

City Image Building from the Perspective of Multimodal Discourse Analysis: A Case Study of the Promotional Video “Follow Wukong to Shanxi”

Yihuan Chen, Yuqiao Chen, Yaohua Du

Beijing Normal-Hong Kong Baptist University, Chinese Culture and Global Communication, Zhuhai, Guangdong, China

Abstract: In the context of globalisation and digital media development, city image construction has increasingly relied on digital communication strategies. However, traditional single-modal texts lack emotional appeal, while multimodal promotional videos which has just gained popularity in recent years remain under-researched in cross-cultural city branding. While “city image” is conventionally examined at the urban level, this study extends the concept to the provincial scale by taking Shanxi as its primary research object. To fill this gap, this study applies the framework of Multimodal Discourse Analysis (MDA), grounded in Systemic Functional Linguistics and visual grammar, to examine how Shanxi’s image is constructed through the promotional video ‘Follow Wukong to Shanxi’. The findings indicate that multimodal formats can integrate sensory elements to create immersive experiences, strengthen cultural identity, and foster cross-cultural communication. In doing so, this paper depicts that MDA offers a valuable analytical framework for understanding and optimising city branding strategies, particularly within the “game + cultural tourism” model.

Keywords: Multimodal Discourse Analysis; City Image; Cross-Culture Communication; Black Myth; Wukong

1. Introduction

1.1 Background

In the context of globalisation and advancements in information technology, constructing a city’s image has become a crucial factor in enhancing urban competitiveness. As a result, cities worldwide increasingly prioritise developing their city images [1][22]. A research highlights that “city image” has emerged as a significant

ideological domain in contemporary China, emphasising the need for effective communication strategies to tell the stories of Chinese cities and enhance cultural identity [12]. Moreover, with the rise of digital tools, particularly the widespread use of social media, multimodal communication-integrating two or more modes-has become a dominant trend [5][20]. This approach facilitates information dissemination and enhances audience engagement and emotional connection through diverse forms of expression. According to some scholars, the “game + cultural tourism” model has proven to be a promising strategy for city image promotion [2][4]. By blending gaming elements with cultural tourism, this model effectively captures a broader audience and strengthens emotional resonance. Its advantages lie in enhancing cultural communication while providing innovative pathways for shaping city images. In globalisation, the challenge for many historically and culturally rich cities lies in protecting their unique cultural characteristics while adopting innovative communication methods to attract a wider audience.

As multimedia elements play an increasingly critical role in urban promotion, the need to analyse these materials systematically has grown more urgent. MDA offers a valuable framework for addressing this need. MDA examines how various semiotic resources, including language, images, sound, and gestures, integrate sensory modes, such as visual, auditory, and tactile, to communicate more effectively [19]. Shanxi, a profound historical and cultural heritage region, faces unique challenges in constructing its regional image in globalisation. Despite its rich cultural resources, effectively conveying its cultural value and gaining widespread recognition remains critical.

Against this backdrop, the success of the domestic game Black Myth: Wukong has introduced innovative applications of the “game

+ cultural tourism” model, offering new avenues for enhancing Shanxi’s cultural appeal and global recognition. While the notion of “city image” traditionally operates at the urban scale, this study extends the concept to the provincial level by taking Shanxi as its primary research object, applying the aforementioned model as an extended practice of city image theory to explore the optimization path of its cultural communication. By employing multimodal communication strategies, Shanxi has the potential to craft a more distinctive image in the global landscape. Utilising the MDA framework can systematically uncover the multidimensional meanings embedded in Shanxi’s cultural heritage, improving its effectiveness and impact on domestic and international communication.

1.2 Purpose of the Study

This study focuses on the promotional video *Follow Wukong to Shanxi*, which is based on the game *Black Myth: Wukong* and adopts the “game + cultural tourism” model to promote Shanxi’s image. Using the framework of MDA, this research explores the processes and strategies involved in shaping Shanxi’s identity as a historically significant province. Through an in-depth video analysis, the study examines how various sensory modes’ interplay enhances cultural identity and fosters cultural confidence. Additionally, it compares the effectiveness of this approach to traditional news reports. Ultimately, the research highlights the advantages of multimodal communication in constructing Shanxi’s image while offering valuable insights and strategies for other cities to enhance their promotional efforts.

