International Journal of Interdisciplinary Research
ISSN(Online): 3090-2959
Vol 2 no 2 (2026): July 2026
https://journal.as-salafiyah.id/index.php/ijir/index
Email: ijireditor7@gmail.com

Stopping The Scroll: Visual Triggers Of User Attention In Short-Video Advertising

[1]Achmad Ridha
[1]*Universitas Negeri Makassar, Indonesia*
[1]*achmad.ridha@unm.ac.id*

**Correspodence Email: achmad.ridha@unm.ac.id*

Abstract: *The proliferation of short-video platforms has fundamentally altered advertising dynamics, compelling marketers to capture user attention within two to three seconds before the scroll moves on. While prior research has explored the experiential dimensions of attention and cognitive flow in TikTok advertising, no study has systematically identified and mechanistically explained the specific visual elements that arrest scrolling behavior and trigger initial attention, a gap that leaves the first causal link in the advertising effectiveness chain theoretically unaddressed. This study employs a qualitative explanatory approach to investigate which visual triggers are most effective in capturing user attention during short-video ad consumption and to explain the cognitive and perceptual mechanisms through which they operate. Drawing on stimulus-based semi-structured interviews with twelve active TikTok users in Makassar City, Indonesia, and guided by visual salience theory, the findings identify four dominant visual trigger categories: (1) dynamic motion and rapid transitions, (2) human facial expression and direct eye contact, (3) high-contrast color compositions, and (4) text overlay and caption placement. The study explains how each trigger operates through distinct perceptual pathways, activating bottom-up attentional capture before top-down cognitive processing can engage. These findings extend Ridha's (2024) phenomenological account of attention experience by moving beyond description to specify the visual mechanisms that initiate the attentional entry phase, a contribution that bridges visual salience theory with digital attention phenomenology in a non-Western short-video market. Practical implications are offered for digital advertisers designing scroll-stopping creative content in the Indonesian short-video market.*
Keywords: *Visual Triggers, Attention Capture, Visual Salience, Short-Video Advertising, Digital Marketing*

INTRODUCTION

In the contemporary digital media landscape, human attention has become the scarcest and most contested resource in marketing communication. The emergence of short-form video platforms, most prominently TikTok, Instagram Reels, and YouTube Shorts, has compressed the window of opportunity for advertisers to an unprecedented two to three seconds before a user's

International Journal of Interdisciplinary Research
ISSN(Online): 3090-2959
Vol 2 no 2 (2026): July 2026
https://journal.as-salafiyah.id/index.php/ijir/index
Email: ijireditor7@gmail.com

thumb moves to the next piece of content (Langan et al., 2022). This behavioral reality constitutes what scholars term the scroll economy: an ecosystem where the decision to stop, watch, or skip is governed not by deliberate reasoning but by near-instantaneous perceptual responses to visual stimuli (Montag et al., 2023). Understanding the specific visual properties that trigger this arrest response is therefore of fundamental theoretical and practical importance.

Despite a growing body of research on digital attention, most existing studies have approached the phenomenon either through quantitative metrics such as view-through rates, skip rates, and eye-tracking data, or through broad experiential frameworks. Ridha (2024), for instance, employed phenomenological inquiry to map the experiential landscape of attention and cognitive flow in TikTok advertising, identifying four layers of attention experience: instant attention, digital distraction, cognitive flow, and emotional connectedness. While this work makes a substantial contribution to understanding the qualitative texture of digital attention, it does not systematically disaggregate the visual elements that initiate the attentional process. The question of precisely which visual triggers cause a user to stop scrolling and why they work remains inadequately addressed in the literature.

This gap is consequential. If attention is the gateway to all downstream advertising effects including brand recall, emotional engagement, and purchase intent, then the visual properties of the attention trigger constitute the first causal link in the advertising effectiveness chain. Visual salience theory (Treisman & Gelade, 1980; Itti & Koch, 2001) provides a foundational framework for understanding how certain visual stimuli involuntarily capture attention through bottom-up perceptual mechanisms, but its application to short-video advertising in the Indonesian context remains limited. Meanwhile, explanatory qualitative research, which seeks not merely to describe but to account for the mechanisms underlying a phenomenon, offers a methodologically appropriate and underutilized approach for investigating this question (Maxwell, 2012).

The Indonesian context adds further specificity to the inquiry. With over 113 million active TikTok users as of 2025 (Statista, 2025), Indonesia represents one of the world's largest short-

International Journal of Interdisciplinary Research
ISSN(Online): 3090-2959
Vol 2 no 2 (2026): July 2026
https://journal.as-salafiyah.id/index.php/ijir/index
Email: ijireditor7@gmail.com

video markets, yet the visual attention preferences of Indonesian users remain understudied relative to Western and East Asian populations. Cultural dimensions of visual processing, including preferences for certain color palettes, facial expressions, and movement styles, may shape which visual triggers prove most effective in this context (Nurhayati & Siregar, 2021).

