BEYOND HOLISTIC REALISM: DIMENSION-WISE GARMENT FIDELITY ASSESSMENT FOR VIRTUAL TRY-ON EVALUATION AND OPTIMIZATION

Anonymous authors
Paper under double-blind review

ABSTRACT

Virtual try-on (VTON) demands not only plausible visual synthesis but, crucially, faithful preservation of reference garment characteristics. However, conventional evaluation metrics—such as PSNR, SSIM, KID, and FID—primarily capture superficial or global realism, failing to reliably reflect multi-dimensional garment consistency. To address this limitation, we introduce **DAT**, a **D**imension-wise **A**ssessment framework for virtual **T**ry-on. DAT disentangles garment fidelity into seven interpretable and critical dimensions: silhouette, color, neckline and sleeve shape, major decoration and structure, material texture, fine-detail fidelity, and logo preservation, formulating each as a dedicated attribute-level prediction task. To effectively train this assessment model, we implement a two-stage learning paradigm comprising scalable weak supervision on 50K paired samples, followed by high-precision refinement on 10K curated annotations distilled via multi-model consensus. Furthermore, a dimensional-weighted cross-entropy loss is introduced to counter severe label imbalances inherent across evaluation categories. Beyond benchmarking, we establish a closed-loop optimization by seamlessly integrating DAT into the reinforcement learning pipeline of Qwen-Image-Edit for VTON, where dimension-wise rewards are adaptively aggregated to prioritize under-optimized visual attributes. Extensive experiments demonstrate that our 8B model achieves state-of-the-art performance across balanced accuracy, SROCC, and PLCC—surpassing prominent frontier proprietary models including Gemini-3.1, Qwen3.7-plus, and GPT-5.5—while serving as an exceptionally effective supervision signal for reward-guided garment generation.

1 INTRODUCTION

Virtual try-on (VTON) aims to transfer a target garment onto a person image while preserving both visual realism and garment fidelity, and has broad applications in e-commerce and digital fashion Cordier et al. (2001); Han et al. (2018); Wang et al. (2018); Hsieh et al. (2019). Although recent image editing models Team (2026); Wu et al. (2025a); Google (2025c) have substantially improved the realism of generated results, realistic synthesis alone is insufficient in practice. A useful try-on image must also faithfully preserve the source garment, including its silhouette, color, local structure, texture, and brand-related details. Even minor deviations in these attributes may significantly reduce the commercial reliability of the generated result. However, garment consistency assessment in VTON remains underexplored. Existing metrics such as PSNR Gonzalez & Woods (2002), SSIM Wang et al. (2004b), FID Heusel et al. (2017), KID Bińkowski et al. (2018), CLIP-I Radford et al. (2021), DiffSim Song et al. (2025) and LPIPS Zhang et al. (2018) are designed for global image quality evaluation and thus fail to capture fine-grained garment attributes. In practice, garment fidelity is inherently multi-dimensional: a try-on result may faithfully reproduce the overall color while distorting the neckline, or preserve the silhouette while losing logos and surface details. Collapsing such diverse attributes into a single scalar score obscures attribute-level discrepancies, limiting interpretability and reducing utility for model diagnosis, result ranking, and downstream optimization.

To address this limitation, we propose **DAT**: a **D**imension-wise **A**ssessment framework for virtual **T**ry-on. Instead of predicting a single overall score, DAT evaluates garment consistency along seven

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

dimensions: *silhouette*, *color*, *neckline and sleeve shape*, *major decoration and structure*, *material texture*, *fine-detail fidelity*, and *logo preservation*. Each dimension is modeled as a discrete attribute-level prediction task. For dimensions whose corresponding attribute may be absent, such as fine-detail fidelity and logo preservation, we introduce an additional not-applicable (NA) category. This formulation enables more informative and interpretable assessment than black-box scalar modeling. Such dimension design is motivated by practical business requirements for garment fidelity assessment, covering both global appearance and identity-critical local details.

Training such assessment model is challenging due to annotation noise and severe score imbalance across dimensions. To tackle annotation noise, we adopt a two-stage training framework. Based on Qwen3-vl-8b model Bai et al. (2025), we first train DAT from 50K weakly labeled samples annotated by Gemini-3.1 Google (2025a), and then refine the model on 10K higher-quality samples constructed through multi-model voting among Gemini-3.1 Google (2025a), Qwen3.7-plus Qwen Team (2025) and GPT-5.5 OpenAI (2025a). To mitigate the severe imbalance in dimension-wise score labels, we employ weighted cross-entropy loss to increase the contribution of underrepresented score categories during training. Moreover, the dimension-wise assessment from DAT can serve as a powerful reward signal to improve the VTON models. We utilize DAT as the reward model and adopt Flow-GRPO Liu et al. (2026) method to perform reinforcement learning (RL) optimization of Qwen-Image-Edit Wu et al. (2025a). Different from simple average of the rewards from multiple dimensions , we introduce an adaptive aggregation strategy that dynamically increases the weights of under-optimized dimensions during training, encouraging more balanced improvement across garment attributes.

Extensive experiments show that DAT outperforms strong proprietary judges, including Gemini-3.1 Google (2025a), Qwen3.7-plus Qwen Team (2025), and GPT-5.5 OpenAI (2025a), across VTON results from multiple sources Wu et al. (2025a); Team (2026); Google (2025c;b); OpenAI (2025b) on balanced accuracy, SROCC, and PLCC. Notably, DAT has only 8B parameters, making it substantially more compute- and resource-efficient than these proprietary large models. Moreover, reward-guided RL optimization based on DAT further improves reward-aligned garment preservation over the original Qwen-Image-Edit Wu et al. (2025a) model and yields qualitatively more faithful try-on results.

Our contributions are three-fold:

- We propose DAT, a multi-dimensional garment assessment framework for VTON that decomposes garment fidelity into seven interpretable dimensions, enabling more informative and interpretable evaluation than conventional scalar metrics.
- We design a two-stage, imbalance-aware training framework for learning the assessment model under noisy supervision, leveraging large-scale weak labels, multi-model-voted refinement data, and weighted cross-entropy to handle annotation noise and severe score imbalance.
- We show that the learned 8B-parameter model is effective for both assessment and optimization: it outperforms strong proprietary judges and can be integrated into reinforcement learning framework as a multi-dimensional reward to improve garment-faithful VTON generation.