1.3 Article Structure

This paper begins with a literature review that revisits the theoretical foundations of MDA and examines its application in shaping regional images and promoting cultural communication. It then identifies gaps in existing research and formulates corresponding research questions. Following this, the methodology is introduced, outlining the data sources and detailing the analytical process. The analysis section explores the communicative effectiveness of the promotional video *Follow Wukong to Shanxi*, focusing on its impact compared to single-modal communication methods, such as traditional news reports. Finally, the paper discusses the study’s significance, contributions, and

limitations, synthesises the key findings, and proposes directions for future research.

2. Literature Review

MDA offers a robust framework for understanding how various semiotic resources collaborate to convey meaning. Its core theories are rooted in Halliday’s Systemic Functional Linguistics (SFL) and Kress and Van Leeuwen’s visual grammar. Halliday identified three primary functions of language: ideational (representing actions and events), interpersonal (expressing emotions and relationships), and textual (organising information) [3]. Building on this foundation, Kress and Van Leeuwen expanded the theory by proposing that visual symbols carry representational, interactive, and compositional meanings [6][12]. By integrating the linguistic functions of SFL with the semiotic features of visual grammar, MDA provides a comprehensive approach to analysing how multiple modes work together to construct meaning and enhance the depth and multidimensionality of communication [10].

In contemporary communication studies, MDA has been widely applied to analyse multimodal texts, such as advertisements and videos, with scholars continuously advancing its methods. For instance, Luca analysed newspaper articles and proposed specific techniques for interpreting media photos and text [7]. Similarly, Ly and Jung explored the importance of images as semiotic resources, particularly their social interpretations in advertising contexts[8]. However, while there has been substantial research on more straightforward forms of multimodal communication, such as news and advertisements, studies on more complex formats, like promotional videos, remain relatively limited [14]. Existing research has primarily focused on the role of videos in city branding and the communicative effects of foreign-language promotional videos. However, inconsistencies persist in how MDA is integrated with narrative storytelling and city branding strategies [11][13][18]. Furthermore, as social media continues to dominate global communication, Wang and Lei highlighted the emerging challenges of multimodal discourse in cross-cultural contexts, particularly in effectively conveying information and cultural values within the framework of globalisation [15].

Despite the progress in multimodal analysis, significant gaps remain. First, promotional

videos have not been fully explored regarding their specific features and communicative effects as complex multimodal texts. Second, most studies have concentrated on local cultural contexts, overlooking the global applicability of MDA in cross-cultural communication and city branding. This study addresses these gaps by examining the multimodal strategies employed in Shanxi's promotional video, *Follow Wukong to Shanxi*, and their role in shaping the region's image.

3. Research Questions

RQ1: What are the similarities and differences between the effects of single-modal and multimodal communication in promoting Shanxi's regional image?

RQ2: How can the MDA framework contribute to constructing regional images and cross-cultural communication?

4. Methodology

4.1 Data

The dataset for this study consists of two primary sources. The first is a news article published by *China Daily*, comprising 591 words, primarily highlighting the cultural value of Shanxi's scenic spots and their visitor experience.

The second is the promotional video *Follow Wukong to Shanxi*, released by New China TV. With 2 minutes and 12 seconds runtime, this video combines virtual scenes from the game *Black Myth: Wukong* with real-life landmarks in Shanxi. It showcases the region's historical sites and unique cultural heritage, blending traditional elements with modern digital storytelling.

4.2 Analysis Procedure

This study applies the MDA framework to examine the textual and audiovisual elements of *China Daily's* news article and the promotional video *Exploring the Real-World Cultural Landmarks of Black Myth: Wukong*. By analysing intertextuality, visual grammar, and sound design, the research aims to uncover how these media construct Shanxi's regional image and cultural narrative.

