CGI: Identifying Conditional Generative Models with Example Images

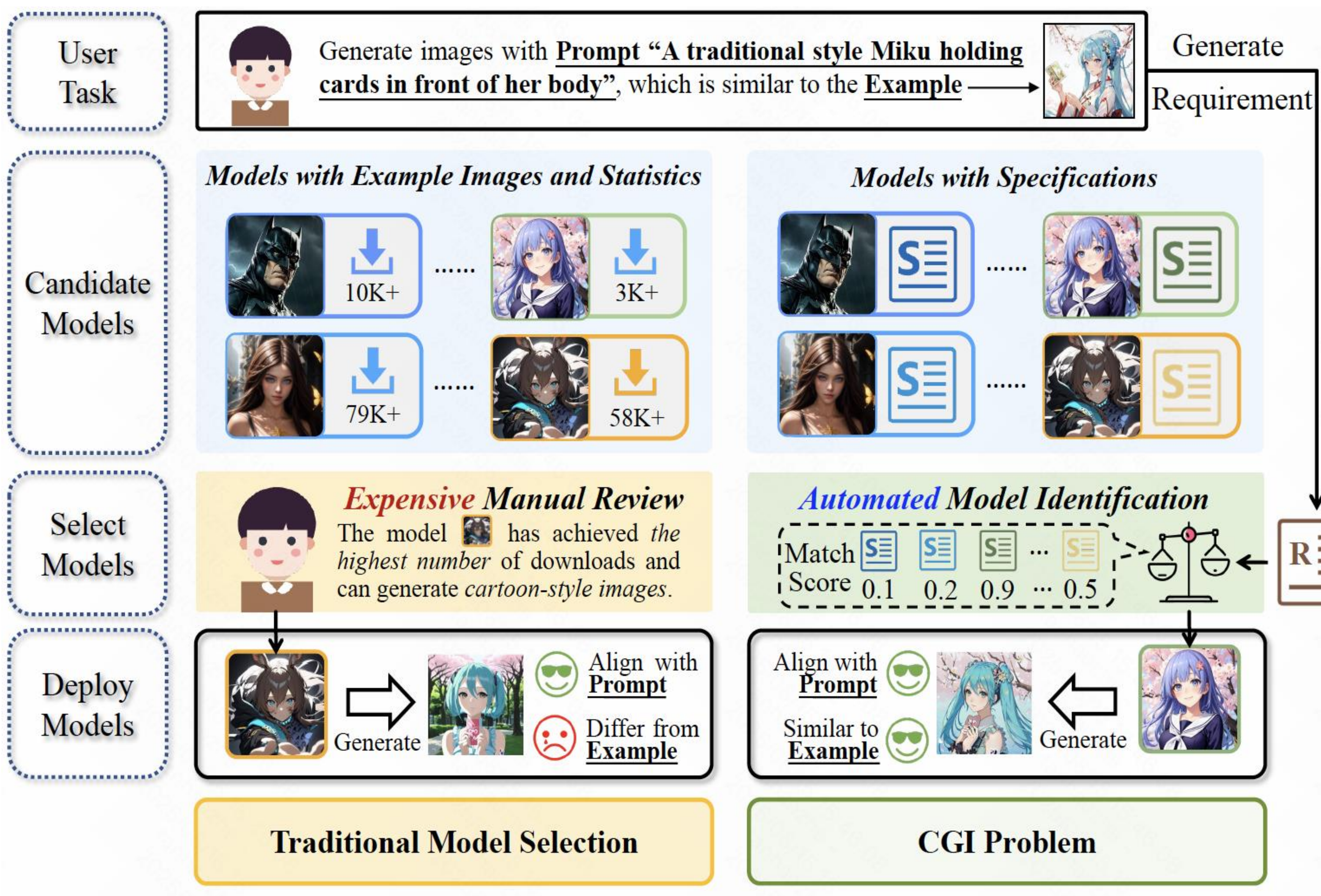
Zhi Zhou*, Hao-Zhe Tan*, Peng-Xiao Song, Lan-Zhe Guo✉

IJCAI International Joint Conferences on Artificial Intelligence Organization

LAMDA Learning And Mining from DatA

TL; DR We investigate a novel setting for identifying generative models and propose an effective solution PMI.