2 Related Work

Virtual Try-On and Its Evaluation Metrics Virtual try-on (VTON) aims to synthesize a person wearing a target garment while preserving both human realism and garment identity Choi et al. (2021); Cui et al. (2021); Chen et al. (2021); Ge et al. (2021). While recent diffusion- and editing-based models Team (2026); Wu et al. (2025a); Yang et al. (2024) have markedly advanced visual realism over early warping-based pipelines Li et al. (2023b); Zhenyu et al. (2023), faithful garment preservation at the attribute level remains under-explored. Conventional metrics like SSIM, LPIPS, and CLIP-I Wang et al. (2004a); Zhang et al. (2018); Ghildyal & Liu (2022); Hessel et al. (2021) emphasize global perceptual realism or pose alignment Ning et al. (2024); Gou et al. (2023), routinely overlooking localized discrepancies. Although recent multi-dimensional benchmarks Wei et al. (2026); Xiaobin et al. (2025) evaluate aesthetics and background fidelity, they still treat garments coarsely. In contrast, our work centers squarely on garment fidelity itself, systematically

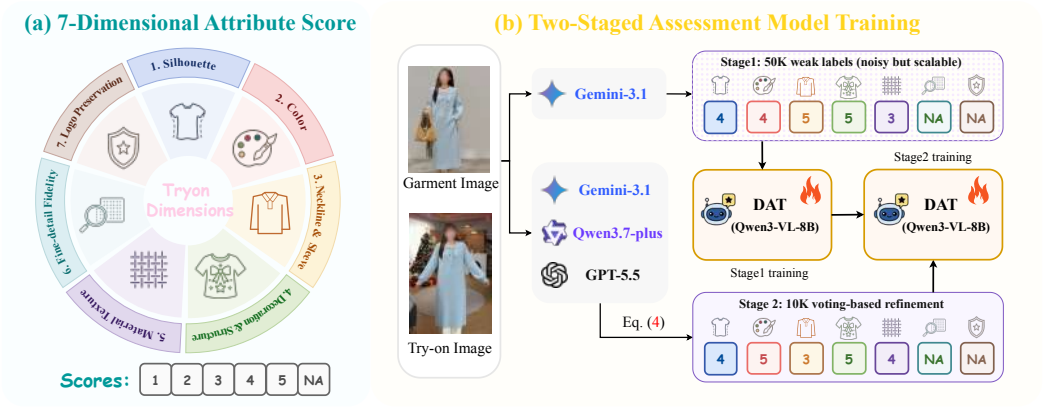

Figure 1: The illustration of the 7-dimensional attribute score (a) and the training pipeline (b) in DAT.

evaluating consistency across multiple decomposed structural components to provide a markedly finer-grained and more rigorous assessment.

Reward Modeling and Preference Learning Reward modeling and preference learning have been widely adopted to align generative models across language, vision, and multimodal domains Christiano et al. (2017); Ouyang et al. (2022); Park et al. (2022); Kim et al. (2023); Wu et al. (2025b); Zhang et al. (2025). However, prevailing approaches predominantly produce a single scalar score Luo et al. (2025); Lai et al. (2026), providing coarse feedback that cannot diagnose which specific attributes are preserved or distorted. While recent vision tasks explore fine-grained attribute analysis Li et al. (2023a); Yang et al. (2026); Cao et al. (2025); Sheng et al. (2026); Zhang et al. (2023), they are neither tailored for virtual try-on nor capable of offering actionable multi-dimensional supervision for downstream policy optimization. In contrast, our method explicitly decomposes garment fidelity into multiple dimensions, predicting fine-grained component-wise rewards to guide virtual try-on models.

3 METHOD

3.1 PROBLEM FORMULATION

Given a garment image I_g, a person image I_p, and a virtual try-on result I_t, our goal is to learn an assessment model that evaluates how faithfully the generated image preserves the garment characteristics specified by I_g. Unlike conventional models that predict a single holistic scalar, we formulate VTON assessment modeling as a fine-grained multi-dimensional assessment problem.

Built upon Qwen3-VL-8B Bai et al. (2025), our model takes the garment image I_g and the try-on result I_t as inputs and predicts attribute-level consistency scores over seven predefined dimensions: silhouette, color, neckline and sleeve shape, major decoration and structure, material texture, fine-detail fidelity, and logo preservation. The details of these dimensions are presented in the supplementary material. Formally, each sample is associated with a label vector

$$\mathbf{y} = [y_1, \ldots, y_7], \tag{1}$$

where

$$y_d \in \{1, 2, 3, 4, 5, \texttt{NA}\}, \quad d = 1, \ldots, 7. \tag{2}$$

We illustrate the proposed 7-dimensional attribute score in Fig. 1 (a). Larger scores indicate better garment consistency, while NA denotes that the corresponding attribute is not applicable or cannot be reliably judged for the current sample.

The assessment modeling task is then to learn a function

$$f_\theta(I_g, I_t) = \{\mathbf{p}_d\}_{d=1}^{7}, \tag{3}$$

where each $\mathbf{p}_d \in \mathbb{R}$ denotes the predicted probability distribution over the label vector for attribute dimension d.

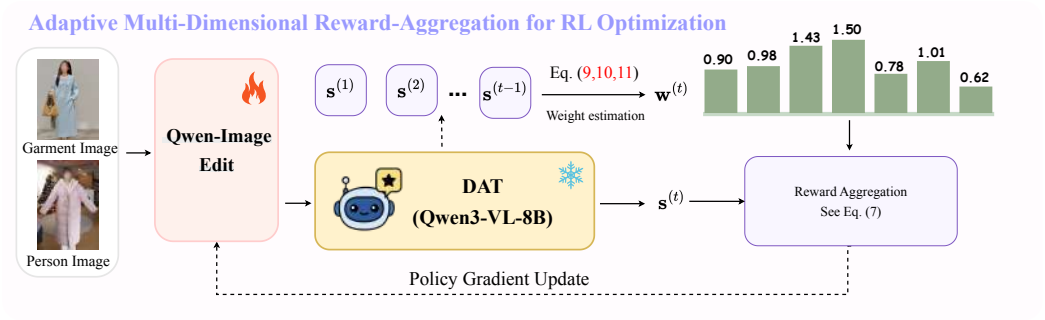

Figure 2: The framework of the adaptive multi-dimensional reward-aggregation RL optimization. We utilize DAT as the reward model to provide the reward signal.

3.2 Assessment Model Training

A key challenge in training multi-dimensional assessment models for virtual try-on lies in the tension between annotation scale and annotation fidelity. Since each sample requires judgments on seven garment-consistency attributes, obtaining large-scale high-quality supervision is prohibitively expensive. Meanwhile, relying solely on a single automatic annotator is cost-effective but introduces non-negligible noise and model-specific bias, particularly on subtle dimensions such as texture, fine details, and logo preservation. In addition, the multi-dimensional score distribution is highly imbalanced, making naive empirical risk minimization prone to overfitting dominant score levels. To address these issues, we adopt a staged supervision strategy together with imbalance-aware optimization. We show the training pipeline of the assessment model in Fig. 1 (b).

Stage 1: Large-scale weak supervision. In the first stage, we use Gemini-3.1 Google (2025a) to annotate large-scale triplets (I_g, I_p, I_t) with seven dimension-wise scores. The purpose of this stage is to construct a coarse but usable supervision source at manageable cost, so the model can learn broad attribute-level correspondences between the reference garment and the generated try-on image across diverse garment categories and generation qualities. Although single-model annotations are inevitably noisy, they provide sufficient scale and coverage for learning a strong initialization.

Stage 2: Consensus-based low-noise refinement. To obtain more reliable supervision, we further build a higher-quality refinement set from a smaller set of samples. For each sample and each attribute dimension d, we utilize Gemini-3.1, Qwen3.7-plus, and GPT-5.5 independently to produce three ordinal scores, denoted by $s_d^{(1)}, s_d^{(2)}, s_d^{(3)}$. We aggregate them into the final label by

$$\tilde{y}_d = \mathrm{median}\left(s_d^{(1)}, s_d^{(2)}, s_d^{(3)}\right). \tag{4}$$

For three annotators, this is equivalent to the rule *"2:1 majority wins, otherwise take the median"*. Such median-based aggregation is particularly suitable for ordinal labels, as it suppresses outlier predictions while preserving the order structure of the score space. We then continue training the stage-1 model on this voting-based refinement set, thereby improving calibration and reducing the impact of annotator-specific noise on difficult attributes.