The textual analysis focuses on how the *China Daily* article uses intertextuality to connect Shanxi's cultural heritage with the game *Black Myth: Wukong*. The article incorporates content written and translated by site managers, creating

a narrative framework that links the historical significance of Xiaoxitian to the game's mythological backdrop. By emphasising the site's cultural value and placing it within a broader context of game culture and heritage, the article successfully positions Shanxi as a tourist destination rich in cultural depth. Additionally, the article employs interactive linguistic strategies, such as quoting site management and second-person pronouns, to create a conversational tone and enhance reader engagement. These strategies address overcrowding and improve visitor management, fostering a stronger perception of enhanced tourist experiences. However, compared to the promotional video, the article demonstrates a notable advantage regarding textual clarity [21]. In contrast, the video's subtitles exhibit weaknesses in logical coherence and the conveyance of background information. The fragmented nature of the subtitles, caused by screen transitions, limits readability and undermines the overall narrative structure. Consequently, while the article excels in presenting a structured and straightforward narrative, the video's subtitles fall short of delivering cohesive storytelling.

The promotional video employs visual grammar to construct a compelling narrative. The storyline is organised around an auction process, with Wukong, the game's protagonist, serving as the narrative centre. Transitions between cultural landmarks connect the story, alternating between the game's virtual scenes and real-life locations in Shanxi. This interplay effectively reinforces Shanxi's cultural image on a visual level. The conceptual representation in the video relies primarily on a classification process, see Figure 1 for an example, showcasing Shanxi's cultural heritage, such as temples, towers, and grottoes. The detailed presentation of these landmarks, accompanied by English subtitles, not only enhances the dissemination of cultural symbols but also provides an accessible entry point for international audiences to understand Shanxi's rich heritage.

Furthermore, the video effectively captures interactive meaning. For instance, a close-up shot of a Buddha statue opening its eyes, fostering an emotional connection between the viewer and the statue while amplifying the mystique of Buddhist culture (See Figure 2). Figure 3 and Figures 4 contrast how long shots and tilted camera angles depict architectural

scenes. These techniques heighten the immersive experience, allowing viewers to deeply appreciate the grandeur and historical significance of Shanxi's ancient structures. Such visual strategies contribute to creating a dynamic and engaging cultural narrative.



Figure 1. Guangsheng Temple



Figure 2. Nanchan Temple



Figure 3. Guanque Tower



Figure 4. Guanque Tower

The sound design in the video plays a pivotal role in enhancing narrative coherence and emotional engagement. The voiceover provides essential background information, filling gaps in the visual narrative while reinforcing connections between the game and Shanxi's cultural landmarks. The video creates a cohesive

storytelling framework by integrating the voiceover with visual elements. The background music, particularly the iconic melody from *Journey to the West*, intensifies emotional involvement during the narrative's climax. This culturally symbolic music enriches the narrative's pacing and deepens the association between the game and Shanxi's cultural heritage through the interplay of auditory and visual symbols. Sound design's diversity and cultural significance amplify the audience's emotional experience[9].

Building on textual, visual, and auditory elements analysis, this study examines the synergy among multimodal components. The findings reveal that these elements work interdependently to construct a unified cultural narrative. First, the collaboration between symbols is evident. Text, visuals, and sound do not exist in isolation; instead, they dynamically interact to convey cultural meaning. For instance, close-up shots of the Buddha statue communicate profound cultural significance, while the voiceover and background music enhance the narrative and emotional resonance of the visual imagery. These elements strengthen the audience's connection to Shanxi's cultural identity. Moreover, multimodal forms engage multiple sensory channels simultaneously, offering a comprehensive experience. Visual details illustrate cultural scenes, sound evokes emotional responses, and subtitles provide supplementary information. This integrated approach may enhance the audience's overall reception of the content, effectively conveying Shanxi's rich cultural heritage.

5. Results

RQ1: What are the similarities and differences between the effects of single-modal and multimodal communication in promoting Shanxi's regional image?

