

Exploring the Visual Narrative Strategies of Chinese Cultural Image in Silk-Themed Reports by Japanese and Korean Media through Multimodal Discourse Analysis

Xuyang Fan

Zhixing College of Hubei University, Wuhan 430011, Hubei, China

Corresponding author: Xuyang Fan (e-mail: dawsonfan@163.com)

ABSTRACT Existing studies have not provided sufficient comparative and critical analysis of how Japanese and South Korean media construct Chinese cultural images through visual narratives. This study develops a multimodal discourse analysis framework to formally decode the visual grammar of Japanese and South Korean silk-themed reports, incorporating the symbolic semantics of traditional textile patterns and fabric textures to examine intermodal dynamics such as synergy and tension. By constructing strategy types that connect silk craftsmanship and garment aesthetics with historical memory and cultural identity, the study reveals the ideological basis of multimodal features in East Asian cultural expression. Experimental analysis shows significant differences between Japanese and South Korean media in the use of “aesthetic othering” and “competitive juxtaposition” as cultural-image construction strategies. In Japanese media, aesthetic othering accounts for 51.7% of strategies, while in South Korean media it accounts for 37.5%. Canonical correlation analysis further indicates that Japan’s aesthetic othering strategy is closely related to tensions between historical memory and visual characteristics, whereas South Korea’s competitive juxtaposition strategy is more closely associated with economic competition. The findings show how the two countries negotiate cultural images through differentiated visual strategies.

INDEX TERMS multimodal discourse analysis, visual narrative, chinese cultural image, silk reports, textile craftsmanship

I. INTRODUCTION

IN the context of increasingly visual and digital global cultural communication, the construction of a national cultural image transcends reliance on single texts and is deeply intertwined with the discourse practice of traditional textile motifs and fabric textures constructed collaboratively by multimodal symbols. These weaving patterns and sartorial elements serve as essential visual anchors that integrate colors and layouts to communicate a cohesive and authentic national identity. Silk, as a core material symbol of Chinese culture and a historical link in cultural exchange, has become a key window for observing others’ perceptions and attitudes towards Chinese culture through its international media representation [1,2]. Focusing on Japan and South Korea, two countries that belong to the Confucian cultural sphere, are geographically close to China, and have a deep cultural and historical connection, the mainstream media’s coverage of the silk theme provides a research field with great tension and comparative value for analyzing the mechanisms of crosscultural visual narrative and image

construction. It is acknowledged that within Japanese and Korean cultures, silk has also developed unique indigenous connotations and historical trajectories. This study, however, focuses specifically on the media representation of silk as a symbol predominantly associated with Chinese culture in the context of cross-cultural reporting. The analysis examines how Japanese and Korean media, as external observers within a shared yet differentiated East Asian cultural sphere, frame this symbol. This focus allows us to trace the discursive strategies employed in negotiating cultural proximity and distance, without negating the complex local meanings silk holds within each domestic context.

Existing research faces several difficulties: In terms of the object dimension, most studies treat “international media” as a homogeneous whole or focus on mainstream Western media, failing to conduct systematic critical comparisons of media from specific neighboring countries such as Japan and South Korea, which are more complex in terms of geopolitics and cultural psychology, resulting in insufficient explanation of the differentiated strategies and motivations

for the construction of cultural images [3,4]. In terms of the method dimension, although multimodal discourse analysis has become the mainstream paradigm, existing applications mostly remain at the level of descriptive listing of visual elements or internal interpretation of single cultural texts, failing to effectively connect formal analysis with ideological critique, and lacking an operational analytical framework that rigorously links micro-modal characteristics (such as shots, colors, and image-text relationships) with macro-national contexts (such as historical complexes and identity anxieties) [5,6]. In terms of the depth of the problem, related studies often stop at identifying the "positive" or "negative" attributes of cultural images, failing to deeply reveal the ideological operation and power negotiation logic hidden behind visual narratives, thus making it difficult to reveal the deep dynamic mechanism of image construction [7,8].

To address these challenges, existing literature has attempted to advance the field through various paths, but all suffer from limitations in addressing the core issues. Research on cultural symbols has focused primarily on the relationship between silk, as a symbol, and the cross-cultural communication history of that relationship (a "civilizational bond"). Some previous research has identified significant similarities in how both cultures value silk and how they express that value through their media representations. However, the existing research does not adequately explore the narrative strategies used by contemporary media producers to convey their ideologically charged images of silk [9,10]. The content analysis approach and framing methodology applied to media image research have confirmed that the shapes of images in various countries differ because of variations in reporting tendencies and theme distributions [11,12]. Thus, neither approach has effectively explored the subtleties of visual rhetoric and multimodal synergy. Kress and van Leeuwen developed the idea of "visual grammar" to provide a more systematic way for researchers to understand how people create and decode visual messages. Several researchers have applied this framework to specific fields (e.g., news media or advertising), allowing those researchers to analyse the meanings created by a viaduct's representations, interactive elements, and composition. However, in applying visual grammar, few researchers have explicitly considered the socio-political contexts within which those visual structures are created and altered [13,14]. As a result, very few studies will provide answers regarding how Japanese and South Korean Media use multimodal discourse (i.e., Pictures and texts) to imagine and define Chinese culture, particularly given the respective ideological perspectives and historical experiences of both Japan and South Korea.

By analyzing the visual narratives surrounding the Silk Road as represented in different media from Japan and South Korea through the use of critical multimodal discourse analysis methods grounded in context, we systematically identify and compare how these two cultures have represented the Silk Road through their respective news outlets while also identifying what ideologies may have motivated

specific representations. The first component of this analysis method entails formally annotating each of the visual samples using Visual Grammar and then examining how the textual and visual components relate with one another (i.e., how images and text work together to convey a particular message). The second component of this analysis method involves an Intermodal Tension Analysis, where we identify systematic patterns of modal synergy, extension, or potential tension and complementarity between images and text and between dominant visual elements and those that served a secondary function, as a method for revealing the ideological underpinnings of representations. Subsequently, it constructs a typology of visual narrative strategies and anchors them within a specific national ideological context constituted by historical memory, cultural identity, and contemporary relations for causal explanation. This paper achieves an empirical comparison of the systematic differences in visual strategies employed by Japanese and South Korean media, integrating the semiotic analysis of traditional textile aesthetics and garment structures to methodologically advance multimodal analysis from formal description to ideological critique. By decoding the symbolic use of fabric textures and weaving motifs, the study provides a new analytical path for understanding the discourse politics and power negotiations of cultural images between neighboring countries.

