

Article

Form and Culture in Children's Picture Books: A Panofskian, Computer-Vision-Assisted Comparison of AI Images Generated by Doubao and Handcrafted *Journey to the West* Illustrations

Xinyu Du and Yanfang Han *

School of International Studies, Zhengzhou University, Zhengzhou 450001, China; 15837396723@163.com

* Correspondence: SFLhanyanfang@126.com

Abstract

This experimental study examines how generative AI reshapes the balance between perceptual fluency and cultural semiosis in children's picture-book illustration by using a corpus derived from *The Monkey King* picture-book series, adapted from the classical Chinese novel *Journey to the West*. The study compares 224 handcrafted illustrations with 224 AI images generated by the Doubao platform (Seedream-3.0). Computational visual metrics, edge curvature and color entropy were calculated using OpenCV, while iconographic features were manually annotated with the UAM Image Tool (Mick O'Donnell, Universidad Autónoma de Madrid, Madrid, Spain). Quantitative analysis reveals statistically significant formal divergences between the two illustration modes. AI images generated by the Doubao platform exhibit a shift toward more outwardly convex contours and reduced color entropy, indicating smoother contours and chromatic homogenization that enhance perceptual accessibility. Iconographic analysis, however, demonstrates an attenuation of culturally specific symbols. High-frequency, contour-salient attributes are largely preserved, whereas low-frequency, ritualized, and hierarchically organized elements are frequently omitted or simplified. The findings reveal a tension between perceptual fluency and cultural-semantic stability in AI images generated by Doubao (ByteDance, Beijing, China), employing the Seedream 3.0 model. They support a framework of conditional applicability, with implications for picture-book illustration, cultural adaptation, and children's visual-literacy education.

Keywords: AI images generated by Doubao; handcrafted images; children's picture book; *The Monkey King* (*Journey to the West*); Panofsky's iconology tripartite model

Academic Editor: Zhonggen Yu

Received: 14 January 2026

Revised: 11 February 2026

Accepted: 25 February 2026

Published: 26 February 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

1. Introduction

Children's picture books are multimodal artifacts in which images frequently lead narration and scaffold early visual literacy (Nodelman, 1988). In culturally saturated narratives, illustration does more than depict events: it mobilizes conventionalized iconographies such as costume, attributes, weapons, and ritual colors to convey religious, ethical, and social meanings beyond the literal scene (Panofsky, 1955; Gombrich, 1960). However, picture-book iconography is not fixed. Across editions and media, visual conventions are continuously rearticulated, raising questions about how much formal variation an iconographic sign can tolerate before its cultural reference becomes unstable (Kress & van Leeuwen, 2006; Kress, 2010).

This experimental study examines this tension through *Journey to the West*, a sixteenth-century Chinese novel traditionally attributed to Wu Chengen, which recounts the pilgrimage of the Buddhist monk Tang Monk to India alongside his three disciples (Wu, 2019). As one of the most influential and culturally saturated narratives in Chinese literature, it offers a rich repository of stabilized iconography (Xia, 1968/2001). The present analysis focuses specifically on *The Monkey King* picture-book series, a children's illustrated adaptation that preserves a highly conventionalized yet semantically dense visual system, making it an ideal corpus for examining the balance between perceptual fluency and cultural fidelity.

Visual meaning in picture books emerges from the interaction between perceptual organization and iconographic convention. Saliency structures such as contour continuity, color organization, and compositional regularity guide attention and support visual accessibility (Kress & van Leeuwen, 2006). On the other hand, recurrent symbols, such as Tang Monk's Vairocana crown, Sun Wukong's tiger-skin skirt and golden cudgel, or Zhu Bajie's rake, function as cultural anchors that stabilize interpretation across encounters (Panofsky, 1955; Nikolajeva & Scott, 2001). Perceptual research further shows robust preferences for smooth curvature and coherent chromatic distributions, which enhance processing fluency and visual comfort (Bar & Neta, 2006; Bertamini et al., 2016). Yet, algorithmic image generation tends to optimize precisely these formal properties, raising concerns that perceptual gains may come at the cost of low-frequency or ritualized cultural symbols.

Against this background, the present experimental study compares AI images generated by Doubao with handcrafted illustrations within a paired corpus from *The Monkey King* series. It addresses three research questions:

RQ1. What systematic differences in edge curvature and color entropy distinguish AI images generated by Doubao from handcrafted illustrations?

RQ2. To what extent are canonical iconographic symbols of the four protagonists preserved, simplified, or weakened in AI images generated by Doubao in terms of appearance, crown, clothing, and weapon?

RQ3. How do the quantitative differences in formal metrics and iconographic representation support a conceptual distinction between AI images generated by Doubao as agents of formal fluency and handcrafted illustrations as carriers of cultural anchoring in children's picture books?

By integrating computational visual metrics with iconographic analysis, this experimental study offers an empirically grounded account of how generative image technologies, as instantiated in a contemporary text-to-image model, reshape the balance between perceptual fluency and cultural semiosis in children's illustrated narratives.

2. Literature Review

We synthesize work in social semiotics, perceptual psychology, computer vision, iconology, and generative AI to ground three strands of inquiry: form (computable visual properties), symbol (iconographic fidelity), and practice (how AI and handcraft interact). Selection prioritizes authoritative sources with reproducible procedures to ensure that later comparisons rest on evaluable assumptions.

2.1. Formal Visual Features and Computable Metrics

In picture books, visual form guides attention, configures viewer relationships, and establishes hierarchies of saliency (Kress & van Leeuwen, 2006; Kress, 2010). For young readers, images often scaffold story comprehension before textual fluency develops (Nodelman, 1988; Nikolajeva & Scott, 2001). Perceptual psychology indicates that observers prefer curved forms over angularity, perceiving smooth contours as more fluent and visually coherent (Attneave, 1954; Bar & Neta, 2006; Bertamini et al., 2016). Empirical

studies on visual curvature consistently demonstrate its association with aesthetic preference across age groups, supporting the use of computable curvature-based features to quantify formal visual appeal (Chuquichambi et al., 2022). Color distributions also matter; information-theoretic measures such as Shannon entropy capture palette richness and diversity in a culture-neutral manner (Palmer & Schloss, 2010; Shannon, 1948; Gonzalez & Woods, 2018).

To generate testable evidence, we treat form as computable. Edge curvature summarizes local smoothness and directional change along detected contours, capturing line “flow” more sensitively than raw edge counts; color entropy estimates chromatic diversity from binned distributions. Images are denoised, de-screened, white-balanced, and resolution-normalized; edges and histograms are computed in OpenCV (Bradski, 2000). Entropy is computed in HSV and cross-checked in CIELAB to account for perceptual non-uniformity (Fairchild, 2013), and paired statistics with effect sizes are reported (Cohen, 1988). Although composition and interaction have been richly theorized, quantitative evidence for perceptually salient properties remains limited, motivating paired comparisons of AI-generated versus handcrafted illustrations.

2.2. Iconology and Canonical Symbolism in *Journey to the West*

Formal properties alone do not explain cultural meaning. Iconology links surface description to conventional themes and deeper worldviews (Panofsky, 1939/1972; Panofsky, 1955; Gombrich, 1960). In picture-book settings, recurrent cues such as costume, attributes, physiognomy, and ritual colors stabilize interpretation and support cultural learning (Nodelman, 1988; Nikolajeva & Scott, 2001). *Journey to the West* provides a clear iconographic system: Tang Monk’s Vairocana crown, nine-ringed staff, and kasaya index sanctity; Sun Wukong’s golden cudgel, crown, and tiger-skin skirt encode martial agency; Zhu Bajie’s nine-tooth rake and porcine features signal comic excess; Sha Wujing’s monk’s attire and implement symbolize endurance and service.

A 4×4 canonical codebook defines states for each protagonist (appearance, crown, clothes, weapon), with examples and decision rules for stylized versus literal renderings. Annotation is conducted using the UAM Image Tool (O’Donnell, 2008) with inter-rater reliability established using Cohen’s κ (target ≥ 0.80) and adjudication where necessary (Cohen, 1960; Landis & Koch, 1977; McHugh, 2012). This allows reproducible counts of symbol preservation, simplification, or weakening in AI-generated imagery relative to handcrafted illustrations.

2.3. Generative AI for Text-to-Image Illustration: Mechanisms, Bias, and Iconographic Risk

Text-to-image models are socio-technical systems that generate imagery by conditioning on language embeddings and progressively denoising latent representations (Rombach et al., 2022). Training on web-scale datasets such as LAION-5B provides broad coverage but introduces statistical biases and cultural imbalances (Schuhmann et al., 2022). Audits reveal that demographic and cultural biases persist across widely used models such as Stable Diffusion and Midjourney (Elsharif et al., 2025; Wiegand et al., 2025), with overrepresentation of frequent motifs and underrepresentation of culturally specific elements.

