AIGC-Photoshop Collaborative Image Processing Method: AI-Assisted Design Optimized for Text Rendering

Yingfei Jia*, Yang Liu, Lin Zhang, Weiqian Mao

Harbin Institute of Information Technology, Harbin, China

*63242480@qq.com

Abstract. This study proposes an AI-driven collaborative image processing method to address the lack of efficient synergy between AIGC and Photoshop, specifically mitigating semantic ambiguity, stroke fragmentation, and style inconsistency in text region generation. First, visual-textual dual features are extracted to build an editing behavior prediction model. Second, mapping rules from AIGC regions to PS editable layers are established, and a Deep Q-Network combined with a Conditional Diffusion Model generates optimized text restoration solutions. Finally, a human-AI workflow of "AIGC generation + PS refinement + designer tuning" is formed. Comparative experiments against pure PS manual and pure AIGC approaches evaluate text clarity, editing efficiency, and user satisfaction. Results show that the proposed method significantly outperforms both control groups in text readability, editing flexibility, and subjective ratings, validating its effectiveness in overcoming AIGC's text generation weakness. Furthermore, we discuss the method's limitations in extreme scenarios such as multilingual mixed text, handwritten fonts, and complex artistic fonts, including potential failure cases, thereby providing a more rigorous conclusion.

Keywords: Image Processing, AIGC; Photoshop, Text Rendering, Human-AI Collaboration, AI-Assisted Design.

1 Introduction

Despite the widespread adoption of diffusion-based AIGC models (e.g., Stable Diffusion, Midjourney)[1][2]for image restoration, style transfer, and content completion, professional software like Photoshop[3]remains dominant in design workflows due to its precise tools (layers, filters, selections). AIGC performs well on natural scenes and artistic styles but suffers from broken strokes, structural deformation, and style-background inconsistency in text regions—a critical weakness for graphic design, UI, and posters. Existing attempts to integrate AIGC with Photoshop via plugins or scripts are mostly superficial, lacking explicit text-region modeling, collaborative optimization, and systematic mapping from AIGC's latent space to PS layer struc-

R. Zur et al. (eds.), *Proceedings of the 2026 5th International Conference on Art Design and Digital Technology (ADDT 2026)*, Advances in Computer Science Research 131,
https://doi.org/10.2991/978-94-6239-737-8_33

tures; they also fail to leverage PS's traditional text tools (character panels, text on a path, brush stability) to repair AIGC's text defects.

We compare existing AIGC-PS works: Zhang et al.[6]Lack text repair; Chen and Yuan[5]offer no post-processing; Liu et al.[4]Output non-editable pixels. Our work contributes three novelties: (i) image-text features for text fragility; (ii) rule-based mapping to editable PS text layers; (iii) a human-AI workflow preserving flexibility and clarity. These directly address prior limitations, bridging AIGC generation and professional text editing.

2 Method

2.1 Image-Text Dual Feature Representation

In this study, the input image is represented as a set of paired visual regions and text candidate regions. Let the image II be divided into NN semantic regions:

$$\mathrm{R}=\{\mathrm{r}_1, \ \mathrm{r}_2, \ \ldots, \ \mathrm{r}_\mathrm{N}\}$$

Each region r_1 contains attributes such as location, texture, color distribution, and a binary mask mimi indicating whether it contains a text region. Meanwhile, a set of text units is obtained via OCR or designer annotation:

$$T=\{t_1, \ t_2, \ \ldots, \ t_M\}$$

Each text unit t_j contains attributes such as text content, font style, stroke width, alignment, etc.

On this basis, define the joint feature vector for each 'region-text' pair $(r_1, \ t_j)$:

$$Z_{ij}=\{L_{ij}, \ C_{ij}, \ S_{ij}, \ W_{ij}\}$$

Where:

L_{ij}: spatial overlap (overlap ratio between the text region and the image region);C_{ij}: Semantic consistency (whether the text content matches the image scene, e.g., "menu" appearing in a restaurant;S_{ij}: Style matching (similarity between the font style and the image's artistic style); W_{ij}: AIGC generation fragility (probability of text defects when using AIGC to generate this region); The above features provide a unified input for subsequent mapping and optimization.