CGI Problem Setting



▲ Figure 1: Comparison between traditional model selection and CGI problem setting

Motivation

Existing generative model hubs provide methods to help users search for models. It is not enough to describe model functionalities by textual and statistical information. *Can we describe generative models functionality for efficient and accurate identification by matching their functionalities with user requirements?*

In this paper, we propose Conditional Generative Model Identification (CGI) problem setting, which aims to provide an effective way to identify the most suitable model using example images rather than requiring users to manually review a large number of models. Figure 1 presents an illustration of the CGI problem.

Problem Setup

Assume the model hub has M conditional generative models, each model is associated with a specification to describe its functionalities for future model identification, there are two stage in the CGI setting:

Submission Stage for developers to upload models then assign a model specification.

Identification Stage for users to select suitable models from the hub for their tasks by uploading example images.

Prompt-Based Model Identification Approach

We propose a novel and efficient approach for the CGI setting called Prompt-Based Model Identification (PMI). As shown in Figure 2, PMI consists of three key modules:

➢ *Automatic Specification Assignment*

$$S_i = \mathcal{A}_s(f_i, \mathcal{P}) = \{Z_i, Q_i\}$$

Generating a specification for each model using a developer-provided prompt set of the model and generated images to describe model functionality within the model matching space.

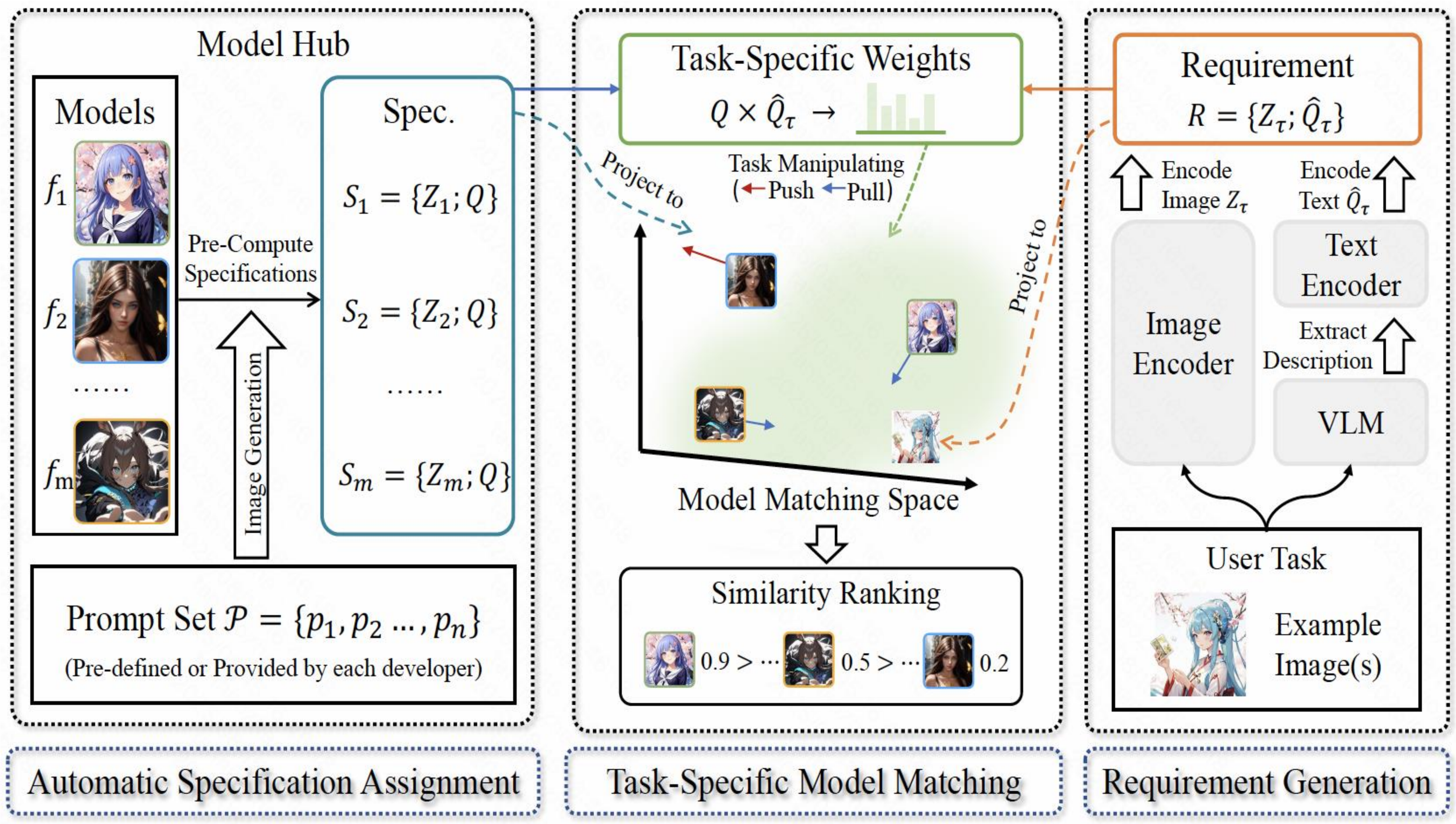
➢ *Requirement Generation*

$$R_\tau = \mathcal{A}_r(X_\tau) = \left\{Z_\tau; \widehat{Q}_\tau\right\}$$

Formulating the requirement for user task by encoding example images and their textual descriptions into the same space.

➢ *Task-Specific Matching*

$$\mathcal{A}_e(S_m, R_\tau) = \frac{1}{N_\tau}\sum_{i=1}^{N_\tau}\left\|\frac{1}{N_m}\sum_{j=1}^{N_m}\frac{q_j^m \hat{q}_i^\tau}{\|q_j^m\|\|\hat{q}_i^\tau\|}k(z_j^m,\cdot) - k(z_i^\tau,\cdot)\right\|_{\mathcal{H}_k}^2$$



▲ Figure 2: The overall illustration of PMI approach

Adjusting the specification in the matching space according to the requirement and identifies the most suitable model with the highest similarity score. The weight in the function measures the similarity between the specification prompts for model and the textual descriptions of example image, transforming original specifications to task-specific specifications.

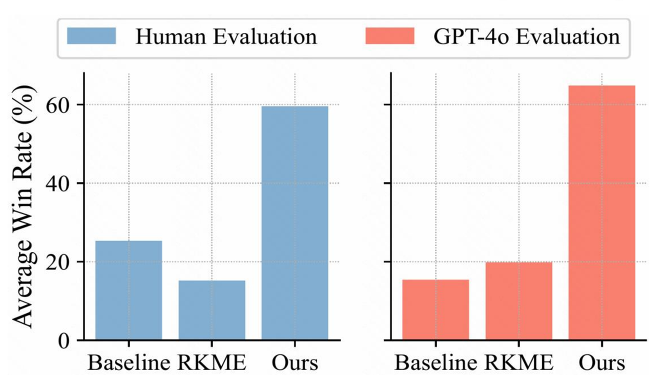
Experiments

Methods	Acc.(↑)	Top-2 Acc.(↑)	Top-3 Acc.(↑)	Top-4 Acc.(↑)	Top-5 Acc.(↑)	Avg. Rank(↓)	FID Score(↓)
Baseline	1.5%	3.0%	4.6%	6.1%	7.6%	33.000	23.44
RKME	3.1%	4.6%	6.2%	7.7%	9.3%	32.014	25.47
PMI	**69.2%**	**78.1%**	**82.8%**	**85.8%**	**88.0%**	**2.874**	**18.42**

