Human Factors in Design, Engineering, and Computing, Vol. 238, 2026, 2831–2841
https://doi.org/10.54941/ahfe1008507

Beyond Formal Likeness: A Four-Tier Evaluation Framework for Stylistic Fidelity in AI Generated Chinese Ink Videos

Yilin Wang

Tsinghua University, Beijing, 100084, China

ABSTRACT

This study presents a four-tier evaluation framework for assessing AI-generated video style transfer in the context of Chinese ink animation. Developed through a case study of one contemporary artist, the framework draws on iterative co-creation, model training, and reflexive thematic analysis of expert interviews. It identifies four key dimensions: (1) visual genre and structural fidelity, (2) alignment with artistic intent, (3) brush-and-ink consistency, and (4) stylized emphasis of signature features. By integrating perspectives from the artist, art scholars, and AI practitioners, the framework emphasizes the expressive and cultural functions of style beyond surface texture. It serves as a practical tool for evaluating stylistic fidelity in AI-assisted video creation, supporting both artists and developers in achieving more meaningful and expressive outcomes. By operationalizing traditional Chinese aesthetic principles into computational contexts, this research bridges digital humanities and computational aesthetics. The framework also provides a culturally grounded basis for discussing responsible style reproduction in generative media.

Keywords: Chinese ink animation, Generative AI, Style transfer, Computational aesthetics, Digital humanities

INTRODUCTION

Generative AI has made image and video style transfer increasingly accessible, while intensifying concerns about artistic identity, style appropriation, and authorship (Schor, 2024; Wang and Zhang, 2024; Chatterton et al., 2025). Evaluating such outputs therefore requires more than measuring visual similarity; it also requires examining whether an artist's visual language and expressive intent are preserved.

This challenge is pronounced in Chinese ink painting, whose aesthetics emphasize expressive vitality, brush-and-ink relations, blank space, and culturally situated principles rather than pictorial realism. Conventional technical metrics and generalized aesthetic assessments can capture image quality or structural similarity, but are less suited to judging these culturally specific qualities in moving images.

Accordingly, this study asks: (**RQ1**) What dimensions are most meaningful for evaluating AI-based video style transfer in Chinese ink painting? (**RQ2**)

Received September 3, 2026; Revised September 30, 2026; Accepted October 7, 2026; Available online December 1, 2026

© 2026 The Authors. This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 License.
For more information, see https://creativecommons.org/licenses/by-nc-nd/4.0/

What creative tensions and opportunities arise when AI emulates a culturally specific artistic style, and how might these inform digital reinterpretations of traditional aesthetics? Through a qualitative case study, four rounds of iterative co-creation and interviews, and reflexive thematic analysis, we develop a four-tier framework covering visual genre and structural fidelity, artistic intent, brush-and-ink consistency, and stylized emphasis of signature features.

RELATED WORK

Style Transfer Capabilities in AI Video Generation

Style transfer modifies the visual style of an image or video while preserving its content structure. Neural style transfer (NST) separated content and style representations (Gatys et al., 2015), while later approaches such as adaptive instance normalization (AdaIN) enabled more flexible and real-time transfer (Huang and Belongie, 2017). Extending style transfer to video introduces the additional challenge of temporal coherence, as frame-level inconsistencies can cause flickering and disrupt visual continuity (Li et al., 2018).

Evaluation of Style Transfer

Existing style-transfer evaluation approaches include Style-Eval (Wang et al., 2021), ArtFID (Wright and Ommer, 2022), and QTEV (Yeh et al., 2018), alongside quantitative metrics such as PSNR, SSIM, and FID (Kumar et al., 2024). These are often complemented by human evaluation of content preservation, style similarity, visual quality, brushwork, cleanliness, and artifacts (Chen et al., 2023a; Chen et al., 2023b; Kumar et al., 2024). However, such criteria are often either generalized or fragmented, making it difficult to capture an artist's culturally situated style. This limitation motivates more artist-centered evaluation strategies.