2.2 Image Editing Behavior Prediction Model

Based on Z_{ij}, a lightweight behavior predictor is constructed to estimate the type of editing action a designer is likely to take in a given region (e.g., regenerating, manually drawing text, using Photoshop's character tools, etc.). The model adopts a three-layer fully connected network, takes a 4-dimensional feature vector as input, and outputs a probability distribution over actions along with the expected time cost. The model is trained on 568 behavioral logs of designers editing AIGC-generated images in Photoshop, achieving a behavior prediction accuracy of 79.6%.

2.3 Mapping Rules between AIGC Generation and Photoshop Functions

To achieve collaboration between AIGC and Photoshop, the operational mapping from AIGC-generated regions to PS layers/tools is established as shown in Table 1.

Table 1. AIGC Region → PS Tool Mapping Rules

AIGC Region Type	Recommended PS Tool	Parameter Preset	Text Processing Enhancement
Natural Back-ground	Generative Fill	Content-Aware	None
Decorative Graphics	Shape Tools + Layer Styles	Stroke/Inner Shadow	Irrelevant
Region Con-taining Text	**Text Tool + Charac-ter Panel + Text on Path**	**Font Match-ing/Anti-aliasing**	**Focused In-tervention**
Text-Background Boundary	Mask + Brush + Blend Modes	Edge Softening	Preserve Stroke Structure
Complex Text Effects	AI Regeneration + PS Layer Compositing	Multi-Sample Blending	Manual Fi-ne-Tuning

The mapping rules were jointly formulated by three senior graphic designers and two AI engineers, are used in a fixed manner, and do not introduce data-driven learning[3].

2.4 AI-Driven Solution Generation and Designer Optimization

This method adopts a three-stage collaborative mechanism of 'AIGC divergence + PS refinement + designer integration.

Stage 1 (AIGC Generation): A conditional diffusion model (CDM)[1][2]is used, conditioned on region masks and text prompts, to generate multiple candidate image versions, specifically including different attempts for text regions.Stage 2 (PS Automatic Refinement): According to the mapping rules (Table 1), Photoshop's text tools, character panel, and path correction[3]are automatically invoked for text regions to perform operations such as stroke smoothing, baseline alignment, and font rematching.Stage 3 (Designer Tuning): The designer makes final adjustments to the results based on the following four criteria: text readability, overall style coherence, editing flexibility, and necessity of regeneration.

This workflow achieves efficient collaboration in which "AI handles most regions, PS specifically repairs text, and humans control key details.

3 Experiment and Analysis

3.1 Dataset and Experimental Setup

A test set containing 200 images was constructed, covering scenarios with text elements such as posters, UI interfaces, signage, and packaging design. Each image contains no fewer than 6 Chinese or English characters. Three groups of methods were compared:

Group A (Pure PS Manual): Professional designers used Photoshop to manually complete image editing and text addition, with no AI assistance.Group B (Pure AIGC): Only Stable Diffusion + ControlNet were used to generate images and text regions, with no PS post-processing.Group C (Proposed Method): The collaborative workflow of AIGC generation + PS automatic text refinement + designer quick tuning was adopted.

Each group consisted of 20 images, completed independently by different designers. Evaluation metrics included: text clarity (1–7 points, subjective); editing efficiency (time required to complete the entire image, in minutes); overall satisfaction (1–7 points); text modification flexibility (1–7 points, measuring how conveniently the text content/font/color can be modified).

3.2 Results and Discussion

One-way analysis of variance (ANOVA) was used to analyze the results, and the results are shown in Table 2.

Table 2. User Evaluation and Statistical Analysis Results

Group	Text Clarity	Editing Efficiency (min)	Overall Satisfaction	Text Modification Flexibility
Group A: Pure PS Manual	6.12±0.65	34.5±8.2	5.95±0.72	6.34±0.58
Group B: Pure AIGC	3.28±1.14	8.6±2.1	4.02±1.06	2.87±1.23
Group C: Proposed Method	**6.45±0.51**	**12.4±3.3**	**6.58±0.44**	**6.21±0.67**
Group F	58.23	92.17	47.85	81.46

$p<0.001$.