II. A CONTEXT-DRIVEN MULTIMODAL AND IDEOLOGICAL LINKAGE ANALYSIS FRAMEWORK

A. OVERALL ANALYTICAL FRAMEWORK: THEORETICAL LINKAGE AND OPERATIONAL PATH

This paper constructs a three-tiered "context-driven multimodal-ideological linkage analysis framework" to systematically deconstruct the differentiated composition and underlying motivations of visual narrative strategies in Silk Road-themed reporting by Japanese and South Korean media. The structure and process of the three-tiered linkage are shown in Figure 1.

This framework follows a progressive logic from micro-modal feature description to macro-ideological interpretation. The first layer is the multimodal data layer, which systematically collects and structurally preprocesses heterogeneous symbolic resources such as images, text, and layout from original media reports, forming an analyzable basic data set. The second layer is the formal analysis and strategy identification layer. At this layer, multimodal data is formally labeled and encoded strictly according to visual grammar theory. By calculating the semantic relationship strength between text and image modalities and performing cluster analysis, recurring and structurally stable multimodal patterns are identified, and these patterns are then summarized and defined as visual narrative strategy types with clear theoretical orientations. The third layer is the contextual interpretation and motivational association layer. This layer places the identified visual strategies within a specific national context constituted by historical memory, cultural identity claims, and contemporary bilateral relations

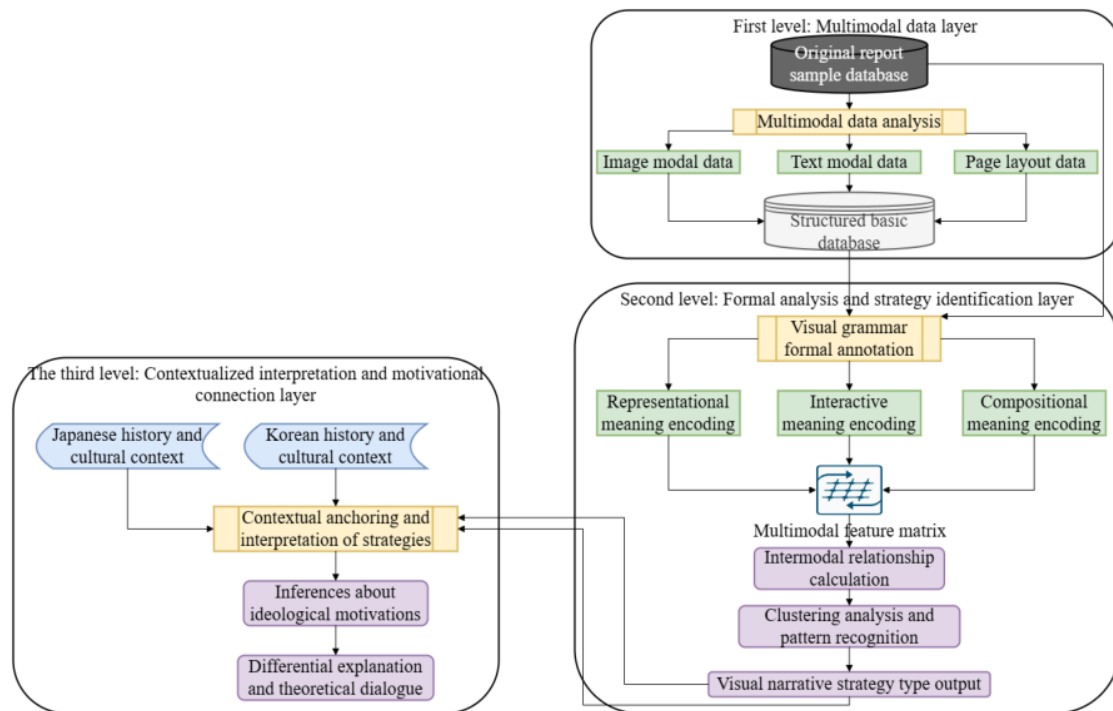


FIGURE 1. Context-driven multimodal-ideology linkage analysis framework.

for critical interpretation. The logical mapping and abductive reasoning between contextual elements and strategy features allow for a better understanding of the underlying ideological operational mechanisms of how strategy selection occurs. The full framework provides a closed-loop analytical pathway from data to interpreting discourse.

B. SAMPLE COLLECTION AND PREPROCESSING

The collection of the sample for this study complied with a systematic approach and took into account both representativeness and balance in order to create a basis for comparison within this analysis, as well as assist with future comparisons by enabling identical data algorithms. The primary focus of the sample was to be “silk” through a view towards the major mass communication channels from Japan and South Korea’s most influential media, with large reach to a large audience, online and traditional, through the use of Japanese Asahi Shimbun, Yomiuri Shimbun, and Japanese NHK [15-17], and South Korea Chosun Ilbo, JoongAng Ilbo, and KBS (Korean Broadcasting System) [18,19]. The time period during which the samples were collected spanned from 2013 through 2023, including the critical first 10 years of the Belt and Road Initiative and its resulting influence on modern media narratives.

The two-tiered standardized evaluation procedure for compiling a collection of papers to create the definitive analysis library for the attached document are done in two parts. The first-tier evaluation screened titles and abstracts to exclude reports where silk was merely a peripheral or inciden-

tal mention, while retaining those where silk functioned as a core topic, a detailed cultural representation, an economic product, a historical heritage, or a central metaphor. The second-tier evaluation then removed brief text-only news flashes or items with missing/compromised primary visual components, ensuring the final corpus consisted of multimedia reports with complete visual-textual elements. For reports employing silk primarily as a metaphor, an additional annotation flag (“Metaphorical Core”) was applied during preprocessing to enable subsequent comparative analysis within the multimodal strategy typology. The combined results of these screening processes will ultimately produce a sample library with a manageable number of articles, focused thematic content, and greatest modal completeness for each article. Additionally, the final analysis will be performed on an equitable basis by controlling for the number of articles included from both Japanese and South Korean journalists.

During the initial preprocessing steps, the selected individual reports become an organized analytical unit. To do this, the metadata (date published, news source, etc.) and raw text are extracted from the report and stored. For core multimodal content, relevant news graphics/images are also extracted, and the pertinent textual explanation/figure titles are combined and placed in alignment with the image(s). It is also important to retain the page layout by storing the relative location and sizing of the image(s) in the HTML (Hyper Text Markup Language) or PDF (Portable Document Format) format. Therefore, all of the raw unstructured media

become structured data objects suitable for formal annotations which will eventually be processed systematically within the analysis layer.

