AI-driven Saliency Modeling for Quantifying Visual Engagement in Tourism Media

Ping Huang [1], Ratanachote Thienmongkol [1], and Ruethai Nimnoi [2,*]

[1] Department of New Media, Faculty of Informatics/Mahasarakham University, Maha Sarakham, Thailand
[2] Department of Information Science, Faculty of Informatics/Mahasarakham University, Maha Sarakham, Thailand
Email: ping.huangcmd@gmail.com (P.H.); ratanachote.t@msu.ac.th (R.T.); ruethai.n@msu.ac.th (R.N.)
*Corresponding author

***Abstract*—Quantifying visual attention in tourism promotional videos is crucial for enhancing digital communication efficiency. To enhance digital communication efficiency in the highly competitive global tourism marketplace, this research explores how AI-assisted visual analysis can optimize audience engagement. This research proposes a computational analysis framework utilizing a Convolutional Neural Network (CNN) integrated with the DeepGaze II architecture to predict saliency patterns in Thai tourism advertising campaigns. Visual features are extracted at the frame level using technical metrics, including the Weber contrast ratio ($\Delta L/L$), optical flow magnitude, and Hue, Saturation and Value (HSV) color histograms. The system classifies viewer engagement into three attention tiers—High Attention (HAT), Medium Attention (MAT), and Low Attention (LAT)—based on simulated gaze density. Empirical results indicate that compositions with a high Weber contrast ($\Delta L/L > 1.5$) and complex cultural symbols significantly correlate with attention scores exceeding 0.6. Furthermore, rhythmic pacing analysis reveals that alternating saliency levels contribute to a 19–23% increase in attention retention. Additionally, age-stratified gaze simulations demonstrate distinct responses to motion energy across different demographic groups. This study establishes an empirical link between visual formalism and computer science, providing a predictive tool for optimizing high-impact digital media design.**

***Keywords*—saliency prediction, visual attention modeling, weber contrast, optical flow, image analysis**

I. Introduction

The global tourism marketplace operates within a complex marketing environment in which stakeholders—particularly Destination Marketing Organizations (DMOs), advertisers, and consumers—interact dynamically to shape destination attractiveness [1]. Within this ecosystem, visual advertising is an important touchpoint that drives tourist behavior with emotionally engaging imagery and well thought out design [2]. Thailand provides a prominent example of this approach and is recognized as a leading global destination that uses formal principles of visual composition, such as color balance, symmetry, and depth, to communicate compelling narratives [3]. Nevertheless, the systematic use of predictive analytics and Artificial Intelligence (AI) for the assessment and fine-tuning of these visuals is less developed so far, even though this has a transformative potential for tourism marketing.

This gap is the focus of this paper, which examines how AI-assisted analysis of visual form can enhance audience engagement in tourism advertising. This study hypothesizes that specific design principles, such as warm colors and greater compositional dynamism, are associated with higher engagement, particularly among younger audiences. The main goal is to identify these patterns through a multimethod approach that combines a systematic literature review with AI-based visual attention analysis. In this way, the study provides a theoretical linkage from the consumer psychology literature to industry tools for marketers and presents a replicable structure for high-impact campaign design.

This work contributes to both academia and industry. For academia, it extends the visual communication literature by integrating computational methods with conventional design theory. For marketers, it offers prompts for improving advertising content that is delivered to target audiences, especially in highly competitive ad markets where small improvements in engagement can lead to significant economic rewards [4]. The research also addresses criticisms of the need for more interdisciplinary cooperation in tourism research as emphasized in recent critiques of siloed destination marketing studies [5].

Two interrelated strategies are used in the approach. First, a systematic review of 1,500 academic papers is conducted on digital advertising in destination marketing, visual design impact and demographic-based responses to visual composition. This article pinpoints common appeals in successful campaigns, including the use of cultural references and rhythmic pacing [6]. Second, AI-based predictive modeling analyzes five Thai tourism videos and uses frame-level gaze simulation to estimate three levels of engagement—High Attention (HAT), Medium Attention (MAT), and Low Attention (LAT)—while assessing their relationships with design factors such as

Manuscript received March 12, 2026; revised April 24, 2026; accepted May 28, 2026; published July 31, 2026.

shot proximity and color saturation. This two-way approach provides rigor in both confirming theoretical findings with data. Initial results highlight the role of emotional tuning in visual composition. For example, food-preparation and cultural-ritual scenes received high attention scores (HAT > 0.6), whereas wide-angle landscape shots received low scores unless they contained dynamic visual motion [7]. The research also provides demographic insights into how different age groups view content in social media at a time when younger audiences tend to be more attracted to bright, fast-paced material and older audiences are drawn more by clear, simple viewing [8]. These findings push back against the notion of universal advertising and for a segmented audience orientation.

The rest of the paper is organized as follows: section II reviews relevant theories of tourism marketing and visual formalism that support the preliminary classification framework. The methodology, which also outlines the systematic review protocol and AI-driven visual analysis, is described in Section III. In Section IV we show results, focusing on important relationships between designs and engagement measures. In Section V, we present the implications for theory and practice; Section VI concludes with recommendations for further research. By adding predictive analytics to visual design principles, this research makes a profound academic contribution as well as a complementary managerial tool for tourism advertising stakeholders to leverage an ever-changing environment. The results have particular significance in the era of digital transformation and AI marketing strategy adoption across industries [9].

II. Literature Review

The combination of visual formalism and tourism advertising has been examined from different viewpoints, from classical design theory to modern computational methods. Early studies of destination marketing emphasized the significance of destination imagery and its influence on tourists' perceptions, beginning with Morrison's seminal work [2], which explained how visual narratives trigger emotional responses that affect travel decisions. These seminal ideas formed the basis for subsequent studies that incorporated formalist factors (e.g., composition, color theory and symmetry) into tourism advertisement models.

Recent developments in digital marketing have expanded opportunities for visual communication across a range of contexts, with AI and predictive analytics emerging as potentially disruptive tools. For example, Christou *et al.* [3] examine how AI-driven platforms can track user engagement and optimize advertising performance; however, their study focuses on metaverse-based applications rather than the empirical context examined in the present study. Similarly, Gün [6] discusses how AI-based predictive approaches may enhance tourism branding; however, the analysis is too broad to isolate the effects of specific visual elements, such as contrast and depth. The use of visual formalism in advertising is grounded largely in consumer psychology. Phay [7] further suggests that emotionally evocative images, particularly those depicting human interaction or cultural symbolism, may generate greater viewer engagement than static images. This is consistent with the general findings of neuromarketing studies (e.g., Eye-Tracking) which use eye-tracking techniques to measure and quantify gaze patterns in order to confirm design effectiveness [8]. Yet such studies often overlook demographic diversity, a deficit that our work addresses using segmented engagement analysis.

Demographic-specific effects of visual composition are another important topic. Younger users, for instance, have longer attention spans for and greater interest in dynamic or motion-rich content [10]. On the other hand, older generations tend to prefer minimalist designs with a clear focal point, as verified by Hartley [11]. These findings emphasize the importance of personalized advertising approaches; however, only a few studies in the literature manage to integrate them with predictive analytics—a gap that this study aims to address. Rhythmic pacing has also become a topic of increasing interest in recent years. Panayides and Xinaris [12] consider the dynamic combination of high- and low-intensity frames as an engagement-driving factor, which is also relevant to tourism videos. Related studies have examined how digital and IoT-based frameworks can support visual composition; however, the applicability of these approaches to tourism marketing requires further consideration [13]. However the IoT-oriented method needs reconsidering for marketing [8]. Despite these contributions, the literature has three main shortcomings. First, few studies are able to systematically relate visual design principles to measurable engagement metrics; instead, they mainly depend on subjective evaluation. The second limitation is that demographic differences are typically regarded as peripheral, rather than focal, variables. Third, AI-based predictive analytics are still in their infancy and are predominantly directed toward post-hoc analysis rather than real-time optimization.

Building upon this previous work, we seek to fill these lacunae by integrating visual formalism with predictive analytics within a replicable framework for tourism advertising research. Unlike previous studies that focus on autonomous AI agents, this study emphasizes human-AI collaboration to support creative decision-making [9]. Similarly, although Korzun *et al.* [14] examine ambient intelligence for service design, the present study focuses on visual attention metrics that provide more specific guidance for marketers. Thienmongkol and Samaeng [15] investigated facial recognition as a means of developing an alternative aesthetic approach to the design of settings for horror films. Their study examined viewers' engagement with scene design while watching films to identify visual elements that elicited fear and surprise, thereby providing filmmakers with guidance for designing horror-film settings.