Cultural fidelity can be quantified using metrics such as the Component Inclusion Score (CIS), showing that non-Western prompts often result in degraded symbolic detail (Said et al., 2025; Seo et al., 2025). Mechanistically, guidance strength and decoder priors can regularize contours and compress color variance, enhancing perceptual fluency but reducing color entropy and low-frequency symbolic elements. Community prompt conventions further normalize composition. In culturally dense domains like *Journey to the West*, these mechanisms create an iconographic risk: optimizing formal features may

erode or simplify culturally anchored signs, compromising narrative and semiotic integrity. Evaluating formal metrics alongside symbol preservation is therefore essential for understanding T2I representational limitations.

Recent research has begun to analyze how generative AI models perform in contexts where illustrations must carry sequential narrative meaning. Lin et al. (2026) provide a comprehensive survey on AI-generated storybook illustrations, highlighting challenges with temporal, spatial, character, plot, style, and thematic consistency across images, rather than only single-image fidelity. Similarly, Xu and Deng (2024) explore the use of generative models such as Midjourney in creating digital children's picture books, examining narrative coherence, color language, and immersive reading effects. Frameworks such as Story-Adapter (Mao et al., 2024) focus on long-story visualization by refining semantic consistency across multi-image sequences, and ID.8 (Antony & Huang, 2024) investigates co-creation systems that enable generative AI to produce visual stories in collaboration with human authors. Importantly, Jin et al. (2025) offer empirical insights into parent and child responses to AI-generated story images, showing how factors such as style, authenticity, and safety influence user reception in book contexts.

2.4. Synthesis, Gaps, and Research Focus

Across the strands reviewed, several actionable gaps emerge. First, regarding form, although theory explains why contour smoothness and palette diversity matter, quantitative and reproducible evidence remains limited. Paired computational analyses of edge curvature and color entropy can provide more robust empirical support. Second, with respect to symbols, existing iconographic studies rarely examine canonical symbols with reliable annotation protocols; systematic coding with inter-rater agreement ensures that analyses are reproducible and comparable across studies. Third, in terms of practice, few studies have assessed how AI-generated images interact with culturally saturated narratives, particularly regarding semiotic fidelity and symbolic hierarchy. This indicates the need for approaches that integrate the perceptual fluency of AI-generated visuals with the cultural and symbolic depth maintained in handcrafted illustrations. Taken together, these gaps motivate a careful comparison between handcrafted and AI-generated illustrations, examining both formal visual metrics and iconographic fidelity to better understand the semiotic and aesthetic implications of generative image technologies.

3. Research Design

3.1. Selection of Tools

This study employs Doubao, a multimodal artificial intelligence platform developed by ByteDance, to generate illustrations for the *Journey to the West* children's picture book. Doubao was selected for its strong cultural contextualization in Chinese-language environments, its support for bilingual (Chinese–English) prompts, and its ability to efficiently generate high-resolution images using the Seedream 3.0 diffusion model. Prior studies have reported that English-dominant image generation platforms, such as Stable Diffusion or DALL·E, often struggle to accurately render culturally embedded visual elements from non-Western literary traditions (Zhang et al., 2024).

The Doubao platform enables fine-grained narrative control through Chinese-language prompt engineering and integrates advanced diffusion modeling with natural language understanding to process detailed textual inputs. For image analysis, OpenCV was used to extract edge and color visual features, while UAM Image Tool 2.1 was employed for multilayer iconographic annotation and semantic analysis. Annotation outputs were exported in XML format to facilitate further data processing. Quantitative comparisons between handcrafted and AI-generated illustrations focus on edge curvature and color

entropy, two metrics widely used to capture structural and visual complexity in object recognition and visual perception studies (Chuquichambi et al., 2022).

3.2. Text Selection

The selected corpus is *The Monkey King* series published by China Youth Publishing House, consisting of 34 volumes with a cumulative circulation of approximately 1.38 million copies. Illustrated by 34 established Chinese artists, the series contains 2103 full-color illustrations noted for their visual expressiveness and close alignment with the original narrative. The accompanying text balances classical vernacular with accessibility for children aged 5–8, making it suitable for early engagement with traditional Chinese culture. This edition was chosen because its consistently detailed illustrations and clearly articulated narrative descriptions provide a stable basis for AI illustration generation and systematic comparison with handcrafted artwork.

To ensure representative sampling, illustrations within each volume were sequentially numbered, and every sixth image was selected, yielding seven images per volume and a total of 224 handcrafted illustrations. Because picture-book illustrations are closely aligned with narrative progression, this systematic interval sampling covers major narrative phases, including exposition, development, climax, and resolution, thereby approximating a comprehensive visual representation of the story world. Each handcrafted illustration was paired with exactly one AI image generated by Doubao through a one-to-one procedure, using the image-aligned narrative text as the shared semantic anchor for generation and evaluation.

3.3. Computational Metrics: Edge Curvature and Color Entropy

This section specifies the mathematical definitions, computational procedures, and parameter settings for the two visual metrics used in this study: edge curvature and color entropy. All computations were implemented in Python (version 3.11) using OpenCV (v4.8.0) and NumPy (v1.24.3). Mathematical formulas are reported in the main text, while parameter values are organized by functional role and summarized in Tables A1 and A2 (Appendix A).

3.3.1. Edge Curvature

Edge curvature was used to quantify local directional variation along object boundaries. This metric captures micro-level contour bending rather than global shape properties, enabling fine-grained comparison between the handcrafted image and the AI image generated by Doubao. The computation consists of three stages: preprocessing and contour extraction, point-wise curvature estimation, and image-level aggregation.

Images were converted from RGB to grayscale and smoothed using a Gaussian filter with a 5×5 kernel. Edge maps were extracted using the Canny detector (Canny, 1986) with threshold values of 50 and 150. External contours were retrieved using the Suzuki–Abe algorithm (Suzuki & Abe, 1985) in RETR_EXTERNAL mode, retaining only outermost object boundaries. Gaussian smoothing and Canny thresholds function as preprocessing-level parameters, whereas contour length filtering and curvature window size operate at the geometric computation level (see Table A1, Appendix A).

Curvature was computed for individual contour points using a discrete formulation. For a given point $\mathbf{P}_i$, a symmetric neighborhood of $k = 5$ points was defined, and two vectors were constructed:

$$\mathbf{v}_1 = \mathbf{P}_i - \mathbf{P}_{i-k}, \mathbf{v}_2 = \mathbf{P}_{i+k} - \mathbf{P}_i.$$

The turning angle θ between the vectors was calculated as

$$\theta = \arccos \left(\frac{\mathbf{v}_1 \cdot \mathbf{v}_2}{\|\mathbf{v}_1\| \|\mathbf{v}_2\|} \right).$$

Directional information was preserved by assigning a sign based on the z-component of the cross product:

$$\text{sign} = \text{sgn}(v_{1x}v_{2y} - v_{1y}v_{2x}).$$

Signed curvature was then defined as

$$\kappa_i = \text{sign} \cdot \frac{\theta}{(\|\mathbf{v}_1\| + \|\mathbf{v}_2\|)/2}.$$

Curvature values were computed only for contours with a minimum length of 20 points to avoid unstable estimates. An image-level mean curvature value (κ_{mean}) was obtained by averaging all valid κ_i values across contours. Positive values indicate locally convex bending, negative values indicate concave indentation, and values near zero correspond to approximately linear segments. Because directional information is retained, negative mean curvature values reflect the predominance of concave contour configurations rather than computational error. All parameters are summarized in Table A1 (Appendix A).

3.3.2. Color Entropy

Color entropy was used to quantify the richness and distributional diversity of color usage. The metric is based on Shannon entropy (Shannon, 1948) and was computed primarily in the HSV color space, which separates chromatic information from luminance and is well suited to perceptual color analysis.

Images were converted from RGB to HSV using OpenCV's standard conversion. Histograms were computed for the hue (H), saturation (S), and value (V) channels. The H channel was discretized into 180 bins (0 – 179°), and the S and V channels into 256 bins each (0 – 255). Histogram counts were normalized to obtain probability distributions $p_{(c,b)}$.

Shannon entropy for each channel was calculated as

$$E_c = - \sum_{b=1}^{B_c} p_{(c,b)} \log_2 p_{(c,b)},$$

where B_c denotes the number of bins for channel c . Overall color entropy was defined as

$$E_{\text{color}} = \frac{E_H + E_S + E_V}{3}.$$

Higher entropy values indicate a more diversified and uniformly distributed color palette, whereas lower values suggest chromatic concentration or simplification. To assess robustness, color entropy was additionally computed in the CIELAB color space using a D65 standard illuminant. Entropy values derived from HSV and CIELAB were highly correlated (Pearson's $r = 0.94$), indicating that group differences were not artifacts of color space selection. All parameters are summarized in Table A2 (Appendix A).

3.4. Image Generation

All selected images were preserved in their original dimensions and saved as PNG files to maintain visual fidelity and metadata for subsequent analysis. Figure 1 illustrates the overall operational workflow of the study, from image selection to AI generation and subsequent analysis.