Aesthetic Assessment

Aesthetic assessment extends beyond formal quality to culturally and contextually situated judgment. Philosophical aesthetics emphasizes the artwork, audience, creator, and cultural context (Levinson, 2005; Jerrentrup, 2024), while empirical aesthetics shows that judgments are shaped by properties of the artwork, individual differences, and context (Nadal and Vartanian, 2022; Graham and Redies, 2010). Computational video aesthetics complements these approaches by modeling motion, color, composition, and temporal features (Niu and Liu, 2012; Li et al., 2022). However, existing approaches rarely address culturally specific aesthetic values or artists' own judgments of stylistic fidelity, motivating a more situated and interpretive perspective.

Cultural and Aesthetic Characteristics of Chinese Ink Painting

Chinese ink painting prioritizes expressive vitality over pictorial realism. Core principles include Xie He's "spirit resonance," intentional blank space, dynamic brushwork, and tonal modulation through the "five shades of ink"

(Li, 1994; Wang, 1981). Blank space structures rhythm and atmosphere, while brushstrokes convey energy and intentional incompleteness rather than merely defining contours. These qualities are difficult to capture through style-transfer methods that emphasize texture, color distribution, or local visual features. More broadly, Chinese ink aesthetics foregrounds relationships between subject and environment, material and void, and action and stillness (Li, 1994; Hung, 1996). Evaluation must therefore consider affective resonance and cultural symbolism in addition to formal similarity.

METHOD

Case Context and Artist Selection

This study adopts a qualitative single-case design centered on the work of contemporary ink painter Lu Xiaobo (Wang, 2016). The case was selected because his crane paintings show a coherent and recognizable visual language, provide clear motifs for cross-output comparison, while his sustained participation enabled iterative feedback. His work is characterized by gestural depictions of cranes in motion, extensive use of blank space, nuanced monochrome ink gradation, and spatial abstraction. These features combine formal consistency with culturally specific expressive qualities, making the case suitable for examining stylistic fidelity across static paintings and AI-generated moving images. Figure 1 illustrates representative characteristics of the artist's work.



Figure 1: Representative crane painting by Lu Xiaobo.

Data Collection

Iterative Co-Creation and Interviews

The data collection process followed an iterative co-creation structure spanning four rounds of video generation, interviews, and refinement. In each round, style-transferred videos were generated using models trained on the artist's ink paintings. These videos were then presented to the interviewees, who were asked to reflect on their stylistic coherence and artistic alignment. Seven interviewees contributed to the study, including the artist himself, four

scholars with backgrounds in art or design, and two developers with expertise in AI systems. Table 1 summarizes the background of the interviewees. Notably, the artist played a central and active role as a collaborator throughout the process. He contributed to dataset selection, critiqued model outputs, and helped articulate stylistic priorities.

Table 1: Interviewee overview.

Interviewees	Count	Expertise
The Representative Artist	1	Artist, senior professor in art and design.
Art or Design Scholars	4	One postdoctoral researcher in design studies, two PhD students (in art and design respectively), and a designer with 10 years of work experience.
AI Practitioners	2	Engineers from tech companies. One algorithm engineer with 5 years of work experience, one AI product architect with 15 years of work experience.

This iterative structure was designed to progressively deepen the understanding of stylistic fidelity. After each interview round, the AI-generated videos were revised to address the issues raised by interviewees. Updated versions were then used as stimuli in the following round. This feedback loop allowed the study to gradually uncover more nuanced dimensions of evaluation, particularly as initial, more obvious flaws were resolved.

Each interview was conducted in person, lasted approximately 30 minutes and followed a semi-structured format. The questions were designed to guide the discussion while leaving space for open-ended insights. Key questions focused on stylistic similarities and differences, satisfactory and unsatisfactory aspects of each version, and potential improvements.

Researcher's Reflexive Notes

Complementing the interviews, the author maintained a reflexive journal documenting interpretive shifts, aesthetic dilemmas, and assumptions emerging during video production and evaluation. These notes provided a secondary data source for tracing changes in the understanding of stylistic fidelity and informing subsequent interviews, while making the researcher's interpretive role explicit.