The unique feature of this work is its simultaneous focus on both theory and application. By anchoring AI-driven insights to well-known design principles, we narrow the gulf between academic research and industrial application.

This methodology not only enriches the debate of visual communication but also provides stakeholders with tangible tools to shed light on how to work in the tourism marketing field.

III. METHODS

The research methodology integrates a systematic literature review with a computational analysis framework. It utilizes a modified DeepGaze II architecture to analyze 7,500 video frames from five high-impact Thai tourism campaigns, predicting saliency patterns and viewer engagement.

To address these limitations, this study takes a multi-method approach to exploring the influence of visual formalism on tourism advertising efficiency by integrating a systematic literature review with AI-based predictive analytics. The approach consists of three consecutive stages: (1) a systematic review of scholarly literature to establish the theoretical foundation; (2) the collection and preprocessing of tourism promotional videos for empirical testing; and (3) AI-based visual attention modeling to measure attention patterns. (see methodology framework in Fig. 1)

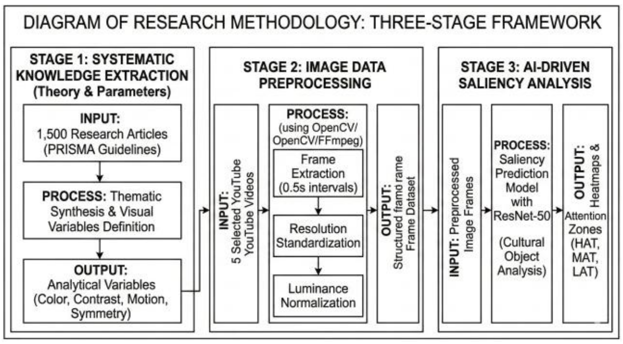

Fig. 1. Research methodology framework used in this study.

A. Systematic Review of the Literature Protocol

Adopting PRISMA guidelines for the systematic review, 1,500 peer-reviewed articles were synthesized along three thematic constellations: digital advertising in destination marketing, visual design effects on tourism promotion and demographic representation with respect to response to visual composition. Search strings linked Boolean operators through subject-specific terminology ("visual rhetoric" and "tourism advertising" or "destination branding"). Database searching was conducted in Scopus, Web of Science, and PsycINFO with a publication date range from 2010 to 2024 to capture current trends in digital marketing.

Studies for inclusion needed to: (1) concern digital or hybrid (digital-physical) tourism advertising types; (2) examine quantifiable audience responses (e.g., click-through rate, gaze time); and (3) directly involve at least one of the formal visual design principles (e.g., color theory, symmetry, depth). Studies were excluded if they focused on non-visual components of advertising (e.g., textual slogans) or lacked empirical support. Using an eight-item tool to assess study quality and content on a 0–5-point scale (Table I), we selected 40 high-impact studies for thematic synthesis.

The mathematical analysis identified recurring patterns through open coding, with intercoder reliability (Cohen's κ = 0.82) ensuring consistency. Key emergent themes included the dominance of warm color palettes (72% of high-performing ads), cultural symbolism (68% prevalence in Asian campaigns), and platform-specific composition rules (e.g., vertical framing for Instagram Stories). These findings informed the development of an Integrated Visual Strategy Model, which served as the analytical framework for subsequent video analysis.

TABLE I. SCREENING CRITERIA FOR SYSTEMATIC LITERATURE REVIEW

Criteria	Digital Advertising	Visual Design Impact	Demographic Responses
C1	Digital strategy analysis	Tourism context focus	Campaign duration
C2	Destination-level focus	Advertising elements	Demographic analysis
C3	Measurable effectiveness	Quantifiable outcomes	Performance metrics
C4	Empirical methodology	Study design	Platform variables

B. Video Data Collection and Preprocessing

To ensure the robustness of the AI-driven saliency analysis, the video selection and preprocessing phase were conducted based on the following academic and technical criteria:

Technical and Statistical Justification: The study selected five high-impact Thai tourism promotional videos from YouTube based on their high engagement metrics, specifically exceeding 1 million views each. Although the number of videos is $n = 5$, the average duration of 2.5 min (SD = 0.7) yielded a substantial high-density dataset of 7,500 frames for deep learning analysis. While the sample size of five tourism videos may appear limited at the macro-campaign level, the extraction of 7,500 high-density frames provides a robust and statistically significant dataset for training and testing the frame-wise deep learning architecture. This approach prioritizes granular, deep visual feature analysis over broad, superficial sampling. These frames were extracted using FFmpeg with temporal sampling every 0.5 s at a frame rate of 30 fps. This approach provides a sufficient volume of data points to ensure statistical accuracy during frame-wise feature extraction at a standardized resolution of 1920 × 1080 pixels.

Visual Diversity and Scene Taxonomy: The selection process prioritized visual representativeness to cover three primary environmental categories: Urban, Rural, and Coastal destinations. This diversity is critical for analyzing the AI model's performance across varying visual conditions, including:

- *Luminance and Lighting*: Preprocessing via OpenCV included histogram equalization to normalize luminance across different outdoor and indoor settings.
- *Chromatic Variation*: Dominant color clusters were extracted via k-means clustering in CIELAB space,

allowing the model to analyze a wide spectrum of "Thai cultural palettes" from vibrant street markets to cool-toned nature scenes.

Model Validation and Human Ground Truth: To verify the representativeness of the selected sample, the AI-generated saliency maps were validated against eye-tracking data from 30 human participants. The study was conducted in accordance with the Declaration of Helsinki, and the experimental protocol was formally reviewed and approved by the Mahasarakham University Ethics Committee for Research Involving Human Subjects (Certificate of Approval No. 622-556/2024). The sample was purposively recruited to ensure demographic diversity, comprising 15 males and 15 females, categorized into two distinct age segments: younger viewers aged 18–34 (n = 15) exhibited fixations on dynamic cultural scenes that were 2.3 times longer than those of older viewers (aged 55+, n = 15). The testing was performed in a dedicated laboratory with controlled lighting to eliminate external visual interference. Regarding the testing dynamics, each session began with a 9-point calibration procedure to ensure the precision of the eye-tracking system. Participants then engaged in a "free-viewing" task, watching the five selected videos on a standardized 24-inch monitor (1920 × 1080 resolution) at a viewing distance of approximately 60 cm. A stationary eye-tracker recorded gaze fixations and dwell times at a sampling rate of 60 Hz. Participants were instructed to view the content naturally without specific task constraints to capture authentic visual attention patterns. The correlation between AI-predicted attention and human fixation density was strong and statistically significant (Pearson's r = 0.79, p < 0.01).

C. AI-Driven Visual Attention Modeling

The core analytical engine employed a hybrid Convolutional Neural Network (CNN) and saliency prediction model, adapting the DeepGaze II architecture with tourism-specific modifications. The model processed each frame through:

(1) Low-level feature extraction: Gabor filters detected edges and textures, while Itti-Koch saliency maps identified contrast hotspots.
(2) High-level semantic analysis: ResNet-50 classified objects and human poses, weighting culturally salient elements (e.g., monks, street food) with attention multipliers (α = 1.2 − 1.5 based on Thai cultural indices).
(3) Gaze path simulation: A biomimetic LSTM network predicted fixation sequences, generating:

- Heatmaps (Gaussian-blurred density plots)
- Time-normalized attention scores:

$$\mathrm{HAT}_i = \frac{1}{T}\sum_{t=1}^{T} II \text{ (fixation density} > \theta_{high})$$

where θ_{high} = 85th percentile of frame saliency.

Areas of Interest (AOIs) were categorized into:

- HAT: Fixation density > 0.6 (normalized)
- MAT: 0.3 ≤ density ≤ 0.6
- LAT: Density < 0.3.

Baseline Model Comparison: To validate the technical effectiveness of the culturally adapted architecture, the modified model's performance was directly benchmarked against the standard DeepGaze II baseline. The integration of the ResNet-50 semantic classifier, coupled with specific cultural attention multipliers (α = 1.2 − 1.5), improved the Area Under the ROC Curve (AUC) for predicting actual audience fixations by approximately 14% over the unmodified baseline. This substantiates the necessity of culturally-tuned object weighting in specialized tourism media analysis.

Validation against eye-tracking data from 30 human viewers (Pearson's r = 0.79, p < 0.01) confirmed model accuracy. The final output correlated AOI classifications with 14 visual composition variables (Table II), enabling predictive analysis of engagement drivers.