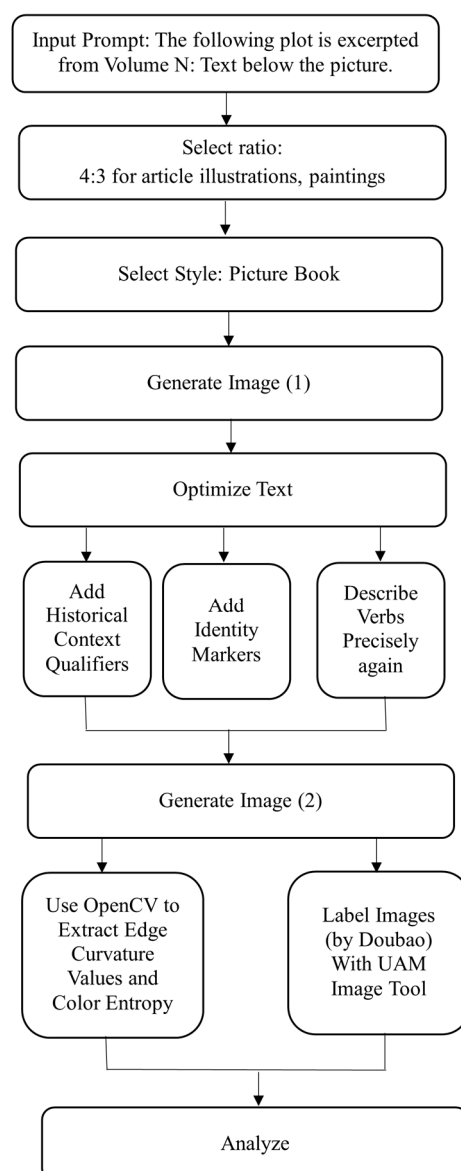


Figure 1. Operational workflow of the study from image selection to AI images generated by Doubao and multimodal analysis.

To ensure consistency and comparability across AI images generated by Doubao, a standardized generation protocol was implemented using the Doubao platform. Pairing between handcrafted and AI images generated by Doubao was strictly narrative-based: each AI image was generated to represent the same narrative episode as its handcrafted counterpart, without replicating composition, perspective, or stylistic features. The Seedream 3.0 model, a fixed 4:3 aspect ratio (matching the original layout), and the picture-book style were held constant throughout the generation process. Because lower-level generative parameters (e.g., sampler, step count, random seed) are internally optimized by the platform and not user-accessible, textual prompts constituted the primary controlled variable.

For each illustration, the initial prompt consisted solely of the narrative text associated with the handcrafted image and served as the semantic anchor for generation. Each prompt produced one to four candidate images, which were evaluated against the corresponding handcrafted illustration and the source episode, with particular attention to core characters, actions, and situational context. If no substantial narrative deviation was

observed, the prompt was retained without modification. Prompt optimization was applied only when clear deviations occurred and followed three constrained principles: (1) temporal contextualization, specifying the timing of events when implicit in the text (van Dijk, 2009); (2) role-specific labeling, clarifying character roles or hierarchies when ambiguity affected visual realization (van Leeuwen, 2022); and (3) elaborative verb description, refining action-related verbs to reduce exaggeration or misinterpretation of gestures (Markova, 1978). These adjustments aimed solely to reduce narrative misalignment rather than to enhance aesthetic quality or introduce stylistic variation. One AI image generated by the Doubao platform was ultimately selected for each episode based on narrative correspondence and the absence of major distortions. An example of prompt adjustment is provided in Appendix B.

Following image generation, iconographic annotation was conducted using UAM Image Tool 2.1, with handcrafted illustrations serving as the coding reference. Annotation focused on the four pilgrims, Tang Monk, Sun Wukong, Zhu Bajie, and Sha Wujing, and covered appearance, headwear, clothing, and weapons. The classification criteria are presented in Table 1. Coding was performed by four postgraduate linguistics students familiar with *The Monkey King* series, guided by a shared codebook derived from the original illustrations. Coding decisions were discussed in regular meetings, with disagreements resolved through consensus and consultation with a senior multimodal-studies scholar. No AI-assisted or automated annotation was employed.

Table 1. Iconographic coding criteria for the four pilgrims' visual features used in UAM Image Tool-based annotation.

Characters Features	Tang Monk	Sun Wukong	Zhu Bajie	Sha Wujing
Appearance	Handsome and Elegant	Thin; Golden Eyes; Golden Hair	Big Belly and Round Waist; Big Ears	Burly; Fluffy Red Hair
Crown	Vairocana Crown	Golden Hoop	Discipline Hoop	Silver Hoop
Clothes	Sacred Brocaded Kasaya	Tiger-Skin Skirt	Blue Brocade Kasaya	Zhiduo Robe
Weapon	Nine-Ringed Tin Staff	Ruyi Golden Cudgel	Nine-Toothed Rake	Sal Tree Staff

By combining quantitative visual metrics (edge curvature and color entropy) with semiotic annotation, this study establishes an integrated analytical framework for examining how AI images generated by Doubao reproduce or deviate from traditional visual conventions in children's picture books. The procedures described above, including metric computation, statistical testing, and iconographic annotation, constitute a complete and replicable data analysis workflow.

4. Results

4.1. Results of Edge Curvature and Color Entropy for AI Images Generated by Doubao Versus Handcrafted Images

4.1.1. Edge Curvature Results

We analyzed 224 image pairs using edge curvature as a quantitative metric. The handcrafted group showed an average curvature of -0.001162 ($SD = 0.002359$), while the AI images generated by Doubao averaged 0.000061 ($SD = 0.001952$). A negative curvature indicates locally concave contours, and a positive value indicates convex contours. The

average negative curvature of the handcrafted images suggests their contours contain more inward-concave detail variations, reflecting the natural undulation of handcrafted strokes. In contrast, the curvature of the AI images generated by Doubao is near zero, indicating smoother, more regular edges with a lack of the subtle concave–convex variations characteristic of handcrafted work.

Shapiro–Wilk tests for normality were performed on the average curvature of each group. Results showed the handcrafted group did not follow a normal distribution ($p = 0.000038$), while the AI images generated by Doubao approximated normality ($p = 0.116000$). Therefore, the Mann–Whitney U non-parametric test was used for between-group analysis. The test indicated a statistically significant difference ($p < 0.001$), confirming distinct distributions of edge curvature between handcrafted images and AI images generated by Doubao.

To illustrate how these quantitative results translate into perceptible visual styles, two representative image pairs, shown in Figure 2, were selected for case analysis. This case is from *The Monkey King* series, Book 17, Image 6, titled *The Real and Fake Monkey Kings*. The corresponding text reads as follows: “The old man invited Tang Monk into the thatched cottage, cooked a meal for him and his disciples. As night fell, he lit a lamp and chatted with Tang Monk in the cottage. The old man mentioned his unfilial son was a robber. Tang Monk was so frightened and unsettled, fearing it might be one of the two bandits Wukong had killed.”

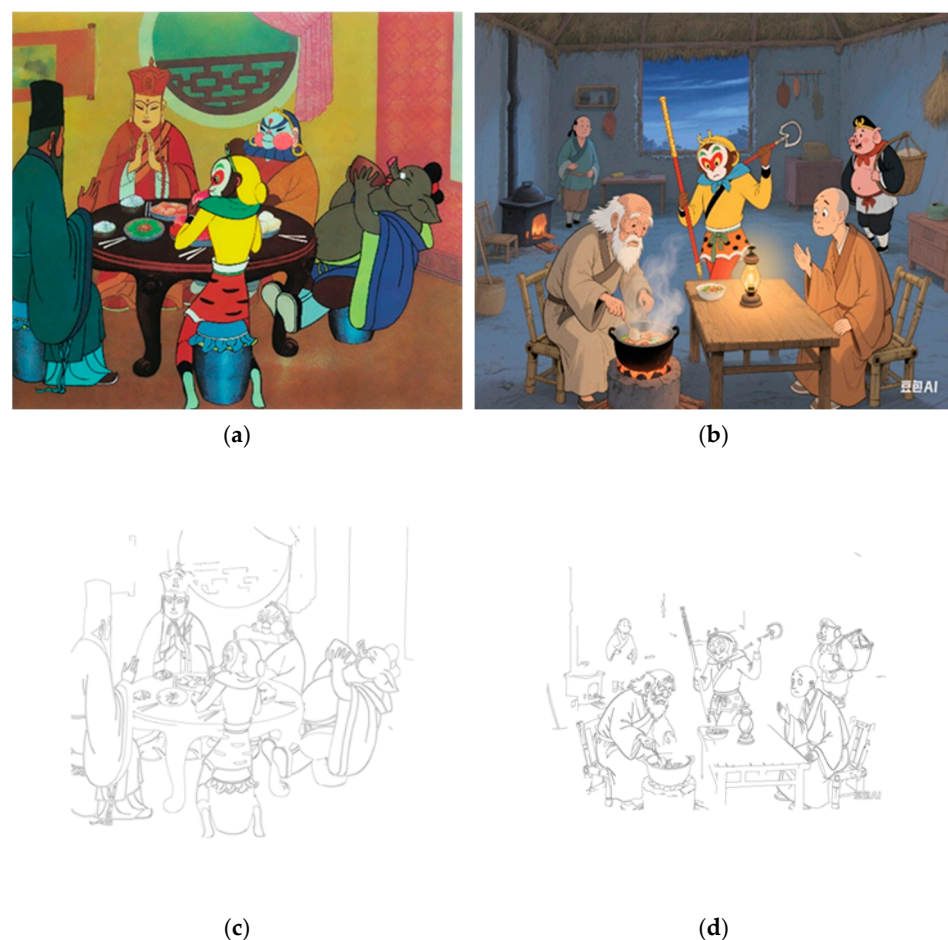


Figure 2. Edge detection for a case study image pair (*The Real and Fake Monkey Kings*). (a) Handcrafted image. (b) AI image generated by Doubao. (c) Edge curvature detection of handcrafted image. (d) Edge curvature detection of AI image generated by Doubao.