This study aims to identify and explain the visual triggers that most effectively arrest scrolling behavior during TikTok advertisement exposure among users in Makassar, Indonesia. It seeks to answer two questions: (1) Which visual elements most consistently capture initial attention in short-video advertisements? (2) Through what perceptual and cognitive mechanisms do these visual triggers operate? The findings contribute to digital advertising theory by bridging visual salience research and attention phenomenology, and offer actionable guidance for creative practitioners in the short-video advertising domain.

LITERATURE REVIEW

Visual Salience and Attentional Capture

The theoretical foundation for understanding visual attention triggers lies in the feature integration theory of visual salience (Treisman & Gelade, 1980), which proposes that certain visual features such as color, motion, orientation, and contrast are processed pre-attentively and in parallel across the visual field, capable of commanding attention involuntarily before conscious processing occurs. This bottom-up attentional capture operates independently of a viewer's goals or intentions, making it particularly relevant to advertising contexts where audiences are not actively seeking engagement with commercial content.

Contemporary computational models of visual salience, most notably those developed by Itti and Koch (2001), operationalize these principles by specifying that attentional priority is assigned to visual locations exhibiting maximal contrast with their surroundings across multiple feature dimensions simultaneously. In the context of digital video, this translates to a preference for sudden motion, high-luminance contrast, and abrupt scene changes as automatic attention

International Journal of Interdisciplinary Research
ISSN(Online): 3090-2959
Vol 2 no 2 (2026): July 2026
https://journal.as-salafiyah.id/index.php/ijir/index
Email: ijireditor7@gmail.com

attractors. Studies applying saliency modeling to video advertising confirm that early-frame visual salience scores are predictive of real-world view completion rates (Wang et al., 2022), validating the theoretical relevance of this framework to short-video advertising.

Visual Attention in Short-Video Environments

Short-video platforms impose distinctive constraints on visual attention dynamics. The infinite scroll interface creates a state of perpetual visual competition, in which each piece of content must overcome the momentum of habitual scrolling behavior. Montag et al. (2023) describe this as a dopamine-driven attentional rhythm that makes sustained focus cognitively costly. Within this environment, the first frame of a video functions as a visual interruption signal: it must register sufficient salience to interrupt the scroll pattern before the user's pre-motor response overtakes perceptual processing (Langan et al., 2022).

Research on visual attention in social media advertising identifies several recurring effective triggers. Kim and Sullivan (2019) found that human faces, particularly those displaying expressive or emotionally congruent expressions, are among the most reliable attention-capturing stimuli in digital video, consistent with the evolutionary prioritization of social information processing. Similarly, Villanueva et al. (2024) demonstrated that visual coherence between motion rhythm, color palette, and narrative tempo significantly enhances scroll-stopping power beyond the contribution of individual elements alone, suggesting that visual trigger effectiveness is partly compositional. Lin and Kim (2023) further documented that text overlay in the first two seconds functions as a cognitive entry point, providing semantic anchoring that transitions viewers from pre-attentive capture to intentional engagement.

From Attention Capture to Cognitive Engagement

The distinction between initial attentional capture and sustained cognitive engagement is theoretically significant and practically actionable. Drawing on Ridha's (2024) phenomenological framework, attention in TikTok advertising operates in layered phases: an instantaneous visual arrest phase, followed by a rapid evaluative phase in which users determine whether to invest

International Journal of Interdisciplinary Research
ISSN(Online): 3090-2959
Vol 2 no 2 (2026): July 2026
https://journal.as-salafiyah.id/index.php/ijir/index
Email: ijireditor7@gmail.com

further attention, and in optimal conditions a flow experience characterized by complete immersive engagement. The visual triggers investigated in the present study operate primarily in the first phase, though their design properties necessarily influence the likelihood of transitioning to deeper engagement phases.

Cognitive load theory (Sweller, 1988; Leppink et al., 2015) offers a complementary explanatory lens. Visual stimuli that create excessive cognitive load through over-complexity, competing focal points, or rapid information density may capture initial attention but prevent processing efficiency, resulting in early disengagement rather than scroll-stopping. Effective visual triggers must therefore balance high salience with processability: they must stand out while remaining legible. This balance is particularly delicate in the compressed temporal window of short-video advertising, where cognitive resources are already taxed by the high-speed scrolling environment.