Imbalance-aware optimization. A further difficulty comes from the highly skewed score distribution across both attribute dimensions and score levels. In principle, one may attempt to rebalance the dataset by resampling underrepresented scores. However, this is unsuitable in our setting: balancing the score distribution of one attribute dimension would generally distort the distributions of other dimensions, since each sample is simultaneously annotated along all seven attributes. We therefore handle imbalance at the loss level rather than the data level, by increasing the training contribution of rare score categories.

Let $q_{d,c}$ denote the empirical probability of score level $c \in \{1, \ldots, 5, \mathrm{NA}\}$ under attribute dimension d in the training set. We define the corresponding class weight as

$$\alpha_{d,c} = \mathrm{clip}\left(\frac{1}{q_{d,c}}, \alpha_{\min}, \alpha_{\max}\right), \tag{5}$$

where $\mathrm{clip}(x, \alpha_{\min}, \alpha_{\max})$ truncates x into the interval $[\alpha_{\min}, \alpha_{\max}]$. We empirically set α_{min} and α_{max} to 1 and 3, respectively. This inverse-frequency weighting amplifies the gradient contribution of underrepresented scores, while clipping prevents extremely rare categories from introducing unstable or disproportionately large updates.

Based on these weights, we optimize the assessment model using a validity-aware weighted cross-entropy loss:

$$\mathcal{L}_{\text{RM}} = \sum_{d=1}^{7} \alpha_{d,y_d}\, \mathcal{L}_{ce}(\mathbf{p}_d, y_d), \tag{6}$$

where $\mathbf{p}_d$ is the predicted score distribution for attribute dimension d, y_d is the corresponding ground-truth label. In this way, the model is trained only on valid supervision while assigning larger learning emphasis to rare but informative score categories. Overall, our model training framework decouples supervision scale from supervision fidelity: it first learns broad attribute-level discrimination from inexpensive weak labels, and then improves reliability through consensus-based refinement under an imbalance-aware objective.

3.3 RL Optimization with Adaptive Multi-Attribute Rewards

DAT provides an effective reward signal for RL optimization in VTON. We use DAT as the reward model and optimize Qwen-Image-Edit Wu et al. (2025a) with Flow-GRPO Liu et al. (2026) as the policy optimization framework. The key problem is how to aggregate the seven attribute-level scores into a scalar reward for RL optimization. A fixed average or sum would treat all dimensions equally and may cause the policy to under-optimize persistent weak attributes.

To address this problem, we propose an adaptive multi-dimensional reward aggregation mechanism. We illustrate the training pipeline in Fig. 2. Let $\mathbf{s}^{(n)} = [s_1^{(n)}, \ldots, s_7^{(n)}]$ denote the reward model scores for the n-th sampled result, and let $\mathcal{V}^{(n)}$ be the set of valid dimensions. We define the scalar reward as

$$R^{(n)} = \frac{\sum_{d \in \mathcal{V}^{(n)}} w_d^{(t)}\, s_d^{(n)}}{\sum_{d \in \mathcal{V}^{(n)}} w_d^{(t)}}, \tag{7}$$

where $w_d^{(t)}$ is the aggregation weight for dimension d at iteration t.

We maintain an EMA of batch-wise scores for each dimension:

$$\bar{s}_d^{(t)} = \begin{cases} \mu_d^{(t)}, & t = 1, \\ \lambda \bar{s}_d^{(t-1)} + (1-\lambda)\mu_d^{(t)}, & t > 1, \end{cases} \tag{8}$$

where $\mu_d^{(t)}$ is the mean valid score of dimension d in the current batch. Based on these running scores, we compute the short-board gap

$$g_d^{(t)} = \max_j \bar{s}_j^{(t)} - \bar{s}_d^{(t)}, \tag{9}$$

and define the adaptive weight as

$$a_d^{(t)} = \min\left(\tau,\, 1 + \gamma g_d^{(t)}\right). \tag{10}$$

To optionally encode prior importance, we further introduce a prior weight π_d and use

$$w_d^{(t)} = \pi_d\, a_d^{(t)}. \tag{11}$$

This mechanism automatically upweights underperforming dimensions and encourages the policy to focus on current weaknesses.

Given scalar rewards $\{R^{(n)}\}_{n=1}^{G}$ for a sampled group of size G, we compute group-normalized advantages as

$$A^{(n)} = \frac{R^{(n)} - \mu_R}{\sigma_R + \epsilon}, \tag{12}$$

where μ_R and σ_R are the mean and standard deviation of rewards within the group. The subsequent policy optimization follows the standard Flow-GRPO objective.

Table 1: Quantitative comparison of different assessment models. B-Acc. denotes Balanced Accuracy. We report B-Acc., SROCC, and PLCC on both in-domain sources and unseen out-of-domain sources. The best results are highlighted in **bold**.

	In-domain			Out-of-domain			Overall		
Method	B-Acc. ↑	SROCC ↑	PLCC ↑	B-Acc. ↑	SROCC ↑	PLCC ↑	B-Acc. ↑	SROCC ↑	PLCC ↑
Qwen3-vl-8b	57.46%	0.2209	0.3506	58.86%	0.3461	0.3769	58.16%	0.2835	0.3637
Qwen3.7-plus	69.76%	0.5186	0.5758	69.10%	0.3916	0.5299	69.43%	0.4551	0.5528
GPT-5.5	74.97%	0.5258	0.5457	73.47%	0.4117	0.5440	74.22%	0.4688	0.5448
Gemini-3.1	**76.35%**	0.5744	0.5830	75.36%	0.4444	0.5595	75.85%	0.5094	0.5713
Ours	76.07%	**0.5884**	**0.6174**	**77.03%**	**0.5048**	**0.6126**	**76.55%**	**0.5466**	**0.6150**

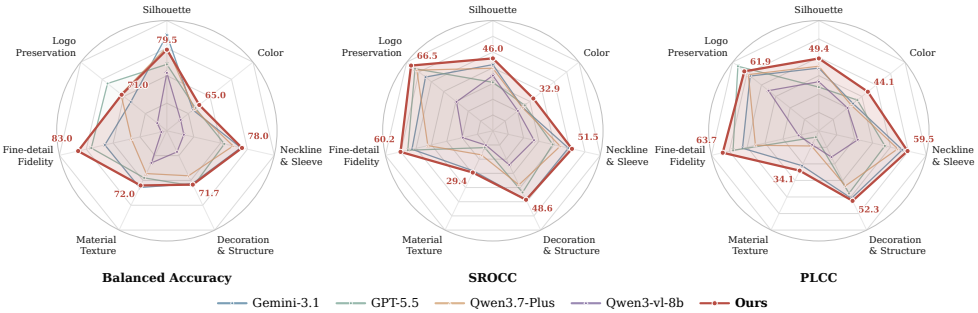

Figure 3: Dimension-wise comparison of different assessment models in balanced accuracy, SROCC, and PLCC. Our method achieves a more balanced and consistent performance across fine-grained garment-consistency dimensions.