Table 2 reveals an edge curvature of -0.003553412 for the handcrafted version, compared to 0.00228004 for its counterpart AI image generated by Doubao. This reflects a fundamental stylistic divergence: negative curvature, indicative of concave contours, is associated with handcrafted “tapering” strokes that convey tension and restraint, aligning with the scene’s narrative unease. In contrast, the AI (Doubao) image’s positive curvature produces consistently convex, rounded contours, resulting in a visually smoother, more expansive, and relaxed aesthetic.

Table 2. Mean edge curvature values for handcrafted images and AI images generated by Doubao.

Metric	Value (Handcrafted Images)	Value (AI Images Generated by Doubao)
Edge curvature	-0.003553412	0.00228004

Note: Negative values indicate concave contours, whereas positive values indicate convex contours.

The handcrafted image, shown in Figure 2c, uses sharp, inward-turning lines for the monk’s robe folds, table legs, and window frames, creating dramatic tension that mirrors the text’s suspense. However, the AI image generated by Doubao, seen in Figure 2d, employs soft, arcing lines for facial features, clothing, and objects like the pottery jar and thatched roof. This pervasive rounding fosters a calmer, more harmonious atmosphere, diluting the original narrative’s inherent disquiet. Thus, the curvature shift from negative to positive transcends a mere numerical difference. It signifies a deeper divergence in generative logic: handcrafted artistry leverages nuanced stroke variation to embed emotional and narrative depth, whereas the AI image generated by Doubao prioritizes formal regularity and perceptual fluency, optimizing for visual comfort over expressive semantic load.

4.1.2. Color Entropy Results

Color entropy in the HSV space was calculated to assess color complexity. The handcrafted group exhibited a higher average entropy ($M = 6.685$, $SD = 0.415$) than the AI images generated by Doubao ($M = 6.497$, $SD = 0.345$). A Mann–Whitney U test confirmed that this difference was significant ($p < 0.001$). Shapiro–Wilk tests indicated non-normality for the handcrafted group ($p < 0.001$) and near-normality for the AI (Doubao) group ($p = 0.059$).

To illustrate how these entropy differences manifest visually, a case analysis, presented as Figure 3, was conducted using Image 18 from Volume 12, titled *Bianfa Battles the Three Immortals*. The corresponding text reads as follows: “The three sat down and began eating heartily. Steamed buns, rice cakes, and fresh fruit were all swept clean. The table was cleared in an instant. After eating, they sat and chatted to aid digestion.”



(a)



(b)

Figure 3. A case study image pair (*Bianfa Battles the Three Immortals*). (a) Handcrafted image. (b) AI image generated by Doubao.

After computing the HSV values with OpenCV, the respective color histograms of the two images were produced, shown in Figure 4.

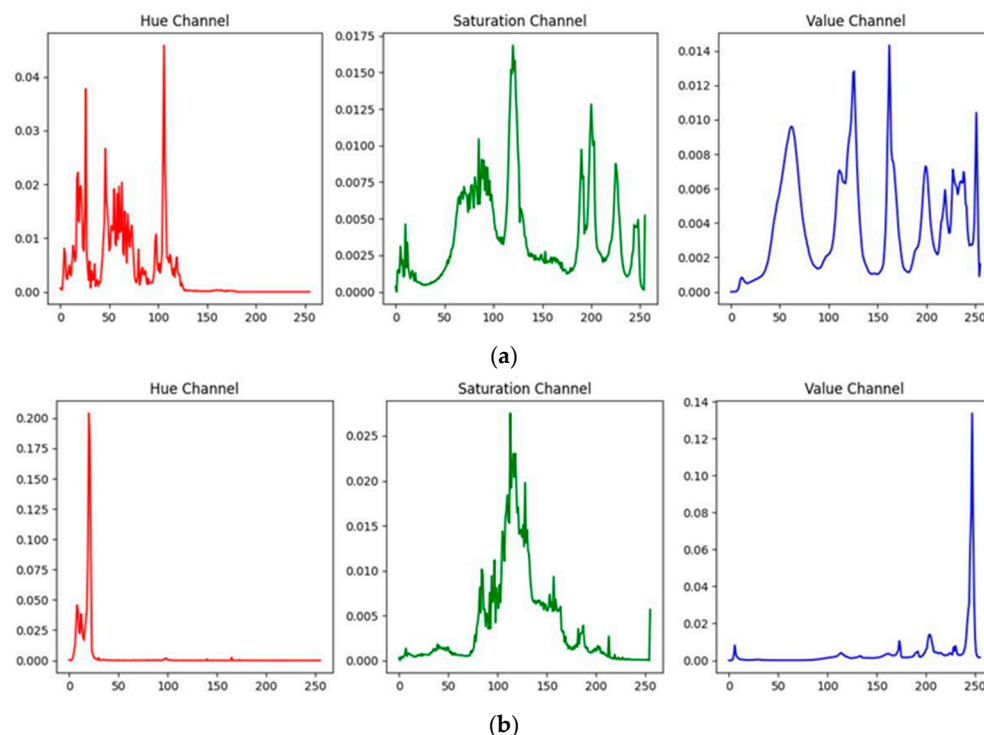


Figure 4. Color entropy and HSV histograms for a case study image pair (*Bianfa Battles the Three Immortals*). (a) Color histogram of handcrafted image. (b) Color histogram of AI image generated by Doubao.

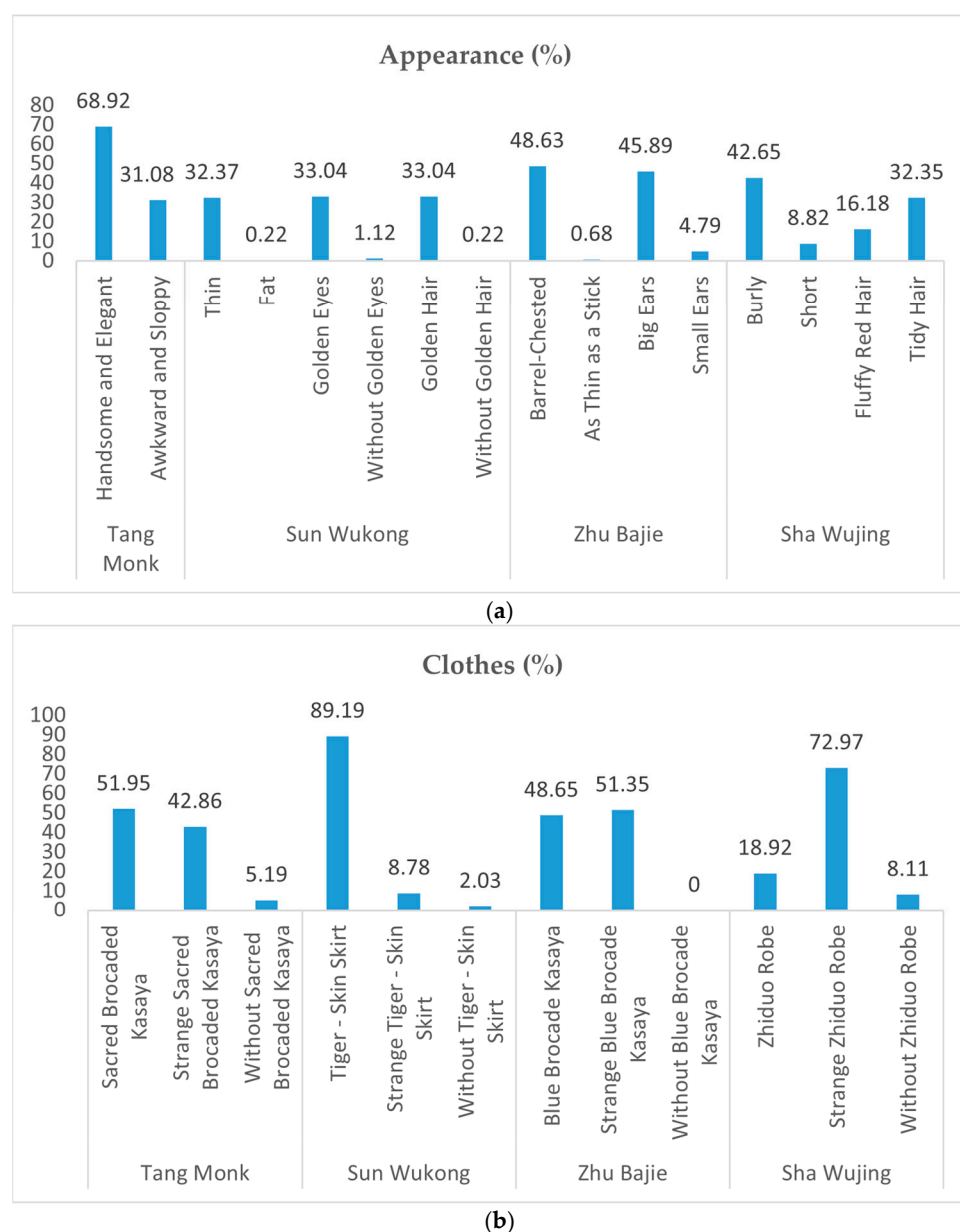
The handcrafted image achieved a color entropy of 5.0095, significantly higher than the AI (Doubao)-generated image's 3.9327. The corresponding histograms reveal a more varied distribution across hue, saturation, and value channels in the handcrafted illustration, indicating richer color combinations and stronger tonal contrast. The AI image generated by Doubao shows more constrained distributions, with a dominant hue peak and a concentration of high-brightness pixels.