C. FORMAL ANNOTATION AND VISUAL GRAMMAR DECODING OF MULTIMODAL DATA

Based on the visual grammar theory of Kress and van Leeuwen, this study constructs a systematic multimodal feature coding system to formally annotate and quantify the images, layouts, and image-text relationships in the samples [20,21]. The core annotation covers the representational meaning of the image (participants and vectors in processes such as actions and reactions), interactive meaning (contact, social distance, perspective, and modality), and compositional meaning (information value, framing, and salience). The logical semantic relationship between the image and text is determined based on the status relationship and logical connection type. The complete coding scope, operation definitions, and rules are shown in Table 1.

The labeled qualitative data were transformed into a numerical feature matrix. Categorical variables were encoded using one-hot encoding, while continuous variables such as visual area proportion were directly recorded [22,23]. The semantic association strength between images and text was quantified by calculating the cosine similarity between the image keyword vector V_i and the text keyword vector T_j , using the formula [24,25]:

$$\text{Sim}_{\cosine}(V_i, T_j) = \frac{V_i \cdot T_j}{\|V_i\| \|T_j\|} \quad (1)$$

As shown in equation (1), each V_i and T_j are derived from a pre-trained word embedding model; therefore, the resulting similarity value (i.e., Simcosine) serves as an objective metric quantifying the degree of semantic alignment (synergy) or divergence (tension) between the image and text modalities. Using this metric provides a foundation for recognizing tension across modalities. Each sample will then have a multidimensional feature representation that serves as input for strategy recognition analysis.

D. TYPOLOGICAL CONSTRUCTION OF VISUAL NARRATIVE STRATEGIES

Using the individual annotations and an aggregate multimodal Feature Matrix as the basis for determining stable types of visual narrative strategies through induction from a multi-modal Feature Matrix, combining a theoretical model through the data-driven approach. Standardized feature matrix clustering utilised a Silhouette coefficient and elbow rule methodology in determining a value for k , then utilised the K-means++ method to partition clusters; the K-means++ method attempts to minimise the total within-cluster Euclidean distance squared from within-cluster samples to each cluster centroid c_j [26,27], with an objective function defined as:

$$J = \sum_{j=1}^k \sum_{x \in C_j} \|x - c_j\|^2 \quad (2)$$

In formula (2), x represents the feature vector of a sample, and c_j represents the j -th cluster. This formula divides the samples into k -th cohesive groups through iterative optimization, and each group represents a potential multimodal pattern. Subsequently, the clustering results are theoretically interpreted. By analyzing the salient features of each cluster center in terms of representation, interaction, composition, and text-image relationships, they are associated with and named according to established critical discourse theories. For example, one cluster is interpreted as an “aesthetic othering” strategy due to its high social distance, conceptual representation, and detailed text-image relationships; another cluster is defined as a “competitive juxtaposition” strategy due to its high visual contrast and juxtaposed representational features, thus constructing a strategy typology for subsequent comparative analysis.

E. ANALYSIS OF THE RELATIONSHIP BETWEEN CONTEXTUALIZED INTERPRETATION AND IDEOLOGICAL MOTIVATIONS

To reveal the underlying motivations behind the choice of visual narrative strategies, this study analyzes the correlation between the strategy types output at the second level and the specific macro-sociocultural contexts of Japan and South Korea. In this study, ‘intermodal tension’ is formally defined as a state of semantic divergence, incongruity, or contradiction between the visual and textual modalities within a multimodal discourse, which may serve to reveal underlying ideological positions. This analysis aims to go beyond strategy description and establish a causal explanation from micro-discourse practices to macroideology. The specific analysis employs abductive reasoning to construct a correlation model between strategy features and contextual elements. First, based on historical documents, diplomatic policy texts, and cultural identity research, key contextual variables for both countries are extracted and quantified, forming a contextual feature vector $C = [c_1, c_2, \dots, c_m]$. Simultaneously, each strategy type is represented as a vector of its mean features across all dimensions of visual grammar F_p . The strength of association is quantitatively assessed by calculating the canonical correlation between the strategy feature vector F_p and the context feature vector C . The formula for calculating the canonical correlation coefficient ρ is:

$$\rho = \max_{a,b} \frac{a^T \Sigma_{12} b}{\sqrt{(a^T \Sigma_{11} a)(b^T \Sigma_{22} b)}} \quad (3)$$

Σ_{11} and Σ_{22} are the covariance matrices of F_p and C , respectively, and Σ_{12} is their cross-covariance matrix. The weight vectors a and b are obtained by solving the eigenvalue problem that maximizes the linear correlation

TABLE 1. Multimodal Feature Coding System Table

Analytical Dimension	Core Category	Operational Definition and Coding Rule	Quantification/Categorization Method
Representational Meaning	Narrative Process	Code for the presence of a “vector” (eg, gaze, gesture) forming an action line. Identify “Actor”, “Goal”, and “Circumstance”.	Categorical Coding: Process Type: Action / Reaction / None; Participant Role: Actor(A), Goal(G), None(N)
Representational Meaning	Conceptual Process	Code for images presenting classification, analytical, or symbolic meaning. Identify “Carrier”, “Symbolic Attribute”, and relational types.	Categorical Coding: Process Type: Classification / Analytical / Symbolic / None; Relation Type: Possessive / Implicational
Interactive Meaning	Contact	Determine if figures directly gaze at the viewer, establishing a “demand” or “offer” relation.	Binary Coding: 0 = Offer (no direct gaze); 1 = Demand (direct gaze)
Interactive Meaning	Social Distance	Determine the virtual interpersonal distance based on shot type (close-up, medium shot, long shot).	Ordinal Coding: 1 = Intimate (close-up); 2 = Social (medium shot); 3 = Public (long shot)
Interactive Meaning	Perspective	Determine if the camera angle is horizontal (eye-level) or vertical (high/low), affecting power dynamics.	Categorical Coding: Horizontal: Eye-level; Vertical: High-angle / Low-angle
Interactive Meaning	Modality	Assess the credibility and affective tone by evaluating color saturation, palette, and contextual detail.	Scale Rating: 1 (Low modality) - 5 (High modality)
Compositional Meaning	Information Value	Determine the informational salience (Given/New, Ideal/Real) based on element placement (left/right, center/margin).	Positional Code: Left / Right / Center; Top / Bottom
Compositional Meaning	Framing	Determine the degree of connection/separation within/between images and text via dividing lines, color blocks, or space.	Categorical Coding: Strong Framing / Weak Framing / No Framing
Compositional Meaning	Salience	Determine the relative visual weight of elements based on size, foreground/background contrast, and color contrast.	Relative Area Ratio: Subject Area / Total Area; Contrast Value: RGB Standard Deviation
Image-Text Relation	Status Relation	Determine if the relation is equal (complementary) or unequal (illustration, anchorage).	Categorical Coding: Equal / Image-dominant / Text-dominant
Image-Text Relation	Logical Semantics	Determine if the relation is one of Elaboration, Extension, Enhancement, or Projection.	Categorical Coding: Elaboration / Extension / Enhancement / Projection

