Crafting Engaging Ads: A Visual Feature Analysis of Video Advertisements Leveraging Computer Vision

Roelof Blommaert

Student number: 505391

Supervisor: Gui Liberali

Co-reader: Ana Martinovici

9-06-2024

Marketing Management MSc Thesis

Data Science and Machine Learning

Erasmus University/Rotterdam School of Management

Word count: 23,154

Abstract

This paper aims to find the effect of visual video features on online engagement, operationalised as views, likes and comments, of Super Bowl advertisements on YouTube. Computer vision and object detection are leveraged to analyse visual features and determine Visual Complexity and Variety. Visual Complexity is a measure aiming to capture the static visual features of advertisements, while Visual Variety measures the overall dynamism of advertising videos. Overall, 132 Super Bowl advertisements were analysed using the YouTube API. The study results show that Visual Complexity and Visual Variety are, to some extent, predictors of online engagement, especially for more active forms of engagement, such as liking and commenting on videos instead of viewing a video. Colour Complexity has a positive relationship with the number of views and comments, and Visual Variety also has an overall positive relationship with likes and comments. Finally, Edge Density has a positive relationship with comments, while Luminance Entropy and Irregularity of Object arrangement significantly negatively influence comments and views respectively. The relationship between Colour Complexity and comments shows significant evidence for a U-shaped relationship, meaning extreme levels of Colour Complexity have the most positive influence on comments.

On the other hand, the relationship between Irregularity of Object Arrangement and views manifests itself as a significant inverted U-relationship, meaning videos with medium levels of Irregularity garner the most views. The findings suggest that high levels of Variety, or dynamism, and extreme levels of Colour Complexity positively influence online engagement. In contrast, high lighting complexity negatively influences the online engagement metrics views and comments. This study is one of a few studies leveraging computer vision in an advertising context on video materials, especially using multiple measures of online engagement. More research needs to be performed to determine the individual effects of visual features on online engagement, primarily to determine the complex non-linear influence of these visual features on attention and engagement in an advertising context.

Keywords: Engagement, Advertising, Computer Vision, Visual Complexity, Visual Variation, Social Media, YouTube, Object Detection, Super Bowl

Preface

Contents

1. Introduction

Throughout the years, online video-sharing platforms such as YouTube, TikTok and Vine have significantly impacted online marketing practices. Consumers can be connected with engaging video content from every corner of society with a couple of clicks, taps, scrolls or swipes; however, the rise of online video-sharing platforms does not only affect user-to-user interactions. The rise of platforms such as YouTube and TikTok has created television advertising alternatives that are constantly accessible to consumers and recommended to the right audience through complex user-specific algorithms. Due to brands massively jumping on the opportunity that online video-sharing platforms offer, the sheer volume of content flowing through video platforms makes for an oversaturated market. Due to this oversaturation, engaging content has become a cornerstone of strategic planning (Ferguson, 2008). Deciphering the secret sauce behind engaging content is one of the most critical tasks of content marketers, digital markers, and marketing analysts, as they need to gain consumers' limited attention. Capturing the attention of a vast audience and translating this into substantial market influence and brand awareness is the final goal for many.

In a more traditional marketing context, brands pay for television advertisements, web advertising or other exposure, with many academics researching the intrusiveness of these types of advertising to the consumer. While marketing is evolving past this more traditional manner of marketing to the use of online algorithms for organic growth, there is a shift in marketing content. Advertisements used to focus more on informational appeals and were much more product-focused, whereas emotional and aesthetic appeals are becoming more and more commonplace.

For years, consumer behaviour research has been interested in the aesthetic appeals of advertising content and its implications on consumer behaviour through subjective measures. The recent rise of computer vision and accessible machine learning applications has created novel ways of objectively measuring and quantifying aesthetic appeals. These automated objective measures offer a framework for a deeper understanding of the drivers of consumer engagement and aid the field of digital marketing in the strategic planning of video marketing content (Nanne et al., 2020).

The Super Bowl is a prime example of the change in advertisement focus throughout the years since it has become the most prominent advertising spectacle worldwide. Many viewers tune in to watch the Super Bowl for the advertisements rather than for the actual sporting event. The advertisements shown during the event even receive prizes for how well-liked the video advertisements are, which showcases the importance of engagement in marketing materials (Hartmann & Klapper, 2018) rather than simply showcasing product features or increasing brand recognition through repeated contact with brands. While the Super Bowl advertisements are initially used for television advertising, the advertisements are also created to generate online engagement. Many of the advertisements are uploaded onto online video-sharing platforms by the companies to utilise the buzz generated by their

Super Bowl spot to capture a broad audience. Given the importance for brands to engage consumers on online platforms to capture their attention and the importance of visual appeals in brand-generated content, the central question in this research paper is:

"How do visual video features of Super Bowl advertisements influence online engagement on YouTube?"

This research question makes for compelling research since it uses the advancements of computer vision technology to quantify the visual features and what effect these visual features have on a consumer's engagement in a real-world context. Consumer engagement is specifically interesting for brand-generated content since capturing a consumer's attention makes it possible to send clear messages, leading to a desired brand perception and influencing purchase decisions. Responses to video advertising are a well-documented field of study, incorporating well-established psychological models and new-age marketing innovations (Tellis et al., 2019; Nikolinakou & King, 2018). Much of this research focuses on studies in controlled environments; however, with the emerging advancements in computer vision in both the industrial and academic fields (Nanne et al., 2020), the effects of visual content features can now be extracted in real-world scenarios.

Since audiovisual and storytelling components and quality are becoming more critical in the field of marketing advertisements and brand-generated content (Li et al., 2019), investigating this construct further and using visual features to predict engagement, makes for a complete analysis, which adds to the existing body of research on engagement with real-world marketing application. Super Bowl advertisements make it possible to analyse high-profile advertisements with differing levels of online success. The quality of the advertisements is high, while video features significantly differ across ads. Super Bowl advertisements are also created for online advertising, while all the advertisements receive equal initial exposure through participation in the event before being posted online.

The data used to answer the research question is gathered from YouTube, the largest video platform. YouTube has active users in 100 countries, with 500 hours of content uploaded every minute (YouTube for press, n.d.). Since YouTube is such a big platform with such rich personal content suggestion algorithms, brands choose this platform to publish their advertising content to create engagement. YouTube is such a well-known platform for advertisers and brand-generated content that YouTube even publishes a yearly global ad leaderboard for the most-watched and best advertisements (YouTube Culture & Trends - data and cultural analysis for you, n.d.). YouTube uses three main symbolic measures for engagement: views, likes and comments. Given the importance of online engagement for brands, it is crucial to understand the context of engagement, visual research in marketing and how visual features and engagement interact based on existing literature. This research analyses the visual features of advertising videos using computer vision and how they influence online engagement. The study is structured as follows: Section 2 reviews the existing literature on engagement and visual research, including the specific theoretical background; Section 3 introduces the methodology, containing information on the sampling, design and measurements; Section 4 describes the data analysis procedure; Section 5 reviews the results and discusses the limitations; Section 6 discusses results and puts the results into perspective.

2. Literature Review

This section discusses the literature on engagement and visual research before theories are developed based on existing literature to guide the exploratory analysis. Engagement is a broad term with various implications depending on the context. In section 2.1, engagement is defined before existing literature on engagement and its prerequisites are reviewed in online and marketing contexts. Section 2.2 presents the existing literature on visual research for dynamic and static visual features before examining popular visual theories and presenting existing literature on visual features and their methodologies in a marketing context. Section 2.3 develops theories about the influence of visual complexity (Pieters et al., 2010; Overgoor et al., 2022) and visual variety (Zhou et al., 2021; Li et al., 2019) on online engagement based on visual stimulation theories (Hebb, 1955; Berlyne, 1970).

2.1 Engagement and its Prerequisites

2.1.1 Engagement across Contexts

Despite the popularity of the term "engagement", the term is not frequently clearly defined in academics (Brodie et al., 2013). Due to this broad and unclear definition, it varies significantly per area of research, such as education, psychology, and marketing. Therefore, understanding the difference in definition per context is essential. In education the construct is divided into three aspects: behavioural, emotional and cognitive engagement. The construct is extensively researched due to the assumed malleable nature of engagement in education (Fredricks et al., 2004). Behavioural engagement refers to actions like participation, presence, timely submission of assignments, and following rules and regulations (Fredricks et al., 2004). Emotional and cognitive engagement are more about relationships and willingness to exert oneself, respectively (Smith et al., 2005). In Psychology, engagement refers to an individual's active involvement in activities and is often associated with concepts such as flow, motivation, well-being and commitment (Schaufeli et al., 2002). Engagement's behavioural, emotional and cognitive aspects are also frequently used to refer to certain behaviours or relational factors influencing active involvement or well-being (Kahn, 1990). Engagement in the context of marketing frequently refers to brand engagement or consumer engagement. Both types of engagement revolve around creating valuable direct interactions between a brand and the consumer (Brodie et al., 2011). Brand engagement refers to the aggregate attachment of consumers toward a

particular brand, while consumer engagement, on the other hand, refers to the interactions between the consumer and content.
The studies regarding engagement in marketing have mostly been theoretical and exploratory, basing their findings on the large body of psychological engagement theories in various contexts. In contrast, empirical research has remained relatively limited (Hollebeek et al., 2014). Due to the rise of social media popularity and the concurrent shift of marketing budget allocation to online marketing practices, marketing-related empirical engagement research has been increasing. Like many studies in a social media context, this study focuses on the behavioural aspect of engagement in online marketing.

2.1.2 Social Media Engagement

Marketing practitioners and academics define online engagement as actions on a platform that can include liking, disliking, commenting, viewing, sharing, and uploading (Khan, 2017). Khan (2017) distinguishes between two levels of social media engagement: consumption, such as viewing and reading, and participation, such as liking, disliking, sharing, commenting, or uploading.
When measuring engagement on social media, many use the earlier-mentioned symbolic activities like rating, discussing, sharing and viewing content (Xu et al., 2019). While these symbolic measures capture the behavioural side of engagement, which is more directly linked to marketing outcomes, they do not reflect the depth of the emotional and cognitive aspects of engagement as outlined by Hollebeek (2011). The measurement of constructs like cognitive and emotional engagement is frequently handled using surveys, as these constructs involve the amount of attention, interest and mental resources assigned to the activity in which they engage (Marks, 2000). Another manner of researching cognitive and emotional engagement in an online marketing and advertising context is biometric measurement, such as eye-tracking (Wedel & Pieters, 2006) and EEG (electroencephalogram) (Pozharliev et al., 2015). Considering the direct link between behavioural social media engagement measures and marketing outcomes, along with the frequent use of these engagement metrics as Key Performance Indicators (KPIs), these measures seem more appealing due to the availability of this social media data.