The Indonesian Short-Video Advertising Context

Indonesia's unique media landscape shapes the visual attention dynamics of its TikTok users. TikTok has become integral to digital marketing strategy for both large brands and small-medium enterprises (UMKM) in Indonesia, functioning as a primary vehicle for visual brand storytelling and consumer engagement (Hayati & Sudradjat, 2022; Novita et al., 2023). Nurhayati and Siregar (2021) document that Indonesian social media users exhibit heightened sensitivity to emotional authenticity and social relatability in advertising content, preferring visual narratives that reflect everyday communal life over aspirational or hyperstylized aesthetics. Firamadhina and Krisnani (2021) further establish that Generation Z users in Indonesia treat TikTok not merely as entertainment but as a socially embedded medium in which content credibility is assessed through visual sincerity and perceived proximity to real-life experience. Husna and Mairita (2024) found that Gen Z's content consumption on TikTok is strongly mediated by visual appeal, whereby content that succeeds in the first peripheral glance, before active evaluation begins, is significantly more likely to be watched to completion. These contextual factors suggest that visual trigger

International Journal of Interdisciplinary Research
ISSN(Online): 3090-2959
Vol 2 no 2 (2026): July 2026
https://journal.as-salafiyah.id/index.php/ijir/index
Email: ijireditor7@gmail.com

effectiveness in Indonesia may not be reducible to universal salience principles alone, but is modulated by culturally situated perceptual dispositions.

METHOD

Research Approach

This study employs a qualitative explanatory design. Unlike purely descriptive or exploratory qualitative research, qualitative explanatory research proceeds from identified phenomena toward theoretical accounts of the mechanisms that produce them (Maxwell, 2012). The goal is not merely to document which visual triggers attract attention, but to reconstruct the perceptual and cognitive pathways through which they operate, drawing on participants' conscious reflections on their own attention processes. This methodological orientation distinguishes the present study from Ridha's (2024) phenomenological description of the attention experience and positions it as a mechanistic extension of that foundational work.

An interpretive phenomenological framework guides data collection and analysis, ensuring that explanatory claims remain grounded in the lived experience of informants rather than imposed from theoretical preconceptions (Smith et al., 2009). This approach is epistemologically appropriate given that attentional capture is both a neurobiological event and a meaningfully interpreted experience, and that the mechanisms of visual trigger effectiveness are partly accessible to reflective introspection, particularly when participants are exposed to concrete stimuli and invited to articulate their real-time responses.

Participants and Sampling

Twelve active TikTok users in Makassar City, Indonesia, were recruited via purposive sampling. Inclusion criteria required: (1) minimum daily TikTok usage of one hour; (2) exposure to in-feed and brand partnership advertisements; (3) age between 18 and 35 years, representing the digital native demographic (Gen Z and early millennials); and (4) demonstrated capacity for reflective self-report about attention and visual experience. An equal gender distribution was

International Journal of Interdisciplinary Research
ISSN(Online): 3090-2959
Vol 2 no 2 (2026): July 2026
https://journal.as-salafiyah.id/index.php/ijir/index
Email: ijireditor7@gmail.com

maintained (6 male, 6 female) to account for potential gender-related differences in visual processing preferences. Informants are identified by codes R1–R12 to ensure confidentiality in accordance with research ethics protocols.

The sample size of twelve informants is consistent with qualitative adequacy norms for interpretive phenomenological research, where theoretical saturation and narrative depth are prioritized over statistical representativeness (Creswell & Poth, 2018). Saturation was assessed through iterative analysis during data collection; no substantially new visual trigger categories emerged after the ninth informant, confirming sample adequacy.

Data Collection

Data were collected through stimulus-based semi-structured interviews. To ensure comparability across informants and to provide concrete referents for visual reflection, all participants were shown a standardized set of six TikTok advertisements. The six advertisements were selected using a purposive stimulus selection protocol guided by three criteria: (1) representational diversity, ensuring the set included at least one advertisement in which each of the four anticipated trigger types (motion-dominant, face-dominant, color-dominant, and text-overlay-dominant) served as the primary visual feature; (2) ecological validity, ensuring each advertisement had been published by an active Indonesian brand or content creator and had accumulated at least 100,000 organic views, reflecting real market exposure conditions; and (3) content neutrality, excluding advertisements for products in categories likely to elicit strong a priori affective responses (e.g., political content, health supplements). The final set was reviewed by a visual communication researcher to confirm that each advertisement's dominant trigger type was identifiable prior to exposure. This stimulus standardization addresses a methodological limitation in Ridha's (2024) study, where participants reflected on self-selected ad experiences, potentially confounding attention trigger effects with content relevance effects.