between the two sets of variables. The range of the canonical correlation coefficient ρ is [0, 1]. The larger the value, the stronger the overall linear correlation between the visual strategy pattern and the contextual features of the target country, providing statistically based motivational evidence for the differentiated selection of strategies. Ultimately, by combining highly relevant quantitative evidence with qualitative discourse analysis of specific historical events and cultural debates, the causal logic of specific strategies is constructed, achieving a leap from formal analysis to social explanation.

III. RESEARCH DESIGN: PROCESS, RELIABILITY, VALIDITY, AND ASSESSMENT

A. OVERVIEW OF EXPERIMENTAL SAMPLES AND DATA

Based on the established collection and screening criteria, this study constructed a balanced sample database for empirical analysis. The database contains 240 valid reports, 120 each from Japanese and South Korean media, with the reports evenly distributed between 2013 and 2023. Both Japan and South Korea have three of the largest mainstream news outlets as a sample to establish both the credibility and broad reach of the sources utilized in this project. The detailed breakdown of these samples from both countries is found in Table 2.

The distribution of samples coming from the media organisations and dates from both countries can be seen in Table 2. Both Japan and South Korea had equal samples for both time spans and total number of samples; therefore any biases that may have been introduced from utilizing an unequal number of samples or a sample concentrated over a single-time-span have been removed from the data available for further comparison between the two countries. Additionally, since the three largest mainstream media in each country

were equally used to obtain the analysis. As such, the results of the analysis will demonstrate the overall tendency of the mainstream media in each country, as opposed to the unique viewpoint of each individual media organisation used for the analysis.

B. ANALYSIS PROCESS AND RELIABILITY AND VALIDITY ASSURANCE MECHANISM

This study has established and followed a formal annotation process for the purpose of guaranteeing objectivity and reliability. A predetermined analysis protocol was established through an agreed-upon analysis approach. The specific analysis process used in this research is described as follows: First, 30 random coding samples were independently coded by two systematically trained coders using the coding system. The inter-coder reliability was assessed based on the initial coding descriptors, and all discrepancies were discussed and calibrated between the coders prior to establishing consensus on the coding manual. Finally, the two coders performed independent formal coding of all 240 samples. In order to quantitatively evaluate the coding consistency of the two coders, the Coen Kappa Coefficient was utilized as a measure of coding consistency [28,29]. This coefficient indicates the consistency achieved by the two coders while controlling for random coding agreement; therefore, it indicates high reliability of the coding results and provides a sound basis of reliable data for subsequent analyses.

C. CORE ANALYSIS DIMENSIONS AND EVALUATION INDICATORS

In order to systematically compare the various visual narrative strategies used in Japanese and South Korean media this study has defined four primary analytical areas and created

TABLE 2. Detailed Distribution of Experimental Samples

Country	Media Outlet	Number of Reports	Time Span	Sample Proportion
Japan	The Asahi Shimbun	40	2013-2023	33.3%
Japan	The Yomiuri Shimbun	40	2013-2023	33.3%
Japan	NHK News Web	40	2013-2023	33.3%
Subtotal (Japan)		120	2013-2023	100%
South Korea	The Chosun Ilbo	40	2013-2023	33.3%
South Korea	The JoongAng Ilbo	40	2013-2023	33.3%
South Korea	KBS News	40	2013-2023	33.3%
Subtotal (South Korea)		120	2013-2023	100%
Total		240	2013-2023	—

a set of measurable, actionable and verifiable evaluation indicators for these elements.

1) Strategy distribution dimension

The frequency and proportion of various visual narrative strategies were statistically analyzed, and the chi-square test was used to evaluate the significant differences in the distribution between Japan and South Korea [30,31].

2) Modal relation dimension

By encoding and intensity calculation of the logical semantic relations of textimage pairs and comparing the distribution of their text-image relation types, the differences between the two countries in discourse integration are quantified.

3) The contextual association dimension

The contextual association dimension is used to quantitatively assess the correlation strength between strategy selection and macro-level motivations by calculating the canonical correlation coefficient ρ between visual strategy feature vectors and national context feature vectors and testing its significance.

4) Overall validity dimension

By calculating the average silhouette coefficient of the two countries' samples in the multimodal feature space, the overall separability of their visual narrative patterns is assessed, and the systematic nature of the differences is confirmed [32,33].

IV. RESEARCH FINDINGS: STRATEGY DISTRIBUTION, MODAL TENSION, AND MOTIVATION ARE RELATED

Before analyzing the results, this study tested the reliability of the multimodal feature coding. The Coen- Kappa coefficients for the two coders across all core analysis categories indicated a good level of consistency among the coders.

A. NATIONAL DISTRIBUTION AND DIFFERENCES IN VISUAL NARRATIVE STRATEGIES

To empirically examine the systematic differences in visual narrative strategies in Japanese and South Korean media,

this chapter focuses on four core strategies identified in the study: “aesthetic othering,” “deconstruction of modernity,” “competitive juxtaposition,” and “symbiotic connection,” conducting a quantitative distribution analysis. Through a statistical analysis of the frequency of the strategies that were used, as well as determining how many times those strategies were used within that nation (and dividing that number by the total number of occurrences of that strategy across all nations), it's possible to determine what the common preference was in relation to how each nation constructs its image (discourse practice). In order to analyse the differences between the two countries' distributions, a chi-square test was conducted, and the comparative results are presented in Figure 2.

