{h}}\|_2 \|\tilde{\mathbf{m}}\|_2}. \quad (12)$$

Point Score for Poster Generation: For the **Dense Generation**, **Composition Generation**, and **Intention Generation** tasks, we use the *Point Score* to evaluate whether the

generated image covers all required key elements. Each generated poster is associated with multiple checkpoints (e.g., required objects, layout cues, or semantic intentions), and a MLLM judge determines for each checkpoint whether it is correctly realized in the image.

Concretely, we use GPT-5 as the automatic judge for Dense Generation and Intention Generation, and Gemini-2.5-Pro as the judge for **Composition Generation**. The Point Score is then computed as in Eq. (10), i.e., as the fraction of images whose generated content is judged to cover all annotated key points.

Score for Style Generation. In the **Style Generation** task, the generative model is instructed (via a textual prompt) to produce a poster in a specified target style. A MLLM is then asked to classify the generated poster into one of the predefined style labels. We compare the predicted style label with the ground-truth target label and compute a style generation score as the accuracy over all evaluated samples:

$$\text{Style Score} = \frac{1}{N} \sum_{i=1}^N \mathbb{1}(\hat{s}_i = s_i^{\text{gt}}), \quad (13)$$

where N is the number of generated posters, \hat{s}_i is the style label predicted by the MLLM (GPT-5) for the i -th poster, s_i^{gt} is the corresponding ground-truth style label, and $\mathbb{1}(\cdot)$ is the indicator function.

Font Richness for Font Generation. In the **Font Generation** task, the generative model is prompted to produce posters with diverse typography, where the prompt explicitly specifies the target text and encourages the use of varied font styles. After generation, we ask an MLLM with strong font understanding ability (GPT-5) to describe the typography of each poster using a fixed vocabulary of font attributes (e.g., *modern*, *playful*, *serif*, *italic*, etc.).

Let \mathcal{A} be the set of all font attributes (e.g., the 37 attributes in our implementation), and $|\mathcal{A}| = M$. For a batch of N generated posters, we define a binary indicator $x_{i,a} \in \{0, 1\}$ that equals 1 if GPT-5 assigns attribute $a \in \mathcal{A}$ to the i -th poster, and 0 otherwise. For each attribute a , we first compute its coverage ratio over the batch:

$$R_a = \frac{1}{N} \sum_{i=1}^N x_{i,a}, \quad (14)$$

which measures how frequently attribute a appears across generated posters.

The overall *Font Richness Score* is then defined as the average coverage ratio over all attributes:

$$\text{Richness} = \frac{1}{M} \sum_{a \in \mathcal{A}} R_a = \frac{1}{NM} \sum_{a \in \mathcal{A}} \sum_{i=1}^N x_{i,a}. \quad (15)$$

Intuitively, this metric reflects how widely the generator explores the font attribute space: higher values indicate that a

broader range of font attributes is realized across the generated posters.

3. Human Evaluation

To verify the reliability of our automatic evaluation, we conduct a series of human studies on both understanding and generation tasks. For several understanding tasks that rely on LLM-based textual judgments (e.g., **Composition Understanding** and **Intention Understanding**), we compare the decisions of the automatic judge with those of human annotators. On a subset, the agreement between the LLM judge and human evaluation reaches approximately 92%, indicating that our LLM-as-judge protocol is largely consistent with human judgments.

For the **Generation** tasks, we employ MLLMs to repeatedly assess the quality and faithfulness of generated images. Specifically, in **Dense Generation**, **Composition Generation**, and **Intention Generation**, the judge verifies whether multiple key pieces of information are correctly rendered in the image, while in **Font Generation** and **Style Generation**, the judge directly assigns font or style labels to each poster. To validate these automatic scores, we randomly sample 30 generated images per task and obtain human ratings under the same criteria. We observe that the relative ranking of generative models remains largely consistent across different MLLM judges and human annotators, suggesting that our automatic evaluation provides a stable and trustworthy proxy for human assessment.

4. Annotator Guideline

We adopt a multi-stage pipeline for data collection and annotation. First, we gather poster images from free sources that explicitly permit research use. All raw images are manually cleaned to remove samples with blurry content, severe artifacts, or copyright concerns. For OCR-related understanding tasks, we rely on reliable digital sources as ground-truth text. Human annotation is mainly required for layout-related tasks, advanced understanding tasks, and the overall rating task.

For each such task, at least three expert annotators independently label every sample. A senior annotator (the *leader*) then cross-checks all submissions, resolves disagreements, and filters out ambiguous cases, retaining only samples with high inter-annotator agreement. Below we summarize the concrete annotation guidelines for representative tasks.

Empty Space Task. We begin from partially edited poster designs, where some design elements have been intentionally removed from the original PSD files. The resulting posters are rendered with an overlaid grid, and the grid patch indices are visible to annotators. Each poster is sent to three annotators with the following instruction: “*This*

is an unfinished poster. New design elements need to be added. Please identify all patch IDs that you consider suitable empty regions for placing new content.” The leader aggregates the proposed patch sets and retains only those samples whose recommended regions achieve more than 90% agreement across annotators, making a final decision when minor discrepancies occur.

Composition Understanding Task. We collect posters that exhibit strong visual reconstruction or structural composition (e.g., displacement, nesting, segmentation). Each poster is distributed to multiple annotators with the instruction: “*Using concise natural language, list the visual design techniques used in this poster (such as displacement, nesting, segmentation, extension, focus, mirroring, cut-out, arrangement, etc.). Describe only the necessary composition cues in bullet points.*” The leader reviews and consolidates all descriptions, and keeps only those posters for which different annotators provide highly consistent composition cues.

Intention Understanding Task. We curate posters that contain clear visual metaphors or conceptual designs. Each poster is assigned to several annotators with the instruction: “*First, carefully read the content in the poster. Then, search for the original source or explanation of this poster online. If the external explanation aligns with your own understanding, keep this sample and decompose its core metaphor or concept into several key pieces of information. If the external explanation conflicts with your interpretation, discard this sample.*” The leader then collects and refines the key-intention annotations, merging overlapping items and removing noisy or inconsistent samples.

Overall Rating Task. For the overall quality assessment, we distribute each poster to multiple annotators with the instruction: “*Please rate the overall design quality of this poster on a scale from 0 to 10, where 0 is the worst and 10 is the best. Consider font properties, layout, textual communication, and creative concept in your score.*” Because different annotators may use different scoring ranges, we first standardize their score distributions (zero-mean and variance normalization), and then discard posters whose inter-annotator score range exceeds a predefined threshold. The remaining posters, which exhibit high rating consistency, are averaged to obtain a stable ground-truth score used in our benchmark.