Following each advertisement viewing, participants were asked: which element first drew their eye; what made them stop scrolling or continue watching; whether they could describe their

International Journal of Interdisciplinary Research
ISSN(Online): 3090-2959
Vol 2 no 2 (2026): July 2026
https://journal.as-salafiyah.id/index.php/ijir/index
Email: ijireditor7@gmail.com

visual experience in the opening moments; and whether they noticed any specific visual feature as particularly arresting. Interviews lasted between 35 and 50 minutes, conducted in person or via video call, and were recorded and transcribed verbatim with informed consent. Field notes captured non-verbal cues and expressive reactions during stimulus viewing.

Data Analysis

Data were analyzed using reflexive thematic analysis (Braun & Clarke, 2019), a method well-suited to explanatory qualitative research because it permits the analyst to move between inductive pattern identification and deductive theoretical interpretation. The analysis proceeded through six stages: familiarization with the data, generation of initial codes, search for themes, review of themes, definition and naming of themes, and production of the final analytical narrative. Codes were initially generated inductively from the transcript data; visual trigger categories were then mapped against visual salience theory and cognitive load literature to generate explanatory accounts.

Trustworthiness was established through member checking (sharing preliminary findings with four informants for validation), peer debriefing with a colleague in visual communication research, and reflexive journaling to monitor researcher positionality. Negative case analysis was conducted to test the robustness of identified themes against disconfirming instances in the data.

RESULT AND DISCUSSION

Research Findings

Analysis of interview data revealed four dominant visual trigger categories that consistently arrested scrolling behavior across informants. Table 1 presents a synthesis of these findings, including representative informant quotations. The following subsections discuss each trigger category in depth, integrating informant accounts with theoretical explanation

Table 1. *Visual Trigger Categories and Representative Informant Accounts*

Informan	Visual Trigger Category	Described Visual Element	Representative Quotation

International Journal of Interdisciplinary Research
ISSN(Online): 3090-2959
Vol 2 no 2 (2026): July 2026
https://journal.as-salafiyah.id/index.php/ijir/index
Email: ijireditor7@gmail.com

R1	Dynamic Motion	Fast-cut opening transition	*"The first cut was so fast I almost missed it, but that was exactly why I stopped."*
R2	Human Facial Expression	Direct-to-camera eye contact	*"She looked straight at me. I felt like she was talking to me personally."*
R3	Color Contrast	Bright color against dark background	*"The yellow text on black just jumped out of the screen."*
R4	Text Overlay	Large caption in opening frame	*"I read the text first before I even registered what the video was about."*
R5	Dynamic Motion	Camera zoom creating kinetic energy	*"It felt like something was rushing toward me instinctively I stopped."*
R6	Human Facial Expression	Exaggerated surprise expression	*"The face was so expressive I was curious what she was reacting to."*
R7	Color Contrast	Saturated product color isolated on neutral	*"Nothing else in the frame competed with it my eye went there immediately."*
R8	Text Overlay	Question-format caption	*"When it said 'Have you ever felt this way?' I had to know what came next."*
R9	Dynamic Motion	Unexpected visual discontinuity	*"It was like a visual surprise my brain registered something was different."*
R10	Human Facial Expression	Warm, authentic smile	*"The smile felt real. Not like a model smile. That's what made me watch."*
R11	Color Contrast	Complementary color scheme with rhythmic motion	*"The colors moved with the music it felt balanced and beautiful."*
R12	Text Overlay	Bold keyword in first frame	*"I saw the word 'FREE' and immediately looked closer."*

Source: Primary data, Interviews (2026)

Trigger 1: Dynamic Motion and Rapid Transitions

The most consistently cited visual trigger across informants was dynamic motion, encompassing rapid editing transitions, sudden camera movements, and kinetic compositional energy. Informants R1, R5, and R9 independently described the experience of a fast visual cut or unexpected motion as producing an involuntary cessation of scrolling, characterized by a sensation of perceptual surprise that preceded conscious evaluation of content. R1 described the fast-cut opening as creating an arrest response before deliberate attention could engage, while R5 articulated an experience of something rushing toward them that triggered an instinctive stop.

International Journal of Interdisciplinary Research
ISSN(Online): 3090-2959
Vol 2 no 2 (2026): July 2026
https://journal.as-salafiyah.id/index.php/ijir/index
Email: ijireditor7@gmail.com

This finding aligns precisely with the motion salience mechanisms described in computational visual attention models (Itti & Koch, 2001), wherein temporal luminance change across frames generates high salience values that automatically draw attentional resources. Within the short-video scroll context, this mechanism operates as an interruption signal: motion discontinuity creates a prediction error in the visual system, compelling involuntary orienting toward its source (Kim & Sullivan, 2019). Importantly, informants distinguished between continuous background motion, which they described as losing its salience through habituation within seconds, and punctuated or unexpected motion events, which retained their arrest capacity precisely because of their surprise quality.