As Figure 2 illustrates, the data offers a detailed view of the absolute and relative distributions of the various strategy types across countries. The large discrepancy between the Japanese media strategy of “aesthetic othering” (51.7% of the total strategies) compared to the South Korean media strategy of “aesthetic othering” (23.3%) shows a large concentration of the former as compared to the latter (62 vs. 28). This high concentration of “aesthetic othering” is due to the fact that within Japanese culture, a long-standing notion of “cultural diversion” has historically been linked to reporting on the history of Chinese silk culture; therefore, Chinese culture has been positioned as an “other” to be aesthetically appreciated. In contrast to the Japanese media, the distribution of South Korean media strategy is much more varied; the prominent strategy type is “competitive juxtaposition” (37.5% of the total strategies), which has occurred 45 times. In South Korea, reporting on Chinese elements is performed within the conceptual framework of comparing them to contemporary South Korean design and/or industries. The rationale for this reporting style is derived from the South Korean sense of identity in the cultural realm as a competitor and position itself as independent and modern while recognizing the historical and cultural connections through visual juxtaposition with China. Therefore, the discrepancy in the distribution of media strategies between the Japanese and South Korean media is an extension of differing historical and psychological perspectives on how to negotiate their images of China.

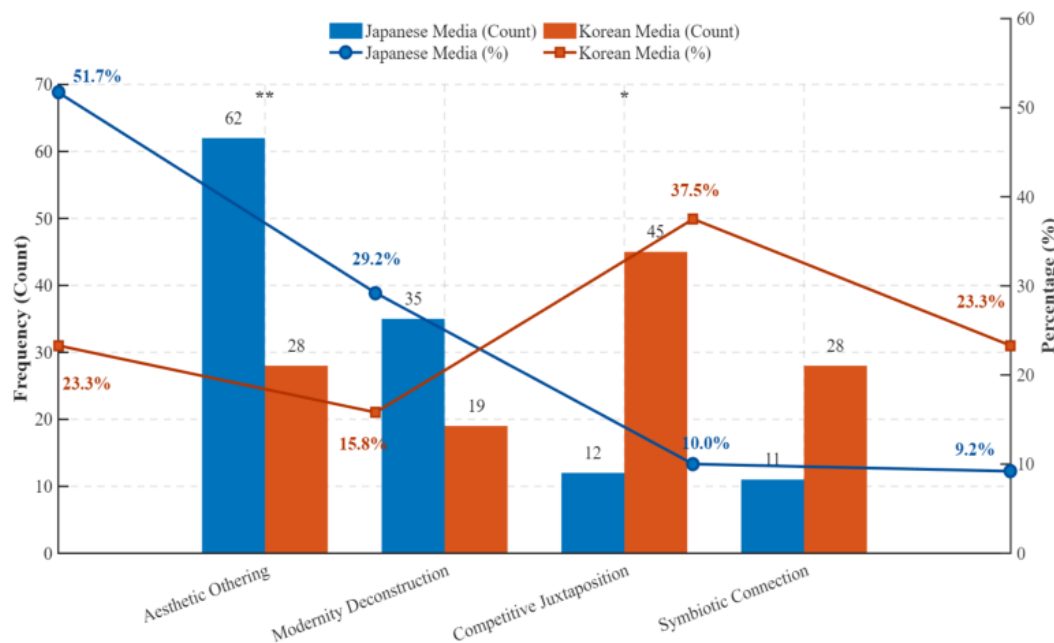


FIGURE 2. Comparison of the distribution of major visual narrative strategies in Japanese and South Korean media. “*” indicates that the result is statistically significant at the 0.05 significance level, and “****” indicates that the result is more significant at the 0.01 significance level.

B. IDENTIFICATION OF INTERMODAL TENSION PATTERNS AND ANALYSIS OF IDEOLOGICAL TRACES

After looking at the big picture regarding how visual narrative strategies are distributed, now the next step is to look closer into the micro-level workings of how these visual narrative strategies operate through discourse. Building on the strategy-level analysis, this section examines the micro-level intermodal dynamics by quantitatively assessing the logical-semantic relationships between image and text modalities. This analysis systematically identifies patterns of semantic synergy, extension, and potential tension, moving beyond a binary view of conflict to reveal the nuanced mechanisms of meaning construction. By encoding and numerically calculating the degree or strength of the logical semantic relationship for each image-text modal combination examined, systematic patterns of modal synergy or tension are identified between Japan and South Korea. The study includes four types of image-text relationships: “complementary reinforcement,” “extension and expansion,” “contrast paradox,” and “unidirectional dominance.” This analysis ultimately forms a quantitative description of the discourse integration strategies of the two countries, and its core comparative results are shown in Figure 3.

Figure 3 presents the analysis results. Figure a shows the frequency of occurrence of the four types of imagetext relationships, and Figure b shows the average intensity index of the corresponding relationships. Data shows that in the “extension and expansion” relationship, South Korean media

used it 68 times with an intensity index of 0.72, both significantly higher than the 45 times and 0.58 of Japanese media. The proactive use of image-text intertextuality in South Korean media allows for the creation of more complex, multidimensional narratives, which aligns with their greater emphasis on participatory and competitive cultural discourse strategies, whilst Japan’s use of image-text intertextuality has historically taken on a “contrast and paradox” logical approach. Based on the relative “intensity index scores” from the two “contrast and paradox” relationships, both countries show a relative low to medium intensity (Japan = 0.41; South Korea = 0.38). Furthermore, this analysis elucidates the differing modal integration logics employed. Japanese media tend toward a relatively stable and anchored meaning construction, often utilizing clear visual-textual correspondences. South Korean media, in contrast, demonstrate a more proactive and expansive discourse construction, frequently employing extension and elaboration to build complex, multilayered narratives. This aligns with their respective overarching strategic preferences identified earlier.

C. ANALYSIS OF THE CORRELATION BETWEEN NARRATIVE STRATEGIES AND CONTEXTUAL MOTIVATIONS

After completing strategy identification and modal relationship analysis, this study needs to causally link the micro-level visual patterns with their macro-level socio-cultural roots. This section reveals the underlying motivations be-

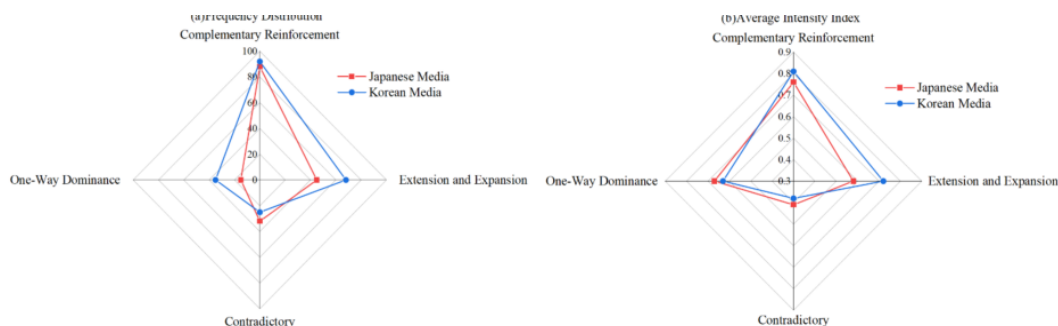


FIGURE 3. Distribution and Intensity of Text-Image Relationships in Japanese and Korean Media (a) Frequency of Occurrence (b) Average Intensity Index.

hind the differentiated visual strategy choices of Japanese and Korean media through quantitative statistics and cross-verification of typical cases, thereby moving beyond mere phenomenon description to achieve an explanatory understanding of the image construction mechanism. The analysis comprehensively utilizes canonical correlation analysis and qualitative evidence chain construction methods, and its core correlation pattern and argumentation path are shown in Figure 4.