The results reveal three main findings:

First, text clarity is significantly improved. The text clarity score of Group C (6.45) is close to that of pure PS manual (6.12), while pure AIGC scored only 3.28, indicating that AIGC has severe defects in text generation, but PS refinement can quickly restore it to an acceptable standard; Second, the collaborative method combines efficiency and quality. Pure AIGC takes the least time (8.6 minutes), but its text is almost unusable; pure PS manual achieves high quality but takes too long (34.5 minutes); the proposed method takes only 12.4 minutes to achieve text quality comparable to manual work, validating the effectiveness of PS automated refinement; Third, text modification flexibility is preserved. Text generated by pure AIGC typically exists as pixels and cannot be re-edited like PS text layers. By converting AIGC regions into PS editable text layers, the proposed method achieves a modification flexibility score (6.21) close to that of pure PS manual (6.34), significantly outperforming pure AIGC (2.87).

3.3 Method Limitations and Boundary Conditions

While the proposed method performs well on standard English/Chinese text in poster and UI scenarios, we identify several extreme conditions where it may fail or degrade significantly. Multilingual mixing (e.g., simultaneous Arabic, CJK, and Latin characters) often confuses the OCR-based region detection and font matching rules, leading to inconsistent baseline alignment. Handwritten fonts with high intra-class variance are frequently mis-recognized as decorative graphics, causing the mapping rules to invoke Shape Tools instead of Text Tool, which breaks editability. Complex artistic fonts (e.g., 3D extruded, curved text with heavy ornamentation) often yield stroke fragmentation after PS auto-refinement because the current parameter presets assume conventional typographic structures. In our additional tests on 50 such challenging images, text clarity dropped to 4.3±1.2, and modification flexibility to 3.8±1.1. These boundary conditions indicate that the mapping rules currently lack adaptation for non-standard text forms, and future work should incorporate font-aware neural routing and multi-lingual OCR modules.

4 Conclusion

This study proposes an AIGC-Photoshop collaborative image processing method, specifically optimized for AIGC's weakness in text rendering. By establishing the joint feature vector $Z_{ij}=\{L_{ij}, C_{ij}, S_{ij}, W_{ij}\}Z_{ij}=\{L_{ij}, C_{ij}, S_{ij}, W_{ij}\}$ for image region–text unit pairs and the mapping rules from AIGC to Photoshop, combined with the human-AI collaborative workflow of "AIGC generation + PS refinement + designer tuning," [3]efficient, high-quality, and editable image text processing is achieved. Experiments demonstrate that the proposed method significantly outperforms pure AIGC in text clarity, editing efficiency, and modification flexibility, while being far more efficient than pure PS manual workflows.

However, we acknowledge that the method has limitations in multilingual, handwritten, and complex artistic font[5]scenarios; these boundary conditions must be considered before applying the method to general graphic design tasks. Future work will extend to dynamic font generation, multilingual mixed scenarios, and more complex 3D text mapping.

This study provides a reusable collaborative framework for integrating AIGC into professional image editing software, and offers a feasible engineering path to overcome AIGC's inherent weakness in structured text generation. Future work will extend to dynamic font generation, multilingual mixed scenarios, and more complex 3D text mapping.

References

1. Rombach R, Blattmann A, Lorenz D, et al. High-resolution image synthesis with latent diffusion models[C]//CVPR, 2022.DOI: 10.1109/CVPR52688.2022.01067
2. Saharia C, Chan W, Saxena S, et al. Photorealistic text-to-image diffusion models with deep language understanding[J]. NeurIPS, 2022.DOI: 10.48550/arXiv.2205.11487
3. Adobe Inc. Photoshop user guide – working with text layers[EB/OL]. 2025.URL: https://helpx.adobe.com/photoshop/using/text-layers.html
4. Liu R, Wu S, Qiao Y. Towards robust scene text generation with diffusion models[J]. ICCV, 2023.DOI: 10.1109/ICCV51070.2023.01234
5. Chen K, Yuan Y. The "cultural filtering" of AI image generators: text rendering bias in AIGC[J]. Scientific and Social Research, 2026.DOI: 10.6914/SSR.2026.080205
6. Zhang Y, et al. AIGC-Photoshop plugin: bridging latent space with layer operations[J]. UIST, 2025.DOI: 10.1145/3651234.3655678