The explanatory implication is that effective motion triggers must create what might be termed salient discontinuity: a visual event that departs from the expected temporal flow of the scrolling interface itself. Smooth, gentle motion fails to meet this criterion because it merges perceptually with the ambient scroll motion; abrupt, directional, or rhythmically irregular motion succeeds because it registers as a distinct interruption event. This distinction refines earlier descriptive accounts of motion as a general attention trigger by specifying the perceptual conditions under which motion actually functions as a scroll-stopper.

Trigger 2: Human Facial Expression and Direct Eye Contact

The second dominant trigger category was human facial expression, with particular salience assigned to direct eye contact with the camera and emotionally expressive face configurations. Informants R2, R6, and R10 each described distinct facial trigger experiences: R2 referenced the experience of perceived personal address through direct eye contact, R6 cited curiosity aroused by an exaggerated emotional expression, and R10 emphasized the authenticity of a natural smile as the distinguishing quality that motivated continued viewing.

The preferential processing of human faces in visual attention is well-established in perceptual psychology, attributable to the evolutionary primacy of social information and the existence of specialized neural processing pathways (the fusiform face area) devoted to face

International Journal of Interdisciplinary Research
ISSN(Online): 3090-2959
Vol 2 no 2 (2026): July 2026
https://journal.as-salafiyah.id/index.php/ijir/index
Email: ijireditor7@gmail.com

detection (Kanwisher et al., 1997, as cited in Kim & Sullivan, 2019). In the short-video advertising context, this biological attentional bias is amplified by the social-interactive affordances of the platform: TikTok's direct-address aesthetic, in which creators speak directly to the camera, normalizes the experience of personal engagement through face-to-face visual contact, making advertisements that adopt this convention feel interpersonally relevant rather than commercially impersonal.

The authenticity dimension of facial expression, reported by informants and most explicitly by R10, adds a culturally inflected explanatory layer. Consistent with findings on Indonesian digital media behavior, users distinguish sharply between stylized or performative expressions (associated with advertising artifice) and genuine, unguarded expressions (associated with personal relevance and credibility) (Nurhayati & Siregar, 2021; Husna & Mairita, 2024). Firamadhina and Krisnani (2021) similarly found that Indonesian Gen Z users rate the perceived realness of a video creator's facial affect as a primary trust signal, preceding evaluation of product claims or brand identity. This suggests that in the Indonesian context, the effectiveness of face-as-visual-trigger depends not only on the biological salience of the face per se but on the perceived authenticity of its affective signal, which constitutes a culturally conditioned modulation of a universal attentional mechanism.

Trigger 3: High-Contrast Color Composition

Color contrast emerged as the third primary visual trigger, described by informants R3, R7, and R11 as producing immediate visual saliency that directed gaze before conscious processing occurred. R3 described a specific luminance contrast effect, namely yellow text on a dark background, as literally 'jumping out' of the screen, while R7 noted that visual isolation of a high-saturation color against a neutral background eliminated perceptual competition and directed focal attention in a single saccade. R11 introduced a more complex version of this trigger, describing a synesthetic experience in which color harmony and rhythmic motion combined to produce an aesthetically compelling visual event.

International Journal of Interdisciplinary Research
ISSN(Online): 3090-2959
Vol 2 no 2 (2026): July 2026
https://journal.as-salafiyah.id/index.php/ijir/index
Email: ijireditor7@gmail.com

From the perspective of visual salience theory, color contrast attracts attention through two distinct but complementary mechanisms: chromatic contrast (hue difference between figure and ground) and luminance contrast (brightness difference). Both mechanisms operate pre-attentively, meaning they generate attentional priority signals before figure-ground segregation is consciously processed (Treisman & Gelade, 1980). In the scroll environment, where the interface background is predominantly white or black and the visual content of most videos is spectrally diverse, advertisements that deploy a single, strongly contrasted color element create a salient discontinuity that registers in peripheral vision even before the user's central gaze moves to the new content.

R11's account of the synesthetic interplay between color and rhythm points toward a more holistic explanatory mechanism, aligning with Villanueva et al.'s (2024) finding that visual trigger effectiveness is partly compositional, arising from the coherent integration of multiple sensory channels rather than from any single element in isolation. This suggests that color contrast functions most powerfully as a scroll-stopper not in isolation but when temporally aligned with other salient features, creating what might be termed a multi-channel salience event that maximally taxes the attentional system's capacity to ignore.