5. Task Illustration

Logo OCR

Please extract text from the image, and return only the plain text without any punctuation or symbols.

ACANA

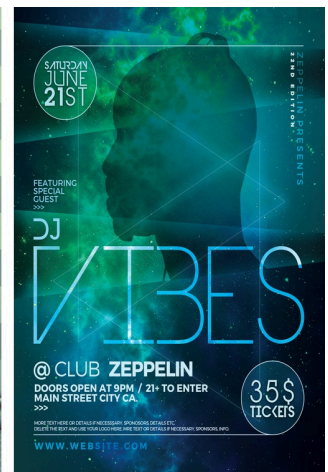
ACUTAS

ADORNEE

Adorrgon

Poster OCR

Please extract text from the image, and return only the plain text without any punctuation or symbols.



Simple OCR

Please extract text from the image, and return only the plain text without any punctuation or symbols.

minor ROOT increase show off across
concerned goodbye shape used to aware sweat IT
scientist TWENTY WOUNDED DUTY
APPROVING again mixture speed
written cure forecast surface crowded train

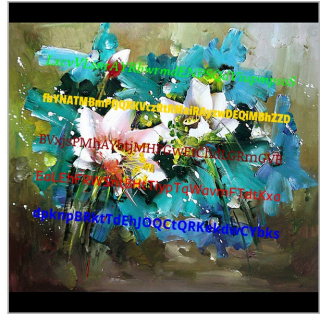
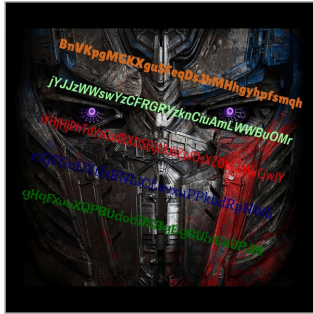
valid stand up for REVISION INSULT
weather tank reduction available
SWOLLEN WASH AWAY humorous
NOSE within NINETIETH evil SIGHT
contact separately EDUCATE

tuesday SMOOTHLY farther FIFTEEN
seven skillful means add EIGHTH pull in
together BILLION hang around with BLOW OUT
stripe APPLY CUT DOWN challenge
MACHINE joint BRAND POPULATION

EVERYBODY rapidly lead dot firmly
INCLUDING CONCERT WARMTH INFECTED BACK
level HOLLOW bed activity REASONABLY
greatly beautifully look at stand back
bake artistic PRESIDENT executive

Hard OCR

Please extract text from the image, and return only the plain text without any punctuation or symbols.



Font Size OCR

Please extract text from the image, and return only the plain text without any punctuation or symbols.

| | | | |
|---|---|---|---|
| <p>DFjBJYfjPRiABdjGKXBMe VaofwYvDyJYmwSKYQGO PBHrfcCdeSxMUhUUJdFn VRwmgtXcUoUxRqEFUDx FNuAbRowWyNNnDodArv CuGRKkDeZAJTMKXKmHk</p> | <p>DFjBJYfjPRiABdjGKXBMe VaofwYvDyJYmwSKYQGO PBHrfcCdeSxMUhUUJdFn VRwmgtXcUoUxRqEFUDx FNuAbRowWyNNnDodArv CuGRKkDeZAJTMKXKmHk</p> | <p>DFjBJYfjPRiABdjGKXBMe VaofwYvDyJYmwSKYQGO PBHrfcCdeSxMUhUUJdFn VRwmgtXcUoUxRqEFUDx FNuAbRowWyNNnDodArv CuGRKkDeZAJTMKXKmHk</p> | <p>DFjBJYfjPRiABdjGKXBMe VaofwYvDyJYmwSKYQGO PBHrfcCdeSxMUhUUJdFn VRwmgtXcUoUxRqEFUDx FNuAbRowWyNNnDodArv CuGRKkDeZAJTMKXKmHk</p> |
| <p>DFjBJYfjPRiABdjGKXBMe VaofwYvDyJYmwSKYQGO PBHrfcCdeSxMUhUUJdFn VRwmgtXcUoUxRqEFUDx FNuAbRowWyNNnDodArv CuGRKkDeZAJTMKXKmHk</p> | <p>DFjBJYfjPRiABdjGKXBMe VaofwYvDyJYmwSKYQGO PBHrfcCdeSxMUhUUJdFn VRwmgtXcUoUxRqEFUDx FNuAbRowWyNNnDodArv CuGRKkDeZAJTMKXKmHk</p> | <p>DFjBJYfjPRiABdjGKXBMe VaofwYvDyJYmwSKYQGO PBHrfcCdeSxMUhUUJdFn VRwmgtXcUoUxRqEFUDx FNuAbRowWyNNnDodArv CuGRKkDeZAJTMKXKmHk</p> | <p>DFjBJYfjPRiABdjGKXBMe VaofwYvDyJYmwSKYQGO PBHrfcCdeSxMUhUUJdFn VRwmgtXcUoUxRqEFUDx FNuAbRowWyNNnDodArv CuGRKkDeZAJTMKXKmHk</p> |
| <p>DFjBJYfjPRiABdjGKXBMe VaofwYvDyJYmwSKYQGO PBHrfcCdeSxMUhUUJdFn VRwmgtXcUoUxRqEFUDx FNuAbRowWyNNnDodArv CuGRKkDeZAJTMKXKmHk</p> | <p>DFjBJYfjPRiABdjGKXBMe VaofwYvDyJYmwSKYQGO PBHrfcCdeSxMUhUUJdFn VRwmgtXcUoUxRqEFUDx FNuAbRowWyNNnDodArv CuGRKkDeZAJTMKXKmHk</p> | <p>DFjBJYfjPRiABdjGKXBMe VaofwYvDyJYmwSKYQGO PBHrfcCdeSxMUhUUJdFn VRwmgtXcUoUxRqEFUDx FNuAbRowWyNNnDodArv CuGRKkDeZAJTMKXKmHk</p> | <p>DFjBJYfjPRiABdjGKXBMe VaofwYvDyJYmwSKYQGO PBHrfcCdeSxMUhUUJdFn VRwmgtXcUoUxRqEFUDx FNuAbRowWyNNnDodArv CuGRKkDeZAJTMKXKmHk</p> |
| <p>DFjBJYfjPRiABdjGKXBMe VaofwYvDyJYmwSKYQGO PBHrfcCdeSxMUhUUJdFn VRwmgtXcUoUxRqEFUDx FNuAbRowWyNNnDodArv CuGRKkDeZAJTMKXKmHk</p> | <p>DFjBJYfjPRiABdjGKXBMe VaofwYvDyJYmwSKYQGO PBHrfcCdeSxMUhUUJdFn VRwmgtXcUoUxRqEFUDx FNuAbRowWyNNnDodArv CuGRKkDeZAJTMKXKmHk</p> | | |