Data Analysis: Reflexive Thematic Analysis

The collected data, including interview transcripts and researcher reflections, were analyzed using reflexive thematic analysis, following Braun and Clarke's six-phase framework (2006, 2019). Initial codes were derived from expressions of evaluation, comparison, and aesthetic judgment found in the interviews and reflections. Through successive rounds of refinement, themes were reviewed, merged, or differentiated based on their relevance and specificity. Rather than aiming for saturation, the process prioritized conceptual clarity and interpretive resonance. The resulting themes were not merely descriptive, but

structured around the evaluative logics interviewees used when interpreting AI-generated imagery. These themes were ultimately organized into the four-tier evaluative framework presented in the Findings section.

Ethical Considerations

All procedures were approved by the Tsinghua University Science and Technology Ethics Committee (Humanities, Social Sciences and Engineering) (Approval No. THU-04-2026-0142) and conducted in accordance with the Declaration of Helsinki. All participants provided written informed consent; the artist additionally consented to being identified and to the use of his paintings for model training and publication. The author's dual role as researcher and interpretive agent is acknowledged.

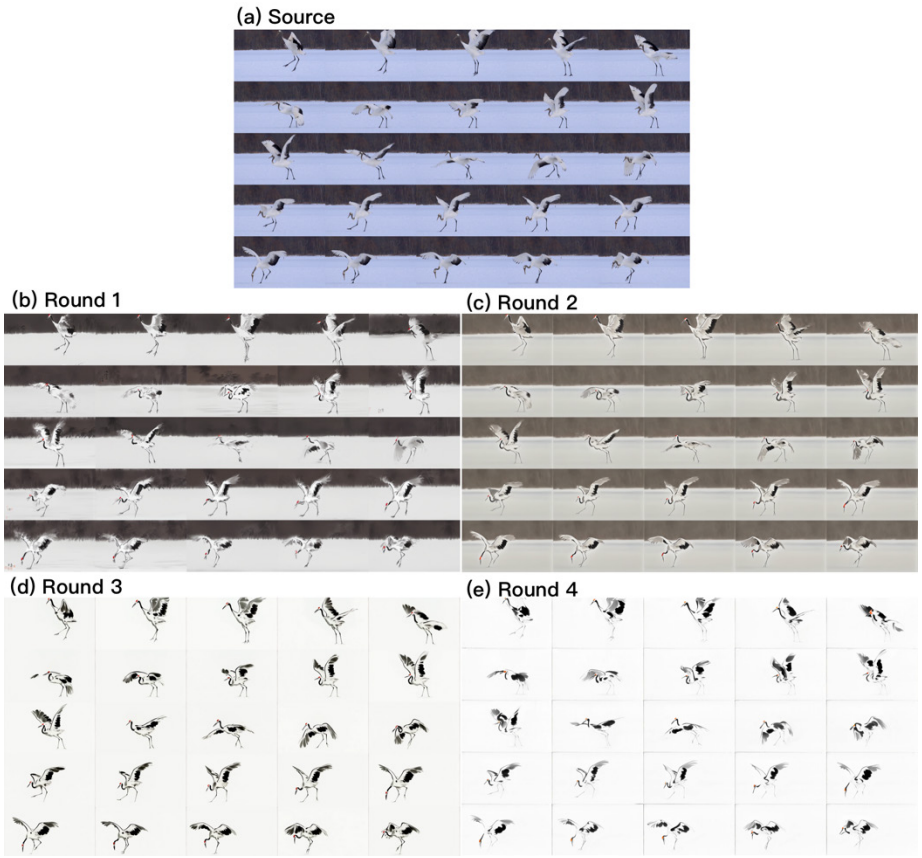


Figure 2: Evolution of the AI-generated Chinese ink video across four iterative refinement rounds: (a) source footage; (b–e) results from Rounds One to Four.