Operationalisation of symbolic measures of engagement on YouTube

Considering the variance in symbolic measures used to operationalise engagement per platform, it is important to highlight which engagement metrics are frequently used for research done using YouTube engagement, as this is the platform relevant for this paper.
In the study by Vraga et al. (2014), views and ratings as a sum of likes and dislikes and the number of comments were used to measure engagement on YouTube for 801 and 365 videos of two social protest movements. Ratings were frequently utilised as a form of engagement before 2021, as this

essentially presents the overall sentiment score of the video. Khan (2017) used likes, dislikes, views and comments to showcase consumption and interaction habits with content despite not using web data directly from the website. A clear distinction between various engagement metrics must be made to analyse the direct influence of certain predictors on engagement or consumption. Munaro et al. (2021), just like Vraga et al. (2014), used the YouTube API and extracted views, likes, dislikes and comments for 11,177 videos to measure engagement. These three highlighted papers were all published or written before 2021, as they still include dislikes as a measure of engagement. YouTube has decided to discontinue dislikes on the platform as of 2021, which makes likes the only engagement metric left to express sentiment about a video in one click (Youtube Team, 2021). The papers mainly outline the operationalisation of the full range of passive and active forms of engagement by Kahn (1990). In the context of YouTube consumption, this comes down to the measurement of views, likes and comments, where the number of views is the most passive form of engagement, and the number of comments is the most active form of online engagement.

Other factors influencing social media engagement

The aspect of time must be remembered when speaking of engagement and the effect of other content features. Social media data is on the web for some time, during which the content starts gaining traction. Some measure engagement at a certain point, while others focus on the fluctuation of engagement over time through time series or longitudinal analyses (Brodie et al., 2013). Many studies in the field of marketing have found that the time a video is online significantly impacts online engagement, according to research by McParlane et al. (2014). Other research looking into the effect of visual features on online engagement decides to take the time since posting or the time of posting (e.g. morning, afternoon or evening) into account (Xiao et al., 2023; Overgoor et al., 2022). Xiao et al. (2023) found that the time since a video is posted, or time, significantly affected engagement metrics on TikTok.

Other content factors, like video duration, are frequently controlled for in various studies examining the effects of video content characteristics. Li et al. (2019) found a significant positive linear effect of video duration on Kickstarter project success and a significant negative impact for the quadratic equivalent of video duration, signifying an inverted U-effect. For social media engagement, Xiao et al. (2023) found the same effect for TikTok videos, where video duration significantly negatively influenced how many likes a video would amass on the platform. The channel, creator or source from which content is posted also significantly influences online engagement. Xiao et al. (2023) found that the number of followers significantly positively affected likes and views of TikTok videos. Gross and Von Wangenheim (2022) also found that the effect of page following on Instagram significantly positively influenced both likes and comments for informational and affective posts.

2.1.3 Online Engagement in Brand- and User-generated Content

Since the rise of social media marketing, the largest source of brand-related content has shifted from brand-generated to user-generated content. Brand-generated content highlights the more traditional marketing content brands produce to reach consumers. User-generated content has become an area of focus in marketing research since the confluence of social media with a consumer's day-to-day life. It refers to peer-to-peer brand-related content (Roma & Aloini, 2019).

User-generated brand-related content is more related to talking about a brand rather than a brand promoting itself. The source of brand-related content significantly affects how the content engages consumers. Consumers find user-generated content to be more authentic and trustworthy, as highlighted by the study by Jahn and Kunz (2012). Jahn and Kunz (2012) studied three brand fan pages in the automobile industry. They found that the perceived authenticity in user-generated content fosters trust, leading to more consumer engagement by analysing 661 posts during one year. Hajli (2014) also found a significant positive influence of user-generated brand-related social media content on trust. In turn, trust favourably influenced intentions to buy and the perceived usefulness of the provided information. Hautz et al. (2014) found that trustworthiness only seems more prevalent among user-generated content when technical quality is low through their analysis, where they manipulated content's perceived source and video quality. Thus, trustworthiness does not differ significantly for user- and brand-generated content when the technical quality of the videos is high (Hautz et al., 2014). Hautz et al. (2014) manipulated video quality by creating less smooth transitions, increasing graininess and aspect ratio and reducing audio quality. Although user-generated content is seen as more relatable, leading to more consumer trust, most advertising spending worldwide is allocated to brand-generated content initiatives.

Many brand-generated content features can drive online social media engagement. Brand personality features such as humour and emotion are associated with higher levels of social media engagement in terms of likes, views, comments, shares or click-through rates (Lee et al., 2018). Through the analysis of 106,316 Facebook messages for 782 companies, Lee et al. (2018) found that emotion, humour, and informativeness of content significantly predict social media engagement for brand-generated content. In line with this finding, Hudson et al. (2015) showcased that emotional responses triggered by marketing efforts play a dominant role in explaining engagement for brand-generated advertisements. These content features driving social media engagement do not just adhere to textual content since the exact effects of emotional visual content features on engagement are found by Rietveld et al. (2020). The research on the determinants of online engagement for brand-generated content is extensive for many forms of content. Still, despite visual content, especially video, being particularly engaging (O’Brien, 2017), the research on video engagement for brand-generated content is relatively small due to the recent extensive focus on peer-to-peer marketing. Therefore, an overview of the literature on the effects of visual features will be presented.

2.2 Visual Analytics and its Outcomes

In this section, the research on static and dynamic video content features will be presented, some visual theories and their implications will be outlined, and the research leveraging video analysis in a marketing context will be highlighted. Visual features of content can strongly impact consumer attitudes, especially emotional aspects (Bhandari et al., 2019). With the rise of data mining, computer vision and machine learning, much more seems possible regarding visual analysis of online video content and advertisements in general. Computer vision and its applications for low-level video analytics are appealing, particularly for low-level features that are easier to quantify and involve less subjectivity, making for more reproducible studies and precise interpretations. While high-level features might be easier to pinpoint, individual low-level features significantly affect visual content's aesthetic appeals despite being more indistinguishable in isolation (Shin et al., 2020).

2.2.1 Static versus Dynamic Visual Features

Static Visual Features

Recently, scholars have started giving more attention to visual features of marketing content, which until recently was mainly focused on static visual features of images, billboards, logos or videos and the effects this has on consumer attitudes and brand-related outcomes. Matz et al. (2019) used object detection, composition, texture and colour analysis for static visual analysis. In the analysis of 1040 professional Shutterstock images, Matz et al. (2019) found that average saturation, brightness and level of detail were positively correlated with general appeals while negatively correlated with the colour white, use of light directionality and number of people in the image. Pieters et al. (2010) analysed 249 randomised full-page magazine advertisements. They found that high and low levels of visual complexity negatively influenced the attention to the advertisement, the brand in the ad and the attitude toward it. Shukla et al. (2018) found that general object orientation or raw scene structure is more directly related to affective encoding rather than individual objects and conspicuous background changes. In line with this finding, Oğuz et al. (2023) found that key light and colour variance was significantly positively linearly related to consumer engagement and both colour variance and motion features were significantly non-linearly associated with engagement. These findings by Oğuz et al. (2023) and Shukla et al. (2018) support using low-level visual frameworks to determine affective or behavioural responses to visual content.

Dynamic Visual Features

There are many theories regarding motion-based aspects of online videos and advertisements, which are all related to stimulus and, therefore, the engagement or attention of consumers regarding the video content. Slow motion, for example, is often used in advertisements to create the perception of product quality or luxury. Slow motion is essentially a technique to decrease the amount of motion within a specific timeframe, lowering the amount of perceived motion or processing complexity. Despite the frequent use of slow motion, Yin et al. (2021) found that consumers felt less engaged with slow motion videos, operationalised as views, likes and dislikes, in 57% of 100 top search results for food advertisements. In the second study done by Yin et al. (2021), they found that the attitude toward advertisements with slow-motion aspects was less favourable than the ads at normal speed based on their analysis of actual advertisements, which 199 university students assessed. In stark contrast to this finding, Stuppy et al. (2023) found that slow motion positively affected the liking and exposure of short-form videos. Other studies conducted in the same paper by Stuppy et al. (2023) showed that valence and movement complexity moderated this relationship between slow motion, where positive content and videos with low movement complexity elicited higher engagement in slow motion than at regular speeds. The study by Farace et al. (2019) showcases the importance of regularity in motion. They found that advertisements with regular motion patterns received significantly higher average product evaluation for 156 respondents than those with irregular motion patterns linked with engagement and stimulus by the researchers.

Some academics, however, disagree with this theory and suggest that this relationship of engagement and attention is highly situationally dependent. For example, Lee and Ahn (2012) found through their analysis of 118 subjects using eye-tracking that in the context of banner ads, static banner ads garnered more attention than animated banner ads with a motion component, and animations even seemed to harm brand attitude. In contrast, the research conducted by Simons et al. (2003) found the exact opposite. Simons et al. (2003) presented 25 subjects with 60 still or moving images from television programmes and movies, and EEG was used to find the effect of image motion and emotion of attention. Simons et al. (2003) found that motion achieves greater sustained attention, and the relationships of emotional valence on attention were more significant for moving imagery.

Generally, although situationally dependent, both static and dynamic visual features significantly affect the attention and engagement of viewers of banners, television programmes, online content, movies and advertisement videos. The static and dynamic visual content must not be overwhelming for brand-generated video content to engage viewers optimally. As discussed by Germeys and D'Ydewalle (2005), visual complexity must be captured by capturing the individual features of separate scenes (momentary), as well as the pacing of scenes (dynamism). Practical film- and marketing-related findings outline the importance of static and dynamic visual information for video content.

2.2.2 Visual Theories

Considering the impact of visual information on Psychology research, Business, Art, Culture and many more contexts, there is an abundance of theories regarding the grouping of visual features and their effects. Apart from separating visual variables into static and dynamic features, as discussed above, many use other frameworks to group certain aspects of visual information or link visual information to specific outcomes. These theories include Visual Complexity, Gestalt Principles, Aestheticism and Visual Variety.

Visual Complexity is a theory related to visual information perception, attention and processing. It measures how much one needs to exert oneself, or the cognitive load imposed, to interpret and process the presented visual information (Donderi, 2006). Visual Complexity is frequently used for the design of websites (Tuch et al., 2009), other user interfaces (Miniukovich et al., 2018), magazines (Pieters et al., 2010), images (Overgoor et al., 2022) and videos (Song et al., 2021). Gestalt principles focus on how people perceive visual elements as whole structures rather than just a collection of parts, using concepts like proximity, similarity, continuity, closure, and figure-ground (Wagemans et al., 2012). The theory is one of the backbones for many other visual scales and attention theories, like Visual Complexity, Visual Variety, Aestheticism, and Visual Stimulation. It is used in contexts ranging from architecture to education. Visual Variety is a theory very similar to Visual Complexity and is heavily underlined by the Gestalt Principles as it focuses more on motion and fluency (Hung & Tzeng, 1981). Visual Variety, or Variation, is primarily used in video context but gets various names like shot length or scene variation (Liu et al., 2018; Zhu et al., 2024) or even Motion Complexity (Stuppy et al., 2023). The theory is mainly utilised in an educational (Zhou et al., 2021) or marketing context (Liu et al., 2018; Li et al., 2019). Aesthetics, although related to all of the measures above, is about measuring expected perceived satisfaction or liking of the presentation of visual information (Lavie & Tractinsky, 2004; Bhandari et al., 2019). Visual aesthetics heavily depend on Gestalt Principles, Visual Processing and thus, Visual Complexity theory.