The rows and columns of the matrix on the left side of Figure 4 represent four national context variables and four visual narrative strategies, respectively. The cell color depth and numerical value represent the correlation strength ρ . Data shows a strong correlation between the “aesthetic othering” strategy and Japan’s historical memory tension and cultural diversion consciousness, with ρ values of 0.82 and 0.78, respectively. This statistical relationship stems from Japan’s historical psychological structure formed during its modernization process, which viewed Chinese culture as a “classical other” to establish its own modern identity. The competitive Juxtaposition strategy correlates strongly with two of South Korea’s most contemporary forms of Economic Competition and Cultural Sovereignty: 0.76 and 0.71 respectively for the former and the latter. The underlying impetus behind South Korea’s relative comparisons is an active definition of Cultural Uniqueness and Economic Competitiveness via comparative narratives in the context of a Globalised and Regionally-dynamic World. The important/substantial connection between these two types of most statistically significant regressions has been clearly articulated on the right side of this image as a result of selected findings as stated below: Aesthetic Othering Strategy (Japan Historical Memory Tension) versus Competitive Juxtaposition Strategy (South Korea’s Competition with Japan/Cultural Sovereignty claims). Additionally, as stated above, the selection of these visual strategies and the statement that you have referenced “is rooted in” are further contextualized within and draw upon the logic of National Identity Negotiation and Geopolitical Culture and are not merely other forms of Media Behavior. As a result of this significance of both the historical development and contemporary aspirations of National Identity of South Korea and Japan, the visual strategies of both their respective media

cannot be understood as random selections.

D. COMPREHENSIVE COMPARISON AND IN-DEPTH DISCOURSE MECHANISM EXPLANATION

The results of the macro-pattern analysis discussed above will be used as a basis for an overall macro-pattern analysis to verify that all systematic and statistically robust differences and similarities between Japanese and Korean visual narratives are consistent with the above reported findings. To facilitate a comprehensive analysis of all Japanese and Korean visual narratives within the Multimode Feature Space (together with an evaluation of their cluster quality), I will conduct a Dimensionality Reduction Visualisation using Principal Component Analysis (PCA) and Silhouette Coefficient (SC) calculations on the two datasets. Results for the overall separation evaluation of Japanese and Korean samples are given in Figure 5.

Samples from Japan and South Korea cluster into separate areas of the feature space defined by the first two principal components (see Figure 5(a)). Most of the Japanese samples are located near the negative end of PC1. This clustering is due to the systematic and structural differences in how the two countries have chosen to use visual grammatical elements and create relationships between text and images. Figure 5 (b) shows that the mean silhouette coefficients of the two countries’ samples are 0.652 and 0.614, respectively, with an overall mean of 0.633, and significant differences in distribution. These high silhouette coefficient values originate from the high degree of consistency within their respective strategies. Specifically, Japanese media tend to consistently use specific strategy sets such as “aesthetic othering,” while Korean media consistently prefer strategy sets such as “competitive juxtaposition,” thus being identified as two well-cohesive clusters at the algorithmic level. Therefore, Japanese and Korean media have constructed two overall distinguishable and internally highly consistent visual narrative modes in their Silk Road-themed reporting, and their differences possess statistical systematicity and robustness.

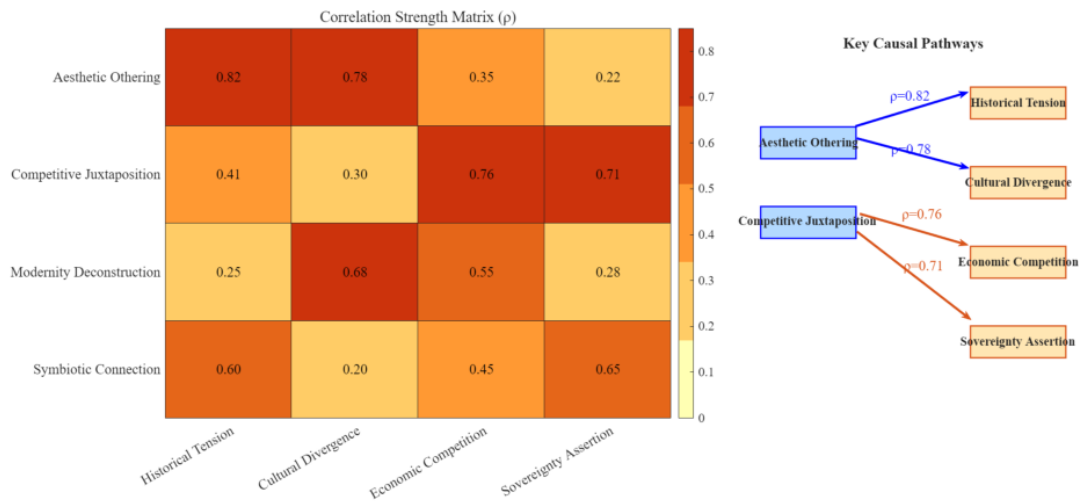


FIGURE 4. Strategy-context association strength matrix and critical path.