ITERATIVE CO-CREATION AND FRAMEWORK FORMATION

The study proceeded through four cycles of generation, interview, and refinement (Figure 2). In **Round One**, a LoRA model (Hu et al., 2021) trained on the artist's paintings was applied to crane footage using a diffusion-based video workflow (Rombach et al., 2022). Although the result

broadly resembled Chinese ink painting and matched the lively subject matter, temporal jitter and frame-level errors limited stylistic evaluation. **Round Two** introduced improved temporal-generation workflows and stronger structural controls (Guo et al., 2024; Hu and Xu, 2023; Ye et al., 2023). Motion became substantially smoother, but the artist and art/design participants identified excessive background detail, insufficient blank space, and mismatches in stroke size, shape, and line relationships. This contrasted with the more positive assessments from AI practitioners. **Round Three** refined the training dataset and removed background information before style transfer. The resulting video better matched the artist's use of blank space and overall brush structure. However, frame-by-frame inspection exposed subtler inconsistencies in ink gradation, stroke scale, flow, and wing-line composition that were less noticeable during normal playback. **Round Four** further optimized the LoRA model and tonal parameters. Participants preferred the final version for its richer ink variation and more expressive brushwork. Across the four rounds, feedback shifted from basic genre resemblance, temporal coherence, and content preservation toward artistic intent, brush-and-ink consistency, and signature stylistic features. Reflexive thematic analysis of this progression produced the four-tier framework presented in the Findings section.

FINDINGS

Reflexive thematic analysis produced four evaluative tiers (Figure 3), progressing from basic visual and structural fidelity to culturally and artistically specific judgments of stylistic fidelity.

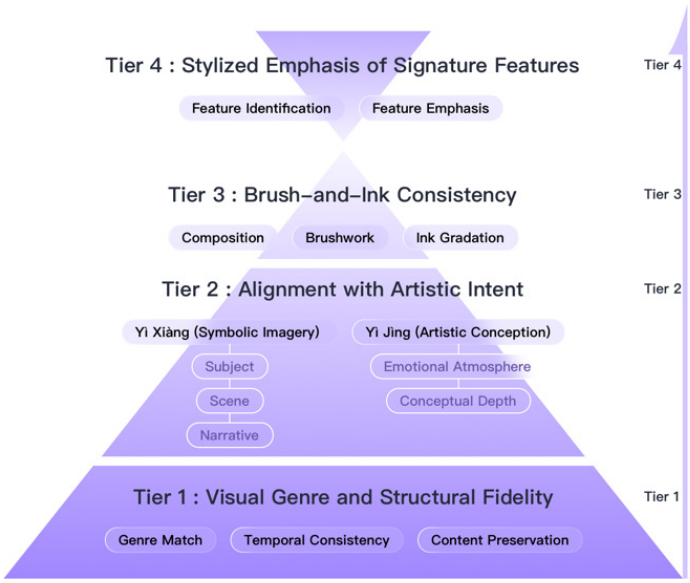


Figure 3: Four-tier evaluation framework for assessing stylistic fidelity in AI-generated Chinese ink videos.

Tier One: Visual Genre and Structural Fidelity

This tier establishes the basic threshold for evaluating whether a video can be plausibly recognized as consistent with the Chinese ink-painting tradition. It encompasses three core dimensions: genre match, temporal consistency, and content preservation.

Genre Match: This criterion assesses whether the overall visual impression aligns with the genre of traditional Chinese ink painting. For example, appropriate visual structure may include multi-point perspective rather than Western linear perspective, and the intentional use of white space.

"This version doesn't feel like ink painting. The entire frame is filled in with no blank space at all. Can we remove the background?" — Interviewee A, artist

Temporal Consistency: This criterion assesses whether motion remains visually coherent across frames without flickering, jitter, or other disruptions.

"The flickering has a kind of futuristic AI vibe, but it feels more like a horror film than a Chinese ink painting. It's jarring." — Interviewee G, AI product architect

Content Preservation: This refers to the degree to which the structural content, such as objects, anatomy, and motion, remains intact during the style transfer process. While some abstraction is expected in ink painting, misalignment in form can produce visual noise.