2.2.3 The Literature on Visual Feature Analysis and Engagement

Considering the field of computer vision in marketing is relatively new, not many papers have been published directly investigating the effect of advertisements on engagement metrics or in a social media or advertising context. Table 1 aims to illustrate academic papers that leverage visual feature analysis for brand-generated advertisements. The table distinguishes between the type of data being either video or imagery, and makes a distinction between using static, dynamic or both types of visual features. Dynamic features are motion-related, such as movements' speed, amount, or frequency throughout frames. In contrast, static features are similar to computer vision applications for images since an average of variables like colourfulness can be calculated over the number of frames for an entire video. Finally, the table also highlights the type of dependent variable used in the study, which is somewhat related to (negative) engagement in most of the outlined studies. The table aims to showcase the spread in engagement operationalisation related to visual features and reveal the study's novelty. Finally, the table should motivate the exploratory approach of the study.

Table 1: Overview of the literature using visual analysis to relate visual features to dependent variables related to engagement, showcasing the novelty of the proposed study

Reference	*Context*	*Type of content*	*Type of visual variable (static or dynamic)*	*Dependent variable*
Teixeira et al. (2010)	Visual Television advertisements' effect on attention and avoidance	Video	Static and Dynamic	Retention in terms of zapping behaviour and gaze patterns (avoidance)
Liu et al. (2018)	Predicting intentions to watch and box office success of movie advertisements	Video	Static	Happiness after watching advertisements
Campbell et al. (2017)	Effect of advertisement features on skipping behaviour for pre-roll ads	Video	Static and Dynamic	Skipping behaviour for pre-roll ads
Pieters et al. (2010)	Impact of visual advertisement features on attention and attitude	Image	Static	Reported liking of the ad and happiness after watching an ad
Overgoor et al. (2022)	Effect of visual complexity on liking behaviour on Tumblr posts	Image	Static	Number of likes on social media posts
Oğuz et al. (2023)	Influence of low-level visual features of ad videos on consumer engagement	Video	Static and Dynamic	User Engagement Scale-Short Form (UES-SF) survey
Zhu et al. (2024)	Influence of visual features on engagement for pro-environmental videos	Video	Dynamic	Live comments on social media post
Song et al. (2021)	Effect of advertisements' visual video features on consumer's click behaviour	Video	Static and Dynamic	Clicking behaviour (click or no click)
This study	Low-level visual features of advertisements' impact on online engagement	Video	Static and Dynamic	Views, likes and comments on advertisements

Table 1 highlights a gap in the literature, as there is a lack of studies investigating the effect of visual features extracted with computer vision on engagement, specifically online engagement metrics. Both subjective and objective engagement metrics have been used to find the relationship between visual information and engagement, but not many have used social media metrics to operationalise engagement or specifically analyse advertisement. The study by Song et al. (2021) is quite similar to the current study as it investigates video advertisements in an online context; however, the study does not test for quadratic relationships like the ones found by Overgoor et al. (2022), Pieters et al. (2010), Teixeira et al. (2010), Zhu et al. (2024) and Oğuz et al. (2023) for both image and video data. Overgoor et al. (2022) and Zhu et al. (2024) both investigate the effect of visual features on online engagement metrics but are both not related to advertisements and, in the case of Overgoor et al. (2022), do not use video data. Despite the differences between the described studies and the proposed study, their findings will be combined with psychological theories to develop a theoretical background.

2.3 Theoretical Background

This section will explore the underlying theories and concepts related to Visual Video Features and (Online) Engagement. Considering that Visual Complexity and Variety rest on psychological theories and have complex relationships with attention and engagement, this section will not aim to develop specific hypotheses but rather outline overall theories, relate these to existing practical outcomes found in research, and implement this into an exploratory framework used as a backbone for the analysis. This model aims to visualise the concepts in the paper.

2.3.1 Visual Variation

Visual variety is an interesting concept for capturing the visual motion features of videos since this variable is associated with visual stimulation and relates to consumer psychology (Berlyne, 1970). Visual variety can best be described as the amount of visual change throughout a piece of motion-based content, where a video with separate scenes and orientation has high levels of variety and content with little scene changes or changes in orientation has low visual variation (Li et al., 2019). The description above makes visual variation a fitting measure of general visual motion since it captures the changes on a frame-to-frame, pixel-to-pixel basis and changes in a video occur via motion.

Visual stimulation has been found to have an inverted U-relationship to consumer engagement, where visuals that are not complicated nor simple to process lead to the most satisfying viewing scenario (Hebb, 1955). Visual variation has been researched in marketing and educational contexts as well, where visual variation of video content has been found to have a significant relationship with fundraiser success by Li et al. (2019) and engagement in online educational content by Zhou et al.

(2021), where lower levels of visual variety have significant adverse effects on the respective project success and engagement metrics. Zhu et al. (2024) found a significant inverted U-relationship between visual variety measures, operationalised as motion variety of objects and video shot length variety, and engagement. Engagement was measured as live comments in 44 social media videos related to pro-environmental tourism. The study done by Stuppy et al. (2023) also highlights this effect; however, the authors call for a measure of motion complexity, which is similar to visual variety. Stuppy et al. (2023) found that for high levels of motion complexity, slow motion videos positively affected the liking and views of short-form videos. This interaction effect states that when visual complexity is high, slowing the video down, thus lowering the perceived complexity or variety of movement in a certain period, is associated with a better attitude toward the video and garnering more likes. Another study using the name motion complexity for a very similar variable to Visual Variation is by Song et al. (2021). Song et al. (2021) found that much visual complexity leads to higher levels of engagement, measured as higher click-through rates, than videos with lower motion complexity.

Liu et al. (2018) found that the number of changes in scenes and the length of scenes particularly affected the happiness of 122 participants, who were asked to watch one of two trailers for 100 movies. The trailer with more scenes, which means the camera angle, background, and foreground change rapidly and frequently made consumers significantly less happy than those with fewer scenes. This finding by Liu et al. (2018) can also be translated into findings related to visual variety since the algorithm used for scene detection uses the frame-to-frame distances to calculate where frame boundaries are located. A high visual variety measure, as defined by Li et al. (2019), is closely related to when one scene starts and another one ends. Some findings combined can all be related to visual stimulus theory and how overstimulating and understimulating leads to less satisfaction (Liu et al., 2018; Hebb, 1955) and engagement (Li et al., 2019; Stuppy et al., 2023; Zhou et al., 2021). The overall linear negative (Zhou et al., 2021; Liu et al., 2018), linear positive (Song et al., 2021; Zhu et al., 2024; Stuppy et al., 2023) and inverted U-relationships (Li et al., 2019) found between visual variationa and engagement, attention or satisfaction showcase the disparity between findings in academic results and the need for additional research in various contexts.

2.3.2 Visual Complexity

Most visual marketing research implementing computer vision to analyse unstructured data uses static measures of images or videos. For example, these variables measure colour, light, or quality features (Zhang et al., 2017). Research reviewed in the literature on static visual features includes objects, orientational clutter, lighting, and colour-related factors. Pieters et al. (2010) proposed a measure for Visual Complexity, which measures static features that are frequently researched, like file size, objects, arrangement of objects, and symmetry of visuals.

Visual complexity, defined in section 2.2.2 as the ease of interpreting and analysing visual content, is often associated with Gestalt Principles (Kusumasondjaja & Tjiptono, 2019). This connection is consistent with most theories and operationalisations of Visual Complexity. Additionally, as discussed by Li et al. (2019) in section 2.3.1, the observed effects are commonly related to visual stimulation theories (Berlyne, 1970; Hebb, 1955).

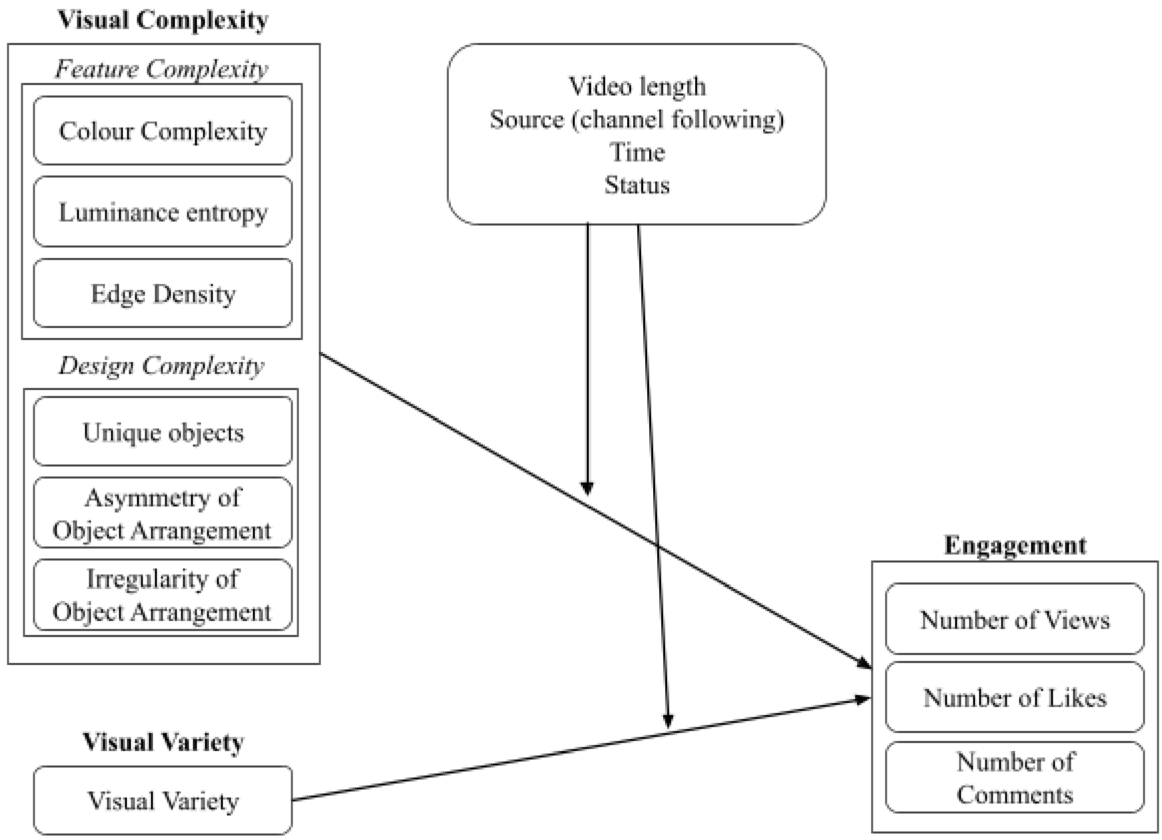
There are two theories regarding how advertisers can best attract attention to visual content where one side is under the impression that simplicity is key (Book & Schick, 1997), while the other side conveys that complexity might be best (Pieters et al., 2010; Song et al., 2021). Overgoor et al. (2022) state that visual complexity has both a positive and negative effect on the attitude toward offline advertisements and comprehension of the ad. One of the earlier measures of visual complexity is the file size of images or marketing prints. As outlined by Calvo & Lang (2004) and Teixeira et al. (2010), file size is a significant predictor of the perception of visual complexity. With the rise of computer vision, Pieters et al. (2010) proposed a well-supported manner of operationalising visual complexity while still using file size for some individual complexity measures. Pieters et al. (2010) suggest a breakdown of visual complexity in basic perceptual features (feature complexity) and visual complexity in design (design complexity). Song et al. (2021) even expand these two features described by Pieters et al. (2010) and Overgoor et al. (2022) into four aspects of Visual Complexity, Feature Complexity, Motion Complexity, Design Complexity and Object Complexity to try to capture more video information through the lens of Visual Complexity theory.