Trigger 4: Text Overlay and Caption as Cognitive Entry Points

The fourth trigger category, text overlay and caption placement, operated through a distinct mechanism from the preceding three. Where motion, faces, and color captured attention through bottom-up perceptual pathways, text overlay appeared to engage a rapid top-down processing stage in which semantic content immediately oriented cognitive evaluation. Informants R4, R8, and R12 each described text as functioning as the initial object of focused attention before the video content itself was registered: R4 read the caption before identifying the video's subject matter; R8 was drawn into continued viewing by a question-format text that created a semantic anticipation structure; R12 was arrested by a high-value keyword ('FREE') that triggered an automatic relevance evaluation.

International Journal of Interdisciplinary Research
ISSN(Online): 3090-2959
Vol 2 no 2 (2026): July 2026
https://journal.as-salafiyah.id/index.php/ijir/index
Email: ijireditor7@gmail.com

This trigger mechanism is explained by the distinction between pre-attentive and attentive processing stages in models of selective attention (Treisman & Gelade, 1980; Leppink et al., 2015). While the presence of text in a visual field is not, in itself, a salient feature under conditions of high visual complexity, the semantic content of highly legible, large-format text can bypass the perceptual competition stage and engage the language processing system directly, constituting a form of attentional capture driven by propositional rather than perceptual salience. In the context of short-video advertising, this mechanism is particularly powerful because language processing activates a different attentional modality from visual-spatial processing, effectively doubling the user's engagement commitment within the first second of exposure.

The question-format text described by R8, functioning through what might be termed semantic curiosity induction, extends the trigger function beyond capture to the activation of a need for cognitive closure (Kruglanski & Gigerenzer, 2011, as cited in Lin & Kim, 2023), a motivational state that increases the cost of scrolling away. This mechanism bridges the gap between initial attentional capture and sustained engagement, explaining why text overlay functions not only as a scroll-stopper but as a scroll-reversal trigger: some informants reported deliberately scrolling back to re-read a caption that had registered in peripheral attention during a scroll past.

Cross-Trigger Patterns and Integrative Discussion

Analysis of cross-case patterns reveals that the most memorable scroll-stopping experiences reported by informants typically involved the simultaneous activation of two or more trigger categories within the same opening frame. Informants who described the strongest arrest responses, characteristically using language of involuntary stopping or instinctive pause, reported visual events in which, for example, a direct-gaze face appeared against a high-contrast background (R2, R10), or a rapid transition revealed a bold text overlay (R4, R9). This pattern suggests a multiplicative rather than additive model of visual trigger effectiveness: co-occurring

International Journal of Interdisciplinary Research
ISSN(Online): 3090-2959
Vol 2 no 2 (2026): July 2026
https://journal.as-salafiyah.id/index.php/ijir/index
Email: ijireditor7@gmail.com

salient features across multiple attentional channels overwhelm the visual system's capacity to sustain the scroll pattern, compelling a stop response.

Figure 1 presents a conceptual framework synthesizing the two-stage attention model that emerges from these findings. In Stage 1 (Visual Trigger Phase), each of the four trigger categories activates bottom-up perceptual pathways, specifically motion salience, face detection, chromatic contrast, and propositional salience, either independently or in combination, generating an involuntary arrest response within the first 1.5 seconds. In Stage 2 (Cognitive-Emotional Engagement Phase), the arrested viewer enters the evaluative-to-flow continuum described by Ridha (2024), in which narrative coherence, emotional resonance, and cognitive manageability determine whether initial capture translates into sustained viewing. The visual trigger is necessary but not sufficient: it creates the attentional opening through which content experience must then make its case.

Figure 1. *Two-Stage Model of Visual Attention in Short-Video Advertising*

STAGE 1: VISUAL TRIGGER PHASE (0–1.5 seconds)

Dynamic Motion | *Facial Expression* | *Color Contrast* | *Text Overlay*

Bottom-up perceptual salience → Involuntary scroll arrest

▼

STAGE 2: COGNITIVE-EMOTIONAL ENGAGEMENT PHASE (1.5+ seconds)

Evaluation → *Emotional Resonance* → *Cognitive Flow*

Top-down processing → Sustained viewing engagement (Ridha, 2024)

This finding has direct implications for the relationship between initial visual triggers and the deeper cognitive engagement described in Ridha's (2024) attention phenomenology. The present study identifies the visual mechanisms that initiate the attention entry phase; Ridha's framework maps the experiential landscape that unfolds if that entry is successfully achieved.