4 Experiments

4.1 Experimental Setup

4.1.1 Implementation Details.

We initialize DAT from the pretrained Qwen3-VL-8B and fine-tune it with LoRA Hu et al. (2022) under our two-stage supervision strategy. We train the model for up to 16 epochs using AdamW Loshchilov & Hutter (2019) with a batch size of 8 and a learning rate of 1×10^{-4}. For reward-guided optimization, we choose the 4-step distilled Qwen-Image-Edit lightx2v (2026.07) as the base model for fast training and leverage DAT to provide the reward signal for policy optimization via Flow-GRPO Liu et al. (2026). We perform LoRA-based policy optimization with a learning rate of 1×10^{-4} for five epochs. All experiments are conducted on 16 NVIDIA H20 GPUs.

4.1.2 Datasets and Evaluation for Assessment Model.

For DAT training, we collect 50K virtual try-on samples from an e-commerce platform, where each sample consists of a garment image, a person image, and a try-on result. The try-on results are generated by three virtual try-on systems: Nano Banana Pro Google (2025c), Qwen-Image-Edit-2511 Wu et al. (2025a), and an internally developed try-on model.

Following our proposed annotation protocol, we leverage Gemini-3.1 Google (2025a) to obtain seven dimension-wise garment consistency scores for each sample, providing large-scale supervision signals for assessment model learning. We further curate 10K high-quality annotations through multi-model voting among Gemini-3.1 Google (2025a), Qwen3.7-plus Qwen Team (2025), and GPT-5.5 OpenAI (2025a).

To evaluate assessment models, we collect real-world garment and person images from an e-commerce platform and construct a test set with 1.6K samples. The try-on images are generated by six systems: Nano Banana Pro Google (2025c), Qwen-Image-Edit-2511 Wu et al. (2025a), the internally developed virtual try-on model, FireRed Team (2026), Nano Banana 2 Google (2025b) and GPT-Image2 OpenAI (2025b). The first three sources overlap with the training data, while the remaining three are completely held out during training and are used to evaluate cross-source gen-

Table 2: Quantitative comparison on Tryon-RLEval across the seven garment-consistency scores and FID.

Method	Silhouette ↑	Color ↑	Neckline / Sleeve ↑	Decoration / Structure ↑	Material Texture ↑	Fine-Detail Fidelity ↑	Logo Preserv. ↑	FID ↓
Qwen-Image-Edit	4.04	3.68	3.63	3.08	3.61	3.05	2.23	92.27
+ RL Optimization	**4.24** (+0.20)	**4.01** (+0.33)	**3.85** (+0.22)	**3.44** (+0.36)	**3.88** (+0.27)	**3.52** (+0.47)	**2.80** (+0.57)	**90.91** (-1.36)

Table 3: Comparison on the StreetTryOn Cui et al. (2025) across four settings: Shop-to-Street, Model-to-Model, Model-to-Street, and Street-to-Street (Top). We adopt FID as the evaluation metric, where lower values indicate better performance.

Method	Shop-to-Street	Model-to-Model	Model-to-Street	Street-to-Street (Top)
Qwen-Image-Edit	46.77	**11.30**	42.65	38.84
+ RL Optimization	**46.35**	11.61	**41.92**	**38.03**

eralization. To obtain reliable ground-truth annotations, each test sample is independently evaluated by five expert annotators along the same seven garment-consistency dimensions, and the final labels are determined by majority voting.

For evaluation metrics, we use balanced accuracy Brodersen et al. (2010), where scores of 1–3 are treated as negative and scores of 4–5 as positive. To account for class imbalance, we average the accuracies on positive and negative samples:

$$\mathrm{BA} = \frac{1}{2}\left(\mathrm{Acc}_{\mathrm{pos}} + \mathrm{Acc}_{\mathrm{neg}}\right). \tag{13}$$

We additionally report SROCC and PLCC to measure correlation with human annotations. For our seven-dimensional predictions, both metrics are computed separately for each dimension and then averaged across dimensions.

4.1.3 Datasets and Evaluation for Reward-guided Optimization.

For reward-guided optimization, we collect 10,000 training samples from an e-commerce platform and introduce Tryon-RLEval, a benchmark with 300 randomly sampled, disjoint images for evaluation. For evaluation metrics, we report our proposed seven-dimensional scores to show the effectiveness of reward-guided optimization. Furthermore, we include the commonly used FID metric to measure distribution-level realism between person images and generated images, and conduct blind user studies to provide a reliable and primary evaluation of try-on quality. We further evaluate our method on the out-of-domain StreetTryon Cui et al. (2025) dataset and follow its evaluation protocol by employing the FID metric under four different try-on settings.

4.2 Main Results on Assessment Modeling

We compare our assessment model with the baseline model Qwen3-VL-8B and several leading proprietary models, including Gemini-3.1 Google (2025a), Qwen3.7-plus Qwen Team (2025), and GPT-5.5 OpenAI (2025a). Quantitative results on the evaluation test set are shown in Tab. 1. Overall, our method achieves the best performance among all competing approaches. Compared with the baseline Qwen3-VL-8B model, our model obtains nearly 20% improvement while maintaining the same 8B-scale architecture. Moreover, it consistently outperforms proprietary models, demonstrating the effectiveness of our approach for fine-grained garment-consistency evaluation. On in-domain sources, our method achieves the highest SROCC and PLCC while remaining competitive in balanced accuracy. More importantly, on out-of-domain sources, our method performs best across all metrics, indicating that the learned assessment model captures robust and transferable garment-consistency signals.

We provide Fig. 3 to illustrate dimension-wise performance across balanced accuracy, SROCC, and PLCC. Compared with proprietary methods, our method achieves the most balanced performance across all dimensions. While competing models excel on individual attributes, ours maintains consistent performance across fine-grained dimensions. This demonstrates that our multi-dimensional formulation captures garment consistency more comprehensively than global visual similarity.

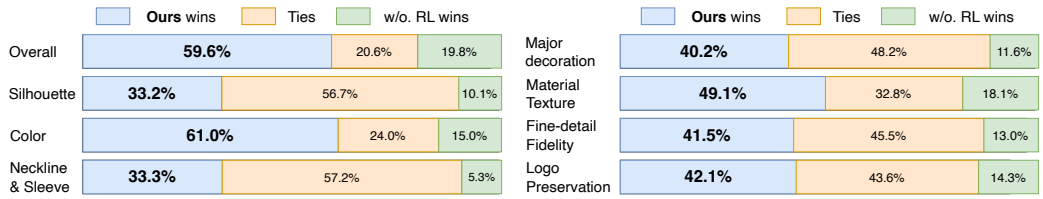

Figure 4: Blind human study comparing the reward-guided model (*Ours*) and the base model (*w/o. RL*) across overall and seven-dimensional garment-fidelity scores.



Figure 5: Qualitative comparison of virtual try-on results without and with reward-guided RL optimization. RL optimization improves garment fidelity, leading to results that better match the reference garment in both global structure and local details.

Table 4: Ablation study on assessment model design. Stage 1 (S1) uses coarse labels. Weighted CE (WCE) assigns higher weights to underrepresented dimensions and hard samples. Stage 2 (S2) refines the model with higher-quality labels from multi-model voting.

