GOT-ALIGN: LEARNING GENERATIVE OPERATION TOKENIZERS FOR PREFERENCE-ALIGNED IMAGE ENHANCEMENT

Anonymous authors
Paper under double-blind review

ABSTRACT

Photo retouching is inherently subjective: the same photograph can admit multiple plausible renditions, while different users may prefer different appearances. Existing methods either learn toward selected expert retouches or rely on explicit controls and instructions, making it difficult to capture implicit and spatially varying preferences. We propose GOT-Align, a preference-aligned generative photo retouching framework that models retouching transformations in a shared spatial operation space. Specifically, a Generative Operation Tokenizer (GOT) discretizes bilateral affine transformations into reusable local operation tokens, whose spatial compositions represent diverse retouching transformations while preserving the underlying image content. A conditional autoregressive Transformer then models the image-conditioned distribution over operation-token maps to generate multiple plausible candidates for the same input. For preference adaptation, human or vision-language model judges compare candidates over local image regions, preserving regional preference evidence that would otherwise be collapsed into a single image-level judgment. These regional preferences are attributed to spatially associated operation tokens and used for GRPO-based policy adaptation. Once adapted to a preference source, the model directly generates preference-aligned retouches for new inputs without requiring per-image controls or textual instructions. Experiments on standard photo retouching benchmarks demonstrate diverse generation and effective adaptation to different preference sources.

1 INTRODUCTION

Photo retouching aims to improve the visual appearance of photographs through adjustments to exposure, contrast, saturation, color balance, and tone while preserving the original image content and structure (Bychkovsky et al., 2011; Gharbi et al., 2017; Hu et al., 2018; He et al., 2020; Moran et al., 2020; Zeng et al., 2022). The same photograph often admits multiple plausible and visually appealing renditions, guided by diverse aesthetic intents and individual preferences (Kang et al., 2010; Kim et al., 2020; Song et al., 2021; Wang et al., 2024; Duan et al., 2025). Consequently, photo retouching is fundamentally a subjective task rather than a deterministic mapping from an input image to a fixed ground truth.

Historically, photo retouching paradigms have evolved through several stages. Early methods relied heavily on hand-crafted tone curves or adjustment functions (Kang et al., 2010; Bychkovsky et al., 2011); however, their limited mapping capacity hindered the achievement of high-quality visual results. With the rise of deep learning, subsequent approaches shifted toward learning complex, parameterized mapping functions (Gharbi et al., 2017; He et al., 2020; Moran et al., 2020; Zeng et al., 2022). Early learning-based methods typically optimize toward a deterministic ground-truth (GT) target image (Yang et al., 2022; Ouyang et al., 2023; Lou et al., 2025). Nonetheless, acquiring massive one-to-one paired supervision is exceptionally labor-intensive. Moreover, due to the non-linear nature of human visual perception, fitting a single target image often fails to faithfully capture the true distribution of high-quality imagery. To bypass these deterministic constraints, recent studies have increasingly adopted generative models, such as diffusion-based frameworks, supervised by collections of high-quality images, framing retouching as a domain-to-domain mapping task (Duan et al., 2025). While generative paradigms capture diverse distributions, they often incur heavy com-

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.

putational overhead and high inference latency, making them poorly suited for practical application. Furthermore, such heavy set-based distribution learning lacks fine-grained guidance for individual image instances, making it difficult to precisely align with a user's specific retouching intent.

In this paper, we propose GOT-Align, an efficient generative photo retouching framework paired with a localized preference-alignment training mechanism. Instead of adopting heavy generative architectures typically used for domain mapping, we build our enhancement framework upon the efficient operator-predicting mechanism of HDRNet (Gharbi et al., 2017). Concretely, we introduce a Generative Operation Tokenizer (GOT) to learn a dataset-level vocabulary of reusable local operation primitives. GOT enables efficient generative retouching by representing each complete retouch as a spatial composition of these discrete operation tokens, where a lightweight, conditional autoregressive Transformer models the image-conditioned distribution over these operation compositions. Furthermore, sampling from this generative distribution yields diverse retouching candidates, which provide a direct interface for preference elicitation, allowing human users or vision-language models (VLMs) to offer comparative feedback. To further capture fine-grained preferences that may vary across image regions, we partition each image into locally coherent superpixel regions and perform candidate selection at the region level. We then exploit the spatial organization of the operation-token representation to project superpixel-level comparative feedback onto spatially associated operation tokens, yielding localized preference signals for GRPO-based post-training (Shao et al., 2024) of the autoregressive policy. This effectively adapts the generative distribution using spatially localized preferences, preserving regional preference information that would otherwise be collapsed by a single image-level reward.

Our main contributions are summarized as follows:

- We propose GOT-Align, a generative photo retouching framework that models diverse retouching distributions and aligns them with preferences.
- We develop the Generative Operation Tokenizer (GOT). Built upon efficient operator-predicting mechanisms, GOT models retouching as the spatial composition of discrete operation tokens.
- We introduce a regional preference alignment strategy. By projecting superpixel-level feedback from human or VLM judges onto associated operation tokens, our method enables precise credit assignment for GRPO-based policy adaptation.

2 Related Work

2.1 Deep Photo Retouching

Deep photo retouching has been studied through a variety of transformation representations. Early learning-based approaches predict enhanced appearances either directly in image space or through compact parametric operators, including bilateral grids, local filters, region-specific transformations, and 3D lookup tables (Gharbi et al., 2017; Moran et al., 2020; Zeng et al., 2022; Ouyang et al., 2023; Lou et al., 2025). These structured representations provide efficient and content-preserving transformations, but most are trained to reproduce a selected expert target and therefore produce a largely deterministic output for each input.