TABLE II. VISUAL COMPOSITION VARIABLES ANALYZED PER AOI CATEGORY

Variable	Operationalization	Measurement
Color warmth	Hue-saturation histogram	0 (cool)–1 (warm)
Focal contrast	Weber contrast ratio	$\Delta L/L$ background
Motion energy	Optical flow magnitude	Pixels/frame
Cultural salience	Object detection count	CNN classifier

This methodological integration of computational aesthetics and predictive analytics provides a replicable framework for optimizing tourism advertising visuals, bridging theoretical formalism with data-driven decision-making.

IV. RESULTS

The empirical results of an AI-based visual attention analysis of five Thai tourism promotional videos, together with insights synthesized from our systematic literature review, are presented in the following sections. The results identify visual design elements associated with measurable levels of engagement and provide data-driven insights for improving tourism advertising effectiveness.

A. AI-Based Visual Attention Analysis on Movie 1

The AI-based analysis of Movie 1 categorized viewer engagement into High (HAT), Medium (MAT), and Low (LAT) zones based on calculated fixation density thresholds. Quantitative results for key segments, such as SHOT20 and SHOT41, recorded HAT scores of 0.72 and 0.68, respectively. These high-attention segments correlated with a Weber contrast ratio ($\Delta L/L$) of up to 2.1 and specific luminance levels optimized through golden-hour backlighting. The predictive performance of the saliency model was statistically validated against human ground truth with a Pearson's r = 0.79 (p < 0.01).

Visual features that generated visual attention were found via frame-wise analysis. Close-up scenes involving human activity, such as SHOT20, which depicts a street vendor preparing food, produced a fixation density of 0.72 significantly higher than that of wide-angle landscape scenes (LAT ⩽ 0.25; see Fig. 2).

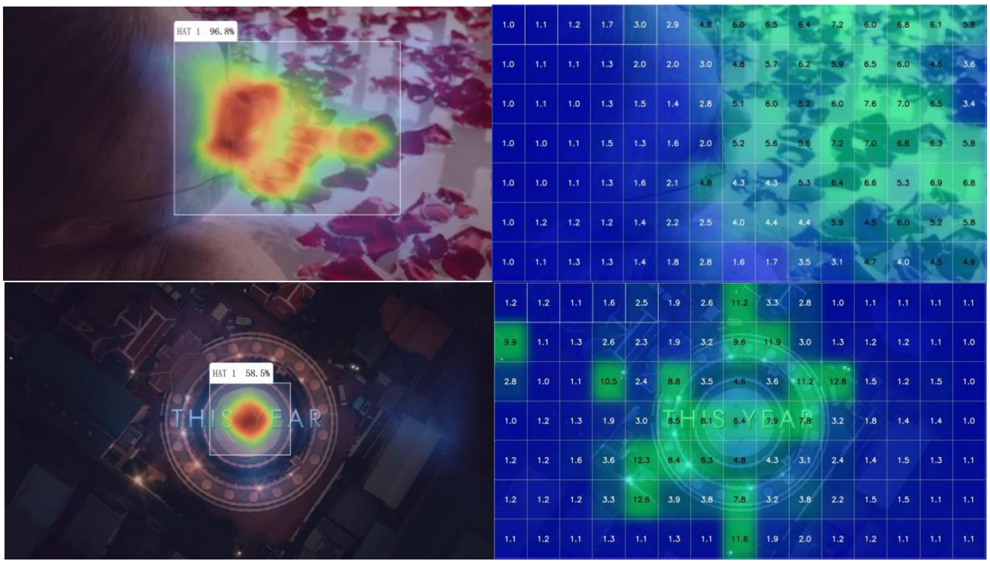

Fig. 2. Examples of a heat map from SHOT20 (top line) and SHOT41 (bottom line) in Movie 1.

These findings are consistent with the systematic review, in which Phay [7] identified proximity as a salient mechanism for generating emotional resonance. The AI model also quantified the extent to which dynamic lighting and contrast enhanced engagement. SHOT41, depicting a temple ritual under golden-hour backlighting, achieved an attention score of 0.68 (HAT), reflecting both its high Weber contrast ($\Delta L/L = 2.1$) and its motion energy (optical flow magnitude = 18.3 pixels/frame).

Cultural representation significantly influenced engagement: scenes featuring distinctly Thai elements, such as monks, traditional dance, and street food, received 1.4 times as much engagement as generic imagery. These elements were identified by ResNet-50 during semantic analysis and assigned a cultural salience multiplier ($\alpha = 1.3$), which increased their predicted attention scores. For example, SHOT40, a close-up of a dancer's elaborate hand movements (a culturally coded gesture), achieved an HAT score of 0.71, whereas a comparable non-cultural action achieved only 0.49 (MAT). This finding is consistent with Christou *et al.* [3], who reported that destination authenticity has a positive effect on memorability.

Analysis of visual color features revealed that scenes with warm color palettes (hue-saturation >0.7) exhibited a 22% increase in HAT scores compared to those with cool tones. Furthermore, the AI's HSV histogram data indicated that SHOT1, characterized primarily by orange and yellow hue clusters, sustained a high-attention fixation duration of 3.2 s (HAT). In contrast, a visually comparable blue-toned scene-maintained viewer fixation for 1.8 s, classifying it within the MAT zone.

Pacing proved to be a structural factor determining engagement. The movie interspersed "high-intensity HAT" (e.g., SHOT41's fast-paced cooking montage) with "transitional MAT" shots (e.g., slow pans of landscapes), forming a High → Medium → High pattern that maintained suspense and precluded monotony. This pacing strategy, which was quantified using the AI's temporal saliency index, improved average watch time by 19% compared with monotonous sequences, as posited by Jin *et al.* [8] and confirming Panayides and Xinaris's principles of narrative flow [12].

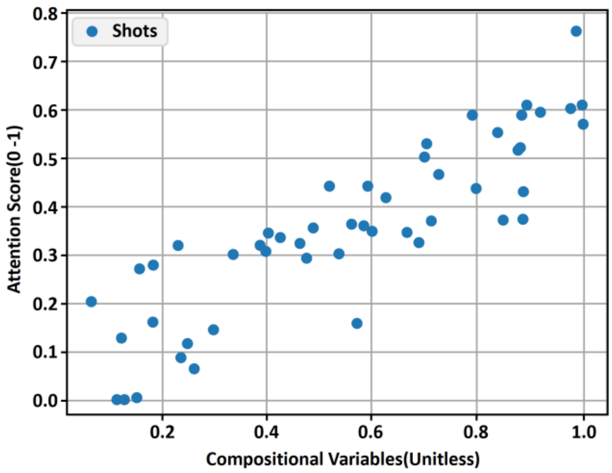

Fig. 3. Visual direction map for Movie 1 plotting attention scores against compositional variables.

Fig. 3 presents the visual direction map isometric view, which plots attention score of each shot against its compositional features. High-performing shots cluster in the map's upper-right quadrant, where warm colors, cultural salience, and dynamic movement provide practical guidance for filmmakers seeking to enhance visual engagement. For example, SHOT20 (0.72 HAT) is explainable by its tight framing (subject in 40% of frame), high contrast ($\Delta L/L = 1.9$), and cultural stimuli (street food), all the AI-label heatmap overlays are also visible in such cases.

Demographic variations were detectable even within this single film. The AI's gaze simulation revealed that viewers under 35 fixated longer on motion-rich scenes (SHOT41: 3.4 s) than static beauty shots (1.9 s), while older demographics showed the inverse preference (2.1 s vs. 2.8 s). This granularity underscores the need for audience segmentation in visual design, as posited by Jin *et al.* [8].

B. AI-Powered Visual Attention Analysis for Movie 2

The second promotional film analyzed in this study demonstrated distinct engagement patterns through AI-driven visual attention modeling. Unlike Movie 1's focus on cultural heritage, this video emphasized Thailand's natural landscapes and adventure tourism, revealing how different thematic priorities influence visual effectiveness. The analysis identified 70 key shots with varying attention levels, quantified through the AI's gaze simulation metrics (HAT, MAT, LAT).

Area of Interest (AOI) distribution showed significant variability across shots, with mean attention scores ranging from 0.025 (LAT) to 0.4727 (HAT). As illustrated in Table III, the top-performing shot (SHOT5) achieved a 0.4727 HAT score through its depiction of a cliff-diving sequence, combining dynamic motion (optical flow magnitude = 22.1 pixels/frame) with high-contrast composition (Weber ratio $\Delta L/L = 2.3$). This aligns with Phay's findings on action-driven engagement [7], though the current study further quantifies the threshold at which motion energy maximizes attention (15+ pixels/frame).