In this study, single-modal communication forms, such as news articles, excel in providing detailed cultural and historical context through structured narratives and intertextuality. This approach is well-suited for in-depth understanding. However, it lacks the sensory appeal necessary to engage audiences on an emotional level. The text-heavy format primarily presents factual descriptions, often missing emotional elements and vivid details. As a result, audiences may remain at a cognitive level of understanding without forming a deeper

emotional connection to the city's image. In contrast, multimodal communication formats, such as promotional videos, integrate visual, textual, and audio elements to evoke emotional resonance and foster cultural identity. The combination of virtual characters and real-world landmarks delivers a powerful visual impact. This dynamic approach showcases Shanxi's ancient culture and merges it with contemporary technology and gaming culture, highlighting Shanxi's modernisation and unique urban charm. Despite shortcomings in the coherence of subtitles, the video's rich visuals, interactive visual elements, and culturally symbolic background music may enhance its appeal, particularly to international audiences. In summary, single-modal communication emphasises depth, while multimodal communication focuses on breadth and emotional engagement. Together, they form a more comprehensive strategy for effectively promoting Shanxi's regional image.

RQ2: How can the MDA framework contribute to constructing regional images and cross-cultural communication?

The MDA framework demonstrates significant advantages in constructing cohesive and immersive narratives by integrating multiple symbolic resources, including text, visuals, and audio. Using the promotional video as an example, the MDA framework effectively combines game-inspired visuals with real-world landmarks, blending traditional cultural elements with modern aesthetics to enhance Shanxi's global appeal. Additionally, the framework leverages universally understandable visual storytelling and multilingual support, such as English subtitles, to foster cross-cultural understanding and reach international audiences. Beyond this, the MDA framework facilitates collaboration between the tourism and creative industries. For instance, partnerships between tourism management departments and game developers can create authentic and emotionally engaging city narratives. As a result, the MDA framework provides a scalable model for promoting culturally rich cities like Shanxi on the global stage, offering an effective solution for enhancing their visibility and cultural impact.

6. Discussion

This study analyses Shanxi's regional image construction and compares it with other Chinese cities that have successfully built cultural

tourism brands, such as Xi'an and Harbin. This comparison provides a broader perspective on the role of MDA in regional image design and cross-cultural communication.

Xi'an's use of multimodal discourse in cultural tourism promotion is particularly noteworthy. According to Wang and Feng, Xi'an has successfully told its city story by releasing a series of cultural promotion videos on the Douyin platform [17]. These videos combine visual, auditory, and textual modes with popular cultural elements. Not only do they showcase historical artefacts and cultural landscapes, but they also blend traditional culture with modern pop culture. For instance, the promotional videos incorporate popular artists and music, attracting a younger audience and transforming what was once considered a traditional regional image into a more inclusive and vibrant one.

Harbin's case shows that multimodal communication extends beyond visual and auditory modes to enhance the perception of touch and taste, further increasing the appeal of a city's image. Harbin has effectively promoted its "Snow World" theme on social media to attract southern tourists, integrating the city's characteristics into a multisensory experience that includes social atmosphere and food culture. Furthermore, Harbin's tourism videos highlight the city's welcoming personality. Wang points out that Harbin has worked hard to create a "differentiated" brand that attracts southern tourists to experience the cold northern climate and taste a diverse range of cuisines, combining tactile and olfactory experiences to create a distinctive cultural tourism brand [16].

The successful experiences of Xi'an and Harbin demonstrate that multimodal communication not only effectively reaches local audiences but also enhances cross-cultural communication through the international expression of multisensory elements and symbols. Xi'an attracts viewers from different cultural circles by using English subtitles and international symbols, while Harbin facilitates cross-cultural communication by promoting its unique northeastern culture to southern tourists. These cases offer important insights into Shanxi's regional image promotion. Although Shanxi's video attempts to reach an international audience by including English subtitles, the simplicity of the subtitles and the lack of background information limit its effectiveness in cross-cultural communication. Moreover, the video focuses primarily on the

visual representation of cultural imagery, neglecting the universality of symbols, which makes it difficult for international audiences unfamiliar with Chinese culture to understand its cultural significance fully. In the future, Shanxi could further optimise its cross-cultural communication strategy by enhancing multimodal integration, improving narrative coherence, and adapting symbols for broader international appeal, thereby expanding its cultural attraction.

