

Supplementary Materials: Product2Img

Anonymous Authors

A IMPLEMENTATION DETAILS OF IDR-LMM

System Prompt:

You are a helpful e-commerce assistant, possessing extensive expertise in deep learning and a specialization in refining text-to-image model iterations. Your current objective is to examine a dataset containing product names and their corresponding images, with a focus on the product image backgrounds. Your role is to assess these backgrounds and assign a rating on a scale from 1 to 5 for their suitability as training data. You are to judge these backgrounds on their aesthetic appeal, the pertinence to the product, their ability to keep the product in focus without distraction, their natural and realistic look, overall image quality including the absence of white borders, and the exclusion of human figures that could complicate model training.

Rate the images using the following scale:

- 5: The image excels in all criteria, demonstrating excellent quality and is ideal for training. Assign a rating of 5 with caution.
- 4: The image may possess minor imperfections; however, its inclusion in the dataset is still deemed beneficial.
- 3: The image quality is acceptable, with only minor mismatches between the background and product. Score such images as 3.
- 2: The image shows subpar quality or a clear mismatch between the product and the background. Assign a rating of 2.
- 1: The image has significant flaws, such as irrelevant backgrounds or the presence of white borders, making it unsuitable for training data.

Apply your expertise to ensure that each image is meticulously rated, facilitating the refinement of our text-to-image model with only the most exemplary training data.

User Prompt:

Here are some examples:

<Start Of Examples>

{few shot cases}

<End Of Examples>

If first round:

Here is the sample to analyze and score, your scoring should be strict, with appropriate deductions for any point of poor performance:

Else:

It is now round {round} of iteration, and the scoring should be more rigorous. Feedback from the last iteration process was obtained: {feedback from last iteration}

In this round, greater attention should be paid to the feedback obtained from the previous round, while the focus on prior guidelines can be appropriately reduced. Here is the sample to analyze and score:

Product Name: {product name}

Product Image: {product image}

Figure 1: p_R for IDR-LMM (Data Selection).

In Iterative Data Refinement with Self-improved LMM, we utilize GPT4-Vision as the LMM with a temperature of 0.0 to ensure the accuracy of scoring and reproducibility of results. For ineffective responses (those without a score in the reply), the temperature is increased to 0.1. The prompt p_R used are illustrated in Figure 1 and 2, and the distribution of scores over three iterations is shown in Figure 3. The distribution of scores becomes more concentrated around higher values with each iteration, demonstrating the consistency of the LMM’s multi-round scoring and the effectiveness of Data Refinement.

System Prompt:

You are a helpful e-commerce assistant, possessing extensive expertise in deep learning and a specialization in refining text-to-image model iterations. Your current objective is to evaluate the model from the last iteration, which is designed for generating product-relevant backgrounds. Your task involves assessing the quality of the backgrounds produced and providing feedback to refine the training data for the next iteration. Please conduct your evaluation based on the following criteria:

1. Overall aesthetic appeal of the generated background.
2. Relevance of the background to the product or its effectiveness in showcasing the product.
3. Attention to detail within the background, such as shadows, lighting, and the realism of certain elements.
4. Adherence to common sense, ensuring items do not appear to be floating unnaturally, with exceptions made for creative concepts such as products intentionally suspended over water.
5. The background should complement the product without drawing attention away from it.
6. Judge realism: The image should look as true-to-life as a real photograph.

For example, the background of this image is quite plain, so in the next round of data filtering, we should avoid backgrounds that are relatively plain; or the current background is not very relevant to the product, so next time, attention should be paid to enhancing the relevance of the image to the product.

Your feedback should not be limited to the aspects mentioned above. Feedback in any other direction is encouraged, so please exercise your initiative.

User Prompt:

Here is the sample to analyze:

Product Name: {product name}

Product Image: {product image}

Please provide the briefest feedback, within 20 words:

System Prompt:

You are a helpful e-commerce assistant, possessing extensive expertise in deep learning and a specialization in refining text-to-image model iterations. Your current objective is to evaluate the model from the last iteration, which is designed for generating product-relevant backgrounds. Your task is to summarize feedback from all previous samples and provide concise and constructive feedback for the next round of data cleaning. Specifically, your feedback should be considerations for selecting training data in the next round. Feedback that has a low probability of occurring should be ignored, concentrating on summarizing feedback that occurs more frequently.

User Prompt:

Here is the feedback from all previous samples; ignore positive feedback and feedback with a low occurrence probability: {all feedbacks}

Figure 2: p_F (Top) and p_S (Bottom) for IDR-LMM (Model Evaluation).

B SUPPLEMENTAL EXPERIMENTS

B.1 The impact of λ

We explore the influence of the hyperparameter λ on performance by experimenting with different values while maintaining stability in other parts. Optimal performance is achieved at $\lambda = 0.01$, as illustrated in Table 1. Setting λ too high skews learning towards the auxiliary task, while too low a value diminishes its benefits. Hence, we select $\lambda = 0.01$ as the default for our experiments.

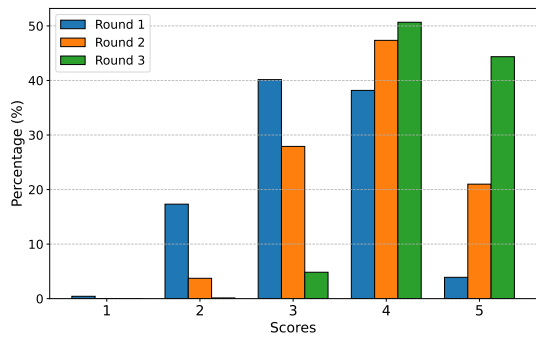


Figure 3: Score Distribution of Three Rounds.

Table 1: Hyperparametric exploration experiments, the training data are the data of the first round of data iteration. Setting $\lambda = 0$ indicates the absence of the CBA.

λ	FID ↓	Aesthetics ↑	CLIP-c-B ↑
0	2.72	5.24	56.10
0.001	2.67	5.25	57.15
0.01	2.67	5.25	57.30
0.1	2.72	5.24	56.64

B.2 The impact of γ

Table 2: The impact on performance with varying parameter γ in the first round of IDR-LMM.

γ	Data Size	FID ↓	Aesthetics ↑	CLIP-c-B ↑
3	19680	2.66	5.24	57.03
4	10071	2.67	5.25	57.30
5	933	3.62	5.26	57.67

We also conducted experiments on the impact of threshold values γ in the IDR process. As shown in Table 2, excessively high thresholds lead to an extremely limited amount of data, which can result in overfitting and thus increase the FID score. Conversely, thresholds that are too low yield marginal improvements in terms of aesthetics as well as the relevance between the background and the product.