Font Matching

Please select the font that matches the target font from the options below

The quick brown fox jumps over a lazy dog

- A. *Pack my box with five dozen liquor jugs.* B. *Pack my box with five dozen liquor jugs.* C. *Pack my box with five dozen liquor jugs.*
- D. *Pack my box with five dozen liquor jugs.* E. *Pack my box with five dozen liquor jugs.* F. *Pack my box with five dozen liquor jugs.*
- G. *Pack my box with five dozen liquor jugs.* H. *Pack my box with five dozen liquor jugs.* I. *Pack my box with five dozen liquor jugs.*

From the nine options (A–I), select the one that matches the font of the target text. Please output a single answer letter directly, without any other explanation or output

Please select the font that matches the target font from the options below

The brown fox jumps

- A. *GLADLY VACAY: I ZIGZAG DAILY.* B. *Gladly vacay, I zigzag daily.* C. *Gladly vacay, I zigzag daily.*
- D. *Gladly vacay, I zigzag daily.* E. *Gladly vacay, I zigzag daily.* F. *Gladly vacay, I zigzag daily.*
- G. *Gladly vacay, I zigzag daily.* H. *Gladly vacay, I zigzag daily.* I. *Gladly vacay, I zigzag daily.*

From the nine options (A–I), select the one that matches the font of the target text. Please output a single answer letter directly, without any other explanation or output

Font Attribute

Please select the font that matches the attribute "angular" from the options

A.

The quick brown fox jumps over a lazy dog

B.

The quick brown fox jumps over a lazy dog

Please select the font that matches the attribute 'angular' from the options. Please output a single answer letter directly, without any other explanation or output.

Please select the font that matches the attribute "attention-grabbing" from the options

A.

The quick brown fox jumps over a lazy dog

B.

The quick brown fox jumps over a lazy dog

Please select the font that matches the attribute 'attention-grabbing' from the options. Please output a single answer letter directly, without any other explanation or output.

Font Effect 1

Please select the option with a shadow font effect from the choices below.

- A. **Hello World!** B. **Hello World!**
- C. **Hello World!** D. **Hello World!**

Please select the option with a shadow font effect from the choices (A–D). Reply only with the letter, no additional output.

Please select the option with a highlight font effect from the choices below.

- A. **Hello World!** B. **Hello World!**
- C. **Hello World!** D. **Hello World!**

Please select the option with a highlight font effect from the choices (A–D). Reply only with the letter, no additional output.

Font Effect 2



This image displays stylized text. Please select, from the options below, the color and effects that match the primary text color and the artistic font effect. color options: ['azure', 'beige', 'black', 'blue', 'colorful', 'gray', 'green', 'indigo', 'orange', 'pink', 'purple', 'red', 'reddish-brown', 'silvery', 'white', 'yellow'] effects options: ['blue and purple gradient light', 'bubble material', 'colorful background', 'composed of balloon', 'composed of coral', 'composed of flame', 'composed of lava rocks', 'composed of legos', 'composed of rainbow', 'composed of roses', 'composed of sand', 'composed of stars and nebulae', 'covered by frost', 'covered by snowflakes', 'covered with foam', 'crystals', 'cyberpunk neon light tube', 'daisies', 'dynamic splash', 'fireworks', 'flame', 'fluorescent', 'frost texture', 'frosty texture', 'furry', 'glass material', 'glossy', 'glossy finish', 'glowing', 'glowing particles inside', 'grasslands', 'icy texture', 'leather', 'lighting', 'metallic texture', 'outline', 'pebble-colored spots', 'pink and purple gradient', 'plants', 'porcelain', 'reflection', 'rose background', 'scattered with colored powder', 'smoke', 'translucent', 'using stars and nebulae', 'water droplets', 'wood grain texture'] Please select the possible answers from the options and output them directly.



This image displays stylized text. Please select, from the options below, the color and effects that match the primary text color and the artistic font effect. color options: ['azure', 'beige', 'black', 'blue', 'colorful', 'gray', 'green', 'indigo', 'orange', 'pink', 'purple', 'red', 'reddish-brown', 'silvery', 'white', 'yellow'] effects options: ['blue and purple gradient light', 'bubble material', 'colorful background', 'composed of balloon', 'composed of coral', 'composed of flame', 'composed of lava rocks', 'composed of legos', 'composed of rainbow', 'composed of roses', 'composed of sand', 'composed of stars and nebulae', 'covered by frost', 'covered by snowflakes', 'covered with foam', 'crystals', 'cyberpunk neon light tube', 'daisies', 'dynamic splash', 'fireworks', 'flame', 'fluorescent', 'frost texture', 'frosty texture', 'furry', 'glass material', 'glossy', 'glossy finish', 'glowing', 'glowing particles inside', 'grasslands', 'icy texture', 'leather', 'lighting', 'metallic texture', 'outline', 'pebble-colored spots', 'pink and purple gradient', 'plants', 'porcelain', 'reflection', 'rose background', 'scattered with colored powder', 'smoke', 'translucent', 'using stars and nebulae', 'water droplets', 'wood grain texture'] Please select the possible answers from the options and output them directly.