S1	WCE	S2	Bal. Acc. ↑	SROCC ↑	PLCC ↑
			58.16%	0.2835	0.3637
✓			74.28%	0.4872	0.5520
✓	✓		75.62%	0.5185	0.5632
✓	✓	✓	**76.07%**	**0.5884**	**0.6174**

4.3 Main Results on Reward-Guided Optimization

We evaluate whether DAT serves as an effective reward signal for RL optimization, comparing the base Qwen-Image-Edit with its reward-guided counterpart from three perspectives: (1) As shown in Tab. 2, our seven-dimensional garment-consistency scores on Tryon-RLEval yield an average gain of 0.35, demonstrating the effectiveness of reward-guided training. Meanwhile, the improved FID confirms enhanced garment fidelity without sacrificing overall image quality. (2) To assess perceptual quality, we conduct a blind human study on Tryon-RLEval with 10 expert annotators. For each sample, annotators compare outputs from the base and reward-guided models against the reference garment across eight aspects: overall similarity, silhouette, color, neckline/sleeve shape, decoration/structure, material texture, fine-detail fidelity, and logo preservation. As shown in Fig. 4, reward-guided optimization achieves consistent gains across all dimensions, validating that the learned reward effectively promotes garment-faithful generation. (3) Furthermore, on the public StreetTryOn Cui et al. (2025) dataset (Tab. 3), reward-guided optimization consistently improves FID over the base model across different settings.

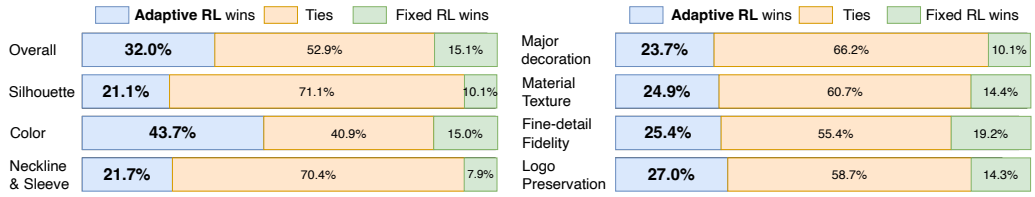

Figure 6: Blind human study comparing adaptive-weighted RL (*Adaptive RL*) and fixed-weighted RL (*Fixed RL*) across overall and seven garment-fidelity dimensions.

Table 5: Ablation study on different reward weighting strategies on Tryon-RLEval.

Strategy	Silhouette ↑	Color ↑	Neckline / Sleeve ↑	Decoration / Structure ↑	Material Texture ↑	Fine-Detail Fidelity ↑	Logo Preserv. ↑	FID ↓
Fixed	4.15	3.87	3.75	3.26	3.78	3.30	2.55	91.92
Adaptive	**4.24** (+0.09)	**4.01** (+0.14)	**3.85** (+0.10)	**3.44** (+0.18)	**3.88** (+0.10)	**3.52** (+0.22)	**2.80** (+0.25)	**90.91** (-1.01)

Qualitative analysis. Fig. 5 shows representative examples before and after reward-guided optimization. The optimized model better preserves garment characteristics across our proposed seven dimensions, which is consistent with both the human study and quantitative results.

4.4 Ablation Study

Ablation on assessment model design. For the assessment model, we adopt the two-stage training strategy and the weighting scheme. We conduct an incremental ablation study to demonstrate the reasonability of the training paradigm of DAT, with results presented in Tab. 4. Compared with the baseline Qwen3-VL-8B, our Stage 1 model trained with coarse labels already achieves substantial improvements. Adding weighted CE consistently improves performance, highlighting the importance of imbalance-aware optimization for learning fine-grained garment consistency. Introducing Stage 2 brings additional gains, as higher-quality labels generated by multi-model voting reduce annotation noise and provide more reliable supervision than coarse labels alone. The full model achieves the best performance, confirming that the proposed two-stage training and weighting scheme are complementary and jointly enhance garment consistency assessment.

Ablation on RL design. We conduct detailed experiments to investigate the effect of different reward weighting strategies in reward-guided optimization, including adaptive weighting and fixed weighting (average across all dimensions). First, we report dimension-wise reward scores in Tab. 5. Compared with fixed weighting, adaptive weighting achieves higher reward scores across all seven garment-consistency dimensions while maintaining better FID performance. We further present a blind user study comparing adaptive and fixed weighting in Fig. 6. The results show that adaptive weighting consistently improves try-on quality, validating its effectiveness. We attribute this to the fact that fixed weighting tends to bias optimization toward easier dimensions, whereas our adaptive strategy emphasizes lagging dimensions based on current performance, leading to more balanced garment-fidelity improvements.

5 Conclusion

In this paper, we propose a fine-grained assessment framework for garment consistency in virtual try-on. Our method decomposes garment fidelity into seven interpretable dimensions, enabling more informative attribute-level evaluation than conventional assessment metrics. Experiments on a manually annotated test set show that our model outperforms strong proprietary judges, including Gemini-3.1, Qwen3.7-plus, and GPT-5.5, in both binary accuracy and correlation with human annotations. Moreover, when used as a reward signal for reinforcement learning optimization of Qwen-Image-Edit, our model improves garment preservation and produces more faithful try-on results. These findings highlight the value of domain-specialized, interpretable, and fine-grained assessment for both evaluating and improving virtual try-on systems.

AI USE STATEMENT

In this work, we used generative AI tools to aid or polish writing, including translation and editing the manuscript to improve readability. We have not used generative AI tools for retrieval or discovery (e.g. finding related work), research ideation or execution and proving mathematical claims. Additionally, we have not used generative AI tools to draft sections of the paper. We have used generative AI tools for generating synthetic datasets, including the annotation of the training and evaluation datsets of DAT and the RL optimization of the VTON model.

REFERENCES

Shuai Bai, Yuxuan Cai, Ruizhe Chen, Keqin Chen, Xionghui Chen, Zesen Cheng, Lianghao Deng, Wei Ding, Chang Gao, Chunjiang Ge, et al. Qwen3-vl technical report. *arXiv preprint arXiv:2511.21631*, 2025.

Mikołaj Bińkowski, Danica J. Sutherland, Michael Arbel, and Arthur Gretton. Demystifying mmd gans. In *ICLR*, 2018.

Kay H. Brodersen, Cheng Soon Ong, Klaas E. Stephan, and Joachim M. Buhmann. The balanced accuracy and its posterior distribution. *ICPR*, pp. 3121–3124, 2010.

Shuo Cao, Nan Ma, Jiayang Li, Xiaohui Li, Lihao Shao, Kaiwen Zhu, Yu Zhou, Yuandong Pu, Jiarui Wu, Jiaquan Wang, Bo Qu, Wenhai Wang, Yu Qiao, Dajuin Yao, and Yihao Liu. Artimuse: Fine-grained image aesthetics assessment with joint scoring and expert-level understanding, 2025. URL https://arxiv.org/abs/2507.14533.

Chieh-Yun Chen, Ling Lo, Pin-Jui Huang, Hong-Han Shuai, and Wen-Huang Cheng. Fashionmirror: Co-attention feature-remapping virtual try-on with sequential template poses. In *ICCV*, pp. 13809–13818, 2021.

Seunghwan Choi, Sunghyun Park, Minsoo Lee, and Jaegul Choo. Viton-hd: High-resolution virtual try-on via misalignment-aware normalization. In *CVPR*, 2021.

Paul F Christiano, Jan Leike, Tom Brown, Miljan Martic, Shane Legg, and Dario Amodei. Deep reinforcement learning from human preferences. *NeurIPS*, 30, 2017.