Later work moves beyond a single target by introducing personalization, multiple styles, and explicit control. PIENet learns user-specific preference representations from selected examples (Kim et al., 2020), while StarEnhancer and flow-based approaches model multiple retouching styles or output distributions (Song et al., 2021; Wang et al., 2024). DiffRetouch models expert-retouched distributions with diffusion and supports attribute-conditioned generation (Duan et al., 2025); PerTouch combines semantic-region controls with VLM-based interaction and memory (Chang et al., 2026); and InstantRetouch enables instruction-guided editing through bilateral transformations (Wu et al., 2026). These approaches substantially broaden the set of attainable retouches, but diversity and controllability are different from preference adaptation: the desired variation is still expressed through selected examples, predefined styles or attributes, semantic controls, or textual instructions. In contrast, our method represents retouching as spatial compositions over a shared discrete operation vocabulary and adapts the resulting generative distribution directly from comparative preference feedback, without requiring a predefined control interface for each output.

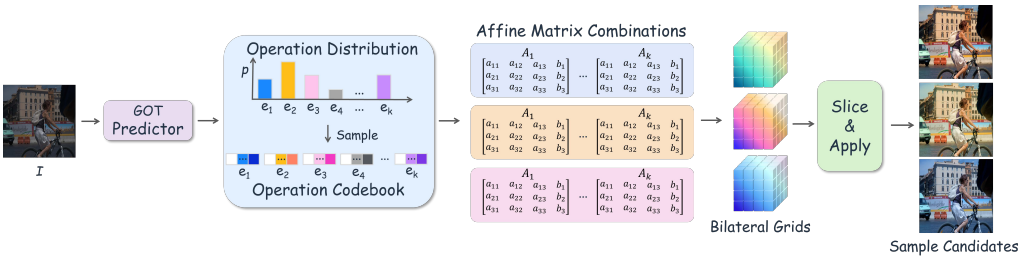

Figure 1: Generative operation modeling with a shared operation codebook. Conditioned on the input image, the GOT predictor models a distribution over spatial compositions of reusable operation primitives from the shared codebook. Sampled compositions are mapped to combinations of local affine transforms that assemble diverse bilateral affine grids, which are sliced and applied to the input image to generate multiple retouching candidates.

2.2 Preference Learning and Alignment for Visual Generation

Preference supervision provides an alternative to explicitly specifying a desired output. In visual enhancement and generation, such supervision has been incorporated through both learned preference evaluators and direct policy optimization. HiLLIE learns an image-quality evaluator from pairwise human rankings and uses it to guide low-light enhancement (Zhao et al., 2025), while ImageReward learns a human-preference reward model for text-to-image generation (Xu et al., 2023). These approaches demonstrate that relative judgments can provide useful supervision when perceptual quality is difficult to describe with a fixed reconstruction objective.

More recent work directly integrates preference signals into generative policy optimization. GRPO performs group-relative policy optimization using rewards from multiple sampled outputs (Shao et al., 2024), Flow-GRPO extends this paradigm to visual generation (Liu et al., 2025), and AceTone applies GRPO to autoregressive LUT generation for conditional image grading (Ma et al., 2026). Most such approaches nevertheless optimize candidate-level rewards, which summarize the quality or preference of an entire generated result. ViPO further addresses spatial credit assignment by redistributing sample-level scalar feedback into structure-aware pixel-level advantages using perceptual representations (Ni et al., 2026). Our setting differs in that the supervision itself is localized at its source: human or VLM judges directly compare candidate retouches over image regions, and the resulting regional preferences are projected onto the spatially associated operation tokens. This preserves the spatial origin of preference evidence and enables explicit region-to-token credit assignment for preference adaptation in the retouching transformation space.

3 Method

We first learn a spatially structured discrete operation space with the Generative Operation Tokenizer (GOT). Given an input–retouch pair (I, I^r), GOT encodes the underlying retouching transformation into a two-dimensional map of discrete operation tokens and decodes it into a bilateral affine grid for reconstructing the retouched image. We then train a conditional autoregressive predictor to model the image-conditioned distribution over operation-token maps. Conditioned only on the input image, the predictor can sample multiple operation compositions, which are decoded through the frozen GOT decoder to produce diverse retouching candidates. Finally, we adapt this generative distribution to a preference source using localized comparative feedback. Regional preferences provided by human or VLM judges are projected onto spatially associated operation tokens, yielding localized credit for GRPO-based post-training of the autoregressive policy.

3.1 Generative Operation Modeling

Existing photo retouching methods typically either learn a deterministic mapping toward a selected expert retouch or introduce output variation through predefined editing attributes, style conditions, or textual instructions. The former primarily treats expert retouches as image-specific prediction

targets, while the latter restricts variation to an externally specified control space. In preference-aligned retouching, however, the desired appearance is not known in advance and may vary across users. Rather than treating each expert retouch only as an endpoint to be reproduced, we view an input–retouch pair as an observation of a plausible retouching transformation. This motivates us to organize the transformations observed across the training data into a shared operation space, in which diverse retouches can be represented through different combinations of reusable operations.

Learning a Shared Operation Space. Photo retouching primarily modifies photometric appearance, such as exposure, color, and tone, while preserving the underlying image content. Such adjustments are also spatially varying, since different image regions may require different types or strengths of transformation. We therefore represent retouching in the transformation domain rather than directly in the image domain. Specifically, we use bilateral affine grids (Gharbi et al., 2017) to parameterize spatially varying photometric transformations. Given an input image $I \in \mathbb{R}^{H\times W\times 3}$ and its retouched reference I^r, the pair provides an observation of the transformation from the original appearance to the retouched one. We represent this transformation with a bilateral affine grid $G \in \mathbb{R}^{16\times 16\times 8\times 12}$, where the first three dimensions define the bilateral grid and each grid cell contains the 12 coefficients of a local 3×4 affine color transformation.