Text Localization



You are a vision-language model assistant for text detection. Given an image and a list of text elements, return a Python list of normalized bounding boxes in the format [[xmin, ymin, xmax, ymax], ...]. Each coordinate should be: \n 1. Expressed as decimals relative to the image's width (x-axis) and height (y-axis) \n 2. Precise to exactly 3 decimal places \n 3. Ordered as [left, top, right, bottom] in normalized coordinates. \n eg. [[0.123, 0.456, 0.789, 0.901],[0.050, 0.112, 0.950, 0.188],[0.001, 0.923, 0.999, 0.987]] \n Return only the list (empty if no matches). No explanations. Text Elements to locate: \n ['REVOLUTION', 'Rock', 'Show', 'FRIDAY, 28 FEB.', '@ THE INN', 'PERFORMERS', 'DOORS OPEN AT 9PM', 'MAIN STREET YOUR CITY', 'LIVE', 'WWW.WEBSITE.COM 555 666 444', 'SLAYMOORE', 'ROBSHOTS', 'THE OWLZ', '& SPECIAL GUEST', 'TICKETS', 'FACEBOOK', 'YOUTUBE', 'TWITTER', 'VIMEO']



You are a vision-language model assistant for text detection. Given an image and a list of text elements, return a Python list of normalized bounding boxes in the format [[xmin, ymin, xmax, ymax], ...]. Each coordinate should be: \n 1. Expressed as decimals relative to the image's width (x-axis) and height (y-axis) \n 2. Precise to exactly 3 decimal places \n 3. Ordered as [left, top, right, bottom] in normalized coordinates. \n eg. [[0.123, 0.456, 0.789, 0.901],[0.050, 0.112, 0.950, 0.188],[0.001, 0.923, 0.999, 0.987]] \n Return only the list (empty if no matches). No explanations. Text Elements to locate: \n ['LiveSupport band name', 'MoustacheParty', 'Got Mo?Get Free Enter', 'Your Place Name, Street 12/Dwww.movemberparty.com', '30.Nov.15']

Text Positioning

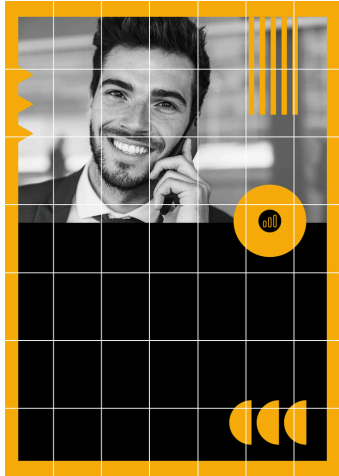


Please examine the orientation of the text in the image and choose one of the following rotation options: [clockwise rotation, no rotation, counterclockwise rotation]. If there is a rotation, the angle will not exceed 90 degrees. Please select the correct rotation direction. Output only the answer, without any additional explanation.

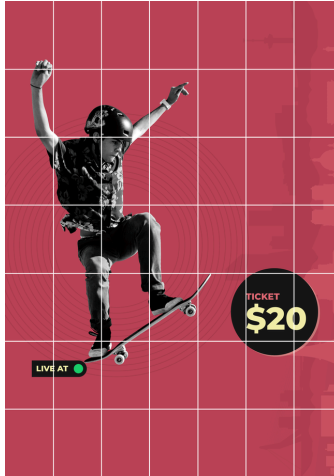


Please observe the text alignment and choose one of the following alignment options: [left-aligned, center-aligned, right-aligned]. Output only the answer in this format, without any additional explanation, for example: ['center-aligned']

Empty Space



This draft poster is overlaid with a 7x7 white grid, dividing it into 49 equally sized patches numbered 0 to 48 in reading order (top to bottom, left to right):\nRow 1: 0, 1, 2, 3, 4, 5, 6\nRow 2: 7, 8, 9, 10, 11, 12, 13\nand so on through Row 7: 42-48.\nPlease identify visually "empty" or under-utilized areas—for example, large background regions with no text or graphics. Taking into account the poster's overall symmetry and the spatial relationships between patches, select 24 patches that are most suitable for adding new design elements. Return only the 24 selected patch IDs as a list, with no additional output. \nExample answer: [35, 36, 37, 38, 39, 40, 41, 43, 44, 45, 46, 47]



This draft poster is overlaid with a 7x7 white grid, dividing it into 49 equally sized patches numbered 0 to 48 in reading order (top to bottom, left to right):\nRow 1: 0, 1, 2, 3, 4, 5, 6\nRow 2: 7, 8, 9, 10, 11, 12, 13\nand so on through Row 7: 42-48.\nPlease identify visually "empty" or under-utilized areas—for example, large background regions with no text or graphics. Taking into account the poster's overall symmetry and the spatial relationships between patches, select 27 patches that are most suitable for adding new design elements. Return only the 27 selected patch IDs as a list, with no additional output. \nExample answer: [35, 36, 37, 38, 39, 40, 41, 43, 44, 45, 46, 47]

Layout Comprison

Which image, A or B, has a more visually appealing layout?



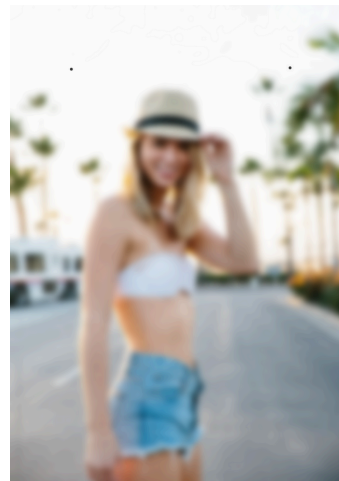
Which poster image, A or B, has a more visually appealing layout? Please output A or B directly

Which image, A or B, has a more visually appealing layout?



Which poster image, A or B, has a more visually appealing layout? Please output A or B directly

Layout Generation



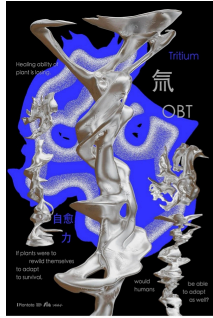
Natural-language Descriptions: The text "New arrival" vertically dominates a significant portion of the right side, taking up much of the vertical space and giving a sense of prominence in the design. "Outfita Instagram Stories Template" is discreetly placed at the bottom left corner of the design, nestled within a darker band, and occupies a smaller area, offering a subtle foundation. The phrase "SHOP NOW" is centrally aligned within an orange circular element on the right side, near the top, creating a focal point that draws attention without dominating the space. Lastly, "DAILY LOOK" is positioned at the top left, within a narrow horizontal strip, giving a header-like feel with minimal spatial coverage. \nBased on the image layout and the natural-language descriptions, directly generate the positions where the following text elements could be placed in the image. Texts: ["New arrival", "Outfita Instagram\Stories Template", "SHOP\NOW", "DAILY LOOK"] \nOutput only the list of bounding boxes ([x_min, y_min, x_max, y_max]) for each text element, using normalized decimal coordinates. Here is an example output:[[0.123, 0.456, 0.789, 0.901],[0.050, 0.112, 0.950, 0.188],[0.001, 0.923, 0.999, 0.987]]