7. Conclusion

7.1 Research Conclusion

Employing the MDA framework, this study examined the similarities and differences between the promotional video Follow Wukong to Shanxi and traditional single-modal news reports in shaping Shanxi's regional image. It also explored the potential of multimodal communication in cross-cultural contexts. The findings reveal that multimodal forms of communication, by integrating textual, visual, and auditory resources, indicating the effectiveness of regional image promotion. Specifically, multimodal approaches strengthen the presentation of cultural symbols and foster audience engagement and cultural identification through interactivity and emotional resonance. To successfully narrate the story of a Chinese city, it is essential to identify its unique cultural features and leverage popular media platforms to integrate multiple sensory modes-visual, auditory, tactile, and olfactory-for a more impactful outcome. This form of communication plays a crucial role in enhancing the cultural appeal of Chinese cities within a globalised context.

At the same time, the study highlights the distinctive role of MDA in cross-cultural communication. By blending traditional cultural elements with modern aesthetics, the MDA framework effectively activates multisensory experiences, enabling broader international dissemination. However, the findings also indicate that the success of multimodal communication depends on the logical integration and universal representation of symbolic resources. While the Shanxi case demonstrates potential in cross-cultural communication, further improvements are needed in areas such as narrative coherence, the adaptability of symbols, and the diversification

of multisensory experiences to achieve greater cultural resonance.

7.2 Significance of the Study

This study provides valuable insights into the use of multimodal communication with "game + cultural tourism" model for regional image promotion, particularly in the case of Shanxi. By applying the MDA framework, the research highlights how integrating text, visuals, and audio can enhance the effectiveness of cultural storytelling. The findings demonstrate that multimodal communication strengthens the representation of cultural elements and fosters emotional connections with diverse audiences, making it a powerful tool for city branding.

Furthermore, this research contributes to the practical understanding of how cities like Shanxi can refine their strategies for cross-cultural communication. By comparing Shanxi's efforts with successful examples from Xi'an and Harbin, the study emphasises the importance of incorporating local cultural uniqueness, improving narrative coherence, and adapting symbolic representations to appeal to international audiences. These insights provide practical recommendations for enhancing the global appeal of culturally rich cities, supporting their efforts to balance tradition with modernity while reaching a broader audience.

7.3 Limitations of the Study

This study faced several limitations when applying the framework of MDA. One notable drawback of MDA is its inherent subjectivity in grammatical analysis. Due to the absence of strict grammatical markers and linear relationships, analysts with different cultural backgrounds, knowledge structures, or motivations may interpret the same semiotic resources differently. This variability can affect the consistency and objectivity of the analysis, and the findings of this study are not immune to such subjectivity.

Additionally, the research and application framework of MDA remains underdeveloped. Currently, MDA is primarily led by scholars in systemic functional linguistics. However, as an interdisciplinary field, it requires the integration of methodologies and perspectives from disciplines such as psychology and sociology. Such integration is crucial for comprehensively understanding the complex meanings embedded in multimodal discourse. However, this study

has not systematically incorporated these interdisciplinary approaches, which limits the breadth and depth of its analytical outcomes.