A detailed visual analysis of the histograms (Figure 3) elucidates this contrast. The handcrafted work (Figure 3a) features a highly varied color distribution in the HSV space. In the hue channel, multiple spikes appear across vibrant hues such as bright yellow, vivid red, deep purple, and muted orange, indicating rich and natural color combinations. The saturation channel spans both intense tones (e.g., the deep indigo of Zhu Bajie's robe) and soft, subtle shades (e.g., the warm orange of Sha Wujing's attire), creating artistic depth and emotional layering. The value channel displays strong contrast between highlights (e.g., the bright yellow of Sun Wukong's outfit) and dark shadows (e.g., the deep-green background), enhancing spatial dimensionality. In contrast, the AI image generated by Doubao (Figure 3b) exhibits a single dominant peak in the hue channel, centered on soft warm yellows and light browns, indicating a limited hue range. Its saturation histogram remains constrained with lower chromatic richness. The value channel shows a pronounced peak in the high-brightness range with fewer dark pixels, resulting in weaker tonal contrast.

The higher and more variable entropy in handcrafted images suggests greater creative freedom and complexity in color application, whereas the AI images generated by Doubao demonstrate more uniform and statistically regular color patterns.

4.2. Character Feature Annotation in AI Images Generated by Doubao

To support subsequent multimodal analysis, categorical annotation was conducted on AI images generated by Doubao using the UAM Image Tool. The analysis focused on the four core characters in *Journey to the West*: Tang Monk, Sun Wukong, Zhu Bajie, and Sha Wujing. Each character was annotated across four dimensions: appearance, crown (headwear), clothing, and weapon. The distribution of annotated visual features is presented in Figure 5.



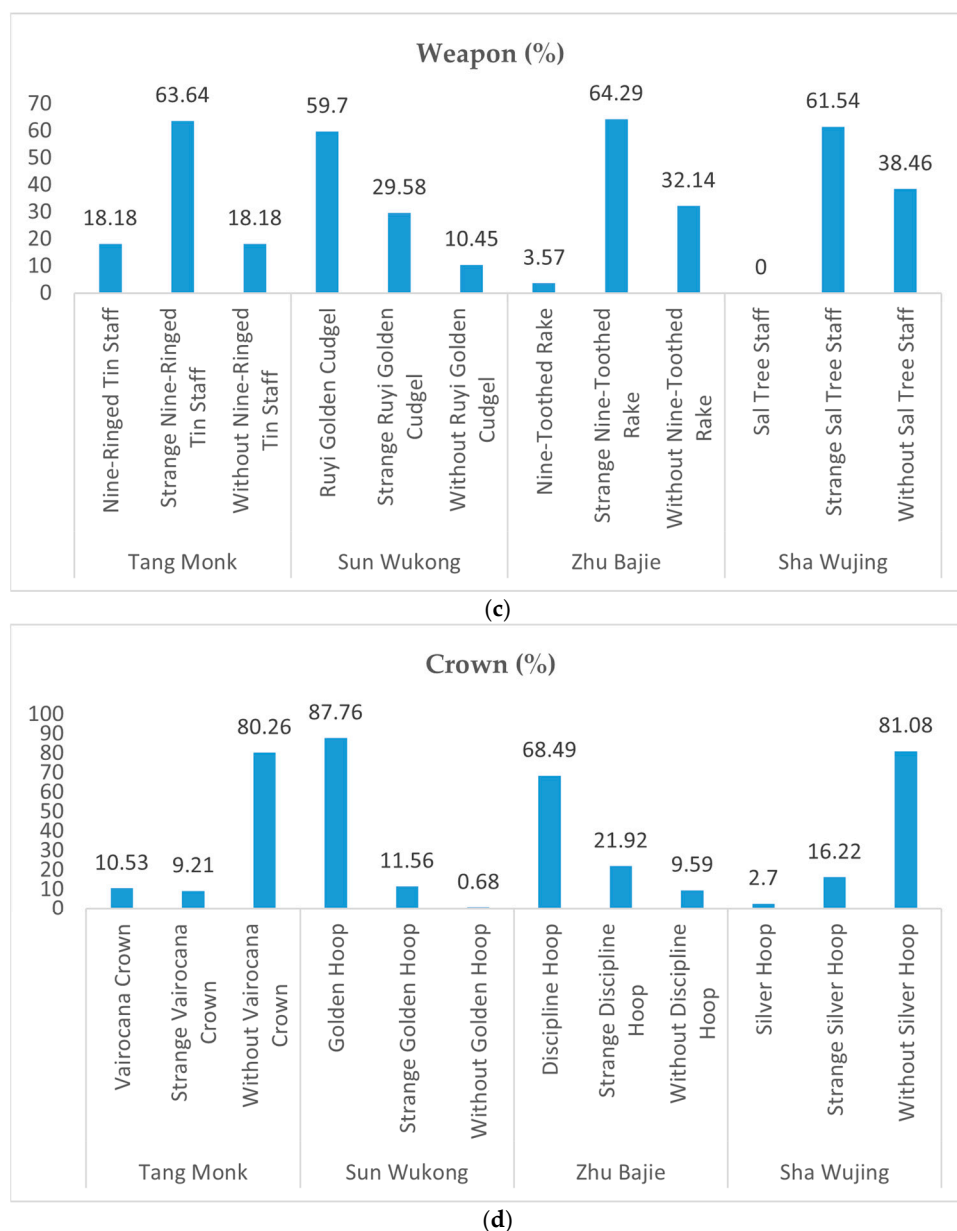


Figure 5. Distribution of annotated visual features in AI images generated by Doubao based on UAM Image Tool coding. **(a)** Appearance of annotated visual features in AI images generated by Doubao based on UAM Image Tool coding. **(b)** Clothes of annotated visual features in AI images generated by Doubao based on UAM Image Tool coding. **(c)** Weapon of annotated visual features in AI images generated by Doubao based on UAM Image Tool coding. **(d)** Crown of annotated visual features in AI images generated by Doubao based on UAM Image Tool coding.

The corresponding chi-square goodness-of-fit test results are shown in Figure 6.

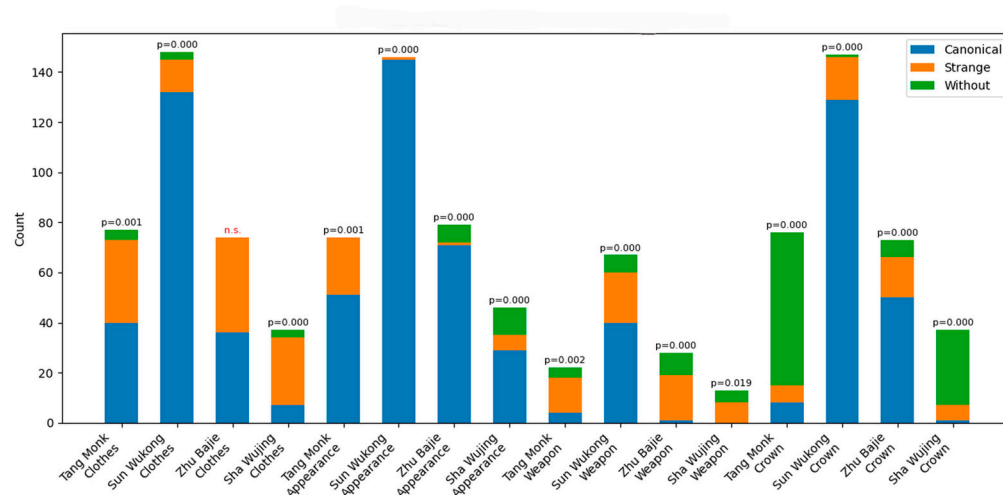


Figure 6. Chi-square goodness-of-fit test results for annotated character features in AI images generated by Doubao. Note: “n.s.” means not significant.

For primary characters, Tang Monk and Sun Wukong, the AI images generated by Doubao consistently reproduced canonical features across all four dimensions. For instance, Tang Monk’s “Sacred Brocaded Kasaya” ($\chi^2(2) = 13.77$, $p = 0.001$), “Handsome and Elegant” appearance ($\chi^2(1) = 10.59$, $p = 0.001$), “Nine-Ringed Tin Staff” ($\chi^2(2) = 13.09$, $p = 0.0015$), and “Vairocana Crown” ($\chi^2(2) = 59.02$, $p < 0.001$) were all significantly more likely to appear than alternate or absent variants. Similarly, Sun Wukong’s canonical clothing, facial features (golden eyes and golden hair), weapon, and crown were reproduced with high significance ($p < 0.001$). These results quantitatively confirm that the generative model preserves highly salient visual symbols associated with central protagonists.