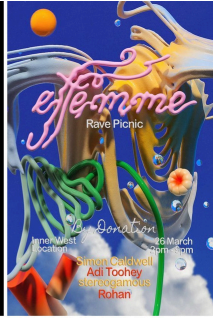
Natural-language Descriptions: The text "Mamma Mia!" occupies a prominent position near the top of the layout and stretches broadly across the upper part of the design. It is centrally placed, giving it significance and drawing immediate attention. Below it, there's the phrase "Love is companionship," which also spans widely but covers a slightly smaller area. This text is positioned just under "Mamma Mia," continuing the thematic engagement across the upper section. Lastly, the word "beautiful" is horizontally oriented toward the bottom of the grouping of text elements, occupying an area slightly larger than the previous text. It features a flowing script style, creating an elegant touch as it is positioned nearer to the middle of the layout, giving balance to the overall design. Together, these text elements create a harmonious overlay above the lower central image, effectively blending with the floral motif around the edges. \nBased on the image layout and the natural-language descriptions, directly generate the positions where the following text elements could be placed in the image. Texts: ["beautiful", "Mamma Mia", "Love is companionship"] \nOutput only the list of bounding boxes ([x_min, y_min, x_max, y_max]) for each text element, using normalized decimal coordinates. Here is an example output:[[0.123, 0.456, 0.789, 0.901],[0.050, 0.112, 0.950, 0.188],[0.001, 0.923, 0.999, 0.987]]

Style Understanding

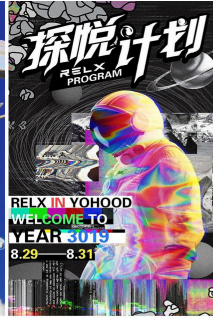
You are a professional visual design analyst. Task: Given an input poster image, identify its "dominant visual style" based on composition, color palette, typography, and artistic features. Return only one style name from the following list: [Flat Design, Illustrative Style, Minimalist Style, Japanese Style, New Chinese Aesthetic, Japanese Style, Cinema 4D Style, Retro Style, Diffuse Glow Style, Acid Graphics, Papercut Style, Pixel Art, Pop Art, Vaporwave Style, Cyberpunk Style, Glitch Art, Memphis Style, Typographic Minimalism] Guidelines: Do not add explanations or probabilities. Output must exactly match one of the items in the list.



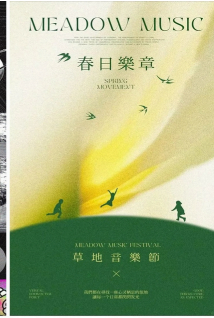
Acid Graphics



Cinema 4D Style



Cyberpunk Style



Diffuse Glow Style



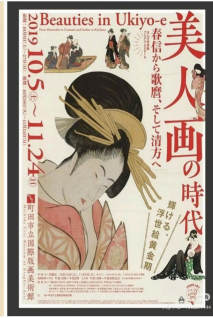
Flat Design



Glitch Art



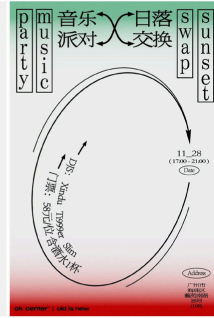
Illustrative Style



Japanese Style



Memphis Style



Minimalist Style



New Chinese Aesthetic



Papercut Style



Pixel Art



Pop Art



Retro Style



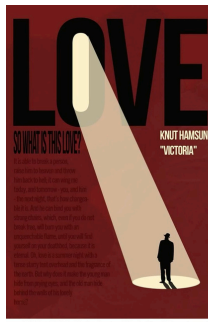
Typographic Minimalism



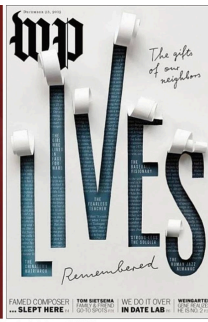
Vaporwave Style

Composition Understanding

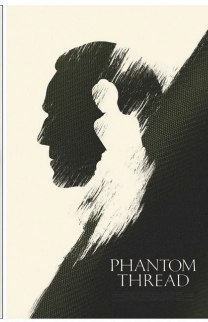
Please describe the poster in detail, including its composition, visual hierarchy, spatial relationships between elements, typography.



- An off-white spotlight starting from the upper left to the lower right
- Dark blocks of contrasting color



- A handcrafted papercraft 3D effect



- The dark area extends past the dividing line to the left, forming the silhouette of a man in profile.
- Within the man's profile, negative space reveals a half-length profile of a woman.
- A distinct color gradient line runs from the lower left corner to the upper right.
- Positive and negative space interface.



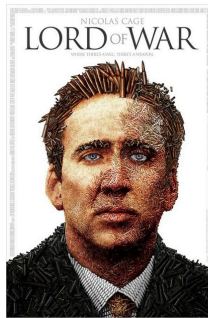
- Copied images of a face in different poses
- Each square holds a cropped image of the face in a different pose
- Three by three grid of squares



- Within the triangle appears a white-like female figure in white, wearing a black lower hat, her head bowed, gazing downward.
- One hand grips a sharp dagger that thrusts beyond the triangle, its blade aimed straight at the viewer.
- Note the base of the larger triangle, within it, there is another triangle containing a white female statue.



- Towering city skyscrapers all around
- Low-angle perspective
- The fractured edges of the buildings usually outline a clear but silhouette



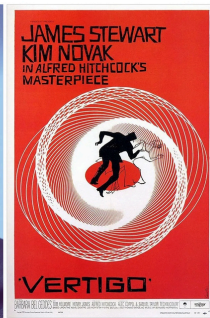
- An upper body silhouette of a man in a suit.
- The top and edges of the head are filled with hundreds of copper-colored bullets and shells.
- The suit jacket is densely laid out with dark shell casings.
- The start color is nearly black with silver-gray bullets to create a pleated button.
- The tie knot is composed of brass-colored casings.



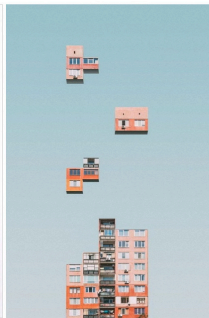
- Diverse green characters and symbols rain-down vertically



- The face is split into sections
- A downward-pointing triangle
- Three color bands



- At the center of the composition is a spiral vortex composed of fine white lines, lightening inward upon its base.
- At the heart of the vortex are two silhouetted figures.