From a practical standpoint, these findings argue for a front-loaded visual design strategy in TikTok advertising: concentrating the highest-salience visual events in the opening 1.5 seconds, deploying at least two trigger categories simultaneously, and ensuring that the trigger is continuous

International Journal of Interdisciplinary Research
ISSN(Online): 3090-2959
Vol 2 no 2 (2026): July 2026
https://journal.as-salafiyah.id/index.php/ijir/index
Email: ijireditor7@gmail.com

with, rather than discontinuous from, the core message and emotional register of the advertisement. This aligns with practitioner evidence from the Indonesian context showing that TikTok content achieving the highest engagement consistently front-loads its most distinctive visual element, whether a color pop, expressive face, or bold text, within the opening frame (Hayati & Sudradjat, 2022; Novita et al., 2023). Sopari and Alawiyah (2024) further confirm that visual content quality and early-frame aesthetic decisions are the dominant predictors of engagement rate in Indonesian social media advertising, outperforming caption text and posting frequency as determinants of scroll-stopping effectiveness.

Comparative consideration of findings from non-Indonesian contexts suggests both convergence and divergence with the present results. The primacy of motion and facial expression as scroll-stopping triggers is broadly consistent with findings from Western and East Asian studies (Kim & Sullivan, 2019; Lin & Kim, 2023; Wang et al., 2022), suggesting that these categories reflect universal perceptual mechanisms rooted in evolutionary attentional biases. However, the authenticity dimension of facial expression, which informants in the present study identified as a decisive moderator of face-trigger effectiveness, appears more pronounced in the Indonesian context than in studies conducted in the United States and South Korea, where stylized and aspirational aesthetics retain greater attentional efficacy (Villanueva et al., 2024). This divergence is consistent with broader findings on cultural variation in digital media trust and credibility assessment (Nurhayati & Siregar, 2021; Astuti & Prasetyo, 2024), and suggests that while the taxonomy of visual trigger types may generalize across populations, the relative effectiveness of specific trigger configurations is modulated by culturally situated perceptual values. Future cross-cultural studies should examine whether the authenticity moderation effect replicates in other collectivist-cultural contexts, such as those found in parts of Africa and South Asia, or whether it is specific to Indonesian media consumption norms.

International Journal of Interdisciplinary Research
ISSN(Online): 3090-2959
Vol 2 no 2 (2026): July 2026
https://journal.as-salafiyah.id/index.php/ijir/index
Email: ijireditor7@gmail.com

CONCLUSION

This study set out to identify and explain the visual triggers that most effectively arrest scrolling behavior during short-video advertising exposure among TikTok users in Makassar, Indonesia. Through qualitative explanatory inquiry grounded in stimulus-based semi-structured interviews, four dominant visual trigger categories were identified: dynamic motion and rapid transitions, human facial expression and direct eye contact, high-contrast color composition, and text overlay as a cognitive entry point. For each trigger, explanatory mechanisms were reconstructed from both informant accounts and theoretical frameworks, moving beyond description to account for why these stimuli work at the level of perceptual and cognitive processing.

The study makes three principal contributions. Theoretically, it bridges visual salience theory and digital attention phenomenology, offering a two-stage model of advertising attention in which perceptual capture precedes cognitive-emotional engagement. This model extends Ridha's (2024) experiential account of TikTok attention by specifying the visual mechanisms that initiate the attentional entry phase, a contribution that addresses a gap not filled by prior quantitative or purely descriptive qualitative work. Methodologically, the stimulus-standardized qualitative explanatory design, with its explicit criteria for advertisement selection and its dual focus on description and mechanism, demonstrates a productive approach to mechanistic inquiry in advertising research. Practically, the findings provide creative practitioners with a theoretically grounded taxonomy of visual trigger types and their operating mechanisms, enabling evidence-based creative decision-making for the Indonesian short-video advertising context.

While the findings are grounded in the Indonesian context, the four-category trigger taxonomy and the two-stage attention model offer potential theoretical generalizability to other large, mobile-first short-video markets. The authenticity moderation effect documented here warrants particular attention in future cross-cultural research, given its implications for how universal perceptual mechanisms interact with locally situated media consumption values. Future

International Journal of Interdisciplinary Research
ISSN(Online): 3090-2959
Vol 2 no 2 (2026): July 2026
https://journal.as-salafiyah.id/index.php/ijir/index
Email: ijireditor7@gmail.com

research should examine whether visual trigger effectiveness varies across demographic subgroups, advertising categories, and platform interfaces. Longitudinal studies tracking changes in trigger salience over time, as users adapt to prevalent visual conventions, would also be valuable, given the rapid evolution of short-video aesthetic norms. Quantitative methods, including eye-tracking and A/B testing, could be productively combined with qualitative explanatory approaches to triangulate the perceptual mechanisms identified here and establish external validity across broader populations.

REFERENCE

Astuti, D. P., & Prasetyo, I. (2024). Kebahagiaan konsumen dalam narasi pemasaran lokal: studi fenomenologis pada UMKM kuliner. Jurnal Kompeten, 9(1), 45–59.