While most academics use visual complexity measures to analyse image content, only a few measured mean static features of video content (Song et al., 2021; Teixeira et al., 2010), despite the importance of video in advertising. Shin et al. (2020) investigate the influence of individual visual characteristics on online engagement metrics on Tumblr (liking and reblogging behaviours), where the effects of their visual complexity measures tend to be positive for image-pixel complexity but negative for object complexity. Overgoor et al. (2022) studied the relationship between individual measures of visual complexity and likes on brand-generated Instagram images, with the same distinction in feature and design complexity as Pieters et al. (2010). Pieters et al. (2010) outline six principles for measuring the design complexity of ads: quantity of objects, irregularity of objects, dissimilarity of objects, detail of objects, asymmetry of object arrangement, and irregularity of object arrangement. In the study by Pieters et al. (2010), all objects are manually coded, and feature complexity is operationalised as the JPEG file size. Shin et al. (2020) utilise an automated aggregate measure of design complexity using computer vision, which only considers the number of objects. Overgoor et al. (2022) use a combined measure at the individual level to measure both design complexity and feature complexity and base these measures on the design complexity principles outlined by Pieters et al. (2010) with some changes. Overgoor et al. (2022) found that both the irregularity of objects and the detail of objects are more closely related to the pixel level, which better fits the category feature complexity, and they found that they are closely related to edge density. Visual complexity can be operationalised as follows, according to Overgoor et al. (2022), based on the two constructs of visual complexity: feature complexity (FC) is made up of Colour complexity, Edge density and Luminance entropy and design complexity (DC) is made up of Objects, Irregularity of object arrangement and asymmetry of object arrangement. While this measure is created for images specifically, the logic proposed by Pieters et al. (2010) and Teixeira et al. (2010) holds up for videos since they are collections of frames with dynamic features and audio. Since the study intends to control for the dynamic, spatial and temporal nature of the video, the measure Overgoor et al. (2022) created inspired by Pieters et al. (2010) would propose a great measure of mean complexity of the frames on a video level for the whole video.

Overgoor et al. (2022) supported the hypothesis that feature complexity had an inverted U-relationship with likes. In contrast, the U shape was not fully supported for design complexity measures since the squared measure for asymmetry of object arrangement negatively influenced the number of likes. Song et al. (2021) specifically used some of the measures handled by Pieters et al. (2010) on video data to see the effects of visual features of advertising videos on engagement. Song et al. (2021) analysed 10000 videos from the Chinese e-commerce platform Taobao and found that design complexity and feature complexity significantly negatively impacted click behaviour but did not test for non-linearity analysis. Song et al. (2021) decided to control for views, likes and comments but used clicking behaviour as a dependent variable. Although clicking behaviour is obviously related to engagement on an e-commerce website, the findings cannot be directly translated into a conceptual relationship with different dependent variables. Teixeira et al. (2010) found a U-shape relationship between visual complexity, operationalised as compressed .GIF file size, and advertisement avoidance. Teixeira et al. (2010) studied Dutch TV commercials using eye tracking, and although the findings for ad avoidance do not exactly align with online engagement, it could indicate the presence of an overall inverted U-relationship with positive engagement metrics.

Figure 1 presents an overview of the higher-order constructs based on existing literature, which helps put the variables and what construct they belong to in perspective.

Figure 1: General overview of constructs and variables in the study

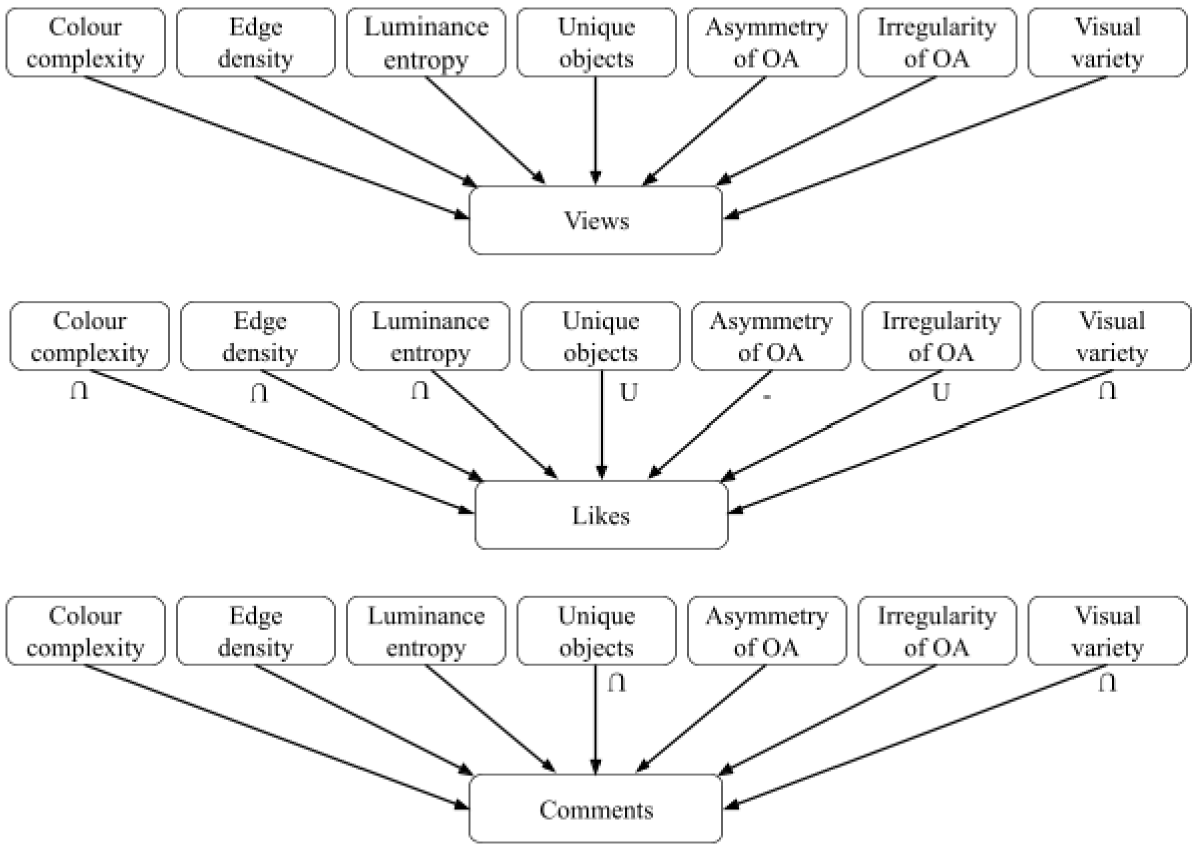


2.4 Exploratory Framework

The relationship between visual features and engagement could be contingent on several factors. These differences in the literature include the type of content (images vs. videos), the context of presentation (educational vs. marketing), and the chosen engagement metric (click-behaviour, likes, views or reblogs) as outlined in Table 1. This paper will explore the relationship between visual video features of advertisements and engagement based on Visual Stimulation Theory and the existing findings of papers exploring Visual Complexity and Visual Variation. Visual features are studied individually, and engagement is operationalised as various engagement metrics. As the independent variables' relation to the exact dependent variables has not been studied in the same context, an exploratory framework is construed instead of a theoretical one. Most visual complexity research that uses the same operationalisation handles image data. In contrast, the ones that use visual complexity on video data do not assess the relationship with the same measures of engagement. Visual variety is a variable only used for video content, and while the literature using this construct is closely related to engagement, the operationalisation of engagement significantly differs.

Figure 2 outlines the expected relationships based on the theories and papers in section 2. While most theories expect (inverted) U-relationships, researchers frequently fail to test these correctly (Haans et al., 2015) or can only partially confirm these due to model complexity or limitations. The figure captures the concepts and relationships based on the research question for similar relationships that have been tested before, while relationships that have not been tested before are left open. Figure 2 showcases the specific findings outlined above. Findings for the engagement construct likes have been based on image data, while the findings for comments are based on video data and live comments as a dependent variable. In the model, a ∩ indicates an inverted U-relationship, or parabola, a U indicates a U-relationship, a + indicates a positive relationship and a - indicates a negative relationship between the construct and the dependent variable.

Figure 2: Examined individual relationships and predicted shapes based on existing literature



**Note: Relationships that do not have a hypothesised direction (+/-/U/∩) are more exploratory due to a lack of concrete prior literature using the same constructs. Instead of presenting inaccurate hypothesised relationships, this paper leaves the relationships open for further exploration.*

An exploratory analysis is used to better understand the effects of visual advertising video features on engagement to avoid having too many complex hypotheses that might have to be rejected due to partial support. The findings by Overgoor et al. (2022) were used as a backbone for the theorised relationships between visual variety and likes despite the research being based on image content. Then, the studies on video content by Song et al. (2021) and Teixeira et al. (2010) using other operationalisations of engagement, such as zapping behaviour or click behaviours, are used to check if these relationships hold or whether the relationships between visual complexity measures and likes by Overgoor et al. (2022) might significantly differ for video content. For the hypothesised relationships between visual variety and engagement metrics, the studies by Li et al. (2019), Fornalczyk et al. (2023), Song et al. (2021), Liu et al. (2018), and Zhou et al. (2021) are used. None of the studies specifically use the online measure of likes; however, some operationalise ad liking as a subjective measure or more general engagement metrics such as the number of gazes or average fixation time. While some (Song et al., 2021; Fornalczyk et al., 2023; Zhou et al., 2021) observe a linear positive relationship of motion complexity or visual variety, others observe an inverted U-relationship (Li et al., 2019; Oğuz et al., 2023). The relationship between visual variety and comments is based on the finding by Zhu et al. (2024), who found that live comments have an overall positive but inverted U-relationship with visual variety or shot length. The measure labels by Zhu et al. (2024) also relate directly to the mean number of objects in a frame, which has an inverted U-relationship with live comments. Other relationships, like those between other visual complexity measures and comments, are likely similar to those outlined with likes due to the relationships with overall engagement metrics or subjective engagement. Despite the similarity, no expected relationships are given due to the more exploratory nature of the relationship of this variable. Considering the small amount of research concerning these relationships, the relationship with all visual features and views is more exploratory.

3. Methodology & Measurement

This section will outline the methodology for measuring the relationships between visual features and online engagement. It will also include the empiric setting, which explores the research context described in section 3.1. Section 3.2 will give some background on the data sample and showcase its characteristics. Section 3.3 explains the research design and includes an overview of each step in the research procedure. Then, section 3.4 provides information about the variable definitions and measurement, along with a visualisation of the sample to conceptualise the measured variables. Section 3.5 proposes models for analysis.