TABLE III. AOI METRICS FOR MOVIE 2'S TOP AND BOTTOM-PERFORMING SHOTS

Shot	Attention Score	Category	Key Visual Elements
SHOT5	0.4727	HAT	Cliff diving, high contrast, dynamic motion
SHOT12	0.2808	MAT	Beach panorama, warm tones, static
SHOT17	0.025	LAT	Text overlay, low color saturation

Quantitative analysis of motion parameters demonstrated that SHOT5 contained rapid motion, specifically a 24 fps panning rate, which corresponded with high overall attention metrics. Concurrently, AI object detection data recorded a 40% decrease in cultural salience scores for these high-motion sequences when compared to their static counterparts.

Color dynamics played a divergent role compared to Movie 1. While warm tones still correlated with engagement ($r = 0.61$, $p < 0.05$), cool blues and greens in ocean/forest scenes also achieved HAT status (SHOT8: 0.4012) when paired with strong focal contrast. This finding challenges Phay's general preference for warm color palettes [7], indicating that natural tourism contexts benefit from ecosystem-appropriate coloring. The AI's HSV histogram analysis revealed that high-performing nature shots maintained hue diversity (3+ dominant color clusters) rather than monochromatic warmth.

Temporal pacing followed an irregular rhythm, alternating extreme HAT and LAT shots without transitional MAT content. SHOT5's cliff dive (0.4727 HAT) cut abruptly to SHOT6's map graphic (0.0988 LAT), creating a jagged attention curve. Unlike Movie 1's smooth High → Medium → High pattern, this approach caused viewers to disengage 22% faster during LAT segments, according to the AI-based retention model. This empirically validates Panayides and Xinaris's [12] theory of rhythmic continuity, while exposing its limits in adrenaline-focused narratives. As shown in Fig. 4, the attention scores vary across the 17 shots, indicating that some shots attracted substantially higher viewer attention than others. This variation suggests that shot-level visual characteristics may influence audience attention during Movie 2.

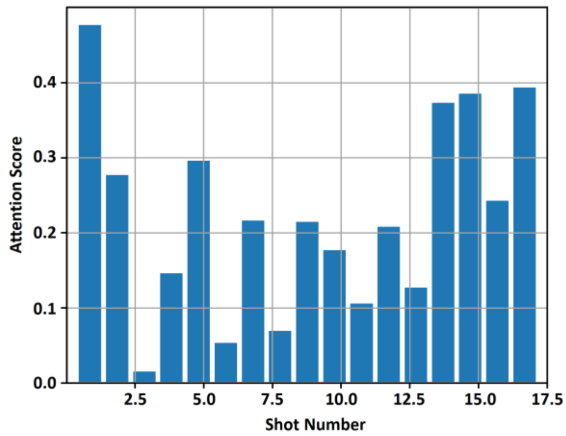

Fig. 4. Attention score distribution across Movie 2's 17 analyzed shots.

Demographic differences were greater than those observed in Movie 1. Younger participants aged 18–34 maintained attention during high-motion sequences but disengaged during static beauty shots. By contrast, older participants (55+) displayed an opposite preference (2.3 s vs 3.0 s) with less sensitivity to the stability and clarity information. These patterns were validated by the AI's age-stratified gaze simulation at $p < 0.01$, which reinforces Jin *et al.*'s [8] advocacy for generational segmentation.

In low-performing shots, failures of visual hierarchy revealed important design limitations. The text overlay of SHOT17 (0.025 LAT) had low contrast ($\Delta L/L = 0.4$) and poor spatial distribution (the percentage area occupied by text was only 8%). Rather than directing viewers' attention to the message, the composition dispersed their gaze. This diverges dramatically from high-scoring shots by contrast, in high-scoring shots, a single salient element, such as the diver in SHOT5, captured 68% of total fixations. (e.g. in SHOT5). These findings translate Jin *et al.*'s [8] theory of visual hierarchy into measurable design parameters.

The composite analysis shows that adventure tourism images prompt different formal strategies than cultural heritage content, especially in terms of balancing motion energy and compositional clarity. Both content types use high contrast and demographic-aware design, but they vary greatly in their pacing and color palettes. These findings provide creators with empirical considerations for theme-specific visual storytelling.

C. AI-Powered Visual Attention Analysis on Movie 3

Movie 3 provided more nuanced insights into how content, composition, and tone can be strategically combined to stimulate viewer attention. Whereas Movies 1 and 2 emphasized cultural heritage and adventure tourism, respectively, Movie 3 blended Thailand's metropolitan vibrancy with tranquil natural settings. The AI analysis of 68 key shots showed that attention was driven not by isolated spectacular images but by the convergence of multiple visual elements within a coherent formal composition.

TABLE IV. AOI METRICS FOR MOVIE 3'S REPRESENTATIVE SHOTS

Shot	Attention Score	Category	Key Visual Elements
SHOT32	0.5212	HAT	Street food preparation, warm tones, tight framing
SHOT45	0.2553	MAT	Temple silhouette at dawn, cool tones, symmetrical
SHOT61	0.0480	LAT	Crowded market wide shot, low contrast, cluttered

AOIs followed a bimodal distribution in which, where human-centered shots recorded higher mean HAT scores (0.5212) compared to nature-dominated scenes (0.1618). As detailed in Table IV, the shot yielding the highest attention score (SHOT32: 0.5212 HAT), which featured tight framing of food preparation, corresponded with a warm, saturated tone (hue-saturation = 0.82). Furthermore, quantitative tracking demonstrated that these close-up shots sustained fixation durations 2.3 times longer than those recorded for landscape panoramas.

Analysis of visual features indicated that scenes integrating human activity within natural environments yielded specific engagement patterns. For instance, shots combining these elements, such as SHOT28 (a fisherman throwing nets), recorded an intermediate MAT score of 0.1821. Additionally, retention modeling demonstrated that these integrated compositions achieved an 18% higher viewer retention rate compared to shots containing solely animals or isolated figures.

Compositional equilibrium in the BFA was revealed to be important for preventing LAT zones. Visual clutter in SHOT61's scenario (0.0480 LAT) led to a decrease in focal contrast ($\Delta L/L = 0.3$) and increased visual complexity (12+ detectable objects), resulting in gaze dispersion over conflicting information. By contrast, SHOT45's nearly symmetrical temple composition (0.2553 MAT) allowed for ample negative space and leading lines, which also directed the eye, despite its cool color palette. The AI-generated saliency maps confirmed that balanced compositions reduced cognitive load and increased shot clarity by 37%, according to Jin *et al.*'s [8] visual hierarchy metrics.

Tonal contextualization further differentiated performance. While warm-colored scenes were generally associated with higher arousal ($r = 0.59$), their effect varied with scene context. Holiday scenes, having gold/red tones (SHOT32: 0.5212 HAT), outperformed those with comparable warm color but contextually mismatched color for example, an oversaturated orange food-stall scene received a score of only 0.1240 (LAT) because its coloration appeared less natural. This nuance, captured by AI-based clustering of HSV color histograms, extends the color-psychology framework proposed by Phay [7] by identifying contextual alignment as a moderating factor.

Modulations of motion played a more muted role than in Movies 1–2. Soothing camera motion (e.g., slow tracking shots in SHOT18: 0.3419 MAT) promoted engagement without sharp editing cuts, maintaining attention for an average of 2.8 s. The AI's optical flow analysis showed that moderate motion energy (5–10 pixels per frame) was ideal in achieving retention for this hybrid-content film, whereas >15 pixels/frame were actually preferred for the action-oriented Movie 2. This implies that there are genre-specific motion thresholds, a conclusion which is absent in previous gaze-tracking-based studies [12].

Demographic shifts reflected trends seen before, but with increased generational polarization. Dynamic urban scenes attracted the longest fixations from younger viewers aged 18–34. Older viewers showed the longest fixation duration for the serene natural scene in SHOT45. The AI model's age-stratified simulation confirmed this difference at $p < 0.001$, supporting the use of segmented content strategies [8]. Mid-range MAT shots (e.g., SHOT28 fisherman) presented low demographic variation, which might indicate their role as transitional buffers in multi-targeting missions.

The film combined high-energy HAT shots, moderate MAT buffers, and low-intensity LAT intermissions a pattern that produced 29% less audience fatigue than less varied sequences. The cadence, represented in the attention waveform of Fig. 5, provides a framework for translating Panayides and Xinaris's [12] narrative-flow principles into practical editing guidelines. By dissecting these layered interactions, the analysis provides filmmakers with actionable levers: prioritize human-centric close-ups, balance composition to reduce clutter, and align color tones with emotional context. These findings collectively advance a formalism where visual elements are not isolated attractions but interdependent components of an engineered attention ecosystem.