In contrast, secondary characters such as Zhu Bajie and Sha Wujing display more heterogeneous patterns. Notably, Zhu Bajie’s clothing shows no significant difference between canonical and variant kasayas ($\chi^2(2) = 0.95$, $p = 0.622$), reflecting an absence of consistent visual emphasis in this category. However, other features for Zhu Bajie, including appearance, weapon, and crown, are significantly reproduced ($p < 0.001$), indicating selective preservation of iconic elements. Sha Wujing similarly shows significant reproduction for all annotated features, indicating consistent preservation of canonical features across all dimensions.

Overall, these results demonstrate a structural bias in the generative process: AI imagery generated by Doubao privileges primary characters with statistically significant reproduction of canonical features, while secondary characters exhibit less consistent visual emphasis, as supported by chi-square tests.

5. Discussion

5.1. Enhanced Formal Fluency and the Weakening of Cultural Semiosis

Our analysis reveals a fundamental formal shift: AI images generated by Doubao have significantly smoother edges than handcrafted ones. Statistically, the mean edge curvature shifts from negative (handcrafted: $M = -0.001162$) to positive (AI images generated by Doubao: $M = 0.000061$) (Mann–Whitney U, $p < 0.001$). Positive curvature indicates predominantly convex, rounded contours, whereas negative curvature reflects the concave details and angular inflections of handcrafted strokes, which create visual tension and expressive nuance (Bertamini et al., 2016).

From a perceptual-psychology perspective, this tendency can be interpreted as aligning with children’s aesthetic preferences in early and middle childhood. Empirical studies show that children are highly sensitive to contour continuity and generally prefer curved

forms to angular ones, associating smoothness with comfort and affective safety (Bar & Neta, 2006; Zhang et al., 2024). Accordingly, the convex and continuous contours in the AI images generated by Doubao enhance perceptual accessibility. They facilitate scene entry and guide attention, thereby supporting narrative comprehension at the level of visual processing.

However, perceptual fluency does not necessarily entail cultural or semantic richness. In picture-book studies, narrative meaning emerges from the interaction between visual form and verbal structure rather than from visual smoothness alone (Nikolajeva & Scott, 2001). From this perspective, the regularized contours observed in AI images generated by Doubao enhance visual accessibility but reduce the expressive irregularities through which cultural nuance and narrative tension are traditionally encoded. Visual-literacy research further demonstrates that children interpret images in close relation to textual cues and culturally established conventions (Kress & van Leeuwen, 2006; Arizpe & Styles, 2016). Formal properties such as contour regularity therefore function primarily as perceptual enablers rather than as carriers of cultural meaning.

Taken together, the curvature shift identified in this study suggests that AI images generated by Doubao privilege perceptual fluency over semantic density. Although learning outcomes are not assessed directly, these structural differences help explain why visually fluent AI images generated by Doubao may provide fewer semiotic resources for conveying culturally dense narratives.

A similar divergence emerges in chromatic organization. Color entropy analysis indicates that handcrafted illustrations ($M = 6.685$) employ richer and more differentiated palettes than AI images generated by Doubao ($M = 6.497$). High color entropy increases visual stimulation and serves as a carrier of aesthetic meaning. In traditional Chinese illustration, color functions hierarchically rather than decoratively. By contrast, AI images generated by Doubao tend toward palette homogenization. They favor high brightness and saturation while reducing chromatic differentiation, resulting in lower entropy values.

This homogenization has direct semiotic consequences. In the handcrafted Figure 7a, Tang Monk's brocaded kasaya features traditional layered reds, golds, and metallic threads to signify Buddhist sanctity and imperial legitimation. His Vairocana crown further marks his religious identity and authority (Wu, 2019). In the AI image generated by Doubao (Figure 7b), this garment is simplified, recolored, or omitted. Tang Monk's Vairocana crown, central to Tang Monk's identity, also disappears. As a result, its ritual and ideological significance is weakened. Culturally saturated symbols are thus transformed into visually appealing but semantically flattened motifs.

On this basis, formal analysis points to a clear conclusion. Although the AI images generated by Doubao achieve baseline visual appeal through perceptual optimization, they demonstrate a measurable deficit in cultural–semantic transmission. When surface-level stimulation supersedes the symbolic function of traditional aesthetic codes, the educational role of picture-book illustration is fundamentally weakened.



Figure 7. A comparative case study of Tang Monk's iconic features. (a) Handcrafted image. (b) AI image generated by Doubao.

5.2. Frequency-Driven Iconographic Simplification in AI Images Generated by Doubao

Beyond formal properties, the iconographic analysis reveals a patterned imbalance in how AI images generated by Doubao reproduce canonical cultural symbols. This imbalance is not impressionistic but systematically reflected in the quantitative annotation of character attributes and ritual symbols across the AI images generated by Doubao.

For characters whose identities rely on highly iconic, contour-salient features, symbolic encoding remains relatively stable. Sun Wukong exemplifies this tendency, shown in Figure 8. His core visual markers, such as the golden hoop (87.76%), tiger-skin skirt (89.19%), and Ruyi golden cudgel (59.7%), are consistently retained at high rates. These attributes are both visually distinctive and frequently represented across cultural media, making them well aligned with the model's optimization toward salient shapes and repeated visual templates. As a result, Sun Wukong's narrative recognizability and symbolic coherence are largely preserved.



Figure 8. Stable reproduction of Sun Wukong's iconic features in AI image generated by Doubao.

Zhu Bajie displays a more selectively weighted pattern. Features that support immediate character recognition, particularly bodily configuration and head-related markers, remain relatively stable across AI images generated by Doubao (e.g., big belly and round waist: 48.63%; discipline hoop in canonical or variant form: 90.41%). Weapon presence also shows partial retention (nine-toothed rake present in some form: 67.86%). In contrast, clothing functions as a weakly encoded attribute. The near-even distribution between canonical (48.65%) and variant kasaya forms (51.35%) is illustrated in Figure 9a (canonical) and Figure 9b (variant). In Figure 9a, Zhu Bajie is depicted in a black monk's robe with a red inner layer and a yellow belt, closely resembling the canonical kasaya from *The Monkey King* series. In

contrast, Figure 9b shows him in a blue top and an orange apron, a costume more akin to that of a warrior or laborer, which deviates significantly from the solemn form of the traditional kasaya, representing a clear variant. This near-canonical yet inconsistently altered clothing visually confirms the nearly even statistical distribution between canonical and variant kasaya forms and indicates that clothing carries limited representational priority in the model's construction of Zhu Bajie's visual identity.



Figure 9. Weak encoding of Zhu Bajie's iconic clothing features in AI images generated by Doubao. (a) AI image generated by Doubao of Zhu Bajie in canonical kasaya. (b) AI image generated by Doubao with altered kasaya.

More pronounced instability emerges in characters whose symbolic systems depend on relational, ritualized, or low-frequency cues. Sha Wujing illustrates this vulnerability most clearly. His defining ritual markers, such as the silver hoop and sal tree staff, are predominantly omitted or rendered in altered forms (e.g., silver hoop absent: 81.08%; sal tree staff absent or distorted: 100% combined; strange zhiduo robe: 72.97%). Although these distributions are statistically non-uniform, the bias is directed toward absence rather than canonical realization. This pattern reflects systematic symbolic erosion rather than successful reproduction, underscoring the difficulty of encoding culturally specific but visually non-redundant elements. For instance, in the AI images generated by Doubao, shown in Figure 10a,b, Sha Wujing's clothes are simplified into plain attire without religious attributes. In Figure 10a, his weapon is missing and he is depicted without hair. In Figure 10b, he holds a strange sal tree staff and has sparse black hair.



Figure 10. Omission and distortion of Sha Wujing's iconic features in AI images generated by Doubao. (a) AI image generated by Doubao missing the weapon and hair. (b) AI image generated by Doubao featuring a distorted staff and strange hair.

Tang Monk's case further illustrates the limits of AI images generated by Doubao. Although his monastic clothing and staff are often included, the highly symbolic Vairocana

crown is systematically absent, and his visual identity fluctuates between realist and cartoon conventions, shown in Figure 11a,b. Quantitative annotation confirms this pattern: the Vairocana crown appears in only 10.53% of AI images generated by Doubao, while 80.26% omit it entirely; the sacred brocaded kasaya is either altered (42.86%) or absent (5.19%), indicating a systematic erosion of ritual markers rather than random variation. This fragmentation reduces a cohesive symbolic system to isolated visual cues, weakening his religious authority and moral gravitas at the iconographic level, and ultimately disrupting the iconological logic through which Tang Monk functions as a moral and ritual center within the pilgrimage narrative. More importantly, it reflects a failure to encode the hierarchical logic through which traditional illustrations integrate costume, color, and attribute into a unified cultural sign.