3.1 Empirical Setting

3.1.1 YouTube

YouTube is a video-sharing platform founded by Steven Chen, Chad Hurley and Jawed Karim in 2005. The platform allows users to upload, view, and share videos. YouTube supports a variety of content from individuals and corporations, making it a comprehensive hub for video content across categories such as education, music, gaming, sports, and more. YouTube accounts are called channels where creators can curate their content. The platform employs complex algorithms to recommend videos to users based on viewing habits and interests. YouTube is so interlinked with online culture that most big corporations are active on the platform, creating content to gain exposure or brand communities with their consumers. YouTube's equivalent of followers on platforms like Instagram and Twitter (X) is called subscribers, which means users essentially request that content from the channel they subscribe to be automatically added to their personalised feed and is more frequently suggested.

Exposure can be gained in various ways on YouTube, such as organic exposure or paid video advertising, which allows users to pay so that their video will be shown as an in-video advertisement for users. Users can engage with content through viewing, liking, and commenting on videos. Since 2022, the dislike button has been removed from YouTube, which has removed an engagement option from YouTube (Youtube [Team Youtube], 2021).

YouTube is the second most visited website worldwide after Google, with 113 billion monthly average visits as of November 2023 (Statista, 2024). The most prevalent visitors to YouTube are from the USA (22 billion), India (10 billion), Brazil (6.6 billion), and South Korea (6 billion) (Statista, 2024a). Apart from being one of the most visited websites, YouTube is also the third most popular platform after Facebook and Instagram concerning advertising.

The dispersion of engagement on YouTube shows that the top 3% of most viewed videos have accrued over 85% of the total views on the platform (Arthurs et al., 2018). 50% of videos uploaded in 2016 gathered less than 89 views (Arthurs et al., 2018). Overall, these statistics show that videos on YouTube are hit or miss, which helps uncover features of popular and non-popular videos, as the difference between engaging and non-engaging content should be explicit.

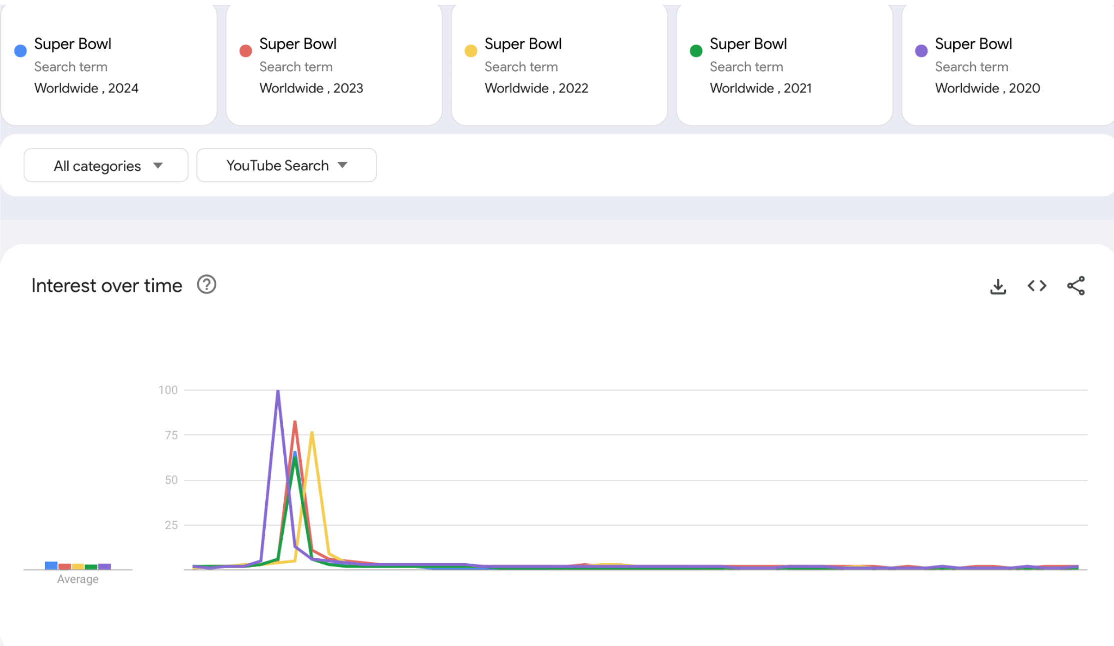
Since advertisements have become less product-focused and more content-focused, brands on YouTube have started sharing advertisements as organic content. Depending on the advertisements, these advertisements can gain enormous amounts of exposure and engagement on YouTube.

3.1.2 The Super Bowl

The sample of advertisements chosen for this study are Super Bowl advertisements. The Super Bowl is the biggest yearly sports and marketing event. The event is a broadcast of the finale of the annual NFL season, where two American Football teams from the American Football Conference and the National Football Conference play against one another. The 2012 rendition of the Super Bowl was the most-watched television broadcast ever, and four of the five most-watched telecasts are renditions of the Super Bowl (Hartmann & Klapper, 2018). Due to the massive audience that tunes in for the Super Bowl each year, corporations have been very eager to spend millions for exposure. A 30-second advertising spot during the 2019 Super Bowl cost nearly 5 million; in 2024, the cost of a 30-second spot has risen to 7 million US Dollars (Picchi, 2024). As corporations pay this much money for their 30 seconds of fame, the advertisements must be engaging, good-looking, shocking, funny or emotional for brands to capitalise on this bought exposure. The Super Bowl is such a deep-rooted event in American culture that three out of four Americans even state they are excited about the advertisements (Picchi, 2024). Despite the efforts brands put into their advertisements, not all ads are a hit each year due to the rising expectations and the history behind the Super Bowl. Most corporations use the buzz around their Super Bowl advertisement by plastering their social media with the original spot, extended cuts, and other variations of this aired Super Bowl ad.

Figure 3 illustrates the relative searches on YouTube for "Super Bowl" in Google Trends (Google, n.d.). The peak search volume is 100%, which is the basis for the relative search percentage. The figure highlights how quickly the Super Bowl buzz starts and dies down again equally over all years for which Super Bowl advertisements are collected.

Figure 3: Relative YouTube searches for "Super Bowl" from 2020-2024



*Note: This graph indicates relative interest and shows the short time frame in which Super Bowl ads are searched for on YouTube, being two to three weeks for all years, as shown in Google (n.d.)

This buzz around the Super Bowl on YouTube lasts 2-3 weeks, as shown in Figure 3. This short buzz indicates there is only a small period of time where the Super Bowl is relevant. One place where Super Bowl advertisements frequently appear is on the brand's YouTube pages. Due to the large scale of Super Bowl advertisements, it is a good sample of high-quality brand-generated videos for visual analytics and its effect on social media engagement. YouTube's API makes for convenient data collection, and the advertisements have varying degrees of organic success on platforms like YouTube despite the buzz and exposure created by their spot during the big game.

3.2 Sampling and Sample Characteristics

3.2.1 Video Sampling

Super Bowl advertisements are remotely sourced from the website "https://www.superbowl-ads.com/". The website functions as an online archive of Super Bowl advertisements over the years and links to each year's uploaded YouTube advertisement or a downloaded version of the advertisement. For this study, Super Bowl advertisements from 2020 to 2024 are included in the sample.

The sampling procedure consists of video sampling for YouTube data collection and frame sampling for video analysis. The video sampling consideration follows the following process:

The embedded video on "https://www.superbowl-ads.com/" is checked for authenticity. This includes:

I. Check if the advertisement is posted on the original channel.
II. Check if the advertisement is of reasonable length (around 30 seconds, 1 minute or even 90 seconds).
    A. Additional web search is done to find the original spot when the advertisement is not near the duration markers.
III. If the advertisement is not on an original channel or of original length, a YouTube search is done to find out if the original spot is posted on YouTube by the official brand channel.
IV. If step c does not garner results, the link, video name and channel name are not gathered for further analysis.

Channels can limit engagement on specific videos by hiding the like count and disabling comments. Videos with disabled comments return a no value in the comment_count column when calling the data API, and they are eliminated from the analysis.

Since not all advertisements from previous years have been successful, some brands have decided to private the videos so they do not show up in the organic YouTube algorithm anymore and are only accessible by URL. In this case, the decision is made to still analyse these videos.

Eliminating these videos from analysis would introduce a performance bias to the sample, where only ads that performed well are taken into account, assuming brands would unlist content that was not successful. For this reason, we will add a dummy variable status to the analysis to control for it and do a sensitivity analysis to see the effect of including and excluding unlisted videos.

3.2.2 Frame Sampling

This research method requires another sampling method, which needs to be addressed: frame sampling. Considering the required processing power for object detection and mathematical calculations of some Visual Complexity measures, a sample of frames must be collected to represent the videos. In most cases, this is done in one of two ways: key frame extraction or interval-based frame selection (Chen et al., 2018).

In this study, frames are sampled across videos on an interval basis so that the sampled frames are evenly spaced throughout the video. This methodology is generally impractical for studies using video analytics to maximise the informativeness or quality of frames for categorisation or machine learning purposes (Chen et al., 2018). Key frame extraction or similar algorithms are developed to maximise frame quality or informativeness based on the degree of variety in frames, colour, or number of actions or objects in frames. Since this study aims to calculate the mean visual complexity and variety, which are low-level visual features, we want the most representative frames for the video without introducing biases like extracting the frames on certain features or qualities. The exact reasons why key frame extraction is a good practice for visual video categorisation make the interval-based extraction of 30 evenly spaced frames useful for averaging visual variation and visual complexity measures on a video level. Measure-level specifications for frame selection and averaging are further explained in section 3.3.

3.2.3 Video Sample Characteristics

The data collection process using "https://www.superbowl-ads.com/" from the previous five years resulted in YouTube data being collected from 159 Super Bowl advertisements.

Of the 159 advertisements, 13 videos have missing values for the like count, and 19 videos have missing values for the comment count, meaning the channel hid this value. Due to some overlap in these missing values, the final cleaned dataset contains YouTube data from 132 advertisements spread over 90 official brand channels. Table 2 showcases the sampled videos and some general advertisement characteristics, like how many videos were published and sampled each year and the status of those sampled videos on YouTube, which can be public or unlisted.

Table 2: YouTube video characteristics of Super Bowl advertisements

Year (uploaded)	*Commercials aired*	*Commercials sampled*	*Status: public*	*Status: unlisted*
2020	70	17	11	6
2021	51	12	9	3
2022	59	16	11	5
2023	70	30	25	5
2024	59	57	56	1

When assessing the data per year in Table 2, there is quite a significant drop-off in ads per year, and it is clear that brands tend to delete content from official brand channels quite frequently. Since there seems to be a significant drop-off in Super Bowl advertisements on YouTube for prior years, it is important to check for survivor bias to see if only very successful advertisements are still available.