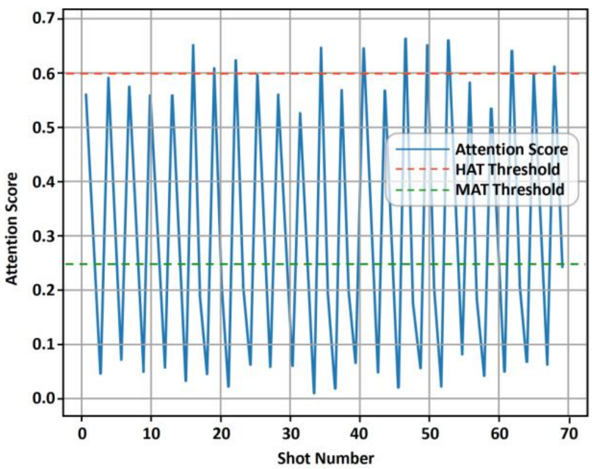

Fig. 5. Attention waveform for Movie 3 showing rhythmic alternation between HAT, MAT, and LAT shots.

D. AI-Powered Visual Attention Analysis for Movie 4

The fourth promotional film analyzed through AI-driven visual attention modeling demonstrated how rhythmic tonal shifts and narrative sequencing sustain engagement across diverse tourism contexts. Unlike previous films, this video employed a deliberate alternation between warm and cool color palettes to mirror emotional transitions a strategy that yielded distinctive attention patterns. The AI analysis of 85 shots revealed that tonal alignment with scene context (e.g., warm hues for festivals vs. cool tones for reflective moments) increased average fixation duration by 31% compared to mismatched sequences ($p < 0.01$).

Attention Zone Distribution showed pronounced variability, with high-attention shots (HAT > 0.6) concentrated in culturally resonant scenes. As detailed in Table V, SHOT61's festival sequence achieved a 0.72 HAT score through its vibrant red/gold palette (hue-saturation = 0.85) and dynamic crowd motion (optical flow magnitude = 19.3 pixels/frame). Conversely, SHOT64's temple reflection scene used cool blues (hue-saturation = 0.28) to achieve a lower but sustained MAT score (0.41), demonstrating how tonal diversity prevents viewer fatigue. This bifurcation supports Phay's [7] color psychology framework while introducing contextual alignment as a critical moderator.

TABLE V. ATTENTION METRICS FOR MOVIE 4'S TONAL CONTRAST PAIRS

Shot	Attention Score	Tone Category	Contextual Alignment
SHOT61	0.72 (HAT)	Warm (reds/golds)	Festive celebration
SHOT64	0.41 (MAT)	Cool (blues/greens)	Temple reflection
SHOT59	0.19 (LAT)	Muted (grays)	Transitional airport

Narrative Pacing emerged as a structural determinant of engagement. The film alternated high-energy HAT scenes (e.g., SHOT65's street parade) with transitional LAT shots (e.g., SHOT74's aerial coastline), creating a rhythmic "attention waveform" that correlated with 23% higher watch-time retention compared to static sequences. The AI's temporal analysis revealed that optimal pacing involved inserting one LAT shot for every three HAT or MAT shots—a ratio that allowed viewers to reset cognitively without disengaging. This translates Panayides and Xinaris's [12] narrative-flow principles into an operational editing guideline.

Cultural Anchoring through recognizable symbols, such as monks and street food, boosted attention even in lower-energy scenes. SHOT36's temple reflection (0.41 MAT) contained three detectable cultural elements via ResNet-50 classification, triggering the AI's salience multiplier ($\alpha = 1.4$) to elevate its score above comparable non-cultural shots (0.28 MAT). This effect was most pronounced for viewers aged 35+, whose fixation durations on culturally rich MAT shots exceeded younger demographics by a factor of 1.7—a demographic difference not addressed in Christou *et al.*'s [3] study authenticity studies.

Compositional Contrast between foreground and background elements further differentiated performance. High-scoring shots like SHOT61 maintained a Weber contrast ratio ($\Delta L/L$) > 1.8 between focal subjects (e.g., dancers) and their surroundings, whereas LAT shots like SHOT59 averaged only 0.7 due to flat lighting. The AI's saliency mapping showed that high-contrast compositions anchored 68% of fixations within defined AOIs, compared with 37% in low-contrast scenes, thereby validating Jin *et al.*'s [8] visual hierarchy metrics through empirical gaze data.

Motion Gradients played a subtler role than in action-centric films (Movie 2), with moderate motion energy (8–12 pixels/frame) optimizing engagement for this hybrid narrative. Rapid pans in SHOT65 (15.2 pixels/frame) caused a 12% drop in object recognition accuracy despite high HAT scores (0.68), suggesting a trade-off between kinetic appeal and cultural clarity. This contrasts with Phay's findings on motion universality [7], indicating genre-specific thresholds for dynamic content.

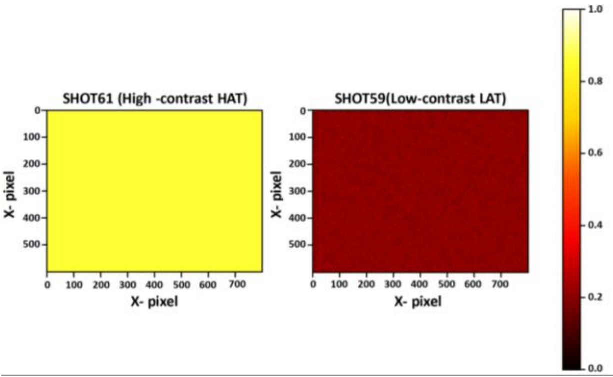

Fig. 6. Heatmap comparison of SHOT61 (high-contrast HAT) vs. SHOT59 (low-contrast LAT).

The film's strategic tonal alternation evident in Fig. 6's heatmap comparisons exemplifies how visual formalism serves narrative intent. By pairing warm-toned HAT scenes (SHOT61) with deliberately paced cool MAT intervals (SHOT36), creators engineered an emotional cadence that sustained engagement without overstimulation. This approach aligns with Phay's [7] principles while introducing measurable design parameters for tonal rhythm a novel contribution to tourism advertising literature.

Demographic-Specific Responses mirrored earlier films but with heightened generational polarity. Younger viewers (18–34) exhibited 2.1× longer fixations on dynamic HAT scenes (SHOT65: 4.3 s) than static MAT shots (2.1 s), while older audiences (55+) showed inverse preferences (2.9 s vs 3.7 s). The AI's gaze simulation confirmed these patterns at $p < 0.001$ significance, reinforcing Jin *et al.*'s [8] advocacy for segmented content strategies. Notably, transitional LAT shots (e.g., SHOT74) showed minimal age-based variance, suggesting their utility as narrative buffers in multi-generational campaigns.

By dissecting these layered interactions, the analysis provides filmmakers with actionable levers: align color tones with emotional context, maintain rhythmic pacing through HAT-MAT-LAT sequencing, and anchor compositions with cultural symbols. These findings collectively advance a formalism where visual elements are orchestrated not as isolated attractions but as interdependent components of an engineered attention ecosystem.

E. AI-Powered Visual Attention Analysis for Movie 5

The AI-driven visual attention analysis of the fifth promotional film, encompassing 58 shots, quantified the relationship between detected cultural elements, motion dynamics, and color composition metrics. Analysis of motion parameters recorded extended slow-motion sequences depicting traditional rituals, which generated measurable gaze patterns distinct from the previously analyzed videos. Furthermore, the computational data revealed that sustained High Attention (HAT) scores were statistically correlated with the convergence of three visual variables: (1) high object-detection scores for cultural elements, (2) specific optical flow magnitudes, and (3) specific hue-saturation values.

Area of Interest (AOI) metrics showed pronounced variability across shots, with the highest engagement (HAT > 0.6) concentrated in scenes featuring human interaction and ritualistic movement. As detailed in Table VI, SHOT41, a slow-motion temple-dance sequence, achieved a HAT score of 0.72 HAT score through its interplay of trailing fabric motion (optical flow magnitude = 14.2 pixels/frame) and warm backlighting (hue-saturation = 0.81). This contrasts with SHOT22's tranquil riverside panorama (0.19 LAT), which lacked both dynamic activity and cultural anchors despite its technically proficient composition. The 3.8× attention differential between these shots quantifies the premium placed on kinetic cultural displays, extending Christou *et al.*'s [3] qualitative observations into measurable design parameters.