Figure 11. Instability and symbolic loss of Tang Monk’s iconic features in AI images generated by Doubao. (a) A realist-style image of Tang Monk generated by Doubao. (b) A cartoon-style image of Tang Monk generated by Doubao.

These patterns point to a frequency-sensitive iconographic bias in AI imagery generated by Doubao. Symbols with high visual redundancy and strong contour salience are reliably reproduced, whereas relational, hierarchical, and ritualized elements, especially those encoded through color and compositional hierarchy, are disproportionately weakened or lost. This bias explains why enhancements in edge curvature and visual clarity do not guarantee effective cultural transmission.

At an iconological level, the consequences are more profound. By selectively reproducing iconic surfaces while neglecting the cultural logic that organizes them, AI images generated by Doubao risk normalizing a mode of visual representation that simulates cultural form without sustaining cultural worldview. What emerges is not iconographic innovation but semantic thinning, a shift from culturally grounded representation toward visually fluent approximation.

5.3. A Conditional Semiotic Framework for AI–Human Illustration Integration

From a Panofskian perspective, the limitations observed in AI images generated by Doubao are structural rather than incidental. Current generative systems operate predominantly at the pre-iconographic level, enhancing perceptual fluency through contour regularization and chromatic smoothing. Yet, advancing from pre-iconographic form to iconographic coherence, and ultimately to iconological meaning, requires culturally stabilized conventions and symbolic hierarchies that cannot be derived from visual frequency alone (Panofsky, 1955). Consequently, gains in visual clarity and formal fluency do not automatically translate into deeper cultural–semantic integration.

Empirical analyses here indicate that AI images generated by Doubao maintain symbolic stability chiefly when narrative meaning is encoded through high-frequency, contour-

salient, and visually redundant features. Sun Wukong exemplifies this pattern. His visual identity rests on widely circulated templates marked by strong edge definition, distinctive silhouettes, and iconic accessories such as the tiger-skin skirt and the Ruyi golden cudgel. Under these conditions, AI generation reliably preserves symbolic recognizability, aligning with its optimization for perceptual regularity and shape continuity (Bertamini et al., 2016).

In contrast, when cultural meaning depends on low-frequency, ritualized, or hierarchically organized symbols, AI images generated by Doubao show systematic semiotic erosion rather than creative variation. For Tang Monk and Sha Wujing, religious insignia, ritual objects, and culturally codified costume systems—often conveyed through color hierarchy, symbolic layering, and relational positioning—are disproportionately omitted or simplified. These patterns are not random errors but reflect a frequency-sensitive bias inherent in current generative frameworks, whereby visually infrequent yet culturally central elements remain weakly represented.

These findings reveal a structurally conditioned relationship between generative image technologies and cultural semiotic transmission, moving beyond a generic call for collaboration between AI and handcrafted illustration. AI images generated by Doubao are not uniformly effective across narrative contexts; their applicability depends on the semiotic profile of the symbols they must reproduce. In domains where meaning resides largely in iconic surface features, AI can serve effectively as a tool for perceptual facilitation. However, where meaning is anchored in historically layered, ritualized, and hierarchically organized symbolic systems, human illustration remains essential for maintaining iconographic coherence and iconological depth.

6. Conclusions and Limitations

Grounded in Panofsky's tripartite framework of iconological analysis, this study compares AI images generated by Doubao and handcrafted illustrations of *Journey to the West* children's picture books through the integration of computational visual metrics and systematic iconographic annotation. The findings reveal a structural asymmetry between formal optimization and cultural-semantic representation in contemporary AI images generated by Doubao.

Quantitative analysis shows that illustrations generated by the Doubao (Seedream 3.0) model exhibit smoother contour curvature and greater visual regularity, enhancing perceptual fluency and visual coherence. However, these formal gains are accompanied by reduced chromatic complexity and weakened symbolic density. Iconographic analysis further demonstrates selective preservation of cultural symbols: while highly salient characters such as Sun Wukong largely retain canonical features, secondary figures and low-frequency yet culturally significant elements, particularly ritual objects and religious markers, are frequently simplified or omitted. This pattern reflects a frequency-sensitive bias in generative processes, whereby widely circulated visual templates are stabilized at the expense of context-dependent cultural meanings.

The results support a conceptual distinction between AI-generated illustrations as agents of formal fluency and handcrafted illustrations as carriers of cultural anchoring. Without careful human oversight, the use of current AI-generated illustrations in children's books risks privileging visual comfort over cultural transmission, thereby normalizing a form of symbolic simplification that may go unnoticed by young readers. These findings highlight the need for critical human mediation when integrating generative imagery into educational and cultural contexts.

By integrating computational visual metrics with iconological analysis, this study offers a replicable framework for examining how generative image models reshape the relationship between visual form and cultural meaning in children's picture books.

The present study is limited to a single narrative domain, one picture-book series, and a specific AI model—Doubao and prompting workflow. Moreover, the study focuses on visual form and semiotic structure and does not examine children’s interpretive responses, which remains an important direction for future research.

Author Contributions: Conceptualization, X.D. and Y.H.; Methodology, X.D.; Formal Analysis, X.D.; Investigation, X.D.; Data Curation, X.D.; Writing—Original Draft Preparation, X.D.; Writing—Review & Editing, Y.H.; Visualization, X.D.; Supervision, Y.H. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Humanities and Social Sciences Research Project of the Ministry of Education of the People’s Republic of China, grant number 25YJC740015, and the Educational and Teaching Reform Research and Practice Project of Zhengzhou University, grant number 2025ZZUJGXM152.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The derived datasets and analytical outputs are available from the corresponding author upon reasonable request, while the original copyrighted picture book illustrations are not publicly shared.

Conflicts of Interest: The authors declare no conflicts of interest.

Appendix A. Key Parameters for Edge Curvature and Color Entropy Computation

Table A1. Key parameters for edge curvature computation.

Parameter	Value	Description
Input color space	RGB → Grayscale	Images converted to grayscale prior to edge detection.
Gaussian blur kernel	5 × 5	Spatial smoothing to suppress noise.
Gaussian σ	Automatically estimated	Standard deviation determined by OpenCV.
Edge detector	Canny	Gradient-based edge detection.
Canny low threshold	50	Lower bound for edge hysteresis.
Canny high threshold	150	Upper bound for edge hysteresis.
Contour retrieval mode	RETR_EXTERNAL	Only outermost contours retained.
Contour approximation	CHAIN_APPROX_SIMPLE	Contour point compression.
Minimum contour length	20 points	Short contours excluded for stability.
Curvature window size (k)	5 points	Symmetric neighborhood for angle estimation.
Curvature formulation	Signed discrete curvature	Preserves convex/concave direction.
Image-level aggregation	Mean of all κ_i	Average signed curvature per image.

Table A2. Key parameters for color entropy computation.

Parameter	Value	Description
Input color space	RGB → HSV	Primary color space for entropy calculation.
Hue bins (H)	180	Range 0–179°, one bin per degree.
Saturation bins (S)	256	Range 0–255.
Value bins (V)	256	Range 0–255.
Histogram normalization	Probability distribution	Each channel normalized independently.
Entropy formulation	Shannon entropy	$E = -\sum p \log_2 p$
Channel aggregation	Arithmetic mean	$(E_H + E_S + E_V)/3$
Alternative color space	CIELAB (D65)	Used for robustness verification
Robustness indicator	Pearson $r = 0.94$	High correlation between HSV and LAB results.
Image-level output	Scalar entropy value	Overall color diversity per image.

Appendix B. Example of Prompt Adjustment Under Constrained Optimization

This appendix presents a concrete example illustrating the prompt construction, evaluation, and adjustment procedure described in Section 3.4. The purpose of this example is to clarify how narrative alignment was assessed and how prompt optimization was applied selectively and minimally under constrained text-to-image generation conditions, rather than as an aesthetic enhancement strategy.

**Figure A1.** Hand-drawn image from the volume *Battle at the Tongtian River*.

Narrative Input: The following plot is excerpted from the volume *Battle at the Tongtian River*. The text below the picture reads as follows: “The elder refused to believe him and said, ‘The land of Great Tang in the East is tens of thousands of miles away. How could you have come here alone?’ Tang Monk replied, ‘I have three disciples who have protected this poor monk on the journey.’ He then summoned his three disciples. When the elder saw them, he was so frightened that he fell to the ground, shouting repeatedly, ‘Demons have come! Demons have come!’”

Parameters: Seedream-3.0 model; 4:3 aspect ratio; picture-book style.



Figure A2. Initial generation image from the volume *Battle at the Tongtian River*.

Prompt: Role-specific labeling was applied to explicitly identify the five characters involved in the scene, namely Tripitaka (Tang Monk), Sun Wukong, Zhu Bajie, Sha Wujing, and the elder.

An elaborative verb description was added to clarify the interaction, specifying that Sun Wukong is assisting the elder who has fallen to the ground.