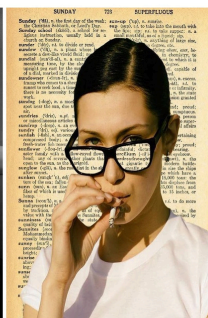
- Vertical
- The building appears to be built upon
- Telescreens falling from the sky
- Assembly and Recalibration



- Three rows by two columns



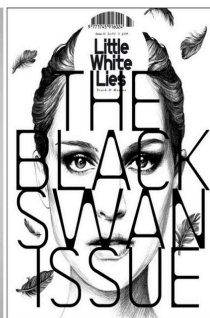
- Misaligned face
- Split face
- Right hand placed beside the head making a finger-gun gesture



- Fingers pinching a cigarette at the tips in a smoking gesture
- The glasses are cut-out, revealing the dictionary page beneath



- Letters stacked from top to bottom
- The letters are hollowed, revealing fragmented scenes and silhouettes, snow, and cracks



- Black-and-white sketch of a woman's face as the background
- Large black lettering
- Lettering overlapping the face and filling the entire frame



- A vertical axis splits the composition in two.
- The four letters are arranged vertically.
- The four vertically stacked letters are cut-out, revealing the face.



- Three triangles arranged from left to right
- Triangular cutouts reveal a woman's profile

Intention Understanding

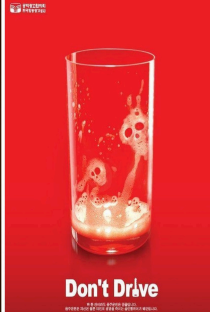
Please provide a detailed description of this poster and explain the design metaphors used in it.



- This is an advertisement poster for Volkswagen.
- The use of a well-known production and reliability, and the design includes subtle as a prominent characteristic of the Volkswagen brand, reinforcing with the text below "Golf" as the "sold out".
- The poster cleverly transforms the iconic VW logo into a metaphor for safety using overlapping circles.
- The poster uses visual metaphors to convey safety and engineering excellence.



- This is a commercial advertisement poster for Buckler non-alcoholic beer.
- The poster uses text to create a tall tower, with the top text placed like beer foam, and the text positioned in another form to mimic the appearance of beer.
- The bold large text itself visually resembles a bottle, humorously deconstructing the idea of drunkenness.
- The text above and below the text plays the top and bottom of the beer glass.



- This is a public service advertisement by the Korea Broadcast Advertising Corporation.
- Against a red background, the text in a nearly empty glass forms the shape of a skull, symbolizing the deadly risk of drinking an empty glass of alcohol.
- The bold text at the bottom, "Don't Drive," highlights the source of the danger, which is drunk driving.
- The "V" in "Drive" is designed in the shape of a key, associating it with using a key.



- This is an environmental public welfare poster from The Times of India.
- In the image, farmers are catching the fish's skull, the head of a plastic bottle.
- The text below asks the audience, questioning "Still using plastic?"
- The poster uses the visual metaphor of "plucking the audience to rethink the use of single-use plastics."
- A QR code invites readers to connect to opposing single-use plastics in a call to action.



- This is a public welfare poster for Feeding America and Food Bank.
- The poster emphasizes different perspectives on the snack, whether it is a snack or dinner, by using the image of a Lay's potato chips bag and target, which is both above and below the image.
- The text at the bottom states that one in six Americans is food insecure, using statistics by the food bank.
- Logos of Feeding America and Food Bank are included to strengthen the call.



- This is a public service poster by the Brazilian anti-smoking organization (ABRCA).
- The poster uses the visual metaphor of a cigarette pack walking towards a coffin-shaped cigarette to depict the deadly consequences of smoking.
- The slogan "A warm welcome to death" emphasizes the lethal effects of smoking.
- The oversized cigarette coffin contrasts with the small figure, highlighting the



- This is a commercial poster for Heinz Spicy Tomato Ketchup.
- The image shows a sofa with red seats in a restaurant, echoing the color of the ketchup.
- The dark red stain on the backrest of the sofa resembles several stains, implying the intense heat of the spicy ketchup.
- The Heinz Spicy Tomato Ketchup logo in the lower right corner reinforces the presence of the advertised product.



- This is a public service poster by IBM.
- The visual design conveys negative space to create and convey dual meanings.
- The silhouette of a woman's face also shows the image of a rooster, symbolizing food and freshness.
- The design expresses the unity between human interaction and advancement in food supply chain technology.
- The vertical use of color blocks attracts on the information and visual.



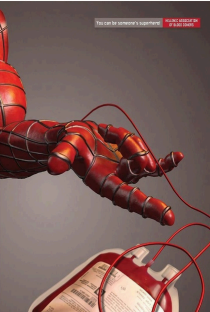
- This is a commercial poster for Cotygate dental clinic.
- The poster mainly consists of three photos of people smiling and coping with gaps in their teeth.
- People's attention is first drawn to the fact that each man's teeth have something in them. However, upon closer inspection, you'll notice that in the first picture, the woman's hand on the man's shoulder has extra fingers, in the second picture, there's an extra phantom arm on the man's shoulder, and in the third picture, the man is missing an ear. This visual



- This is a commercial advertisement poster for Heinz ketchup.
- French fries are depicted vertically, leaning toward the ketchup bottle, reinforcing the slogan "When they can't resist the temptation of ketchup."
- The background is vibrant red, matching the brand's color. The minimalist design highlights the strong brand combination and immediate appeal of Heinz ketchup and fries, while evoking a sense of fun and energy.



- This is a commercial poster for AXIX wet wipes.
- The main visual of the poster is a yellow toilet cap spinning red ketchup halfway, with the liquid divided into two parts, the middle part very clean and endless.
- Below, a hand is depicted pulling an AXIX wet wipe from the packaging, accompanied by the slogan "Nothing's quicker than this," explaining that the speed of the spinning liquid is wiped clean by AXIX wet wipes. It emphasizes the wipe's ability to quickly remove and conveniently complete the cleaning.



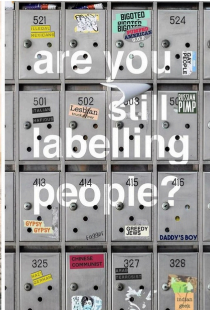
- This is a public service advertisement poster by the Greek Blood donors association.
- The main visual on the poster depicts an arm similar to Spider-Man's costume, metaphorically representing a superhero or hero image. It contrasts the extended blood tube with the blood pack, depicting the scene of donating blood.
- The slogan "You can become someone's superhero" encourages the public to view blood donation as a heroic act, giving individuals the power to save lives.