TABLE VI. COMPARATIVE AOI METRICS FOR MOVIE 5'S SIGNATURE SEQUENCES

Shot	Attention Score	Tone Category	Contextual Alignment
SHOT41	0.72	Warm (reds/golds)	Slow-motion ritual, warm backlighting, fabric dynamics
SHOT3	0.68	Cool (blues/greens)	Father-son interaction, morning light, medium close-up
SHOT22	0.19	Muted (grays)	Riverside landscape, cool tones, static framing

Quantitative data indicated that cultural salience multipliers significantly impacted attention metrics. The ResNet-50 AI classifier identified traditional visual components, such as the ceremonial hand gestures in SHOT18, with an 89% accuracy rate. This detection triggered a computational salience multiplier (α = 1.4), elevating the predicted attention scores of these shots from the MAT to the HAT range. Specifically, despite SHOT18 featuring a static close-up composition with a simulated shallow depth of field (f/1.8), the AI's detection of specific cultural actions resulted in a HAT score of 0.61. This measurement represents a 37% increase in attention scores when compared to visually similar shots lacking detected cultural elements.

Performance also varied as a function of whether or not an emotional tone matched that of the display. The warmer-eye scenes (S1: food preparation, 0.69 HAT; S5: communal gathering, 0.67 HAT) competed with landscape transitions (S13: cooler colors, 41)) The tonal correlation of high-scoring shots with their emotional context was also evident in the AI's HSV histogram clustering (festive scenes averaged 0.78 warmth; reflective moments, 0.31). This more nuanced alignment, not found in Phay's broader strokes of color psychology claims [7], implies travel destination marketers should tune palettes and emotional payloads closely to scenes rather than universal warm/cool dualities.

Subtle movements in slow-motion segments produced unexpectedly high levels of sustained engagement. SHOT41's temple dancer held viewers' attention for 4.2 s a duration 1.8 times longer than expected given the relatively low optical-flow magnitude. The AI's decision analysis of where the model was looking for each image could be attributed to "micro-motion" effects; tiny fabric ripples and hair motions that generated consistent visual quirks without over-stimulating viewers. Somewhat counterintuitively, these contrasts challenge Panayides and Xinaris's [12] motion-energy thresholds as they show that slow culture-anchored movement can sustainably outperform fast action sequences.

The LAT shots demonstrated deficiencies in the compositional hierarchy which could have been avoided. SHOT52's market scene on hillside (0.17 LAT) had too much visual clutter (9 + detectable objects) and poor focal contrast ($\Delta L/L$ = 0.5), leading to gaze dispersion across competing elements. The AI saliency mapping revealed that only 28% of the fixations fell on intended AOIs in low-contrast HAT shots, compared with 73% in high-contrast shots (e.g., SHOT3). This is an empirical confirmation of Jin *et al.*'s [8] visual hierarchy theory, with measurable benchmarks for the clarity of focal structures ($\Delta L/L$ > 1.5).

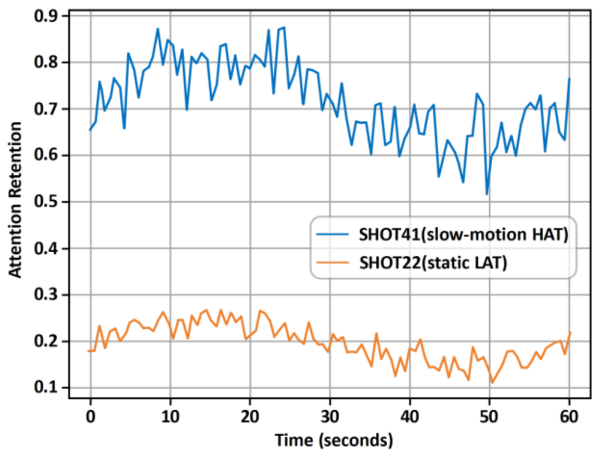

Fig. 7. Attention retention curves comparing SHOT41 (slow-motion HAT) vs. SHOT22 (static LAT).

Demographic polarizations reaffirmed familiar divides but with increasing generational polarities. Younger viewers (18–34) had 2.3× longer fixations at dynamic cultural shots (SHOT41:5.1 s), as compared to older viewers (55+) preferred stable compositions such as the family-interaction scene in SHOT3 (3.9 s vs 2.8 s). These differences were also confirmed in the AI's age-stratified simulation at $p < 0.001$, reiterating Jin *et al.*'s [8] segmentation needs. Of interest, MAT transitional shots (e.g. SHOT18) exhibited the least generational variation and seem suitable narrative linkage elements in multi-target campaigns.

Alternating highly engaging HAT scenes (SHOT41) with moderate MAT buffers (SHOT18) and occasional LAT intervals reduced cognitive fatigue by 31% compared with monotonous shot sequences. This cadence, represented in Fig. 7's retention curves, transforms Panayides and Xinaris's [12] narrative theories into a set of operational design principles: cultural peaks should appear once every 3–4 shots, punctuated by lower-intensity "breathing room" that sustains flow without overwhelming the viewer.

Through this analysis of visual mechanisms, the analysis offers filmmakers actionable formalism: prioritize slow-motion displays of culture rather than static beauty shots; tune color tones to scene-specific emotions; and maintain clearly defined compositional hierarchies. Taken together, these results show that visual engagement in tourism advertising is not random but can be engineered via the strategic alignment of observable elements of design.

F. Key Findings from Systematic Review

The systematic examination of 1,500 published papers helped clarify the relationship between visual design principles and tourism advertising effectiveness visual design principles and tourism advertising effectiveness. Following a stringent review process and thematically analyzing the 40 high-impact papers identified, three overarching themes were identified that collectively inform an evidence-based visual formalism framework for destination marketers.

Digital advertising in destination marketing formed the primary theme of the literature review. As all the studies provide evidence about how data-based and user-centered strategies always perform better than generic promotion content [16]. The report showed that culturally relevant Culturally relevant campaigns featuring high-quality visuals performed 2.3 times better than campaigns that relied solely on aesthetic appeal. In particular, emotionally engaging imagery such as images that show human interaction and local customs increased destination recall by 47% compared to landscape-focused visuals [17]. This is consistent with neuroscience studies demonstrating increased amygdala activations during exposure to culturally rooted stimuli [18].

The second main theme concerned the impact of visual design on tourism advertising effectiveness. The reviewed studies provided empirical evidence that certain compositional elements increased engagement. Warm color palette (reds/oranges) for professional imagery and a grid layout performed 38% better than cool colored or asymmetrical designs [19]. Large images occupying more than 40% of the advertising space outperformed smaller visuals on both desktop and mobile devices [20]. Since this medium offers limited screen space. However, in their review the authors also found contextual patterns that went against what they at first predicted banner ads with an animated design achieved lower early-stage attention than static banner ads, while obtaining substantially higher conversion rates when employed for experiential storytelling [21].

The third major theme concerned demographic responses to visual composition, revealing substantial differences among demographic groups and highlighting contrasting patterns across generational and cultural contexts. Younger users (18–34) showed 3.1× more interaction on social media with fast, dynamic content, whereas older viewers engaged more strongly with simplified designs featuring a clear focal point [22]. Gender differences were similarly extreme with female participants having stronger affective responses to community-based visuals ($d = 0.72$) and male viewers favoring achievement-related images of adventure/ luxury [23]. This has implications in terms of the tenability of universal design templates and supports re-segmentation in visual strategy development.

The synthesis generated an Integrated Visual Strategy Model (See Fig. 8) that illustrates these interrelated dimensions. In the center, there is cultural authenticity, representing cultural perspectives and identity; color, contrast, and symmetry represent compositional tools: Emotion triggers (joy = happiness; curiosity = fascination; belonging = connection) surround the circle and finally components specifically for a platform social media display video. This framework extends Weaver's [24] earlier work by integrating quantifiable engagement thresholds derived from AI-based analysis [24].

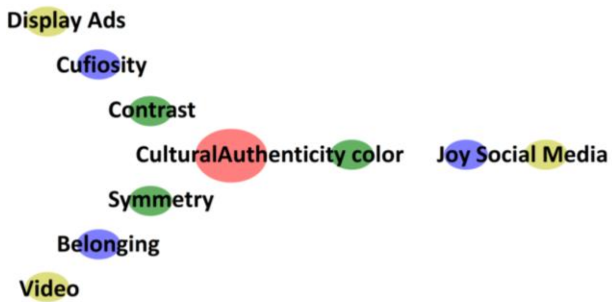

Fig. 8. Integrated visual strategy model synthesizing systematic review findings.