References

- [1] Chen, S. (2015). Shaping of local city image under new media context. *2015 International Conference on Economics, Management, Law and Education*, 207–210. <https://www.atlantispress.com/proceedings/emle-15/25846459>
- [2] Consalvo, M. (2022). *Atari to Zelda: Japan's videogames in global contexts*. MIT Press.
- [3] Halliday, M. A. K. (1978). *Language as social semiotic*. Edward Arnold London. <https://edisciplinas.usp.br/mod/resource/view.php?id=5033119#page=274>
- [4] Hou, Y. (2024). From virtual journeys to real travels: Exploring the impact of Chinese-themed digital games on international players' tourism intentions. *Modern Economics & Management Forum*, 5(5), 910–915. <https://doi.org/10.32629/memf.v5i5.2882>
- [5] Kress, G., & Leeuwen, T. van. (2001). *Multimodal discourse: The modes and media of contemporary communication*. Bloomsbury Academic. <https://zh.z-library.sk/book/3430336/b05f4a/multimodal-discourse-the-modes-and-media-of-contemporary-communication.html>
- [6] Kress, G., & Van Leeuwen, T. (2002). *Colour as a semiotic mode: Notes for a grammar of colour*. *Visual Communication*, 1(3), 343–368. <https://doi.org/10.1177/147035720200100306>
- [7] Luca, I. S. (2020). A multimodal discourse analysis in media. *Romanian Journal of English Studies*, 17(1), 74–80. <https://doi.org/10.1515/rjes-2020-0009>
- [8] Ly, T. H., & Jung, C. K. (2015). Multimodal discourse: A visual design analysis of two advertising images. *International Journal of Contents*, 11(2), 50–56. <http://dx.doi.org/10.5392/IJoC.2015.11.2.050>
- [9] New China TV. (2024, August 22). *Explore the real-world cultural landmarks of "Black Myth: Wukong"* [Video]. YouTube. <https://www.youtube.com/watch?v=q5IBs24dAAg>
- [10] Nurudeen, M. A., Ogungbe, E. O., & Zakariyah, M. (2021). A socio semiotic approach to multimodal discourse of selected Nollywood film advertisement posters. *Studies in Pragmatics and Discourse Analysis*, 2(2), 31–45.
- [11] O'Halloran, K. L., Tan, S., Smith, B. A., & Podlasov, A. (2011). Multimodal analysis within an interactive software environment: Critical discourse perspectives. *Critical Discourse Studies*, 8(2), 109–125. <https://doi.org/10.1080/17405904.2011.558687>
- [12] Oakes, T. (2021). China's urban ideology: New towns, creation cities, and contested landscapes of memory. *Urban China Reframed*. Routledge. <https://www.taylorfrancis.com/chapters/edit/10.4324/9781003147367-4/china-urban-ideology-new-towns-creation-cities-contested-landscapes-memory-tim-oakes>
- [13] Wang, Q., & Zhang, D. (2019). Branding city destinations: A comparative analysis of the multimodal discourse in Beijing and London publicity films. *Chinese Semiotic Studies*, 15(3), 379–410. <https://doi.org/10.1515/css-2019-0022>
- [14] Wang, X. (2022). The persuasive mechanism of Wuhan city image videos based on multimodal discourse analysis: A diachronic study under the background of COVID-19. *Journal of Innovation and Social Science Research*, 2591, 6890.
- [15] Wang, X., & Lei, X. (2023). Discourse analysis of Chinese provincial and regional foreign promotional videos in the framework of Fairclough's three-dimensional model. *International Journal of Languages, Literature and Linguistics*, 9(5), 398–430. <https://doi.org/10.18178/ijll.2023.9.5.438>
- [16] Wang, Y. (2024). From "online star city" to "long-term famous city": Research on the strategy of city image communication in the new media era-Taking Harbin, the "online star city", as an example. *Advances in Education, Humanities and Social Science Research*, 10(1), 136.
- [17] Wang, Y., & Feng, D. (2023). History, modernity, and city branding in China: A multimodal critical discourse analysis of Xi'an's promotional videos on social media. *Social Semiotics*, 33(2), 402–425. <https://doi.org/10.1080/10350330.2020.1870405>
- [18] Yao, Y., & Zhuo, Y. (2018). A multimodal discourse analysis of the promotional video of Hangzhou. *English Language Teaching*, 11(10), 121.

- <https://doi.org/10.5539/elt.v11n10p121>
- [19] Zhou, L. (2020). A multi-modal discourse analysis on D&G's advertisements. *International Journal of Languages, Literature and Linguistics*, 6(2), 99–102.
- [20] Zhu, Y. (2007). Theory and methodology of multimodal discourse analysis. *Foreign Language Research*, 5, 82–86.
- [21] Zhu, X. (2024). Shanxi scenic area offers free return visits. *Explore real-world cultural landmarks of "Black Myth: Wukong."* China Daily. <https://www.chinadaily.com.cn/a/202410/09/WS6705dec2a310f1265a1c6990.html>
- [22] Zukin, S. (2012). Competitive globalization and urban change: The allure of cultural strategies. *Rethinking Global Urbanism*. Routledge. <https://www.taylorfrancis.com/chapters/edit/10.4324/9780203118368-3/competitive-globalization-urban-change-sharon-zukin>