To construct a reusable discrete operation space, we learn a shared codebook $\mathcal{E} = \{e_k\}_{k=1}^{K}$ of recurring local retouching operations, where each entry $e_k \in \mathbb{R}^D$ represents an operation primitive shared across different images and spatial regions. Each primitive captures a reusable local transformation pattern rather than a complete retouching style or a manually specified editing attribute such as exposure or saturation. The resulting codebook therefore serves as a dataset-level vocabulary of local retouching operations.

Given a training pair (I, I^r), GOT encodes the corresponding retouching transformation into a continuous spatial representation $Z_e \in \mathbb{R}^{16\times 16\times D}$. At each spatial location (i, j), the continuous feature is quantized to its nearest operation primitive:

$$z_{i,j} = \arg\min_k \|Z_e(i,j) - e_k\|_2^2, \qquad Z_q(i,j) = e_{z_{i,j}}. \tag{1}$$

This yields a two-dimensional operation-token map $z \in \{1,\dots,K\}^{16\times 16}$ and its quantized representation $Z_q \in \mathbb{R}^{16\times 16\times D}$. A complete retouching transformation is thus represented as a spatial composition of local operation primitives rather than by a single style code, allowing image-specific retouches to be expressed as different configurations over a shared operation vocabulary.

GOT is trained end to end using image-space reconstruction together with a standard vector-quantization objective (van den Oord et al., 2017). The quantized representation Z_q is decoded into a bilateral affine grid $\hat{G}$, which is applied to the input image through bilateral slicing to reconstruct the corresponding retouched result.

Autoregressive Modeling of Operation Compositions. Once the shared operation space has been established, we model diverse retouching by learning the image-conditioned distribution over plausible operation compositions. Specifically, we use a conditional autoregressive model to factorize the distribution over operation-token maps as

$$p_\theta(z \mid I) = \prod_{t=1}^{T} p_\theta(z_t \mid z_{<t}, C), \tag{2}$$

where $z = (z_1, \dots, z_T)$ denotes the operation-token sequence obtained from the two-dimensional token map under a fixed ordering, T is the number of token positions, and C denotes the visual condition extracted from the input image. The model is trained on operation-token sequences obtained from the frozen GOT, learning different spatial compositions over the shared operation vocabulary. At inference time, multiple operation-token maps are sampled from $p_\theta(z \mid I)$, producing different bilateral affine grids and thus multiple plausible retouching candidates for the same input image.

3.2 Localized Preference Alignment

Controllable retouching provides explicit mechanisms for steering the output, but typically requires users to specify the desired appearance for each input. Instruction-guided methods require users to

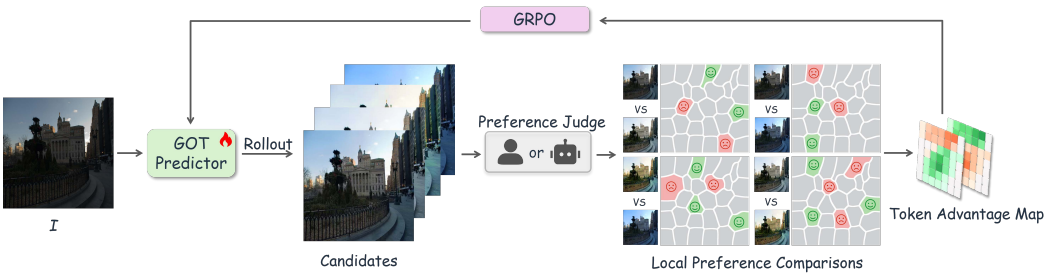

Figure 2: Localized preference alignment. For an input image I, the autoregressive predictor rolls out multiple retouching candidates. A human or VLM judge compares a candidate pair within superpixel regions, and the resulting regional preferences are projected onto the spatially associated operation tokens to form a token advantage map, which drives GRPO updates of the predictor.

formulate an appropriate instruction for every image (Wu et al., 2026), while aesthetic intentions can be abstract and difficult to translate into precise editing descriptions. Attribute-conditioned methods avoid free-form language but restrict control to predefined attribute dimensions and ranges (Duan et al., 2025), further requiring users to translate subjective preferences into numerical control values. Rather than asking users to specify how each image should be edited, we learn what is preferred directly from comparative choices among generated candidates. Moreover, because such preferences may vary across image regions, we preserve their spatial origin throughout the alignment process.

Given the generative distribution learned in Sec. 3.1, we sample N retouching candidates for each input and obtain comparative feedback from a human or VLM preference source. Candidate pairs (a, b) are compared within local image regions, and the evaluator only needs to indicate which candidate is preferred in each region. This comparison-based interface can express subjective and abstract aesthetic preferences without requiring explicit editing attributes or textual descriptions. Importantly, different candidates may be preferred in different regions of the same image, making a single image-level winner insufficient for localized preference learning.

We instantiate the comparison regions using SLIC superpixels (Achanta et al., 2012), yielding a set of content-adaptive regions $\mathcal{R} = \{R_m\}_{m=1}^{M}$. For each candidate n, the pairwise regional comparisons are aggregated into a regional preference score $S_m^{(n)}$. Unlike a global scalar reward, these regional scores retain where the preference originates in the image and thus provide spatially localized evidence for the underlying retouching decisions.