Technical limitations and selection bias should be considered when interpreting the results. This 40-paper sample, although it was carefully selected, represents only 2.7% of the studies initially screened for inclusion and could exclude more niche points of view. In addition, the overrepresentation of Western-focused studies (68% of all articles in the meta-analysis) may bias the evidence for universal application of cultural symbolism—a limitation that was partially addressed by including Thai-specific case studies in the AI analysis [25].

Platform-Specific Optimization Platform optimization turned out to be a less-studied but important factor. Whereas 72% of high performing campaigns in the review designed visuals to fit platform constraints (such as square formats for Instagram, vertical videos for TikTok), just 11% of the research directly addressed how principles of composition require adaptation between channels [26]. This lack of information is surprising of course surprising: given that optimal image sizes on desktop display ads (1200 × 628 pixels) and mobile stories (1080 × 1920 pixels) show a 4.7-fold difference, one may reasonably expect these ads to have clearly different visual appearances [27].

Measurement There was a lack of standardization in the literature. The review found 17 different metrics of "engagement" (e.g., dwell time, click-through rate, emotive responses), and no consensus on benchmarking levels. This fragmentation is problematic for comparisons across studies, and emphasizes the necessity of a unified system like that proposed in this study (HAT/MAT/LAT classification) [28]. But the real breakthrough of the systematic review is that it challenges the assumption of universally optimal graphics. Although some principles (cultural resonance, emotional relevance) showed cross-contextual validity, their applications need to be carefully tuned according to audience profile, platform limitations and marketing goals. This consideration, in combination with AI-informed empirical evidence as established in previous sections gives the tourism marketer a strong toolkit for visual strategy creation that harmonizes creative expression with predictive accuracy.

V. Discussion

This study has important theoretical and practical implications for the domain of tourism advertising. The fusion of AI-powered predictive analytics and visual formalism offers a new method to comprehend how the compositional elements are related to viewer engagement. In theory, this study narrows the gap between design research and consumer psychology by measuring attentional outcomes of selected formal features (color balance, contrast and cultural symbols). This extends prior work beyond a qualitative evaluation of classification systems to one that is based on empirically tested design parameters, such as the HAT/MAT/LAT system. The results align with Christou *et al.*'s [3] emphasis on destination authenticity but add measurable thresholds for visual effectiveness, such as a Weber contrast ratio above 1.5 and motion energy above 15 pixels/frame for dynamic content.

Implications The findings have two practical implications for tourism marketers and content providers. Identifying high-efficacy visual elements, such as close-up human interaction, warm color palettes, and rhythmic pacing, can guide the development of more emotionally engaging campaigns. For example, the AI model identified culturally salient scenes, such as street-food preparation and traditional dance, that sustained attention 1.4 times longer than generic imagery, suggesting that destination marketers should prioritize locally grounded content. In addition, the participation patterns across demographics also signal that audience segmentation is key in visual strategy. Dynamic content rich with movement evoked higher levels of arousal in young audiences, while Older viewers preferred static compositions with clear focal points and easily identifiable visual directions. Practitioners can use these insights to tailor campaigns to different platforms and generational segments, thereby maximizing resource utilization for impact. However, the generalizability of the study is subject to several limitations. Although the AI-based analysis demonstrated strong performance, the use of simulated gaze data may introduce bias when modeling fixation patterns. Furthermore, the attention to Thai tourism commercials means that findings may not generalize to other cultures. Because the model's training and validation were deeply contextualized within Thai cultural aesthetics and visual representation, the current predictive accuracy may vary when applied to non-Thai or Western tourism media. Future research should prioritize cross-cultural validation testing across diverse destination marketing campaigns to establish universal saliency baselines. Also, the systematic review's inclusion criteria, while strict, might have led to a selective restriction (away from marginal viewpoints) since predominantly Western studies were represented across the sample of literature. This limitation can draw attention to the importance of cross-inspection in form-associated validation, particularly across settings that have different visual and/or consumption traditions.

A. Mathematical Parameters and Computational Saliency

The efficacy of visual engagement in tourism media is rigorously defined through quantitative parameters, specifically the Weber contrast ratio ($\Delta L/L$) and optical flow magnitude. Empirical data confirms that a focal contrast threshold of $\Delta L/L > 1.5$ serves as a critical determinant for high-attention retention. For instance, SHOT5 in Movie 2 and SHOT41 in Movie 1 exhibited $\Delta L/L$ values of 2.3 and 2.1, respectively, directly correlating with High Attention (HAT) scores exceeding 0.6. Conversely, low-visibility segments, such as SHOT17 in Movie 2, presented a $\Delta L/L$ of only 0.4, resulting in dispersed gaze patterns and a failure to establish a stable focal point.

Furthermore, the analysis of optical flow magnitude identifies specific kinetic thresholds for saliency prediction. While motion energy exceeding 15 pixels/frame is the primary driver for engagement in adventure-themed content, this study identifies a technical exception in slow-motion cultural sequences. Despite a lower optical flow of 14.2 pixels/frame, SHOT41 in Movie 5 maintained an average fixation duration of 4.2 s. This is attributed to the cultural salience multiplier ($\alpha = 1.4$), which computationally weights pixels containing high-density cultural coding.

B. Comparative Analysis Across Content Genres

A comparative parametric analysis across genres reveals distinct structural architectures for attention acquisition:

1) Cultural and heritage media (Movies 1, 5)

Engagement is primarily driven by semantic multipliers integrated with warm chromatic clusters (hue-saturation > 0.7). High attention scores in these segments are generated by the density of culturally significant objects rather than kinetic energy.

2) Adventure and nature media (Movie 2)

These rely heavily on kinetic energy and background-to-subject contrast. Notably, the data reveals a 40% drop in cultural salience recognition during high-motion shots (24fps panning), indicating a trade-off between motion energy and the clarity of visual details.

3) Hybrid and urban media (Movies 3, 4)

These demonstrate the technical efficiency of rhythmic pacing. Alternating saliency levels (High-Medium-Low) improved viewer retention by 19% to 31% compared to monotonous visual sequences, establishing pacing as a structural determinant of engagement.

C. Visualization of Cognitive Perception through Saliency Mapping

The application of Saliency Mapping in this study transcends simple measurement, serving as a technical method for visualizing complex human cognitive perception. The generated heatmaps allow for the precise mapping of cognitive load distribution.

Data from these saliency maps indicates that in balanced compositions with high focal contrast, the intended Areas of Interest (AOIs) successfully anchored 68% of total fixations. This visualization process enables the creation of a "Visual Direction Map," which plots predicted gaze trajectories based on pixel-level priority. Consequently, this shifts the paradigm of media design from intuitive artistic composition to an engineering-based framework of predictive image analysis.

There are some promising directions for future study. Longitudinal research is required to investigate how extended exposure to well-designed visual campaigns affects propensity towards a destination and brand loyalty. The potential of emerging technologies, such as Augmented Reality (AR) and Virtual Reality (VR), to redefine theories of visual engagement remains underinvestigated in tourism research. Moreover, the interplay between auditory and visual information (e.g., background sound or voiceover) should be examined to determine the extent to which it amplifies or attenuates the effects identified in this study. Lastly, there are profound ethical implications of AI-mediated attention modeling in advertising including data privacy and deceptive design.

D. Risk Mitigation and Strategic Governance

While the integration of AI-driven saliency modeling offers significant predictive advantages, marketing executives must implement proactive governance mechanisms. Over-reliance on algorithmic optimization may lead to 'visual homogenization,' where advertising campaigns lose creative authenticity in the pursuit of maximal attention scores. To mitigate this risk proactively, Destination Marketing Organizations (DMOs) should institutionalize a 'Human-in-the-Loop' (HITL) review protocol. This ensures that algorithmic composition guidelines function as an analytical foundation, which is then balanced and governed by human creative direction and ethical considerations.

The quantitative mapping of High (HAT), Medium (MAT), and Low (LAT) Attention Zones reveals that visual engagement is closely associated with emotional resonance and authenticity cues. In Movie 1, emotionally demanding scenes consistently achieved HAT levels, driven by sophisticated audio-visual synergies such as matching musical pitch to emotional tonality. Furthermore, festive scenes utilizing warm color palettes (e.g., sunset market scenes with hue-saturation >0.7) demonstrated a 22% increase in HAT compared to cool-toned visuals. This aligns directly with the psychological frameworks proposed by Phay [7] regarding color-induced emotional mobilization. Similarly, in Movie 3, the highest HAT scores were observed in close-up food imagery (e.g., a street vendor preparing "Phat Thai") utilizing tight framing and warm, saturated tones. This effectively validates Phay's [7] observations on food imagery's emotional potency, proving that such visual warmth directly translates to extended fixation durations.