Frédéric Cordier, WonSook Lee, HyeWon Seo, and Nadia Magnenat-Thalmann. From 2d photos of yourself to virtual try-on dress on the web. In *People and computers XV—interaction without frontiers: Joint proceedings of HCI 2001 and IHM 2001*, pp. 31–46. Springer, 2001.

Aiyu Cui, Daniel McKee, and Svetlana Lazebnik. Dressing in order: Recurrent person image generation for pose transfer, virtual try-on and outfit editing. In *ICCV*, pp. 14638–14647, October 2021.

Aiyu Cui, Jay Mahajan, Viraj Shah, Preeti Gomathinayagam, Chang Liu, and Svetlana Lazebnik. Street tryon: Learning in-the-wild virtual try-on from unpaired person images. In *WACV*, pp. 1414–1423. IEEE, 2025.

Chongjian Ge, Yibing Song, Yuying Ge, Han Yang, Wei Liu, and Ping Luo. Disentangled cycle consistency for highly-realistic virtual try-on. In *CVPR*, pp. 16928–16937, 2021.

Abhijay Ghildyal and Feng Liu. Shift-tolerant perceptual similarity metric. In *ECCV*, 2022.

Rafael C. Gonzalez and Richard E. Woods. *Digital Image Processing*. Prentice Hall, 2002.

Google. Gemini 3.1 pro. https://blog.google/innovation-and-ai/models-and-research/gemini-models/gemini-3-1-pro/, 2025a. Official Google blog post. Accessed: 2026-07-16.

Google. Nano banana 2. https://gemini.google/overview/image-generation/, 2025b. Official Google Gemini image generation page. Accessed: 2026-07-16.

Google. Nano banana pro. https://blog.google/innovation-and-ai/products/nano-banana-pro/, 2025c. Official Google blog post. Accessed: 2026-07-16.

Junhong Gou, Siyu Sun, Jianfu Zhang, Jianlou Si, Chen Qian, and Liqing Zhang. Taming the power of diffusion models for high-quality virtual try-on with appearance flow. In *ACMMM*, 2023.

Xintong Han, Zuxuan Wu, Zhe Wu, Ruichi Yu, and Larry S Davis. Viton: An image-based virtual try-on network. In *CVPR*, pp. 7543–7552, 2018.

Jack Hessel, Ari Holtzman, Maxwell Forbes, Ronan Le Bras, and Yejin Choi. Clipscore: A reference-free evaluation metric for image captioning. In *EMNLP*, pp. 7514–7528, 2021.

Martin Heusel, Hubert Ramsauer, Thomas Unterthiner, Bernhard Nessler, and Sepp Hochreiter. Gans trained by a two time-scale update rule converge to a local nash equilibrium. In *NeurIPS*, 2017.

Chia-Wei Hsieh, Chieh-Yun Chen, Chien-Lung Chou, Hong-Han Shuai, Jiaying Liu, and Wen-Huang Cheng. Fashionon: Semantic-guided image-based virtual try-on with detailed human and clothing information. In *ACMMM*, pp. 275–283, 2019.

Edward J Hu, Yelong Shen, Phillip Wallis, Zeyuan Allen-Zhu, Yuanzhi Li, Shean Wang, Liang Wang, Weizhu Chen, et al. Lora: Low-rank adaptation of large language models. *ICLR*, 1(2):3, 2022.

Sungdong Kim, Sanghwan Bae, Jamin Shin, Soyoung Kang, Donghyun Kwak, Kang Yoo, and Minjoon Seo. Aligning large language models through synthetic feedback. In *EMNLP*, pp. 13677–13700, 2023.

Jianyu Lai, Sixiang Chen, Jialin Gao, Hengyu Shi, Zhongying Liu, Fuxiang Zhai, Junfeng Luo, Xiaoming Wei, Lujia Wang, and Lei Zhu. Posterreward: Unlocking accurate evaluation for high-quality graphic design generation. In *CVPR*, pp. 7762–7772, 2026.

Chunyi Li, Zicheng Zhang, Haoning Wu, Wei Sun, Xiongkuo Min, Xiaohong Liu, Guangtao Zhai, and Weisi Lin. Agiqa-3k: An open database for ai-generated image quality assessment. *TCSVT*, pp. 1–1, 2023a. doi: 10.1109/TCSVT.2023.3319020.

Zhi Li, Pengfei Wei, Xiang Yin, Zejun Ma, and Alex C. Kot. Virtual try-on with pose-garment keypoints guided inpainting. In *ICCV*, pp. 22788–22797, October 2023b.

lightx2v. Qwen-image-edit-2511-lightning. https://huggingface.co/lightx2v/Qwen-Image-Edit-2511-Lightning, 2026.07. accessed: 2026-07-28.

Jie Liu, Gongye Liu, Jiajun Liang, Yangguang Li, Jiaheng Liu, Xintao Wang, Pengfei Wan, Di Zhang, and Wanli Ouyang. Flow-grpo: Training flow matching models via online rl. *NeurIPS*, 38:40783–40818, 2026.

Ilya Loshchilov and Frank Hutter. Decoupled weight decay regularization. In *ICLR*, 2019.

Xin Luo, Jiahao Wang, Chenyuan Wu, Shitao Xiao, Xiyan Jiang, Defu Lian, Jiajun Zhang, Dong Liu, and Zheng Liu. Editscore: Unlocking online rl for image editing via high-fidelity reward modeling. *ICLR*, 2025.

Shuliang Ning, Duomin Wang, Yipeng Qin, Zirong Jin, Baoyuan Wang, and Xiaoguang Han. Picture: Photorealistic virtual try-on from unconstrained designs. In *CVPR*, pp. 6976–6985, 2024.

OpenAI. Introducing gpt-5.5. https://openai.com/zh-Hans-CN/index/introducing-gpt-5-5/, 2025a. Official OpenAI blog post. Accessed: 2026-07-16.

OpenAI. Gpt-image-2. https://developers.openai.com/api/docs/models/gpt-image-2, 2025b. Official OpenAI API documentation. Accessed: 2026-07-16.

Long Ouyang, Jeffrey Wu, Xu Jiang, Diogo Almeida, Carroll Wainwright, Pamela Mishkin, Chong Zhang, Sandhini Agarwal, Katarina Slama, Alex Ray, et al. Training language models to follow instructions with human feedback. *NeurIPS*, 35:27730–27744, 2022.

Jongjin Park, Younggyo Seo, Jinwoo Shin, Honglak Lee, Pieter Abbeel, and Kimin Lee. Surf: Semi-supervised reward learning with data augmentation for feedback-efficient preference-based reinforcement learning. In *ICLR*, 2022.

José Luis Pech-Pacheco, Gabriel Cristóbal, Jesús Chamorro-Martinez, and Joaquín Fernández-Valdivia. Diatom autofocusing in brightfield microscopy: a comparative study. In *ICPR*, volume 3, pp. 314–317. IEEE, 2000.

Qwen Team. Qwen3.7-plus. https://qwen.ai/blog?id=qwen3.7-plus, 2025. Official Qwen blog post. Accessed: 2026-07-16.

Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, et al. Learning transferable visual models from natural language supervision. In *ICML*, 2021.

Xiangfei Sheng, Xiaofeng Pan, Zhichao Yang, Pengfei Chen, and Leida Li. Fine-grained image quality assessment for perceptual image restoration. In *AAAI*, volume 40, pp. 8914–8922, 2026.