"Why does this part flicker so much? Because in a few frames, the crane's wings are drawn in the wrong spatial order. These misaligned frames cause the entire video to jitter." — Reflexive Notes

Tier Two: Alignment With Artistic Intent

This tier evaluates whether the video captures the artist's expressive intent through two interrelated concepts in Chinese aesthetics: *yì xiàng* (symbolic imagery) and *yì jìng* (artistic conception) (Li, 1994; Wang, 1981). Together, they address whether stylistic transfer extends beyond visual resemblance to expressive coherence.

***Yì xiàng*: Subject, Scene, and Narrative:** *Yì xiàng* concerns whether the subject, setting, and movement correspond to recurring motifs in the artist's visual language. Subject evaluates thematic fit; scene considers spatial and symbolic context; narrative examines whether movement, shot rhythm, and camera choices support the intended expression.

"The original video clip you picked works really well. The crane's movements feel lively. That aligns with Lu's style. The birds in his paintings are always full of spirit, even exaggerated in motion." — Interviewee D, design scholar

***Yì jìng*: Emotional Atmosphere and Conceptual Depth:** *Yì jìng* concerns the atmosphere and conceptual resonance emerging from these elements. It evaluates whether the video conveys an emotional tone and symbolic or philosophical depth comparable to the artist's work.

"It feels like the crane is completely unrestrained, playful, almost willful." — Interviewee C, art scholar

Tier Three: Brush-and-Ink Consistency

This tier focuses on whether the material and visual qualities of brush-and-ink work in the style-transferred video reflect the artist's expressive idiom. Chinese ink painting is a medium rooted in specific tools, gestures, and ink-handling techniques.

Composition: This dimension examines the spatial organization of visual elements, including the structure of negative space, object placement, and proportional layout. It asks whether the overall visual balance resembles the artist's style, including both local structures and the global arrangement of form and void.

"In that version, the crane's belly and thighs just don't have the same sense of structure as in the original painting." — Interviewee A, artist

Brushwork: Brushwork is central to the expressive language of ink painting. Different types of brushes produce distinct line textures. Artists use techniques such as outline, texture stroke, rub, dot, wash and movement gestures like lift, press, pause, drag, turn, break to convey vitality and rhythm. In this dimension, we examine stroke size, shape, strength, and flow across anatomical parts. Even if basic lines are retained, the stroke quality must reflect expressive variation.

"In one version, the brushstrokes feel too broken. In another, they're all the same size. There should be different types. Like the neck and wings should feel more fluid and forceful." — Interviewee C, art scholar

Ink Gradation: Ink in Chinese painting is not a single color but a spectrum. The concept of five shades of ink (scorched, concentrated, dense, diluted, and clear) enables artists to build spatial depth and emotional resonance using only monochrome. These variations emerge from brush pressure, water mixing, and paper absorption. This criterion evaluates not only whether dark and light tones appear, but whether their proportion and layering reflect the original painting's logic.

"This version fails to balance the ink tones. The darkest line should be at the top edge of the wing. The proportion of mid-gray is incorrect." — Interviewee A, artist

Tier Four: Stylized Emphasis of Signature Features

Fine stylistic details can become less perceptible in moving images because of attentional and motion-related perceptual effects (de Pontes Nobre et al., 2020; Simons, 2000; Simons and Levin, 1997; Idrees et al., 2020; Haase et al., 2024). This tier therefore examines whether an artist's signature motifs are first identified and then selectively emphasized so that they remain legible during motion.

Feature Identification: This criterion identifies recurring motifs that define the artist's visual identity, such as the flowing contour of the crane's neck and expressive wing strokes in this case.

Feature Emphasis: This criterion evaluates whether these motifs are selectively amplified in motion to preserve stylistic recognizability without compromising overall fidelity.