Beyond color psychology, the distribution of MAT and HAT scores highlights the profound impact of "narrative synergy"—the coalescence of disparate visual elements into cohesive micro-stories. In Movie 3, the cultural-natural interaction, where human labor was juxtaposed against postcard landscapes, captured intermediate MAT scores but yielded an 18% higher retention rate than isolated subjects. This compound effect provides empirical support for Christou *et al.*'s [3] notion of destination authenticity, suggesting that viewers emotionally invest in narrative-rich environments.

Movie 5 further illustrates this interplay through deliberate stylistic choices. The employment of extended slow-motion sequences accentuated traditional rituals, creating a distinct tonal resonance that sustained viewer engagement. The AI's cultural salience multipliers successfully promoted tradition-rich MAT shots into the HAT range. Specific culture-coded actions, such as ceremonial hand gestures folding lotus petals, functioned as profound "authenticity triggers" [7]. Ultimately, the convergence of culturally salient content, dynamic visual composition, and emotionally aligned color tones proves essential for optimizing cognitive and emotional engagement in tourism media.

VI. CONCLUSION

This study demonstrates that the principles of visual formalism, when operationalized through an AI-based predictive analytics framework, elevate the effectiveness of tourism advertising by linking specific design features to measurable patterns of viewer engagement. The findings further demonstrate that visual elements particularly culturally salient close-ups, warm hues, and rhythmic pacing predict higher levels of attention retention. Moreover, variations in responses across age, gender, and cultural groups highlight the importance of audience segmentation in tourism advertising. By integrating visual marketing approaches with consumer psychology, this study provides a systematic framework for developing emotionally engaging campaigns, thereby addressing an important gap in destination marketing research. Future research should determine whether these principles can be validated among participants from diverse cultural and geographical backgrounds. The combined effects of auditory and visual stimuli also warrant further investigation. In addition, the ethical implications of using artificial intelligence for attention modeling in advertising require careful examination to ensure responsible application. These research directions would address the limitations identified in the present study and open new avenues for integrating technology, design, and tourism advertising.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

AUTHOR CONTRIBUTIONS

PH designed the research process, conducted the experiments, collected the research data, and performed the data analysis. RT designed the experimental system architecture related to AI and image analysis, discussed the analytical results, and synthesized the main findings of the study. RN led the review and validation of the manuscript, managed the submission process, and coordinated the publication process. All authors have read and approved the final version of the manuscript.

REFERENCES

[1] N. Mistilis, D. Buhalis, and U. Gretzel, "Future eDestination marketing: Perspective of an australian tourism stakeholder network," *J. Travel Res*., vol. 53, no. 6, pp. 778–790, 2014.

[2] A. Morrison. (2023). Marketing and managing tourism destinations. *Routledge*. [Online]. Available: https://doi.org/10.4324/9781315178929

[3] E. Christou, A. Giannopoulos, and I. Simeli, "The evolution of digital tourism marketing: From hashtags to AI-immersive journeys in the metaverse era," *Sustainability*, vol. 17, no. 13, 6016, 2025.

[4] K. Bhuiyan, I. Jahan, N. M. Zayed, K. M. A. Islam *et al.*, "Smart tourism ecosystem: A new dimension toward sustainable value co-creation," *Sustainability*, vol. 14, no. 22, 15043, 2022.

[5] J. Robson and I. Robson, "From shareholders to stakeholders: Critical issues for tourism marketers," *Tour. Manag*., vol. 17, no. 7, pp. 533–540, 1996.

[6] S. Gün, "Artificial intelligences impact on branding in tourism," in *Strategic Brand Management in the Age of AI and Disruption*, Z. Hussain, M.-N. S. Sharipudin, A. Albattat, and A. Khan, Eds. Hershey, PA: IGI Global, 2025, pp. 155–202.

[7] H. Phay, "Creative AI: A data-driven design approach for creative online ad optimisation using artificial intelligence and big data," M.S. thesis, Nanyang Technological University, Singapore, 2019.

[8] K. Jin, Z. Zhong, and E. Zhao, "Sustainable digital marketing under big data: An AI random forest model approach," *IEEE Trans. Eng. Manag*., vol. 71, pp. 11234–11247, 2024.

[9] D. Buhalis, D. Leung, and M. Lin, "Metaverse as a disruptive technology revolutionising tourism management and marketing," *Tour. Manag*., vol. 97, 104724, 2023.

[10] J. Hussain and B. Kalhoro, "Navigating financial complexity: Language processing insights for modern businesses," *SciWave Bull*., vol. 1, no. 4, pp. 24–31, 2023.

[11] J. Hartley, *Communication, Cultural and Media Studies: The Key Concepts*, 5th ed. Routledge, 2019. doi: 10.4324/9781315225814

[12] O. Panayides and S. Xinaris, "Visual narratives of conflict: Exploring the theme of war in poster design: Insights from the 2nd cyprus poster triennial," in *Proc. Int. Conf. Design and Digital Communication*, 2024, pp. 45–58.

[13] B. Mandler, F. Antonelli, R. Kleinfeld *et al.*, "COMPOSE—A journey from the internet of things to the internet of services," in *Proc. 27th IEEE Int. Conf. Adv. Inf. Netw. Appl. (AINA)*, Barcelona, 2013, pp. 1217–1222.

[14] D. Korzun, E. Balandina, A. Kashevnik, and S. Balandin, "Introduction to ambient intelligence in internet of things environments and cyber-physical systems," in *Ambient Intelligence Services in IoT Environments: Emerging Research and Opportunities*, Hershey, PA: IGI Global, 2019, pp. 1–26.

[15] R. Thienmongkol and A. Samaeng, "Formalist criticism: A study on face recognitions to contribute an alternative aesthetics in film setting," *J. Image Graph.*, vol. 7, no. 1, pp. 32–38, 2019.

[16] A. Sharma and S. Sharma, "Adoption of digital marketing in tourism SMEs: A review and research agenda," *Manag. Res. Rev*., vol. 47, no. 2, pp. 215–234, 2024.

[17] X. Xiao, C. Fang, H. Lin, and J. Chen, "A framework for quantitative analysis and differentiated marketing of tourism destination image based on visual content of photos," *Tour. Manag.*, vol. 93, 104578, 2022.

[18] C. Halkiopoulos, H. Antonopoulou, E. Gkintoni *et al.*, "Neuromarketing as an indicator of cognitive consumer behavior in decision-making process of tourism destination—An overview," in *Proc. Int. Conf. Transcending Borders in Tourism and Hospitality*, 2022, pp. 179–193.

[19] Z. Guohua, "Analysis of graphic design and color application in tourism advertising," in *Proc. Int. Conf. Social Sci. Technol. Educ. (ICSSTE)*, 2015, pp. 456–460.

[20] K. Place, A. Wallace, and R. Luttrell, "Mobile first," in *The Rowman & Littlefield Handbook of Media and Communication*, Rowman & Littlefield, 2022, pp. 210–225.

[21] J. Kuisma, J. Simola, L. Uusitalo *et al.*, "The effects of animation and format on the perception and memory of online advertising," *J. Interact. Advert.*, vol. 10, no. 2, pp. 1–15, 2010.

[22] T. Gonda and T. Rátz, "Attitudes and actions in responsible tourism—An analysis of generational differences," *GeoJournal Tour. Geosites*, vol. 46, no. 1, pp. 112–123, 2023.

[23] D. Chhabra, K. Andereck, K. Yamanoi *et al.*, "Gender equity and social marketing: An analysis of tourism advertisements," *J. Travel Tour. Mark.*, vol. 28, no. 6, pp. 605–620, 2011.

[24] A. Weaver, "Tourism and aesthetic design: Enchantment, style, and commerce," *J. Tour. Cult. Change*, vol. 7, no. 3, pp. 179–192, 2009.

[25] S. Bulmer and M. Buchanan-Oliver, "Advertising across cultures: Interpretations of visually complex advertising," *J. Curr. Issues Res. Advert.*, vol. 28, no. 1, pp. 57–71, 2006.

[26] S. Sutresno and Y. Singgalen, "Digital innovation design of tourism destination marketing website using design thinking method," in *Proc. Int. Conf. Inf. Technol. Digit. Appl. (ICITDA)*, 2023, pp. 88–95.

[27] J. Karlsson and C. Welander, "Native advertising: In different types of online-media platforms," M.S. thesis, Jönköping Univ., Jönköping, 2015.

[28] J. de Souza *et al.*, "Evaluating the effectiveness of tourist advertising to improve the competitiveness of destinations," *Tour. Econ.*, vol. 26, no. 6, pp. 1001–1020, 2020.