3.3 Definitions and Measurement

3.3.1 Visual Variety Measurement

Li et al. (2019) calculated visual variance by dividing the video into ten equally distanced clips, where a frame is extracted precisely in the middle of each clip as this lowers the computational power needed to analyse the number of videos Li et al. (2019) examined in the study. In this study, a similar methodology is used by extracting a frame when the frame count is exactly at 1/10th of the rounded total frame count. This methodology allows for a thorough video analysis without the limited computational power needed for video analysis and calculations being a concern. Using the measure proposed by Li et al. (2019), colour images are transformed into grayscale to reduce complexity and computational load. For instance, a colour pixel (49,15,14) becomes a grayscale value of 26 by averaging the RGB values. Next, the pixel values are normalised from [0,255] to [0,1] to adjust for exposure differences, ensuring consistency across frames captured at different times. Finally, the pixel-level distance is computed between frames using the Manhattan norm, summing up the absolute distances across all pixel pairs. This approach assumes uniform frame sizes across the video.

Finally, the result of a video-level calculation is a list of 9 summed distance measures from 2 consecutive frame pairs averaged over the nine frame pairs to get the final measure of visual variety on a video level. There are nine frame pairs over ten selected frames because frame one does not have a preceding frame, so frames 1-2, 2-3, 3-4 …. 8-9, and 9-10 are the pairs created. This variable accounts for the dynamic temporal aspect of content within a video. Figure 4 illustrates frame pairs with low (25th percentile), medium (50th percentile) and high (75th percentile) visual variety scores in the sample. The figure shows the preceding and current sampled frames to conceptualise the visual variety scoring. Although the frames are not identical for a low level of visual variety, the colours, arrangement and objects are quite similar, leading to a low variety score. At a high level of visual variety, frames one and two are entirely different. Nothing within frame one can be identified in frame two apart from the colour black in the cinematic black bars in frame one and the background in frame two.

Figure 4: 25th, 50th and 75th percentile of Visual Variety scores visualised between frame pairs

*Note: The 25th, 50th and 75th percentile scores for visual variety were calculated using the distance between each consecutive frame-pair for all of the 132 video advertisements. This method led to 9 frame pair distance measures per video. The distance between frame pairs was calculated to find consecutively sampled frame pairs that align with the 25th, 50th and 75th percentile scores.

3.3.2 Visual Complexity Measurement

Visual complexity is divided into **feature complexity (FC),** consisting of;
Colour complexity, Edge density, Luminance entropy and **design complexity (DC)**, consisting of the Unique number of objects, Irregularity of object arrangement and Asymmetry of object arrangement.

These variables will be aggregated as follows: For each examined frame, each of these measures is calculated per frame and then averaged over the number of frames examined, which results in the average complexity score per video. These independent variables will be measured as follows, following the outline of Overgoor et al. (2022):

1. **Colour complexity** (FC): For each frame, RGB is transformed into CieLab colour space. CieLab is a colour space that, instead of RGB (Red, Green, Blue), uses 2 measures of colour, which are the a-axis, an axis with red or green at either of the extremes and the b-axis, an axis with blue or yellow at either of the extremes. The final measure in CieLab is L*, which is the Chroma, or the perceptual lightness. The lightness and the colour combination based on the values of the a- and b-axes make the colours we perceive as humans. After the CieLab conversion, the mean Chroma (L*) and the standard deviation along the a- and b-axes are calculated. Then, the colourfulness of the frame can be calculated using the most accurate measure proposed by (Hasler & Süsstrunk, 2003):

   $$Colour_i = 0,94 \star \mu_c + \sqrt{\sigma_a^2 + \sigma_b^2}$$

2. **Edge density** (FC): Edges are detected using the edge detector from Canny (1987). Every pixel is classified by a binary value, 0 not being an edge and 1 being an edge. An edge is essentially a contour or a place where the brightness or colour of pixels changes very drastically. These edges can be the outlines of objects but also shadows or details in objects. Edge density is then calculated by dividing the total number of pixels on an edge by the total number of pixels.

   $$Edge_i = (\frac{e_i}{N})$$

3. **Luminance entropy** (FC): Luminance Feature Complexity: To measure this, RGB is first converted into YUV to obtain per-pixel luminance (Y). Then, unique luminance levels and count pixels are identified at each level to calculate luminance variety entropy as follows:

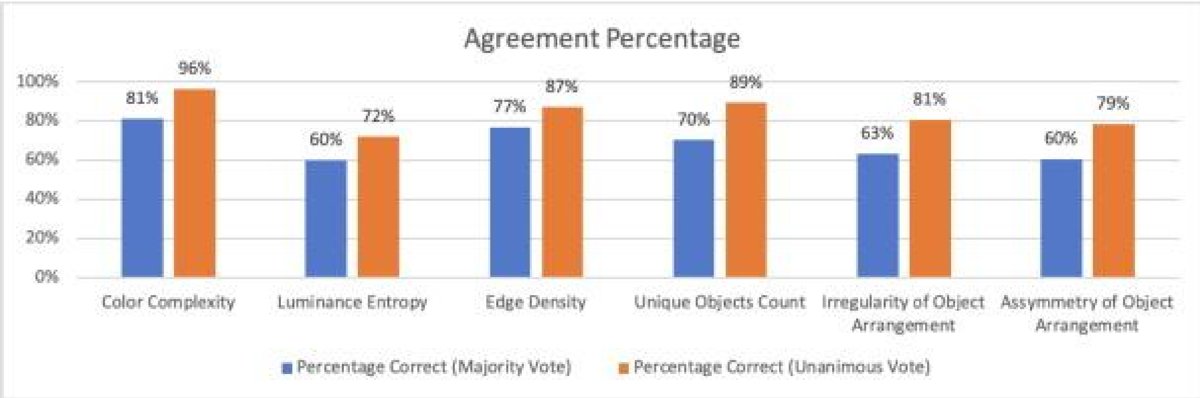
   $$Luminance_i = -\sum_{j=1}^{T} n_j log(\frac{n_j}{N})$$

   Where T is the total number of unique luminance levels, n_j is the count of pixels that contain unique luminance levels j, and N is the number of total pixels.

4. **Objects** (DC): Overgoor et al. (2022) use a pre-trained mask R-CNN, as Nagle and Lavie suggested to be the most accurate (2020). A pre-trained mask R-CNN is also used since the results of Nagle and Lavie (2020) indicate the best object-level performance for visual complexity. Detectron 2, developed by Wu et al. (2019) from META, counts the unique number of objects in the analysed frames.
5. **Irregularity of object arrangement** (DC): Irregularity of objects focuses on visual clutter, or the perceptual organisation, by finding similarities, proximity and grouping. For this, the visual clutter model and code will be used by Rosenholtz et al. (2007a) and Overgoor et al. (2022). Following the approach outlined by Rosenholtz et al. (2007a), the oriented opponent energy is calculated, following the principles set by Bergen and Landy (1991). This calculation produces a pair of values $(\kappa cos(2\theta), \kappa sin(2\theta))$ at each point and scale within the image, which θ signifies the local orientation angle and κ reflects the intensity of a predominant orientation at the specific scale and point. The volume (or area) is encapsulated by an orientation distribution ellipsoid based on the covariance matrix derived from the pair of values to quantify orientation clutter. The overall irregularity in the placement of objects within the frame is then quantified by taking the average of this clutter measure throughout the image.
6. **Asymmetry of object arrangement** (DC): Following the methodology employed by Overgoor et al. (2022), the image is divided into a horizontal and vertical plane to compare the pixels of each opposing plane (left compared to right and top compared to bottom), using the same feature congestion map, or orientation, as the irregularity measure. Finally, the average of vertical and horizontal asymmetry is taken.

Figure 5 showcases the findings by Overgoor et al. (2022), who surveyed 289 students to find the overlap between the objective and perceived measures of visual complexity as a pre-test. The survey consisted of showing each student 35 image pairs randomly sampled from 900 images. Respondents were asked to pick the most complex image out of the two. The scores indicate the overlap between the objectively calculated complexity measures and the perceived complexity assessment. Overgoor et al. (2022) found a Cronbach's alpha of .74 and the objective and subjective measures to unanimously overlap in over 60% of instances for all variables over all image pairs. This finding is in line with the finding by Shin et al. (2020) that human perception and machine calculations are highly related.

Figure 5: Agreement in perceived and machine-calculated complexity (Overgoor et al., 2022)



*From Overgoor et al. (2022, p.648)

Similar to Figure 4, Figure 6 illustrates the low (25th percentile), medium (50th percentile), and high (75th percentile) levels for all measures of visual complexity relative to the 132 sampled advertisements. Visual complexity is calculated one frame at a time as opposed to visual variety. The figure shows how high, medium and low levels of colour complexity, edge density, luminance entropy, asymmetry of object arrangement, irregularity of object arrangement, and number of unique objects differ and what these visual concepts look like to the human eye.

Figure 6: 25th, 50th and 75th Percentile of Visual Complexity scores in the video sample

Low (25th percentile) | Medium (50th percentile) | High (75th percentile)