"When viewers watch video, they don't catch every detail. The artist's unique style gets blurred in motion." — Reflexive Notes

DISCUSSION

The findings suggest that stylistic fidelity in AI-generated art is not a single measure of visual similarity but a layered and culturally situated judgment. Across the four iterations, evaluation progressed from genre recognition, temporal coherence, and content preservation toward artistic intent, brush-and-ink consistency, and signature features. This progression indicates that technical adequacy functions as a prerequisite rather than an endpoint: once obvious generation artifacts were resolved, culturally specific differences became more salient. In particular, AI practitioners tended to accept technically coherent outputs earlier, while the artist and art/design participants continued to identify discrepancies in blank space, brushwork, ink gradation, and expressive intent. This tension highlights the limits of engineering-centered evaluation and the importance of creator expertise and domain knowledge in assessing culturally specific generative media.

The study also reveals a second tension between static fidelity and dynamic legibility. Features that characterize an artist's style in a still painting may become less perceptible in motion. Consequently, faithful video reinterpretation may require selective emphasis rather than literal replication. Stylistic fidelity in moving images can therefore be understood as a process of medium-sensitive adaptation, in which source content and motion, artistic intent, and signature visual features must remain perceptually and expressively coherent. This creates an opportunity for human-in-the-loop generative workflows in which evaluation guides not only error correction but also deliberate artistic adaptation. By making such stylistic decisions explicit, the framework may further support more transparent and responsible practices of style reproduction.

The single-artist case limits the framework's transferability. Future studies should test it across artists, traditions, and media, and combine qualitative evaluation with empirical aesthetics methods, such as audience ratings or physiological measures, to examine its perceptual and emotional dimensions.

CONCLUSION

This study develops a culturally situated four-tier framework for evaluating stylistic fidelity in AI-generated Chinese ink videos. Moving beyond surface resemblance, the framework connects technical coherence with artistic intent, brush-and-ink expression, and the perceptual legibility of signature features in motion. The findings further suggest that faithful cross-media style transfer may require medium-sensitive adaptation rather than literal replication, offering a basis for more creator-centered and culturally responsive evaluation of generative media.

REFERENCES

Braun, V., Clarke, V. (2006). Using thematic analysis in psychology. Qualitative Research in Psychology, 3(2).

Braun, V., Clarke, V. (2019). Reflecting on reflexive thematic analysis. Qualitative Research in Sport, Exercise and Health, 11(4).

Chatterton, E., Zhang, J., Blackford, L. (2025). Another Chinese court finds that AI-generated images can be protected by copyright: The Changshu People's Court and the 'half heart' case. Technology's Legal Edge, March 21.

Chen, F., Ji, N., Zhao, Y., et al. (2023a). Controllable feature-preserving style transfer, in: International Conference on AI-generated Content. Springer, pp. 95–104.

Chen, H., Shao, F., Chai, X., et al. (2023b). Quality evaluation of arbitrary style transfer: Subjective study and objective metric. IEEE Transactions on Circuits and Systems for Video Technology, 33(7), pp. 3055–3070.

de Pontes Nobre, A., de Melo, G.M., Gauer, G., et al. (2020). Implicit processing during inattentional blindness: A systematic review and meta-analysis. Neuroscience & Biobehavioral Reviews, 119, pp. 355–375.

Gatys, L.A., Ecker, A.S., Bethge, M. (2015). A neural algorithm of artistic style. Available from: https://arxiv.org/abs/1508.06576.

Graham, D.J., Redies, C. (2010). Statistical regularities in art: Relations with visual coding and perception. Vision Research, 50(16), pp. 1503–1509.

Guo, Y., Yang, C., Rao, A., et al. (2024). AnimateDiff: Animate your personalized text-to-image diffusion models without specific tuning. Available from: https://arxiv.org/abs/2307.04725.

Haase, T.M., Rich, A.N., Gilchrist, I.D., et al. (2024). Attention moderates the motion silencing effect for dynamic orientation changes in a discrimination task. Journal of Vision, 24(13).

Hu, E.J., Shen, Y., Wallis, P., et al. (2021). LoRA: Low-rank adaptation of large language models. Available from: https://arxiv.org/abs/2106.09685.

Hu, Z., Xu, D. (2023). VideoControlNet: A motion-guided video-to-video translation framework by using diffusion model with ControlNet. Available from: https://arxiv.org/abs/2307.14073.