The regional preference scores are defined in image space, whereas the generative policy makes decisions in the discrete operation space. The spatial structure of the operation-token map provides a direct bridge between the two: preference evidence from an image region can be associated with the operation tokens governing the corresponding spatial area. We therefore project each regional preference score onto its spatially associated token positions, yielding a token-level preference credit $q_t^{(n)}$ for candidate n at position t. This region-to-token projection preserves localized preference evidence instead of collapsing all regional judgments into a single candidate-level reward.

To optimize each sampled candidate relative to the alternative retouches generated for the same input, we define a group-relative localized advantage:

$$A_t^{(n)} = q_t^{(n)} - \frac{1}{N} \sum_{j=1}^{N} q_t^{(j)}. \tag{3}$$

Here, $A_t^{(n)}$ measures the relative preference of the operation selected by candidate n at token position t among the sampled alternatives. Consequently, different operation choices within the same candidate may receive different, or even opposite, optimization signals.

We use these localized advantages to post-train the autoregressive policy with GRPO (Shao et al., 2024), while keeping the learned operation space fixed. The optimization therefore adjusts the probabilities of operation compositions without redefining the underlying operation vocabulary. Locally preferred operation choices are reinforced, whereas less-preferred choices are suppressed at the cor-

responding spatial positions. Through this process, the image-conditioned generative distribution is adapted toward the specified human or VLM preference source.

4 EXPERIMENTS

4.1 EXPERIMENTAL SETUP

Datasets and Protocols. We conduct experiments on the MIT-Adobe FiveK (Bychkovsky et al., 2011) and PPR10K (Liang et al., 2021) datasets. MIT-Adobe FiveK contains 5,000 photographs, each retouched by five experts (A–E). Since existing photo-retouching methods provide pretrained models under different FiveK evaluation protocols, we follow the corresponding protocol for each group of comparisons and report them separately in Sec. 4.2. Expert C is used as the common retouching style for all FiveK comparisons. For methods supporting expert-specific checkpoints or conditions, we use the corresponding Expert-C setting. All FiveK results are evaluated at 340p.

PPR10K contains 11,161 portrait photographs with retouched versions from three experts (A–C). Following the official protocol, we use 8,875 images for training and 2,286 images for testing. Expert A is used as the common retouching style, and all methods are evaluated at 360p. For methods supporting expert-specific checkpoints or conditions, we use the corresponding Expert-A setting. For quantitative method comparisons, GOT and GOT-Align use greedy decoding to produce a deterministic retouching result for each input.

Preference Evaluation. Since photo retouching is inherently one-to-many, similarity to a single expert reference does not fully reflect subjective preference. We therefore use pairwise preference judgments as the primary evaluation for method comparison. For each trial, two retouched results of the same input image are presented side by side with method identities hidden. Each pair is independently evaluated by Qwen3.6-35B-A3B-FP8 (Qwen Team, 2026), GPT-5.6 Luna (OpenAI, 2026), and human participants. For each VLM evaluator, every image pair is evaluated twice with the display order reversed. A win is recorded only when both evaluations consistently prefer the same underlying result; otherwise the comparison is treated as a tie. Human evaluation uses randomized left–right presentation.

For the two VLM evaluators, we report the percentage of pairwise comparisons in which the evaluated method is preferred. For human evaluation, we recruit 10 participants to perform pairwise preference comparisons. For each pair, participants select the visually preferred result or indicate a tie when no clear preference can be made, without access to the method identity. We summarize the collected human judgments using the **Human Mean Opinion Score (HMOS)**, which measures the overall human preference for the evaluated method across the pairwise comparisons.

Implementation Details. GOT uses an operation codebook with $K = 512$ entries and latent dimension $D = 64$, producing a 16×16 operation-token map for each retouching transformation. During training, images are resized to 512×512. GOT is trained for 300 epochs with a batch size of 16 using AdamW with a learning rate of 2×10^{-4}. The autoregressive predictor is trained for 100 epochs with a batch size of 16 using AdamW with a learning rate of 3×10^{-4}.

For preference alignment, we sample five retouching candidates for each input and partition each image into approximately 40 SLIC superpixels. Unless otherwise specified, GOT-Align denotes the model post-trained from GOT using localized preferences provided by Qwen3.6-35B-A3B-FP8. The autoregressive policy is optimized with GRPO using a learning rate of 1×10^{-5} while keeping the learned operation space fixed. The Qwen- and human-aligned variants are optimized for 250 and 50 steps, respectively.

4.2 COMPARISON WITH STATE-OF-THE-ART METHODS

We compare GOT and GOT-Align with representative photo-retouching methods, including AdaInt (Yang et al., 2022), LUTBGrid (Kim & Cho, 2024), BPAM (Lou et al., 2025), TSFlow (Wang et al., 2024), StarEnhancer (Song et al., 2021), and DiffRetouch (Duan et al., 2025). Since existing methods adopt different test splits on MIT-Adobe FiveK, we organize the FiveK comparisons according to the evaluation settings used by the corresponding released implementations or check-

points. Within each setting, all methods are evaluated on the same preprocessed test images, while model-specific normalization and inference procedures follow their official implementations.

Following Sec. 4.1, we report pairwise preference rates evaluated by Qwen, GPT-5.6 Luna, and human participants. Each value denotes the percentage of pairwise comparisons in which GOT or GOT-Align is preferred over the corresponding competing method. A value above 50% indicates that our method is preferred more frequently.

MIT-Adobe FiveK. We first compare against AdaInt (Yang et al., 2022), LUTBGrid (Kim & Cho, 2024), and BPAM (Lou et al., 2025). Under the MIT-Adobe-5K-UPE setting, we further compare against StarEnhancer (Song et al., 2021), TSFlow (Wang et al., 2024), and DiffRetouch (Duan et al., 2025). The results are summarized in Table 1.