Yiren Song, Xiaokang Liu, and Mike Zheng Shou. Diffsim: Taming diffusion models for evaluating visual similarity. In *ICCV*, pp. 16904–16915, 2025.

Super Intelligence Team. Firered-image-edit-1.0 technical report. 2026. URL https://arxiv.org/abs/2602.13344.

Bochao Wang, Huabin Zheng, Xiaodan Liang, Yimin Chen, Liang Lin, and Meng Yang. Toward characteristic-preserving image-based virtual try-on network. In *ECCV*, pp. 589–604, 2018.

Zhou Wang, A.C. Bovik, H.R. Sheikh, and E.P. Simoncelli. Image quality assessment: from error visibility to structural similarity. *TIP*, 13(4):600–612, 2004a. doi: 10.1109/TIP.2003.819861.

Zhou Wang, Alan C Bovik, Hamid R Sheikh, and Eero P Simoncelli. Image quality assessment: from error visibility to structural similarity. *TIP*, 13(4):600–612, 2004b.

Xinyi Wei, Sijing Wu, Zitong Xu, Yunhao Li, Huiyu Duan, Xiongkuo Min, and Guangtao Zhai. Vtonqa: A multi-dimensional quality assessment dataset for virtual try-on. *arXiv preprint arXiv:2601.02945*, 2026.

Chenfei Wu, Jiahao Li, Jingren Zhou, Junyang Lin, Kaiyuan Gao, Kun Yan, Sheng-ming Yin, Shuai Bai, Xiao Xu, Yilei Chen, et al. Qwen-image technical report. *arXiv preprint arXiv:2508.02324*, 2025a.

Jie Wu, Yu Gao, Zilyu Ye, Ming Li, Liang Li, Hanzhong Guo, Jie Liu, Zeyue Xue, Xiaoxia Hou, Wei Liu, et al. Rewarddance: Reward scaling in visual generation. *arXiv preprint arXiv:2509.08826*, 2025b.

Hu Xiaobin, Liang Yujie, Luo Donghao, Peng Xu, Zhang Jiangning, Zhu Junwei, Wang Chengjie, and Fu Yanwei. Vtbench: Comprehensive benchmark suite towards real-world virtual try-on models. *arXiv preprint arXiv:2505.19571*, 2025.

Xu Yang, Changxing Ding, Zhibin Hong, Junhao Huang, Jin Tao, and Xiangmin Xu. Texture-preserving diffusion models for high-fidelity virtual try-on. In *CVPR*, pp. 7017–7026, June 2024.

Zhichao Yang, Jianjie Wang, Zhixianhe Zhang, Pangu Xie, Xiangfei Sheng, Pengfei Chen, and Leida Li. Fine-grained image aesthetic assessment: Learning discriminative scores from relative ranks. In *CVPR*, 2026.

Richard Zhang, Phillip Isola, Alexei A Efros, Eli Shechtman, and Oliver Wang. The unreasonable effectiveness of deep features as a perceptual metric. In *CVPR*, 2018.

Weixia Zhang, Guangtao Zhai, Ying Wei, Xiaokang Yang, and Kede Ma. Blind image quality assessment via vision-language correspondence: A multitask learning perspective. In *CVPR*, pp. 14071–14081, 2023.

Yi-Fan Zhang, Xingyu Lu, Xiao Hu, Chaoyou Fu, Bin Wen, Tianke Zhang, Changyi Liu, Kaiyu Jiang, Kaibing Chen, Kaiyu Tang, et al. R1-reward: Training multimodal reward model through stable reinforcement learning. *arXiv preprint arXiv:2505.02835*, 2025.

Xie Zhenyu, Huang Zaiyu, Dong Xin, Zhao Fuwei, Dong Haoye, Zhang Xijin, Zhu Feida, and Liang Xiaodan. Gp-vton: Towards general purpose virtual try-on via collaborative local-flow global-parsing learning. In *CVPR*, June 2023.

A Appendix

A.1 Details of the Dimension-wise Garment Assessment Protocol of DAT

To obtain a more diagnostic and interpretable evaluation of virtual try-on fidelity, DAT (**D**imension-wise **A**ssessment framework for virtual **T**ry-on) compares each source garment image with its corresponding generated virtual try-on image along seven dimensions: *silhouette*, *color*, *neckline and sleeve shape*, *major decoration and structure*, *material texture*, *fine-detail fidelity*, and *logo preservation*. DAT is designed to assess **garment-level consistency**, rather than overall visual plausibility alone, since a generated result may appear realistic while still failing to preserve product-defining attributes of the original garment. In this section, we present the details of these dimensions.

The seven dimensions capture complementary aspects of garment fidelity. *Silhouette* and *color* measure preservation of the garment's core identity. *Neckline and sleeve shape* and *major decoration and structure* evaluate key design elements that define garment style. *Material texture* assesses whether the perceived fabric characteristics are faithfully rendered. *Fine-detail fidelity* measures the preservation of small but semantically meaningful non-logo elements, while *logo preservation* is treated separately because logos are particularly sensitive to distortion and are of practical importance in commercial settings.

Each dimension is rated on a **five-point scale (1–5)**, where higher scores indicate better consistency between the source garment and the generated result. In general, **5** denotes near-perfect preservation, **4** indicates minor deviation, **3** corresponds to partial consistency with noticeable discrepancies, **2** reflects severe inconsistency, and **1** denotes complete failure. Dimensions that are not applicable to a given sample are marked as **N/A** and excluded from dimension-specific analysis. In particular, *fine-detail fidelity* is evaluated only when the garment contains relevant non-logo fine-grained elements (e.g., buttons, zipper pulls, embroidery, or small labels), and *logo preservation* is evaluated only when the source garment contains a clearly visible logo on the garment itself.

Silhouette. Evaluates whether the overall cut, length, proportion, and garment category are preserved. A score of **5** indicates exact preservation of the garment silhouette; **4** indicates slight proportional deviation that does not affect recognition; **3** indicates noticeable silhouette change while remaining within the same broad garment category; **2** indicates a clear change in garment type or core silhouette; and **1** indicates complete failure or severe structural corruption.

Color. Evaluates consistency of the garment's main color appearance. A score of **5** indicates an almost exact match in hue, saturation, and brightness; **4** indicates minor deviation likely caused by illumination or rendering effects; **3** indicates obvious color difference but within a similar color family; **2** indicates incorrect major color or hallucinated patterns; and **1** indicates chaotic or corrupted color rendering.

Neckline and sleeve shape. Evaluates whether the neckline type and sleeve type are preserved. A score of **5** indicates correct preservation of both; **4** indicates minor deviations in shape or proportion; **3** indicates partial resemblance with noticeable confusion in local design details; **2** indicates an incorrect basic neckline or sleeve type; and **1** indicates severe generation artifacts in the corresponding region.

Major decoration and structure. Evaluates the preservation of important decorative and structural components, such as printed pattern, pockets, belts, visible zippers, lace, paneling, and closure structure. A score of **5** indicates that all key elements are correctly preserved; **4** indicates minor deviations in location, shape, or styling; **3** indicates noticeable replacement or simplification of major elements; **2** indicates missing, hallucinated, or functionally altered structural components; and **1** indicates severe corruption of the relevant region.

Material texture. Evaluates whether the perceived fabric category and texture expression are faithfully rendered. A score of **5** indicates highly consistent material appearance; **4** indicates the correct material category with slightly weakened texture expression; **3** indicates a visually similar but mismatched material impression; **2** indicates clearly incorrect material type; and **1** indicates no meaningful texture expression or strong artificial artifacts.