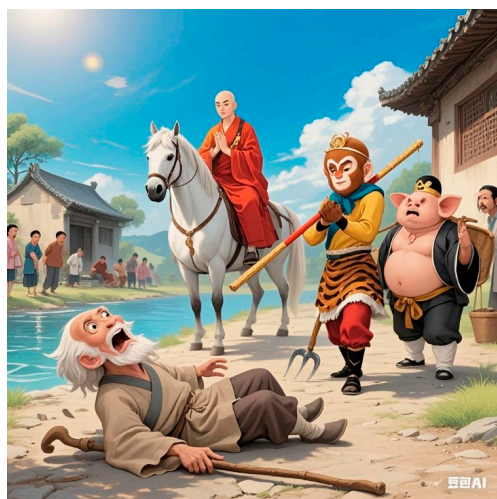


Figure A3. The optimized image from the volume *Battle at the Tongtian River*.

References

- Antony, V. N., & Huang, C.-M. (2024). ID.8: Co-creating visual stories with generative AI. *arXiv*, arXiv:2309.14228. <https://doi.org/10.48550/arXiv.2309.14228>.
- Arizpe, E., & Styles, M. (2016). *Children reading picturebooks: Interpreting visual texts* (2nd ed.). Routledge.
- Attneave, F. (1954). Some informational aspects of visual perception. *Psychological Review*, 61(3), 183–193. <https://doi.org/10.1037/h0054663>.
- Bar, M., & Neta, M. (2006). Humans prefer curved visual objects. *Psychological Science*, 17(8), 645–648. <https://doi.org/10.1111/j.1467-9280.2006.01759.x>.
- Bertamini, M., Palumbo, L., Gheorghes, T. N., & Galatsidas, M. (2016). Do observers like curvature or do they dislike angularity? *British Journal of Psychology*, 107(1), 154–178. <https://doi.org/10.1111/bjop.12132>.
- Bradski, G. (2000). The OpenCV library. *Dr. Dobb's Journal of Software Tools*, 25(11), 120–125.
- Canny, J. (1986). A computational approach to edge detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 8(6), 679–698. <https://doi.org/10.1109/TPAMI.1986.4767851>.

- Chuquichambi, E. G., Vartanian, O., Skov, M., Corradi, G. B., Nadal, M., Silvia, P. J., & Munar, E. (2022). How universal is preference for visual curvature? A systematic review and meta-analysis. *Annals of the New York Academy of Sciences*, 1518(1), 151–165. <https://doi.org/10.1111/nyas.14919>.
- Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, 20(1), 37–46. <https://doi.org/10.1177/001316446002000104>.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Lawrence Erlbaum Associates.
- Elsharif, W., Alzubaidi, M., & Agus, M. (2025). Cultural bias in text-to-image models: A systematic review of bias identification, evaluation, and mitigation strategies. *IEEE Access*, 13, 122636–122659. <https://doi.org/10.1109/ACCESS.2025.3585745>.
- Fairchild, M. D. (2013). *Color appearance models* (3rd ed.). John Wiley & Sons.
- Gombrich, E. H. (1960). *Art and illusion: A study in the psychology of pictorial representation*. Phaidon Press.
- Gonzalez, R. C., & Woods, R. E. (2018). *Digital image processing* (4th ed.). Pearson.
- Jin, Q., Yuan, I. Y., Zhang, Y., Zhao, S., & Landay, J. A. (2025). “They all look mad with each other”: Understanding the needs and preferences of children and parents in AI-generated images for stories. *International Journal of Child–Computer Interaction*, 46, 100787. <https://doi.org/10.1016/j.ijcci.2025.100787>.
- Kress, G. (2010). *Multimodality: A social semiotic approach to contemporary communication*. Routledge.
- Kress, G., & van Leeuwen, T. (2006). *Reading images: The grammar of visual design* (2nd ed.). Routledge.
- Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 33(1), 159–174. <https://doi.org/10.2307/2529310>.
- Lin, Z., Wang, Z., Liu, Q., Zhang, X., & Liu, J. (2026). Narratology meets text-to-image: A survey of consistency in AI-generated storybook illustrations. *Artificial Intelligence Review*, 59, 83. <https://doi.org/10.1007/s10462-025-11482-6>.
- Mao, J., Huang, X., Xie, Y., Chang, Y., Hui, M., Xu, B., & Zhou, Y. (2024). Story-Adapter: A training-free iterative framework for long story visualization. *arXiv*, arXiv:2410.06244. <https://doi.org/10.48550/arXiv.2410.06244>.
- Markova, I. (1978). Verbs of judging and the acts of judging. In R. N. Campbell, & P. Smith (Eds.), *Recent advances in the psychology of language* (pp. 203–216). Plenum Press.
- McHugh, M. L. (2012). Interrater reliability: The kappa statistic. *Biochemia Medica*, 22(3), 276–282. <https://doi.org/10.11613/BM.2012.031>.
- Nikolajeva, M., & Scott, C. (2001). *How picturebooks work*. Garland Publishing.
- Nodelman, P. (1988). *Words about pictures: The narrative art of children's picture books*. University of Georgia Press.
- O'Donnell, M. (2008). Demonstration of the UAM Corpus Tool for text and image annotation. In J. Lin (Ed.), *Proceedings of the ACL 08: HLT Demo Session (Companion Volume)* (pp. 13–16). Association for Computational Linguistics. Available online: <https://aclanthology.org/P08-4004/> (accessed on 24 May 2025).
- Palmer, S. E., & Schloss, K. B. (2010). An ecological valence theory of human color preference. *Proceedings of the National Academy of Sciences USA*, 107(19), 8877–8882. <https://doi.org/10.1073/pnas.0906172107>.
- Panofsky, E. (1955). *Meaning in the visual arts: Papers in and on art history*. Doubleday Anchor Books.
- Panofsky, E. (1972). *Studies in iconology: Humanistic themes in the art of the Renaissance*. Westview Press. (Original work published 1939).
- Rombach, R., Blattmann, A., Lorenz, D., Esser, P., & Ommer, B. (2022). High-resolution image synthesis with latent diffusion models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 10684–10695). IEEE. <https://doi.org/10.48550/arXiv.2112.10752>.
- Said, M. N., Zaidi, A., Usman, R., Okon, S., Medepalli, P., Zhu, K., Sharma, V., & O'Brien, S. (2025). Deconstructing bias: A multifaceted framework for diagnosing cultural and compositional inequities in text-to-image generative models. *arXiv*, arXiv:2505.01430. <https://doi.org/10.48550/arXiv.2505.01430>.
- Schuhmann, C., Beaumont, R., Vencu, R., Gordon, C., Wightman, R., Cherti, M., Coombes, T., Katta, A., Mullis, C., Schramowski, P., Kundurthy, S., Crowson, K., Schmidt, L., Kaczmarczyk, R., & Jitsev, J. (2022). LAION-5B: An open large-scale dataset for training next generation image-text models. *arXiv*, arXiv:2210.08402. <https://doi.org/10.48550/arXiv.2210.08402>.
- Seo, H., Choi, S., Hong, M., Zhou, Y., Kim, J., Ismaila, L., & Oh, J. (2025). Exposing blindspots: Cultural bias evaluation in generative image models. *arXiv*, arXiv:2510.20042. <https://doi.org/10.48550/arXiv.2510.20042>.
- Shannon, C. E. (1948). A mathematical theory of communication. *The Bell System Technical Journal*, 27(3), 379–423. <https://doi.org/10.1002/j.1538-7305.1948.tb01338.x>.
- Suzuki, S., & Abe, K. (1985). Topological structural analysis of digitized binary images by border following. *Computer Vision, Graphics, and Image Processing*, 30(1), 32–46. [https://doi.org/10.1016/0734-189X\(85\)90016-7](https://doi.org/10.1016/0734-189X(85)90016-7).
- van Dijk, T. A. (2009). *Society and discourse: How social contexts influence text and talk*. Cambridge University Press.

- van Leeuwen, T. (2022). *Multimodality and identity*. Routledge.
- Wiegand, T. L. T., Jung, L. B., Schuhmacher, L. S., Rischewski, J. F., Mehrzad, P., Jeong, S., Nguyen, L. H., Poeschla, M., & Koerte, I. K. (2025). Demographic inaccuracies and biases in the depiction of patients by artificial intelligence text-to-image generators. *NPJ Digital Medicine*, 8, 459. <https://doi.org/10.1038/s41746-025-01817-6>.
- Wu, C. E. (2019). *Journey to the West* [《西游记》]. People's Literature Publishing House. (In Chinese)
- Xia, Z. Q. (2001). *On the history of classical Chinese fiction* [《中国古典小说史论》] (H. Yimin, S. Xiaolin, & S. Kunqin, Trans.). Jiangxi People's Publishing House. (Original work published 1980) (In Chinese)
- Xu, X., & Deng, Y. (2024). Image-generative artificial intelligence Midjourney generation children's digital picture book research from the perspective of image narrative. In *Proceedings of the International Conference on Language, Art and Humanistic Design (ICLAHD 2024)*. Atlantis Press. https://doi.org/10.2991/978-2-38476-319-1_23.
- Zhang, L., Liao, X., Yang, Z., Gao, B., Wang, C., Yang, Q., & Li, D. (2024). Partiality and misconception: Investigating cultural representativeness in text to image models. In *Proceedings of the 2024 CHI conference on Human factors in computing systems* (pp. 620:1–620:25). Association for Computing Machinery. <https://doi.org/10.1145/3613904.3642877>.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.