Table 1: Pairwise preference rates (%) on MIT-Adobe FiveK. AdaInt, LUTBGrid, and BPAM are evaluated under the setting used by their released implementations, while StarEnhancer, TSFlow, and DiffRetouch follow the MIT-Adobe-5K-UPE setting. Values above 50% indicate a preference for our method.

Method	Qwen		GPT-5.6 Luna		HMOS	
	GOT	GOT-Align	GOT	GOT-Align	GOT	GOT-Align
AdaInt (Yang et al., 2022)	58.7	61.0	68.3	64.1	62.3	63.4
LUTBGrid (Kim & Cho, 2024)	65.2	67.7	70.3	70.7	65.6	65.8
BPAM (Lou et al., 2025)	59.6	62.1	65.3	64.7	60.7	61.2
StarEnhancer (Song et al., 2021)	61.4	62.7	57.8	57.7	60.2	61.3
TSFlow (Wang et al., 2024)	59.7	59.2	58.6	58.4	59.3	58.9
DiffRetouch (Duan et al., 2025)	59.4	63.5	64.1	66.7	62.8	63.4

PPR10K. On PPR10K, we compare GOT and GOT-Align against AdaInt (Yang et al., 2022), LUTBGrid (Kim & Cho, 2024), BPAM (Lou et al., 2025), and DiffRetouch (Duan et al., 2025). The results are reported in Table 2.

Table 2: Pairwise preference rates (%) on PPR10K. Values above 50% indicate a preference for our method.

Method	Qwen		GPT-5.6 Luna		HMOS	
	GOT	GOT-Align	GOT	GOT-Align	GOT	GOT-Align
AdaInt (Yang et al., 2022)	68.7	66.1	57.6	56.6	56.7	57.1
LUTBGrid (Kim & Cho, 2024)	65.5	69.5	61.2	63.1	62.4	62.7
BPAM (Lou et al., 2025)	66.7	64.9	59.0	59.3	61.1	60.8
DiffRetouch (Duan et al., 2025)	68.7	69.2	67.2	67.4	65.3	66.1

Across both datasets, every method-level comparison in Tables 1 and 2 favors our method: all reported rates lie above 50%, ranging from 56.6% to 70.7%, for both GOT and GOT-Align and under all three evaluators. Human evaluation is directionally consistent with both VLM evaluators across all comparisons, providing complementary evidence that the observed preference is not specific to a single automatic evaluator. Because both the pre-alignment and aligned models are preferred over the competing methods, these cross-method comparisons do not directly isolate the contribution of preference alignment. We therefore evaluate this effect explicitly in Sec. 4.3 by comparing each aligned variant against its pre-alignment counterpart.

Qualitative Comparisons. Figures 3 and 4 present representative visual comparisons on MIT-Adobe FiveK and PPR10K. GOT produces visually plausible retouching results with flexible adjustments in exposure, color, contrast, and local tone, while GOT-Align further adapts these characteristics according to the learned preference.



Figure 3: Visual comparison with state-of-the-art methods on the FiveK dataset for photo retouching.



Figure 4: Visual comparison with state-of-the-art methods on the PPR10K dataset for photo retouching.

4.3 Ablation Study

We ablate the contribution of localized preference alignment by comparing the aligned model against GOT under each preference source. Alignment is performed separately with Qwen- and human-provided localized feedback on FiveK and PPR10K, and every resulting pair is evaluated by Qwen, GPT-5.6 Luna, and human participants. The results are reported in Table 3.

Table 3 directly evaluates localized preference alignment by comparing each aligned variant against GOT. All aligned variants achieve preference scores above 50% under Qwen, GPT-5.6 Luna, and HMOS on both datasets, showing that localized preference optimization consistently improves preference over the pre-alignment model. On FiveK, the preference patterns are also consistent with source-specific adaptation: Qwen-based alignment receives a higher score from Qwen, whereas human-based alignment receives a higher HMOS. Together, these results demonstrate that GOT-Align can adapt the generative distribution using both VLM- and human-provided localized preferences.

5 Conclusion

In this work, we present GOT-Align, a preference-aligned generative framework for photo retouching that models diverse retouching operations and adapts them to different aesthetic preferences. We discretize bilateral-grid transformations into a structured operation-token space and autoregressively model their spatial compositions, enabling multiple plausible retouching results to be generated for the same input. Furthermore, localized comparative feedback is projected onto the corresponding operation tokens and used to optimize the generative distribution, allowing preference adaptation without requiring predefined editing attributes or textual instructions. Experiments on MIT-Adobe FiveK and PPR10K demonstrate the effectiveness of the proposed generative operation modeling and show that localized preference optimization can consistently shift the output distribution toward different preference signals.

Table 3: Preference alignment ablation. Each value is the preference rate (%) of the aligned model over GOT. Feedback denotes the localized preference source used to align the model, while the three rightmost columns denote the evaluators that measure the preference rate.

		Evaluator		
Dataset	Feedback	Qwen	GPT-5.6 Luna	HMOS
FiveK	Qwen	66.7	62.1	58.8
FiveK	Human	61.3	64.2	65.3
PPR10K	Qwen	64.4	64.5	58.5
PPR10K	Human	62.5	57.9	58.6



Figure 5: Qualitative effect of localized preference alignment under Qwen- and human-provided feedback.

AI USE STATEMENT

Generative AI tools were used to assist with language polishing and improving the clarity of the manuscript, but were not used to originate the core research idea, scientific hypotheses, or method design. As part of the studied system and evaluation, Qwen3.6-35B-A3B-FP8 was used to provide localized pairwise preference annotations for preference alignment, while Qwen3.6-35B-A3B-FP8 and GPT-5.6 Luna were used as pairwise preference evaluators in the experiments. All AI-assisted content and model-generated outputs used in this work were reviewed and incorporated under procedures specified by the authors, who take responsibility for the final manuscript and reported results.