Fine-detail fidelity. Evaluates the preservation of non-logo fine-grained elements, such as buttons, zipper pulls, embroidery, and small labels. A score of **5** indicates accurate preservation of number, shape, position, and appearance; **4** indicates minor distortion; **3** indicates noticeable inconsistency or distortion; **2** indicates missing or incorrect elements; and **1** indicates severe corruption of the fine-detail region. This dimension is marked as **N/A** when no relevant fine-grained elements are present.

Logo preservation. Evaluates whether a visible garment logo is preserved in terms of shape, text, color, proportion, and position. A score of **5** indicates faithful and legible reproduction; **4** indicates slight distortion but clear recognizability; **3** indicates approximate preservation with substantial blur or deformation; **2** indicates loss of brand identity despite approximate placement; and **1** indicates complete disappearance or hallucination of an incorrect logo. This dimension is marked as **N/A** when no visible logo is present on the garment.

Overall, this fine-grained protocol provides a more informative benchmark than a single holistic score, as it explicitly reveals which aspects of garment fidelity are preserved well and which remain challenging for current virtual try-on methods.

A.2 Training Data Filtering

Our training data are collected from a large-scale e-commerce platform and consist of two image sources: **garment images** and **person images**. Our garment images are not flat-lay product images (i.e., a single piece of clothing shown on a white background); rather, they are in-the-wild images. In other words, our garment images include a large number of instances where models are wearing the target product. Person images are obtained from the live-streaming rooms of the e-commerce platform channels. To improve data quality and reduce noisy supervision, we apply a multi-stage filtering pipeline before constructing training pairs.

First, we remove corrupted or invalid samples, including damaged garment images and damaged person images that cannot be reliably decoded or displayed. Second, we discard low-quality samples whose image resolution is lower than 720p, as well as images that are excessively blurry. To identify blurred images, we employ a blur detection operator Pech-Pacheco et al. (2000) and remove samples that fall below a predefined sharpness threshold.

Third, for garment images, we use a vision-language model (VLM) to identify and exclude samples that contain multiple distinct products in a single image. Such cases are common in e-commerce data, where one product image may simultaneously display multiple variants, such as different colors or different styles of the same item. Since our goal is to establish an unambiguous correspondence between the source garment and the target try-on result, these multi-product samples are removed from the training set.

Fourth, for person images, we use the face detector provided by MediaPipe, an open-source cross-platform ML framework developed by Google, to filter out images with multiple faces. Samples containing multiple persons are excluded, as they introduce ambiguity regarding which person should be associated with the target garment during training.

After filtering, we obtain a cleaner set of product images and model images suitable for synthetic pair construction.

A.3 Synthetic Pair Construction

After the filtering stage, we randomly pair a garment image with a person image and use an image editing model to synthesize a virtual try-on result. We utilize Nano Banana Pro Google (2025c), Qwen-Image-Edit-2511 Wu et al. (2025a) and an internally developed try-on model to generate the try-on results. The editing instruction for each paired sample is automatically generated by Gemini-

3.1 Google (2025a), which produces a single English prompt describing how the garment in the model image should be replaced by the garment shown in the product image.

The prompt generation is designed to enforce high-fidelity product transfer. In particular, the generated instruction explicitly specifies the object to be replaced in the model image, identifies the target garment in the product image, and emphasizes that the edited result should preserve the target garment's fidelity, while keeping the other parts of the person image unchanged.

To avoid ambiguity in complex scenes, the prompt generator follows several additional rules. If the person in either image is wearing multiple layers of clothing, only the outermost garment is replaced. If the person in the model image is holding a hanger with a garment, only the held garment is replaced. These constraints help reduce editing ambiguity and improve the reliability of the synthesized virtual try-on data. The full prompt template is shown below.

> **ROLE & GOAL**
> You are an expert AI assistant specializing in generating precise, unambiguous English instructions for an image editing model. Your goal is to create a prompt that guides the editing model to perform a product swap in an e-commerce live streaming image.
>
> **TASK**
> Given two images, generate a single, clear text instruction for the image editing model.
>
> **INPUTS**
> 1. **Image 1:** An image of a model/host displaying a product (the source image).
> 2. **Image 2:** A standalone image of the new product to be swapped in (the reference image).
>
> **OUTPUT FORMAT**
> Your output must be a single line of text containing the English editing instruction.
>
> **CRITICAL RULES**
> 1. Identify the object to be replaced in Image 1 and describe the new object from Image 2.
> 2. Explicitly state that the new product appearance should be based on Image 2 or the reference image.
> 3. Clearly instruct the model to replace or swap the items.
> 4. If the person in Image 1 or Image 2 is wearing multiple garments, replace only the outermost garment. If the person in Image 1 is holding a hanger with a garment, replace only the held garment and do not replace the garment worn by the person.
> 5. Emphasize that the material, color, texture, details, and any logos of the new product from Image 2 must be accurately preserved.
> 6. Do not change the person's face, expression, or gesture.

A.4 User Study Protocol

We conduct a human preference study to compare the quality of virtual try-on models with or without RL optimization based on DAT. The goal of this study is to evaluate, from a garment-fidelity perspective, which generated result is more consistent with the garment image.

In each comparison instance, annotators are presented with a triplet of images composed of: (1) the garment image on the left, (2) a virtual try-on result generated by Method A in the middle, and (3) a virtual try-on result generated by Method B on the right. The two try-on results are generated by two different models, and their left-right assignment is randomized to avoid positional bias. Both try-on images are synthesized by image editing models that integrate the garment from the garment image onto a person image.

Annotators are asked to evaluate the two generated results by comparing the garment from each of them against that from the garment image. For each comparison instance, annotators complete eight annotation tasks, each corresponding to a different evaluation dimension. The objective is to determine which method produces a try-on result whose garment is more faithful to the garment image.

To improve annotation reliability, annotators are instructed to inspect each triplet at least twice before making any decision. For each task, annotators must select exactly one label. The available labels are Method A, Method B, and Same, except for dimensions that may be absent in the garment image, in which case an additional N/A-type option is provided.

The eight annotation tasks are defined as follows:

Overall consistency. Which result is more consistent with the garment image overall?
Options: **Method A / Method B / Same**

Silhouette Which result better matches the silhouette or overall garment shape in the garment image?
Options: **Method A / Method B / Same**

Color. Which result better matches the garment color in the garment image?
Options: **Method A / Method B / Same**

Neckline and sleeve shape. Which result better matches the neckline and sleeve shape in the garment image?
Options: **Method A / Method B / Same**

Major decoration and structure. Which result better matches the major decorative and structural elements of the garment in the garment image?
Options: **Method A / Method B / Same**

Material texture. Which result better matches the perceived garment material in the garment image?
Options: **Method A / Method B / Same**

Fine-detail fidelity. If the garment contains fine-grained elements, which result better preserves those details with respect to the garment image?
Options: **Method A / Method B / Same / No fine-grained elements**

Logo preservation. If the garment contains a visible logo, which result better preserves the logo with respect to the garment image?
Options: **Method A / Method B / Same / No visible logo**

This user study protocol enables direct pairwise comparison between different virtual try-on systems and provides fine-grained human judgments on which method better preserves the identity-defining attributes of the source garment.