Colour Complexity

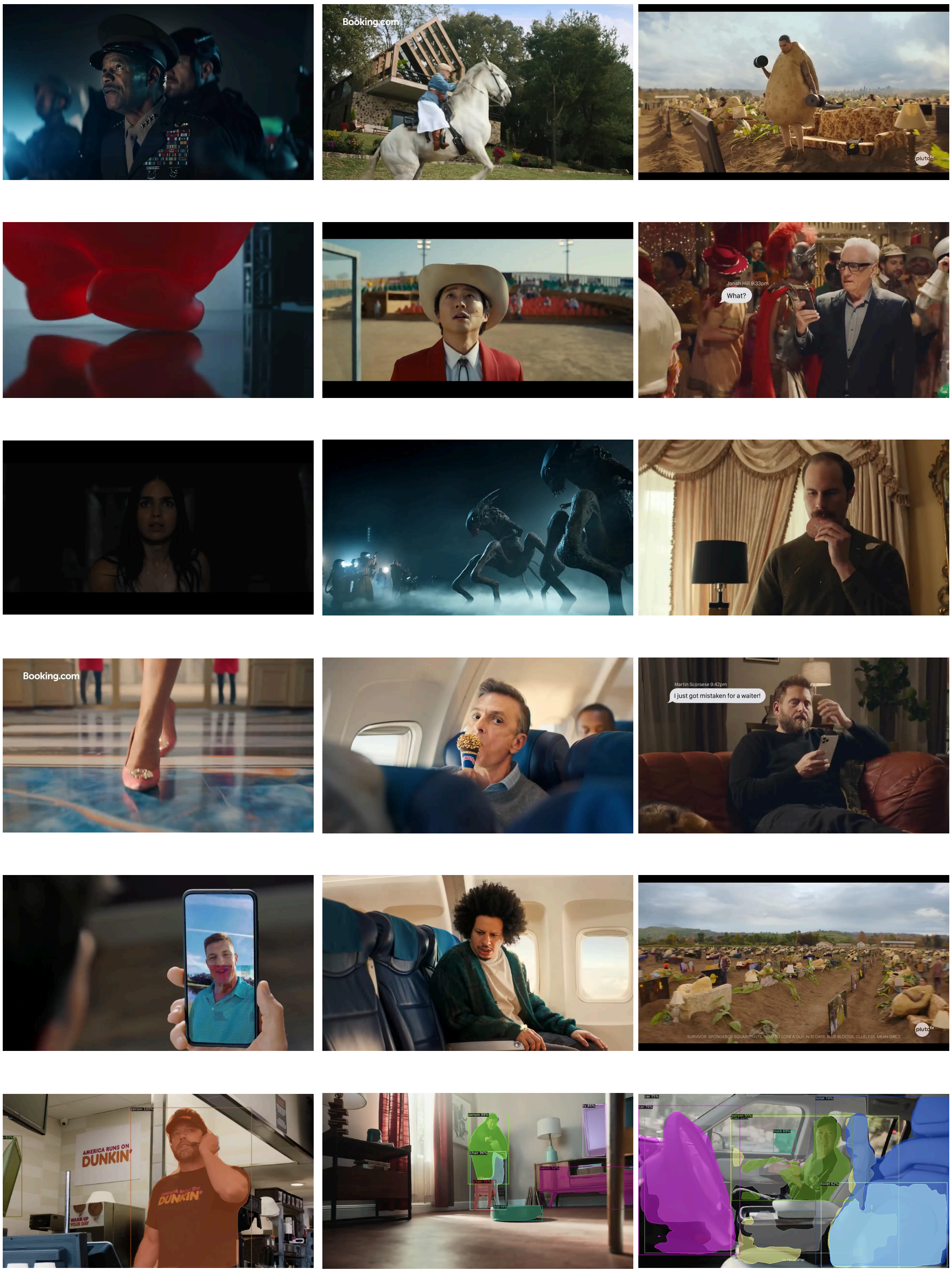
Luminance Entropy

Edge Density

Asymmetry of OA

Irregularity of OA

Unique objects



*Note: The 25th, 50th and 75th percentile scores for each facet of visual complexity were calculated using the scores of the sample of 132 video advertisements. Thirty equally distanced frames were extracted from a random set of 10 advertisements in the sample to find frames closest to each of the percentile scores for each visual complexity variable. So, the frames displayed below are low, medium and high visual complexity scores relative to the sample of 132 Super Bowl advertisements, or the 3960 total frames analysed.

3.3.3 Online Engagement

Views: The number of times a video has been viewed. This variable is retrieved using the YouTube Data API v3.
Likes: The number of likes under the video on YouTube. This variable is retrieved using the YouTube Data API v3.
Comments: The number of comments under the video on YouTube. This variable is retrieved using the YouTube Data API v3.

3.3.4 Control Variables

Number of account subscribers: The number of accounts subscribed to the YouTube channel from which the video was published. This variable is retrieved using the YouTube Data API v3.

Time: The number of days since the video was posted on YouTube. This variable is calculated based on the analysis date (14th of April) and the upload time, retrieved using the YouTube Data API v3.

Length: The length of the YouTube video in seconds. This variable is retrieved using the YouTube Data API v3.

The control variables listed are various YouTube metrics that need to be controlled for. Account subscribers, time posted and video duration are added to control for variables that might influence the traction and engagement of the advertisement. The amount of subscribers a brand channel has reflects the size of the brand community of this specific brand on YouTube. Accounts with more subscribers end up in more users' feeds, which could significantly influence the relationship between visual video features and engagement. The time a video has been online is equally important to control for since videos that have been recently added had less time to accumulate the same amount of traction as other videos. Lastly, the duration of the videos is controlled for since most videos are equal in length due to the format of the Super Bowl, but some are significantly longer. The duration of the advertisement could influence consumer engagement and how visual information is measured. Even though a case can be made for adding personal branding as a control variable, based on the number of brands compared to the sample size, this variable could lead to overfitting the regression model. Other research using visual complexity has also decided to leave individual creator characteristics out of the equation, apart from variables that could directly impact the dependent variable, such as channel or content source (Teixeira et al., 2010; Pieters et al., 2010; Overgoor et al., 2022; Shin et al., 2020; Zhou et al., 2021; Li et al., 2018; Stuppy et al., 2023).

Table 3 summarises the variables relevant to this study, the construct they belong to, the variable's role in the study, and a short description of the variable.

Table 3: All variables in the study

Concept	Variable	Variable type	Description
Feature Complexity	Colour Complexity	Independent variable	*Mean variation and distribution of colours*
Feature Complexity	Edge Density	Independent variable	*The mean number of edges per unit area for each frame, indicating texture or detail*
Feature Complexity	Luminance entropy	Independent variable	*Mean Randomness in brightness levels*
Design Complexity	Number of unique objects	Independent variable	*The mean of distinct object types in every frame*
Design Complexity	Irregularity of object arrangement	Independent variable	*The mean degree of randomness in object placement*
Design Complexity	Asymmetry of object arrangement	Independent variable	*Mean lack of symmetry in object positioning*
Visual variety	Visual variety	Independent variable	*The mean difference between pixel-pairs*
Online engagement	Views	Dependent variable	*Number of views*
Online engagement	Likes	Dependent variable	*Number of comments*
Online engagement	Comments	Dependent variable	*Number of likes*
	Length	Control	*Duration of video in seconds*
	Time	Control	*Time in days since the video was uploaded*
	Subscribers	Control	*Number of subscribers of the YouTube channel*
	Status	Control	*Status of the video on YouTube, being either public or unlisted*

3.4 Research Procedure and Data Collection

3.4.1 YouTube Data Collection

Several data collection and cleaning steps must be completed before creating a feature matrix for the Super Bowl advertisement. From the URLs gathered through the sampling procedure, the video IDs are extracted from the URL of each YouTube video shown in Figure 7. The figure shows the whole YouTube video URL, of which the specific Video ID is marked with red, after which the channel information is presented.

Figure 7: YouTube video ID

"https://www.youtube.com/watch?v=qoER8XCOPaM&t=1s&ab_channel=Homes.com"

With YouTube's Data API v3, the video ID list is passed through the Application Programming Interface (API) with a GET request in a Python script in chunks of 50 and added to a data frame. The following data points were collected with the YouTube Data API that are relevant for further analysis: *id, publishedAt, channelId, title, channelTitle, duration, viewCount, likeCount, commentCount and status.*

Another GET request was made using the channel IDs resulting from the video requests with the 'GET channel' parameter from the YouTube Data API to get channel information. This request results in the final raw YouTube-related variable needed for analysis, which is subscriber count, as shown in Figure 8. The figure shows the first four rows of the YouTube data feature matrix, resulting from the video and channel GET requests. It shows the video and channel information gathered using the YouTube v3 Data API.

Figure 8: First four rows of the YouTube feature matrix as a result of video and channel-level API requests

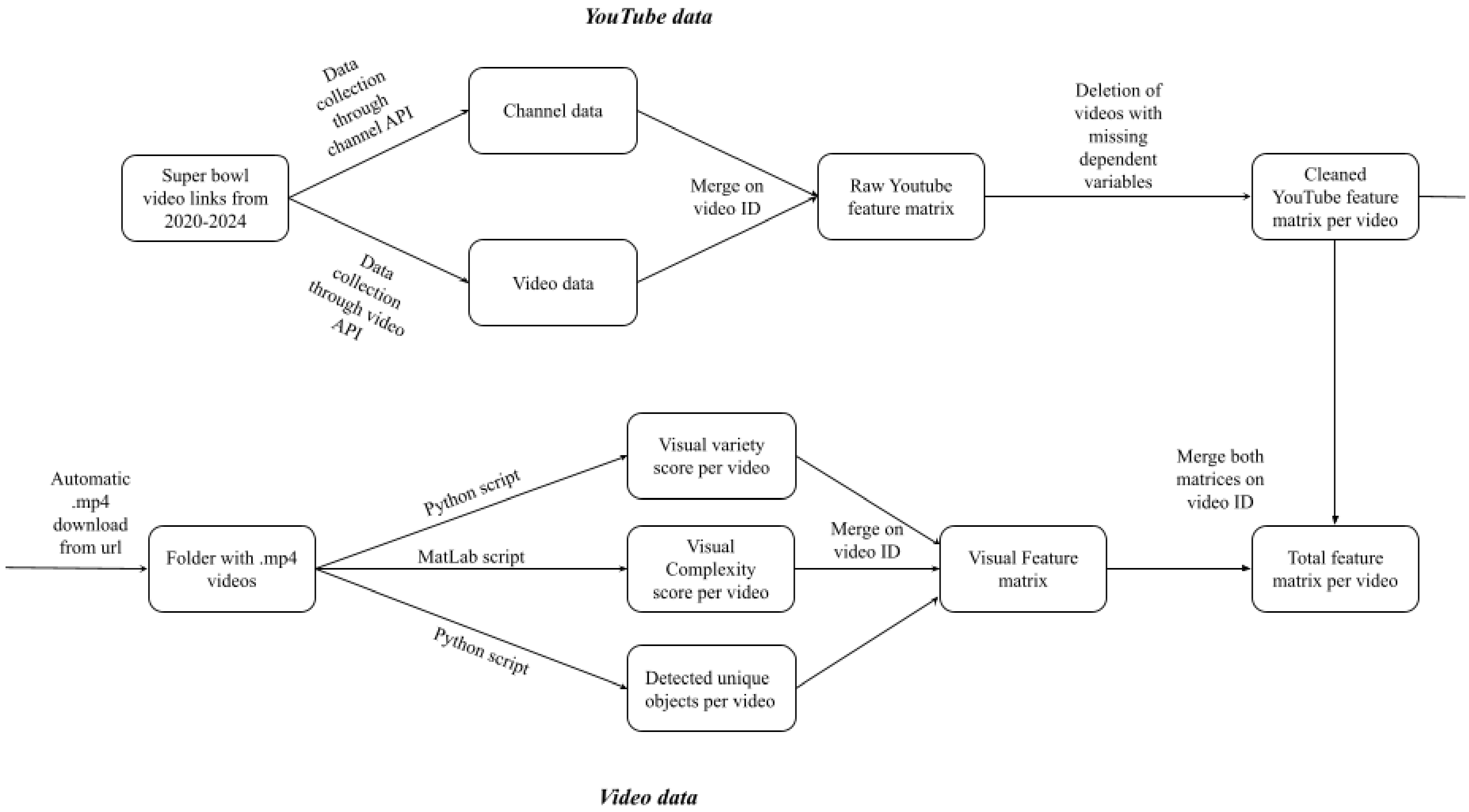
	id	publishedAt	channelId	title	channelTitle	duration	viewCount	likeCount	commentCount	status	subscriberCount
1	vQ88JB-lgtQ	2024-02-12T02:54:16Z	UCIrgJInjLS2BhlHOMDW7v0g	"Well Said" | Big Game Spot | Disney+	Disney Plus	PT31S	2892202	1969.0	309.0	public	1530000
2	AJNUI0WqEdE	2024-02-12T03:21:18Z	UCq0OueAsdxH6b8nyAspwViw	Monkey Man | Elevator	Universal Pictures	PT31S	166043	5151.0	197.0	public	8770000
3	DTLM5o-9hYg	2024-02-12T03:03:52Z	UCXQ7GzszfuW84cEWGpKlFiQ	DoorDash | DoorDash All the Ads - :30	DoorDash	PT31S	210524	995.0	2200.0	public	63700
4	lWVFBkx8_4E	2024-02-11T14:09:08Z	UC4aAdHyFHHMz48k2fNQAUGg	Homes.com | Quick question	Homes.com	PT31S	28388	84.0	16.0	public	4810

Finally, all YouTube-related variables were collected and merged into one large YouTube feature matrix before data cleaning could commence. Before some control variables were cleaned and factorised, videos with missing dependent variables were deleted from the sample. Video duration was re-calculated to contain the duration of the video in seconds, and time was constructed to contain the days since the video was uploaded onto YouTube on April 19th, 2024. After data was cleaned and unusable observations were deleted from the cleaned dataset, a Python script was used, which leveraged the PyTube package to download videos from their URL automatically. This URL is reconstructed based on each video's unique ID.