ETHICS STATEMENT

This work studies subjective photo retouching and therefore involves both human participants and photographs of people. The human preference evaluation reported in Sec. 4.1 collected pairwise aesthetic judgments from ten volunteer participants. Participants only compared pairs of retouched images and were never asked for personal information; no demographic, biometric, or otherwise identifying data were recorded, and the collected judgments cannot be traced back to individuals.

All experiments use the publicly available MIT-Adobe FiveK (Bychkovsky et al., 2011) and PPR10K (Liang et al., 2021) benchmarks under their original licenses and intended research use, and we do not redistribute either dataset. PPR10K consists of portrait photographs, and we use it solely for evaluating photometric retouching; our method adjusts exposure, color, contrast, and local tone without modifying facial geometry, identity, or scene content, and we perform no identity-related analysis.

REPRODUCIBILITY STATEMENT

All information needed to reproduce our results is contained in Sec. 4.1 and the appendix. Sec. 4.1 specifies the datasets and evaluation protocols, including the train and test splits, the expert retouching style used for each benchmark, the evaluation resolutions, and the pairwise preference protocol applied to both the model-based and the human evaluators. It also lists the complete set of architectural and optimization hyperparameters for the generative operation tokenizer, the autoregressive predictor, and the GRPO-based preference alignment stage, together with the superpixel setting used to form localized feedback. Because existing methods adopt different FiveK evaluation splits, Sec. 4.2 states which setting each comparison follows, and every competing method is run from its official implementation and released checkpoint on identically preprocessed inputs.

REFERENCES

Radhakrishna Achanta, Appu Shaji, Kevin Smith, Aurélien Lucchi, Pascal Fua, and Sabine Süsstrunk. Slic superpixels compared to state-of-the-art superpixel methods. *IEEE Transactions*

on Pattern Analysis and Machine Intelligence, 34(11):2274–2282, 2012.

Vladimir Bychkovsky, Sylvain Paris, Eric Chan, and Frédo Durand. Learning photographic global tonal adjustment with a database of input/output image pairs. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 97–104, 2011.

Zewei Chang, Zheng-Peng Duan, Jianxing Zhang, Chun-Le Guo, Siyu Liu, Hyungju Chun, Hyunhee Park, Zikun Liu, and Chongyi Li. Pertouch: Vlm-driven agent for personalized and semantic image retouching. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 40, pp. 2752–2759, 2026.

Zheng-Peng Duan, Jiawei Zhang, Zheng Lin, Xin Jin, XunDong Wang, Dongqing Zou, Chun-Le Guo, and Chongyi Li. Diffretouch: Using diffusion to retouch on the shoulder of experts. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, pp. 2825–2833, 2025.

Michaël Gharbi, Jiawen Chen, Jonathan T. Barron, Samuel W. Hasinoff, and Frédo Durand. Deep bilateral learning for real-time image enhancement. *ACM Transactions on Graphics*, 36(4):118:1–118:12, 2017.

Jingwen He, Yihao Liu, Yu Qiao, and Chao Dong. Conditional sequential modulation for efficient global image retouching. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pp. 679–695, 2020.

Yuanming Hu, Hao He, Chenxi Xu, Baoyuan Wang, and Stephen Lin. Exposure: A white-box photo post-processing framework. *ACM Transactions on Graphics*, 37(2):26:1–26:17, 2018.

Sing Bing Kang, Ashish Kapoor, and Dani Lischinski. Personalization of image enhancement. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1799–1806, 2010.

Han-Ul Kim, Young Jun Koh, and Chang-Su Kim. Pienet: Personalized image enhancement network. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pp. 374–390, 2020.

Wontae Kim and Nam Ik Cho. Image-adaptive 3d lookup tables for real-time image enhancement with bilateral grids. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pp. 91–108, 2024.

Jie Liang, Hui Zeng, Miaomiao Cui, Xuansong Xie, and Lei Zhang. Ppr10k: A large-scale portrait photo retouching dataset with human-region mask and group-level consistency. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 653–661, 2021.

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. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2025.

Junyu Lou, Xiaorui Zhao, Kexuan Shi, and Shuhang Gu. Learning pixel-adaptive multi-layer perceptrons for real-time image enhancement. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 14095–14105, 2025.

Tianren Ma, Mingxiang Liao, Xijin Zhang, and Qixiang Ye. Acetone: Bridging words and colors for conditional image grading. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 25851–25860, 2026.

Sean Moran, Pierre Marza, Steven McDonagh, Sarah Parisot, and Gregory Slabaugh. Deeplpf: Deep local parametric filters for image enhancement. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 12826–12835, 2020.

Ziqi Ni, Yuanzhi Liang, Rui Li, Yi Zhou, Haibin Huang, Chi Zhang, and Xuelong Li. Seeing what matters: Visual preference policy optimization for visual generation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 27260–27269, 2026.

OpenAI. GPT-5.6: Frontier intelligence that scales with your ambition, July 2026. URL https://openai.com/index/gpt-5-6/.

Wenqi Ouyang, Yi Dong, Xiaoyang Kang, Peiran Ren, Xin Xu, and Xuansong Xie. Rsfnet: A white-box image retouching approach using region-specific color filters. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 12160–12169, 2023.

Qwen Team. Qwen3.6-35B-A3B: Agentic coding power, now open to all, April 2026. URL https://qwen.ai/blog?id=qwen3.6-35b-a3b.