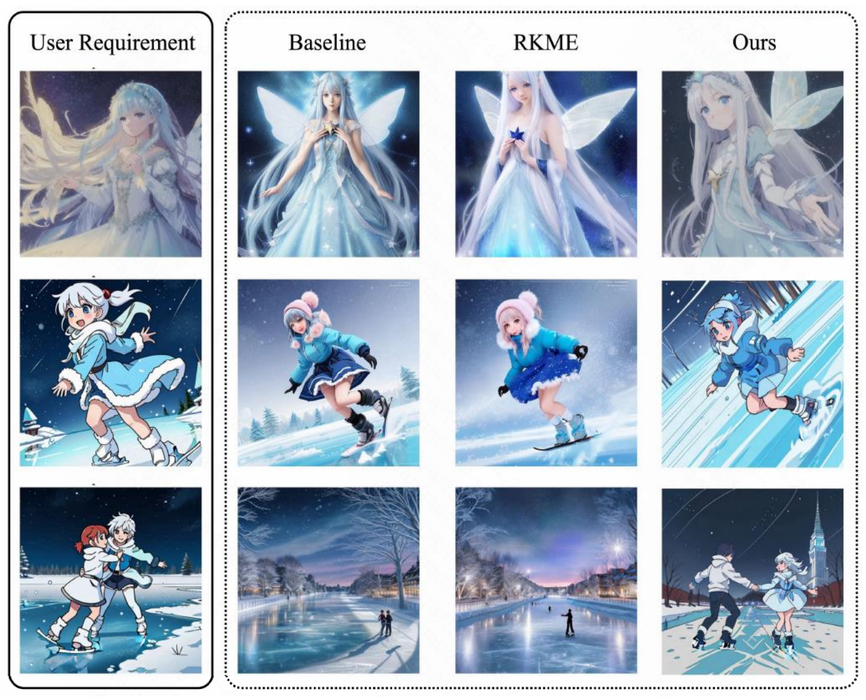
▲ Table 1:Performance of PMI approach

	Accuracy(↑)			FID Score(↓)		
	Baseline	RKME	PMI	Baseline	RKME	PMI
1 image	1.5%	3.1%	**69.2%**	23.44	25.47	**18.42**
2 images	1.5%	3.1%	**84.4%**	23.44	25.53	**18.17**
3 images	1.5%	3.1%	**89.7%**	23.44	25.53	**18.14**
4 images	1.5%	3.2%	**92.7%**	23.44	25.58	**18.21**
5 images	1.5%	3.2%	**94.0%**	23.44	25.61	**18.18**
6 images	1.5%	3.2%	**95.9%**	23.44	25.53	**18.12**

▲ Table 2:Performance with multiple images



▲ Figure 3: Human & GPT-4o Evaluation



▲ Figure 4: Visualization

Model Identification Performance

Table 1 shows that PMI significantly outperforms RKME in both accuracy and rank metrics. Table 2 presents results with multiple example images as user requirements, our PMI approach also surpasses existing methods.

Human & GPT-4o Evaluation

Figure 3 presents the average win rate of generated images by the model selected from each method voted by human users and GPT-4o. The results show that Our PMI achieves the highest win rate with a large margin indicating that our PMI are more consistent with the user requirements.

Visualization

We visualize the generated images from models identified by each method in Figure 4, with example images in the first column, which shows that our PMI successfully identifies models that match the example images' style.

✓ If you are interested in this paper, feel free to contact Zhi Zhou or Hao-Zhe Tan (zhouz@lamda.nju.edu.cn, tanhz@lamda.nju.edu.cn).

✓ This research was supported by the National Natural Science Foundation of China (Grant No. 624B2068, 62306133, 62250069), the Key Program of Jiangsu Science Foundation (BK20243012), Jiangsu Science Foundation (BG2024036), and the Fundamental Research Funds for the Central Universities (022114380023).