- This is an AEA commercial poster.
- The poster creatively associates sleeping with anti-aging. The main visual compares AEA's cream to anti-aging cream, placed in a transparent cream jar labeled "SLEEP," accompanied by a slogan about the most natural anti-aging method being sleep.
- The phrase "Tomorrow begins tonight" at the bottom emphasizes the importance of a comfortable sleep.



- This is a Sincere commercial poster.
- The poster humorously depicts Godzilla walking on a lake surrounded by donuts, accompanied by the slogan "You're not when you're hungry," suggesting that business involves hunger.



- This is a business marketing poster.
- The creative metaphor of categorizing individuals, much like categorizing mail. The question "Are you still labelling people?" challenges the audience to consider how they label.
- This is a powerful call to embrace diversity and reject simplistic categorization.



- This is a Berger paint commercial poster.
- The advertisement features a white billboard against a blue sky background, with an insouciant painter using a roller brush to apply paint that blends seamlessly with the sky behind, symbolizing Berger's soft and "natural fresh color."



- This is a commercial advertisement poster for Kibon ice cream.
- A close-up view of the creamy, smooth ice cream highlights the product's "70% milk" (70% milk content), emphasizing its creaminess. The curled edge in the bottom right corner of the poster reveals a waffle cone pattern, metaphorically representing the ice cream.



- This is a commercial poster for Calcom Broadband.
- The poster depicts a great call illustration at the bottom of the USB plug, showing various online activities, emphasizing the extensive possibilities of a broadband connection, and underscoring the high-speed internet capability of Calcom Broadband.

Overall Rating

Please carefully evaluate the given poster and assign a score from 1 to 10 based on the following three aspects:\n1. Typography Design – clarity, creativity, and consistency of font usage.\n2. Layout Composition – balance, hierarchy, and visual flow of the overall structure.\n3. Visual Metaphor and Aesthetics – how effectively the poster conveys meaning through imagery, symbolism, and color harmony.\nOutput only a single number (1–10) representing your overall score, without explanation or extra text.



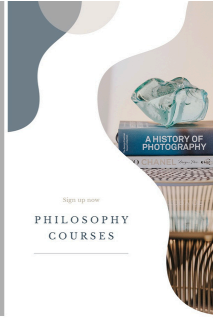
Score: 6.3



Score: 4.1



Score: 5.6



Score: 2.0



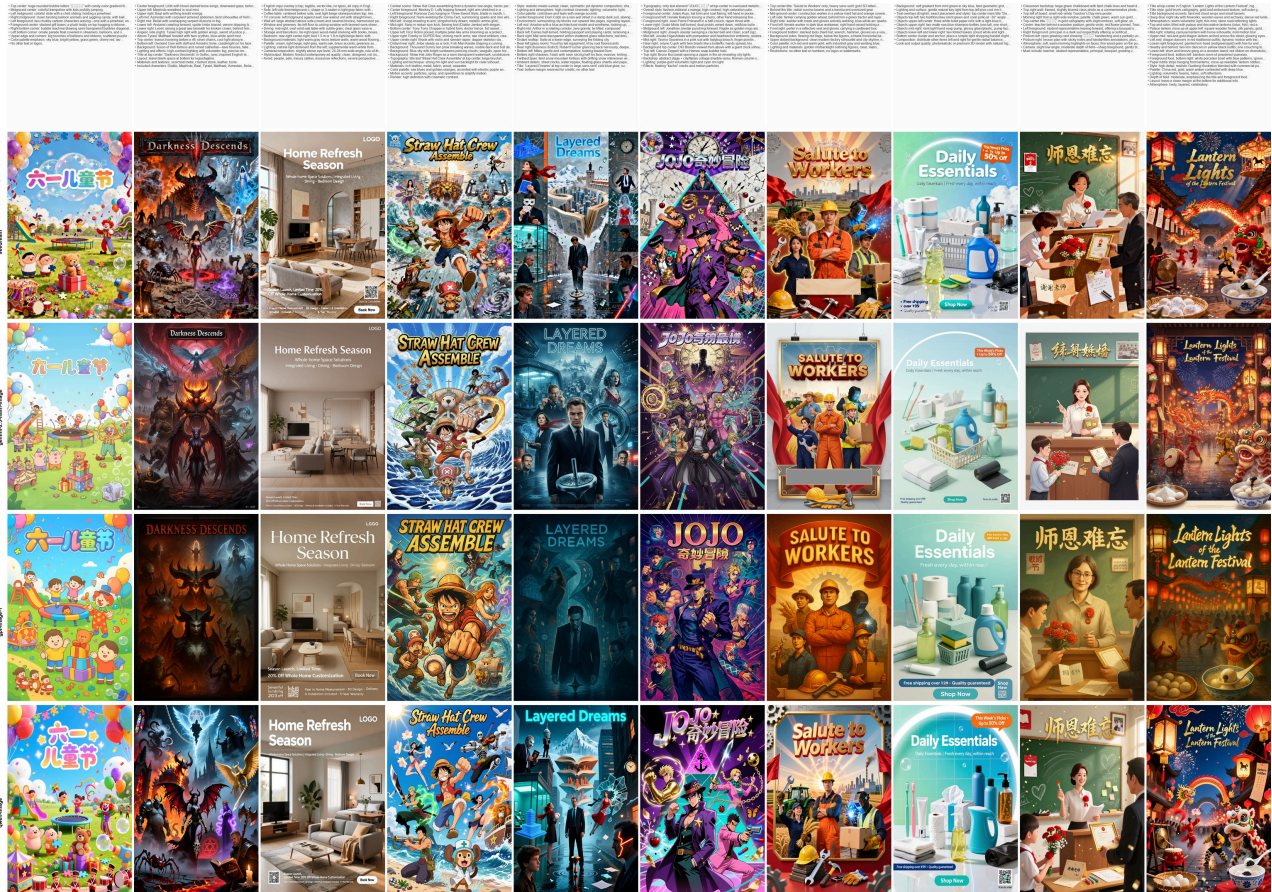
Score: 5.5



Score: 5.4

6. Generation Task Results

Dense Generation Results




































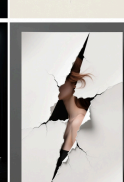




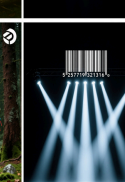


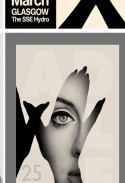
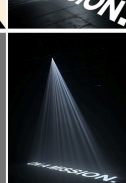