Zhihong Shao, Peiyi Wang, Qihao Zhu, Runxin Xu, Junxiao Song, Xiao Bi, Haowei Zhang, Mingchuan Zhang, Y. K. Li, Y. Wu, and Daya Guo. Deepseekmath: Pushing the limits of mathematical reasoning in open language models. *arXiv preprint arXiv:2402.03300*, 2024.

Yuda Song, Hui Qian, and Xin Du. Starenhancer: Learning real-time and style-aware image enhancement. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 4126–4135, 2021.

Aaron van den Oord, Oriol Vinyals, and Koray Kavukcuoglu. Neural discrete representation learning. In *Advances in Neural Information Processing Systems (NeurIPS)*, pp. 6306–6315, 2017.

Haolin Wang, Jiawei Zhang, Ming Liu, Xiaohe Wu, and Wangmeng Zuo. Learning diverse tone styles for image retouching. *IEEE Transactions on Image Processing*, 33:310–321, 2024.

Jiarui Wu, Yujin Wang, Ruikang Li, Fan Zhang, Mingde Yao, and Tianfan Xue. Instantretouch: Efficient and high-fidelity instruction-guided image retouching with bilateral space. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 8216–8226, 2026.

Jiazheng Xu, Xiao Liu, Yuchen Wu, Yuxuan Tong, Qinkai Li, Ming Ding, Jie Tang, and Yuxiao Dong. Imagereward: Learning and evaluating human preferences for text-to-image generation. In *Advances in Neural Information Processing Systems (NeurIPS)*, pp. 15903–15935, 2023.

Canqian Yang, Meiguang Jin, Xu Jia, Yi Xu, and Ying Chen. Adaint: Learning adaptive intervals for 3d lookup tables on real-time image enhancement. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 17522–17531, 2022.

Hui Zeng, Jianrui Cai, Lida Li, Zisheng Cao, and Lei Zhang. Learning image-adaptive 3d lookup tables for high performance photo enhancement in real-time. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(4):2058–2073, 2022.

Xiaorui Zhao, Xinyue Zhou, Peibei Cao, Junyu Lou, and Shuhang Gu. Hillie: Human-in-the-loop training for low-light image enhancement. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV) Workshops*, pp. 3992–4002, 2025.

APPENDIX

A ADDITIONAL EXPERIMENTAL DETAILS

A.1 LOCALIZED VLM PREFERENCE ANNOTATION

For Qwen-based preference alignment, we use Qwen3.6-35B-A3B-FP8 to provide localized pairwise preference annotations. Given two retouched candidates of the same input image together with a superpixel map, the evaluator independently judges each candidate region and selects only regions exhibiting a clear and reliable local preference. The judgment considers local naturalness, exposure, color, contrast, highlight and shadow detail, and visible artifacts.

To avoid introducing weak or redundant supervision, the evaluator may select no region and returns at most six regions with reliable preference evidence. For each selected region, it reports the preferred candidate, the type and magnitude of the visual difference, its confidence, and a short explanation. The prompt used for localized annotation is provided below.

System Prompt.

You are an expert in LOCAL image-enhancement quality evaluation at the superpixel-region level. You are given two enhanced versions A and B of the SAME photo, plus a superpixel map with numeric region IDs. For each candidate region, decide INDEPENDENTLY whether one version is clearly and reliably better there for human perceptual quality (naturalness, exposure, color, contrast, highlight/shadow detail, artifacts).

Rules:

- Judge each region ONLY by its own pixels. Do NOT judge the whole image and do NOT pick a global winner.
- Do NOT select a region merely because A and B differ by a global tone/color/brightness/saturation/contrast shift. Select only if that shift makes THIS region clearly more (or less) natural, realistic, better exposed, clearer, or less artifacted than the other version.
- Select a region only if the local difference is clear, meaningful, and reliable. It is completely fine to select no regions.
- Evaluate A and B strictly on their own local merits. Do NOT try to balance the number of A-winners and B-winners; do not look for opposite preferences on purpose.
- Select between 0 and 6 regions. If more than 6 regions appear valid, keep only the clearest, most reliable, and non-redundant regions. Prefer selecting fewer regions over including weak or repetitive evidence.
- `region_id` must be one of the given candidate IDs; `winner` is "A" or "B" only.

Output JSON only.

User Prompt.

Two enhanced versions (A, B) of the same photo, plus a superpixel map with numeric region IDs.

Candidate region IDs: {`candidate_ids`}

For each candidate region, select it ONLY if all hold:

(1) A and B clearly differ there,

(2) the difference is meaningful (not weak or purely stylistic),

(3) one version is clearly better locally.

Return at most 6 selected regions. Returning an empty list is valid when no region has a clear and reliable preference.

Output JSON ONLY in this schema (only the selected regions; nothing else):

```
{
  "selected_regions": [
    {
      "region_id": <int>,
      "winner": "A" | "B",
      "difference_type":
        "color_naturalness" | "exposure" | "contrast" |
        "highlight" | "shadow" | "detail" | "artifact" |
        "overall_local_naturalness",
      "magnitude": "large" | "moderate",
      "confidence": "high" | "medium",
      "reason": <str>
    }
  ]
}
```

A.2 VLM PAIRWISE EVALUATION

For quantitative evaluation, Qwen3.6-35B-A3B-FP8 and GPT-5.6 Luna independently compare two retouched versions of the same input image at the whole-image level. Unlike the localized annotations used for preference alignment, this evaluation considers the overall visual quality of the complete image, including naturalness, exposure, color, contrast, highlight and shadow detail, and visible artifacts.