Huang, X., Belongie, S. (2017). Arbitrary style transfer in real-time with adaptive instance normalization, in: 2017 IEEE International Conference on Computer Vision (ICCV), pp. 1510–1519.

Hung, W. (1996). The Double Screen: Medium and Representation in Chinese Painting. Reaktion Books.

Idrees, S., Baumann, M.P., Franke, F., et al. (2020). Perceptual saccadic suppression starts in the retina. Nature Communications, 11(1).

Jerrentrup, M. (2024). Imagine art: The status of works generated by artificial intelligence. International Journal of Cultural Studies, 27(6), 814–830.

Kumar, D.P., Bhadula, S., Al-Farouni, M., et al. (2024). Artistic style transfer using generative adversarial networks: A Pix2Pix implementation, in: 2024 International Conference on Communication, Computer Sciences and Engineering (IC3SE), pp. 1351–1357.

Levinson, J. (2005). Philosophical aesthetics: An overview, in: The Oxford Handbook of Aesthetics. Oxford University Press.

Li, M., Wang, Z., Ren, J., et al. (2022). MVVA-Net: A video aesthetic quality assessment network with cognitive fusion of multi-type feature–based strong generalization. Cognitive Computation, 14(4), pp. 1435–1445.

Li, Y., Fang, C., Yang, J., et al. (2018). Flow-grounded spatial-temporal video prediction from still images, in: Computer Vision – ECCV 2018: 15th European Conference, Munich, Germany, September 8–14, 2018, Proceedings, Part IX. Berlin, Heidelberg: Springer-Verlag, pp. 609–625.

Li, Z. (1994). The Path of Beauty: A Study of Chinese Aesthetics. New York: Oxford University Press.

Nadal, M., Vartanian, O. (2022). Empirical aesthetics: An overview, in: The Oxford Handbook of Empirical Aesthetics. Oxford University Press.

Niu, Y., Liu, F. (2012). What makes a professional video? A computational aesthetics approach. IEEE Transactions on Circuits and Systems for Video Technology, 22(7), pp. 1037–1049.

Rombach, R., Blattmann, A., Lorenz, D., et al. (2022). High-resolution image synthesis with latent diffusion models. In 2022 IEEE/CVF conference on computer vision and pattern recognition (CVPR), pp. 10674–10685.

Schor, Z. (2024). Andersen v. Stability AI: The landmark case unpacking the copyright risks of AI image generators. NYU Journal of Intellectual Property & Entertainment Law Blog, December 2.

Simons, D.J. (2000). Attentional capture and inattentional blindness. Trends in Cognitive Sciences, 4(4), pp. 147–155.

Simons, D.J., Levin, D.T. (1997). Change blindness. Trends in Cognitive Sciences, 1(7), pp. 261–267.

Wang, X. (2016). Integrating art and science, knowing and acting: Professor Lu Xiaobo, Dean of the Academy of Arts & Design, Tsinghua University, and his path of academic exploration. ZHUANGSHI, (10), pp. 58–63.

Wang, Y., Zhang, J. (2024). Beijing Internet Court grants copyright to AI-generated image for the first time. Kluwer Copyright Blog, June 6.

Wang, Z. (1981). Introduction to Aesthetics. China: People's Publishing House.

Wang, Z., Zhao, L., Chen, H., et al. (2021). Evaluate and improve the quality of neural style transfer. Computer Vision and Image Understanding, 207, 103203.

Wright, M., Ommer, B. (2022). ArtFID: Quantitative evaluation of neural style transfer, in: Andres, B., Bernard, F., Cremers, D., et al. (Eds.), Pattern Recognition. Cham: Springer International Publishing, pp. 560–576.

Ye, H., Zhang, J., Liu, S., et al. (2023). IP-Adapter: Text compatible image prompt adapter for text-to-image diffusion models. Available from: https://arxiv.org/abs/2308.06721.

Yeh, M.C., Tang, S., Bhattad, A., et al. (2018). Quantitative evaluation of style transfer. Available from: https://arxiv.org/abs/1804.00118.