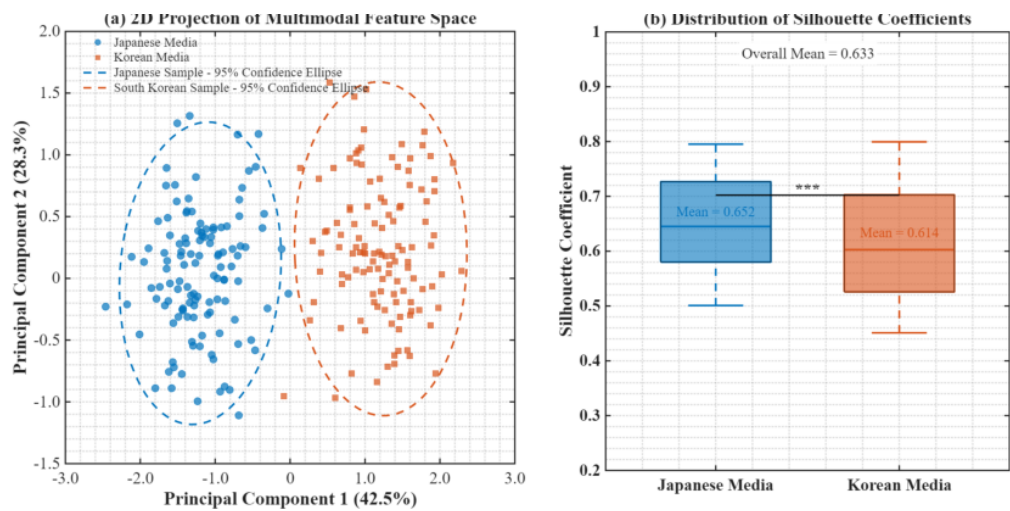


FIGURE 5. Multimodal feature space distribution and clustering effectiveness (a) Two-dimensional projection scatter points of feature space (b) Profile coefficient distribution. “****” indicates that the result is statistically significant at the 0.001 significance level.

V. CONCLUSION

This study constructs a context-driven multimodal-ideological linkage analysis framework, integrating the evolution of silk weaving techniques and textile trade motifs, to systematically reveal the differentiated mechanisms of visual narrative strategies in Silk Road-themed reporting by Japanese and South Korean media. By examining how textile craftsmanship and fabric semiotics shape these narratives, the research uncovers the underlying motivations and cultural motivations behind the representation of historical and contemporary Silk Road identities. Through the combination of visual grammar through formal annotation, a form of cluster analysis that determines a typology of media strategies, and a method for determining correlations between media strategies and the socio-political context using a method of canonical correlation analysis, a causal explanation has been established that relates micro-modal

characteristics to macro sociopolitical motives. Based on empirical research, it can be shown that Japanese media strategy is primarily centred on “aesthetic othering”, accounting for approximately 51.7% of the Japanese media strategy uses. Additionally, visual features of Japanese media strategies correlate with the historical memory tension, as evidenced by a canonical correlation of 0.82. In Korea, approximately 37.5% of the media strategy is based on “competitive juxtaposition” and this media strategy has a strong correlation with economic competition motivation, with the latter two being correlated (correlation value). Furthermore, using principal component analysis as a means to separate the sample of the Japanese and Korean media strategies, it can be concluded that, within the multimodal feature space, the samples are generally separable and have an average silhouette coefficient of 0.633. As such,

the results of this research will provide insight into the relationship between Japanese and Korean culture through the perspective of comparative analysis of cross-cultural visual communication. Additionally, the results of this study will serve as empirical evidence of the complex mentality of “neighborly Orientalism” by examining how traditional textile aesthetics and garment symbolism are utilized in the logic of discourse politics during the cultural image negotiation of neighboring countries. “Neighborly Orientalism” refers to negotiating one’s cultural identity and regional status in the context of modernity through differentiated, hierarchical, or competitive visual representations based on geographical and cultural proximity. By analyzing the semiotic appropriation of weaving patterns and fabric heritage, the research unveils the strategies of international communication that define the power dynamics and material identity of East Asian nations.

AUTHOR CONTRIBUTIONS

Fan Xuyang designed, collected and analyzed the data, and drafted the manuscript. Fan Xuyang conducted the study, critically revised the manuscript for important intellectual content, and gave final approval of the version to be published. Fan Xuyang participated fully in the work, take public responsibility for appropriate portions of the content, and agreed to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

CONFLICTS OF INTEREST

The author declares no conflict of interest.

FUNDING

This research received no external funding.

ACKNOWLEDGEMENTS

Not applicable.