To reduce positional bias and unstable judgments, each image pair is evaluated twice with the display order reversed. A win is recorded only when both evaluations consistently prefer the same underlying retouched result, regardless of whether it is displayed as A or B. If the two evaluations prefer different underlying results, the pair is recorded as a tie. We also record a tie when both evaluations report the difference magnitude as `none`. The resulting aggregated win/tie/loss judgment constitutes one vote for that image pair.

For each evaluation, the model first describes the perceptual differences between the two candidates, reports the magnitude of the difference, and then selects the preferred result. The same prompt and aggregation rule are used for all compared methods and for both VLM evaluators.

System Prompt.

> You are an expert judge of photo retouching quality. You are given two retouched versions of the SAME photo. Judge which one looks better to a human viewer, considering naturalness, exposure, color, contrast, highlight and shadow detail, and absence of artifacts.
>
> **Rules:**
>
> - Look at both images carefully before deciding.
> - Judge only the visual quality. The display order carries no information: do not let the order in which the images appear influence your answer.
> - Fill in "differences" before "winner": name what actually differs between the two versions, concretely and in one or two sentences. Then choose based on what you wrote there. If the two versions look the same to you, say so in that field.
> - Report in "magnitude" how far apart the two versions are: "none" if you cannot see any difference, "subtle" if you can only see it by looking closely, "clear" if it is immediately visible, and "large" if the two look like different edits. Judge the size of the difference honestly and separately from which version you prefer.
>
> Output JSON only.

User Prompt.

> Two retouched versions of the same photo.
>
> The candidate file names are: `<id>_A.png` and `<id>_B.png`
>
> Which one looks better overall?
>
> Output JSON ONLY in this schema, in this key order, and nothing else:

```
{
  "differences":
    "<what differs between the two, 1-2 sentences>",
  "magnitude":
    "none" | "subtle" | "clear" | "large",
  "winner":
    "<id>_A.png" | "<id>_B.png"
}
```

A.3 Human Preference Evaluation

Human evaluation follows a simple whole-image pairwise comparison protocol. Participants are shown two retouched versions of the same input image with the corresponding method identities hidden and the left–right order randomized. They are asked which result is visually more appealing overall and may select either candidate or indicate a tie when no clear preference can be made. Participants are instructed to judge the overall photographic appearance rather than similarity to a particular expert reference.

We recruit ten participants for the human study, consistent with the evaluation protocol described in Sec. 4.1. The collected pairwise judgments are aggregated into HMOS to summarize the overall human preference for each evaluated method.

B Additional Experiments and Results

Operation Tokenizer Reconstruction. We first evaluate whether the learned discrete operation-token representation can faithfully preserve the retouching transformations observed in the datasets. Given an input–retouch pair, the corresponding transformation is encoded into the discrete operation-token space and decoded through the GOT decoder to reconstruct the retouched image.

On MIT-Adobe FiveK, reconstruction quality is evaluated over 2,500 input–retouch pairs, corresponding to 500 test images and all five expert retouches (A–E). Averaged over the five experts, GOT achieves a PSNR of 34.98 dB and an SSIM of 0.9507. On PPR10K, we evaluate 6,858 input–retouch pairs, corresponding to 2,286 test images and all three expert retouches (A–C). Averaged over the three experts, GOT achieves a PSNR of 37.31 dB and an SSIM of 0.9825. These results indicate that the discrete operation-token representation preserves the underlying retouching transformations with high reconstruction fidelity.

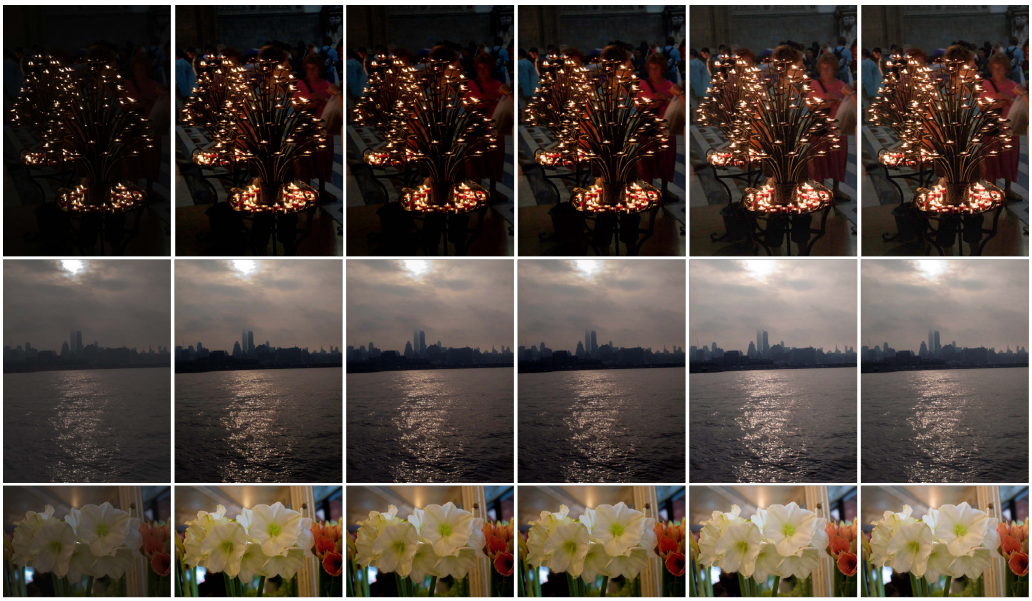

Figure 6: Additional visual comparisons with state-of-the-art methods on the FiveK dataset for photo retouching.



Figure 7: Additional visual comparisons with state-of-the-art methods on the FiveK dataset for photo retouching.



Figure 8: Additional visual comparisons with state-of-the-art methods on the PPR10K dataset for photo retouching.