Astuti, R., & Prasetyo, D. (2024). Social media culture and emotional authenticity among Indonesian youth. Asian Communication Journal, 9(1), 33–47.

Braun, V., & Clarke, V. (2019). Reflecting on reflexive thematic analysis. Qualitative Research in Sport, Exercise and Health, 11(4), 589–597. https://doi.org/10.1080/2159676X.2019.1628806

Creswell, J. W., & Poth, C. N. (2018). Qualitative inquiry and research design: Choosing among five approaches. Sage.

Firamadhina, F. I. R., & Krisnani, H. (2021). Perilaku generasi Z terhadap penggunaan media sosial TikTok: TikTok sebagai media edukasi dan aktivisme. Share: Social Work Journal, 10(2), 199–208. https://doi.org/10.24198/share.v10i2.31443

Hayati, C. S., & Sudradjat, R. H. (2022). Pemanfaatan TikTok sebagai platform digital marketing dalam upaya peningkatan brand awareness Butik Aishable. SEIKO: Journal of Management & Business, 4(3), 415–426.

Husna, A. H., & Mairita, D. (2024). Gen Z dan perilaku konsumsi konten influencer pada TikTok. Jurnal Riset Komunikasi, 7(1), 86–100. https://doi.org/10.38194/jurkom.v7i1.1002

International Journal of Interdisciplinary Research
ISSN(Online): 3090-2959
Vol 2 no 2 (2026): July 2026
https://journal.as-salafiyah.id/index.php/ijir/index
Email: ijireditor7@gmail.com

Itti, L., & Koch, C. (2001). Computational modelling of visual attention. Nature Reviews Neuroscience, 2(3), 194–203. https://doi.org/10.1038/35058500

Kim, Y., & Sullivan, P. (2019). Emotional and cognitive engagement in short-form video advertising. Journal of Interactive Advertising, 19(2), 95–108.

Langan, R., Cowley, E., & Nguyen, T. (2022). The power of the first few seconds: Attention triggers in short-form video marketing. Marketing Science Review, 41(3), 221–239.

Leppink, J., Paas, F., van Gog, T., van der Vleuten, C. P. M., & van Merriënboer, J. J. G. (2015). Effects of pairs of problems and examples on task performance and different types of cognitive load. Learning and Instruction, 30, 32–42.

Lin, Y., & Kim, S. (2023). Beyond clicks and views: Understanding attention quality in short-form video platforms. Journal of Marketing Communication, 29(2), 179–198.

Maxwell, J. A. (2012). A realist approach for qualitative research. Sage.

Montag, C., Yang, H., & Elhai, J. D. (2023). How the TikTok algorithm shapes user attention and dopamine-driven engagement. Computers in Human Behavior Reports, 10, 100279.

Nurhayati, S., & Siregar, L. (2021). Cultural patterns in social media engagement: Evidence from Indonesia. Indonesian Journal of Communication Studies, 8(1), 55–70.

Ridha, A. (2024). Attention span dan cognitive flow dalam konsumsi iklan TikTok: Studi kualitatif tentang pengalaman perhatian audiens. Jurnal TAJAM, 8(2), 191–202.

Smith, J. A., Flowers, P., & Larkin, M. (2009). Interpretative phenomenological analysis: Theory, method and research. Sage.

Statista. (2025). TikTok users in Indonesia from 2020 to 2025. https://www.statista.com

Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. Cognitive Science, 12(2), 257–285.

Treisman, A. M., & Gelade, G. (1980). A feature-integration theory of attention. Cognitive Psychology, 12(1), 97–136.

International Journal of Interdisciplinary Research
ISSN(Online): 3090-2959
Vol 2 no 2 (2026): July 2026
https://journal.as-salafiyah.id/index.php/ijir/index
Email: ijireditor7@gmail.com

Villanueva, J., Kim, D., & Patel, S. (2024). The visual determinants of engagement in social media advertising. International Journal of Advertising, 43(2), 315–334. https://doi.org/10.1080/02650487.2023.2251369

Novita, D., Herwanto, A., Mayndarto, E. C., Maulana, M. A., & Hanifah. (2023). Penggunaan media sosial TikTok sebagai media promosi pemasaran dalam bisnis online. Jurnal Minfo Polgan, 12(2), 2543–2550. https://doi.org/10.33395/jmp.v12i2.13312

Sopari, R. M., & Alawiyah, W. D. (2024). Pengaruh visual content marketing dan copywriting terhadap tingkat engagement pengguna Instagram di Kota Bandung. Jurnal Nasional Manajemen Pemasaran dan SDM, 5(3), 112–124.

Wang, R., Yu, J., & Li, Q. (2022). Visual saliency as a predictor of advertising effectiveness in short-form video. Journal of Advertising Research, 62(3), 256–271.