REFERENCES

- [1] G. Gao, “Diplomatic Communication of Silk Sign: Vehicle, Signification, and Narrative,” *Critical Arts*, vol. 39, no. 4, pp. 71-85, 2025, doi: 10.1080/02560046.2024.2421558.
- [2] M. S. Brown and D. O’Brien, “The silk roads and shared heritage in Europe: Beyond ‘China to Rome’,” *China Information*, vol. 39, no. 3, pp. 279-300, 2025, doi: 10.1177/0920203X241281989.
- [3] J. Zhang, “International Media Analysis: Critical Analysis Comparing News Reports In Listed West/South Digital Media,” *Journal Of Education*, vol. 46, no. 1, pp. 124-127, 2024, doi: 10.54097/4bwn6d38.
- [4] Y. Ni, “Navigating Intercultural Communication in the Media Landscape: Challenging Stereotypes and Cultivating Critical Perspectives—Hall Model as the Framework,” *Digital Society*, vol. 3, no. 2, pp. 36, 2024, doi: 10.1007/s44206-024-00126-x.
- [5] C. Guan, “Multimodal positive discourse analysis of national image publicity video,” *Language and Semiotic Studies*, vol. 8, no. 3, pp. 66-85, 2022, doi: 10.1515/lass-2022-2001.
- [6] M. Singh, “Exploring Multimodal and Critical Dimensions in Discourse Analysis: Theoretical Foundations and Methodological Approaches,” *International Journal of English Literature and Social Sciences*, vol. 10, no. 3, 618153, 2025, doi: 10.22161/ijels.103.59.
- [7] L. Zhang and A. Jocuns, “A multimodal critical discourse analysis of stance-taking in English news reports on the environment,” *Journal of Studies in the English Language*, vol. 18, no. 2, pp. 108-142, 2023.
- [8] W. Liu, C. Du, and F. Liu, “Communicating change: A multimodal critical discourse analysis of China’s poverty reduction posters,” *Frontiers in Communication*, vol. 9, no. 1, 1444032, 2024, doi: 10.3389/fcomm.2024.1444032.
- [9] N. J. Ebele and W. Ahmad, “Intercultural communication and mutual appreciation of civilizations,” *Migration Letters*, vol. 21, no. S14, pp. 814-834, 2024, doi: 10.59670/ml.v21iS14.11300.
- [10] J. Zhang and X. Yang, “Exploring the archaeological treasures of the silk road: Tibetan arts and crafts, traditional decorative patterns, and pattern symbols analysis,” *Mediterranean Archaeology and Archaeometry*, vol. 23, no. 2, pp. 197-212, 2023.
- [11] Z. Xu and M. Zhang, “How news media visually dehumanize victims of humanitarian crises through framing disparities: A quantitative comparative analysis,” *International Communication Gazette*, vol. 86, no. 8, pp. 633-654, 2024, doi: 10.1177/17480485231216583.
- [12] A. Rafiee, W. Spooen, and J. Sanders, “Framing similar issues differently: A cross-cultural discourse analysis of news images,” *Social Semiotics*, vol. 33, no. 3, pp. 515-538, 2023, doi: 10.1080/10350330.2021.1900719.
- [13] M. Tosun and F. Ç. Yıldırım, “A Visual Grammar Analysis of a Time Magazine Cover Photo in the Framework of Multimodal Discourse Approach,” *Dil ve Edebiyat Dergisi*, vol. 20, no. 1, pp. 55-72, 2024.
- [14] B. Anwar, “Unlocking visual grammar theory: A multi-modal discourse analysis of non-product advertisement,” *Pakistan Journal of Society. Education and Language (PJSEL)*, vol. 10, no. 2, pp. 105-115, 2024.
- [15] Y. Zhang, “Analysis of Reporting Characteristics of Japan’s Asahi Shimbun on Belt and Road Initiative,” *International Journal of Frontiers in Sociology*, vol. 6, no. 10, pp. 9-14, 2024, doi: 10.25236/IJFS.2024.061002.
- [16] K. Shin, “The Evolution of Japanese Perceptions of China since 1945,” *Asia-Pacific Review*, vol. 30, no. 2, pp. 148-166, 2023, doi: 10.1080/13439006.2023.2251807.
- [17] R. Nakano, “A geocultural power competition in UNESCO’s Silk Roads project: China’s initiatives and the responses from Japan and South Korea,” *Journal of Current Chinese Affairs*, vol. 52, no. 2, pp. 185-206, 2023, doi: 10.1177/18681026221094054.
- [18] E. E. Song, “South Korea’s Narratives on China: Evidence from Elites, Masses, and Scholars,” *Asian Perspective*, vol. 49, no. 2, pp. 379-396, 2025, doi: 10.1353/anp.2025.a960999.
- [19] E. Tsutaki and S. Hashimoto, “The Role of the Lotus Sutra Exhibition in Inter-Civilizational and Inter-Religious Dialogue,” *The Journal of Oriental Studies*, vol. 31, no. 1, pp. 4-32, 2022.
- [20] J. Qi and Q. Hu, “Analyzing Multimodal Construction of China’s Image Based on the Illustrated Articles in The Economist,” *Theory and Practice in Language Studies*, vol. 14, no. 11, pp. 3641-3653, 2024, doi: 10.17507/tpls.1411.34.
- [21] D. Li, “Multimodal Analysis of Image Texts in College English Textbooks using Visual Grammar Theory for Real-time Embedded Systems,” *Scalable Computing: Practice and Experience*, vol. 24, no. 4, pp. 1137-1144, 2023, doi: 10.12694/scpe.v24i4.2434.
- [22] M. K. Dahouda and I. Joe, “A deep-learned embedding technique for categorical features encoding,” *IEEE Access*, vol. 9, no. 1, pp. 114381-114391, 2021, doi: 10.1109/ACCESS.2021.3104357.
- [23] Y. Sun, J. Zheng, H. Zhao, H. Zhou, J. Li, F. Li, et al., “Modifying the one-hot encoding technique can enhance the adversarial robustness of the visual model for symbol recognition,” *Expert Systems with Applications*, vol. 250, no. 1, 123751, 2024, doi: 10.1016/j.eswa.2024.123751.
- [24] R. Jin, T. Hou, T. Jin, J. Yuan, and C. Du, “A method for image-text matching based on semantic filtering and adaptive adjustment,” *EURASIP Journal on Image and Video Processing*, vol. 2024, no. 1, pp. 23, 2024, doi: 10.1186/s13640-024-00639-y.
- [25] R. Zeng, W. Ma, X. Wu, W. Liu, and J. Liu, “Image-text cross-modal retrieval with instance contrastive embedding,” *Electronics*, vol. 13, no. 2, pp. 300, 2024, doi: 10.3390/electronics13020300.
- [26] J. Meng, Z. Yu, Y. Cai, and X. Wang, “K-Means++ clustering algorithm in categorization of glass cultural relics,” *Applied Sciences*, vol. 13, no. 8, pp. 4736, 2023, doi: 10.3390/app13084736.
- [27] S. Juanita and R. D. Cahyono, “K-means clustering with comparison of Elbow and silhouette methods for medicines clustering based on user reviews,” *Jurnal Teknik Informatika (JUTIF)*, vol. 5, no. 1, pp. 283-289, 2024, doi: 10.52436/1.jutif.2024.5.1.1349.
- [28] G. Rau and Y. S. Shih, “Evaluation of Cohen’s kappa and other measures of inter-rater agreement for genre analysis and other nominal data,” *Journal*

- of English for Academic Purposes, vol. 53, no. 1, 101026, 2021, doi: 10.1016/j.jeap.2021.101026.
- [29] M. T. Eser and G. Aksu, "Comparison of the results of the generalizability theory with the inter-rater agreement coefficients: International Journal of Curriculum and Instruction," 2022; 14(2):1629-1643.
- [30] A. H. Joarder and M. H. Omar, "On the exact distribution of the difference between two chi-square variables," Bulletin of the Malaysian Mathematical Sciences Society, pp. 45(Suppl 1) 505-522, 2022, doi: 10.1007/s40840-022-01373-2.
- [31] K. Chaturvedi, D. K. Vishwakarma, and N. Singh, "COVID-19 and its impact on education, social life and mental health of students: A survey," Children and youth services review, vol. 121, no. 1, 105866, 2021, doi: 10.1016/j.chidyouth.2020.105866.
- [32] Y. Mao, X. Yan, J. Liu, and Y. Ye, "Congmc: Consistency-guided multimodal clustering via mutual information maximin," IEEE Transactions on Multimedia, vol. 26, no. 1, pp. 5131-5146, 2023, doi: 10.1109/TMM.2023.3330093.
- [33] M. S. Jain, K. Polanski, C. D. Conde, X. Chen, J. Park, L. Mamanova, et al., "MultiMAP: Dimensionality reduction and integration of multimodal data," Genome Biology, vol. 22, no. 1, pp. 346, 2021, doi: 10.1186/s13059-021-02565-y.