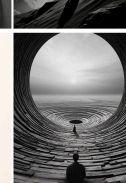



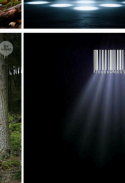
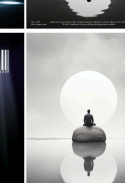
Font Generation Results

The image displays a grid of 36 promotional cards, arranged in 6 rows and 6 columns. Each card represents a different font generation result for a specific event or service. The cards are color-coded and contain various details such as dates, times, and contact information. The events include:

- ADVERTISING THAT GETS RESULTS**: Small pledges. Big impact. Focus on digital marketing strategies.
- HELP US LAUNCH PROJECT NAME**: Small pledges. Big impact. Focus on community support.
- National Day Celebration**: Unity | Heritage | Hope. Celebrating the 100th anniversary of the United States.
- Commencement 2025**: Celebrating the Class of 2025. Focus on graduation ceremonies.
- Corporate Identity**: Define. Align. Amplify. Focus on brand identity and consistency.
- 10th Anniversary Celebration**: A decade of excellence, innovation, and community. Focus on celebrating a decade of success.
- Investment Forum 2026**: Navigate Markets. Build Wealth. Invest with Insight. Focus on financial investment opportunities.
- TechElogyExpo 2026**: Invest the Next. Together. Focus on technology and innovation.
- Children's Day Celebration**: A day of wonder, play, and possibilities. Focus on children's activities and entertainment.
- MAISON ECLAT**: Luxury, distilled. Focus on high-end fashion and accessories.

The cards are designed to showcase the versatility and quality of the font generation process, with each card featuring a unique layout and color scheme.

Composition Generation Results

| | | | | | | | | | |
|--|---|--|---|--|---|--|--|---|---|
| <p>• The right subject is an open palm. • The right is a view of the people being broken into the sea.</p>  | <p>• Silhouette of two hands. • The silhouette of a woman's face inside the silhouette.</p>  | <p>• A hand with a light beam pointing down from above. • A hand with a light beam pointing down from above.</p>  | <p>• A jagged lightning bolt. • The jagged lightning bolt is inside a hole in the wall.</p>  | <p>• From the inside looking out. • A person is looking out from a tunnel.</p>  | <p>• On the right is a silhouette of a human figure. • The silhouette is a collage of images.</p>  | <p>• A man's profile silhouette. • The silhouette is a collage of images.</p>  | <p>• The man's upper body appears 'fused' to the left side of the image. • The body appears to be fused.</p>  | <p>• A hand with a light beam pointing down from above. • The hand is a silhouette.</p>  | <p>• A hand with a light beam pointing down from above. • The hand is a silhouette.</p>  |
| <p>• A hand with a light beam pointing down from above. • The hand is a silhouette.</p>  | <p>• Silhouette of two hands. • The silhouette of a woman's face inside the silhouette.</p>  | <p>• A hand with a light beam pointing down from above. • The hand is a silhouette.</p>  | <p>• A jagged lightning bolt. • The jagged lightning bolt is inside a hole in the wall.</p>  | <p>• From the inside looking out. • A person is looking out from a tunnel.</p>  | <p>• On the right is a silhouette of a human figure. • The silhouette is a collage of images.</p>  | <p>• A man's profile silhouette. • The silhouette is a collage of images.</p>  | <p>• The man's upper body appears 'fused' to the left side of the image. • The body appears to be fused.</p>  | <p>• A hand with a light beam pointing down from above. • The hand is a silhouette.</p>  | <p>• A hand with a light beam pointing down from above. • The hand is a silhouette.</p>  |
| <p>• A hand with a light beam pointing down from above. • The hand is a silhouette.</p>  | <p>• Silhouette of two hands. • The silhouette of a woman's face inside the silhouette.</p>  | <p>• A hand with a light beam pointing down from above. • The hand is a silhouette.</p>  | <p>• A jagged lightning bolt. • The jagged lightning bolt is inside a hole in the wall.</p>  | <p>• From the inside looking out. • A person is looking out from a tunnel.</p>  | <p>• On the right is a silhouette of a human figure. • The silhouette is a collage of images.</p>  | <p>• A man's profile silhouette. • The silhouette is a collage of images.</p>  | <p>• The man's upper body appears 'fused' to the left side of the image. • The body appears to be fused.</p>  | <p>• A hand with a light beam pointing down from above. • The hand is a silhouette.</p>  | <p>• A hand with a light beam pointing down from above. • The hand is a silhouette.</p>  |
| <p>• A hand with a light beam pointing down from above. • The hand is a silhouette.</p>  | <p>• Silhouette of two hands. • The silhouette of a woman's face inside the silhouette.</p>  | <p>• A hand with a light beam pointing down from above. • The hand is a silhouette.</p>  | <p>• A jagged lightning bolt. • The jagged lightning bolt is inside a hole in the wall.</p>  | <p>• From the inside looking out. • A person is looking out from a tunnel.</p>  | <p>• On the right is a silhouette of a human figure. • The silhouette is a collage of images.</p>  | <p>• A man's profile silhouette. • The silhouette is a collage of images.</p>  | <p>• The man's upper body appears 'fused' to the left side of the image. • The body appears to be fused.</p>  | <p>• A hand with a light beam pointing down from above. • The hand is a silhouette.</p>  | <p>• A hand with a light beam pointing down from above. • The hand is a silhouette.</p>  |
| <p>• A hand with a light beam pointing down from above. • The hand is a silhouette.</p>  | <p>• Silhouette of two hands. • The silhouette of a woman's face inside the silhouette.</p>  | <p>• A hand with a light beam pointing down from above. • The hand is a silhouette.</p>  | <p>• A jagged lightning bolt. • The jagged lightning bolt is inside a hole in the wall.</p>  | <p>• From the inside looking out. • A person is looking out from a tunnel.</p>  | <p>• On the right is a silhouette of a human figure. • The silhouette is a collage of images.</p>  | <p>• A man's profile silhouette. • The silhouette is a collage of images.</p>  | <p>• The man's upper body appears 'fused' to the left side of the image. • The body appears to be fused.</p>  | <p>• A hand with a light beam pointing down from above. • The hand is a silhouette.</p>  | <p>• A hand with a light beam pointing down from above. • The hand is a silhouette.</p>  |