Both Python and MatLab were used for video analysis. Python was used to calculate visual variation and for object detection using Detectron2, an open-source pre-trained Mask R-CNN built by Meta. The other variables related to visual complexity were calculated in MatLab while leveraging the code built by Overgoor (2022), Rosenholtz et al. (2007b) and Walthers and Koch (2007). The visual complexity calculations depend on many helper functions, mathematical calculations and computer vision functions. In contrast, visual variety and object detection have been coded using computer vision and object detection applications.

Figure 9 gives a full overview of the data collection procedure. The figure is divided into 2 main branches, one for collecting YouTube data and one for video analysis. The outcome of the data collection procedure is a feature matrix, with one row per video, that has the measures of visual information and YouTube specific information for further analysis.

Figure 9: Overview of the data collection procedure



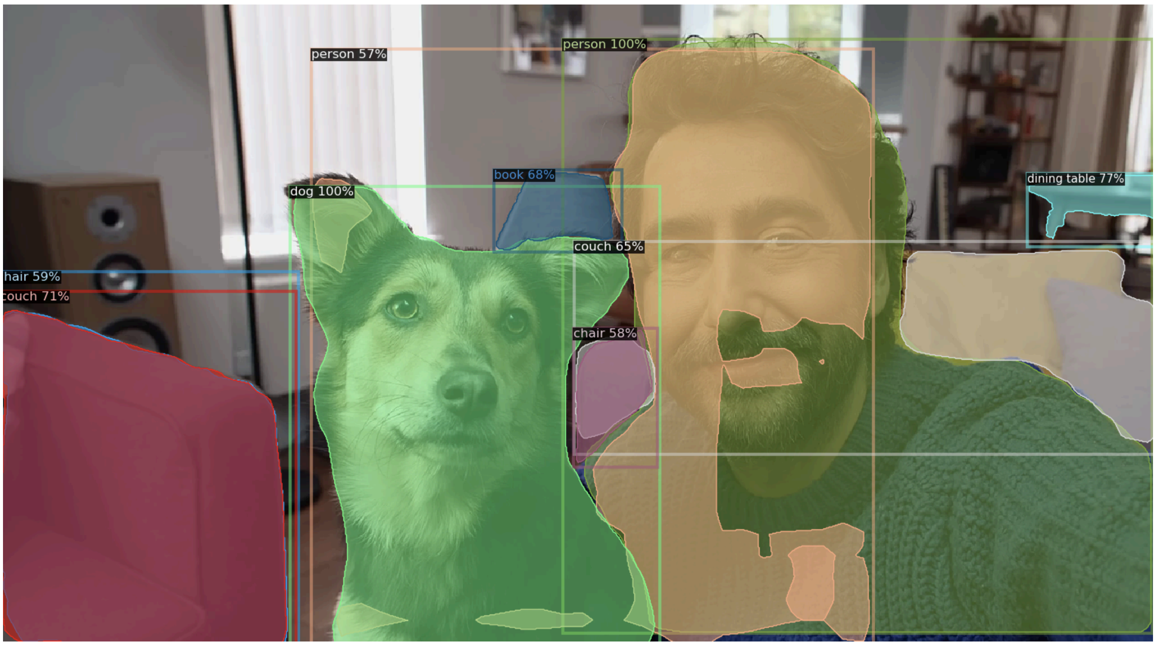
3.4.2 Object Detection Algorithm

Object detection is a computer vision application that allows computers to identify and track objects in an image or video. It involves classifying data (pictures or videos) and locating the objects in the data. This detection is typically achieved through algorithms and models that recognise patterns and characteristics, such as edges, shapes, and textures. Modern object detection systems use deep learning models, particularly convolutional neural networks (CNNs), to detect objects accurately. These models are trained on large datasets with labelled images to learn object features. Once trained, the models can predict object locations in new photos by outputting bounding boxes and associated class labels for each detected object. This research uses a pre-trained Mask R-CNN (Region-based Convolutional Neural Network) for object detection—a faster R-CNN model designed for object detection tasks within images. Mask R-CNN adds a branch for predicting segmentation masks on each Region of Interest (RoI), enabling the model to perform both object detection and instance segmentation simultaneously.

Using a pre-trained machine learning algorithm might be less accurate than training an object detection algorithm using data from the same context. Training a model, however, is much more resource-intensive. For the current application of counting objects over various frames, the highest accuracies and tracking characteristics are not as important since the goal is simply counting objects, not identifying them with individual labels. He et al. (2017) extended the Region-Based Object Detection Model of Girshick et al. (2014) by proposing Mask R-CNN for more accurate object detection. In the research on visual complexity, both Overgoor et al. (2022) and Shin et al. (2019) used pre-trained Mask R-CNN models. Another popular pre-trained model for object detection is YOLO ("You Only Look Once") by Ultralytics (YOLOV8: A New State-of-the-Art Computer Vision Model, n.d.). YOLO is better for high-speed or real-time applications than a Mask R-CNN like Detectron2 but less accurate due to the lack of region-based detection (Girshick et al., 2014) and masking (He et al., 2017).

Figure 10 shows an example of the Mask R-CNN algorithm Detectron2 on one of the sampled frames from a random advertisement in the video sample. The output of the Mask R-CNN shows the object box, the object label, and the confidence with which the model estimates that the label is correct for the visual object as a percentage. Considering that the precise accuracy of the label itself is not highly important, as unique objects are simply counted, the algorithm is set to only detect objects that it can label with more than 50% confidence.

Figure 10: Output of Mask-R CNN using Detectron2 in Python



3.5 Models

3.5.1 Full Linear Model for Effect Estimation

A two-stage analysis will assess the relationship between online engagement and visual features. Due to the limited amount of data and complexity of the conceptualised (inverted) U-relationships by Stuppy et al. (2023), Pieters et al. (2010), Li et al. (2019), and Overgoor et al. (2022), a full linear model will be used first to analyse the individual linear effects of all visual variables on online engagement. In section 3.5.2, the non-linearity testing is discussed, which is the second stage of the analysis.

The full linear models are constructed as follows:

$$Views = \beta_{1,0} + \beta_{1,1} Colour\ Complexity + \beta_{1,2} Edge\ Density + \beta_{1,3} Luminance\ Entropy + $$
$$+ \beta_{1,4} Object\ count + \beta_{1,5} Object\ Arrangement\ Asymmetry + \beta_{1,6}\ Object\ Arrangement\ Irregularity +$$
$$\beta_{1,7} Visual\ Variety + \beta_{1,8} Subscribers + \beta_{1,9}\ Duration + \beta_{1,10} Time + \beta_{1,11} Status + \epsilon_1$$

$$Likes = \beta_{2,0} + \beta_{2,1} Colour\ Complexity + \beta_{2,2} Edge\ Density + \beta_{2,3} Luminance\ Entropy +$$

$$+ \beta_{2,4} Object\ count + \beta_{2,5} Object\ Arrangement\ Asymmetry + \beta_{2,6}\ Object\ Arrangement\ Irregularity + \beta_{2,7} Visual\ Variety + \beta_{2,8} Subscribers + \beta_{2,9}\ Duration + \beta_{2,10} Time + \beta_{2,11} Status + \epsilon_2$$

$$Comments = \beta_{3,0} + \beta_{3,1} Colour\ Complexity + \beta_{3,2} Edge\ Density + \beta_{3,3} Luminance\ Entropy + + \beta_{3,4} Object\ count + \beta_{3,5} Object\ Arrangement\ Asymmetry + \beta_{3,6}\ Object\ Arrangement\ Irregularity + \beta_{3,7} Visual\ Variety + \beta_{3,8} Subscribers + \beta_{3,9}\ Duration + \beta_{3,10} Time + \beta_{3,11} Status + \epsilon_3$$

Here, $\beta_{i,j}$ represents the effect of the explanatory variable j on the dependent variable i for $\in (1, 2, 3)$, j for $\in (1, 2, \ldots, 10, 11)$, where i = corresponds to views, $i = 2$ corresponds to likes and $i = 3$ corresponds to comments. Here, j=0 represents the intercept, $j = 1$ corresponds with Colour Complexity, $j = 2$ corresponds with Edge Density, $j = 3$ corresponds with Luminance Entropy, $j = 4$ corresponds with Object Count, $j = 5$ corresponds with Object Arrangement Asymmetry, $j = 6$ corresponds with Object Arrangement Irregularity and $j = 7$ corresponds with Visual Variety. The explanatory variable $j = 8$ corresponds with Subscribers, $j = 9$ corresponds with Duration, $j = 10$ corresponds with Time and $j = 11$ corresponds with Status. The error terms ϵ_i summarise the effect of variables not in this model on the dependent variable (Heij, 2004).

3.5.2 Non-Linear Models

The non-linearity and (inverted) U-relationships are assessed in a way that should not cause overfitting due to the number of predictors and limited data points. Following the methodology of Overgoor et al. (2022) and Li et al. (2019), quadratic relationships should be tested with at least the linear and squared predictor present in one model, which is in line with the methodology for testing non-linear and (inverted) U-relationships outlined by Haans et al. (2015). Due to the added complexity of using linear and squared predictors of each variable in one model to compare against the full linear models above, a choice is made to create smaller models for each independent variable to reduce the chance of overfitting. If the choice were made to include a full model with a quadratic term for each independent variable, this would grossly violate most rules of thumb related to the number of predictions per observation (Babyak, 2004). Two smaller models are created per dependent variable, where one has the linear independent variable with controls as predictors, and the other has the linear and squared independent variable with controls as predictors. These models are not used to determine the individual effects of visual variables on online engagement or compare with the full models; instead, they serve as a tool to assess the level of non-linearity and test the (inverted) U-relationships.

A pair of these models take on the following general form:

$$DV = \beta_0 + \beta_1 IV + Controls + \epsilon$$

$$DV = \beta_0 + \beta_1 IV + \beta_2 IV^2 + Controls + \epsilon$$

The DV could be views, likes, or comments, and the IV could be Colour Complexity, Edge Density, Luminance Entropy, Unique Objects, Irregularity of OA, Asymmetry of OA, or Visual Variety in the formula above. The controls are subscribers, duration, status, and time in every model. Section 5.2.2 further explains the assessment of non-linearity and (inverted) U-relationships. These models, therefore, make for 7 model pairs per dependent variable. This means that there are 3 times 7 model pairs for non-linearity testing or a total of 42 models.

This testing methodology assumes that excluding other visual variables, linear or squared, from the small models does not significantly confound a non-linear model's non-linearity.

4. Data Analysis

This section will analyse data before the results are interpreted. Section 4.1 will cover descriptive statistics to explore the data. Section 4.2 will contain some steps in the pre-analysis, such as finding the correct regression model type and testing for multicollinearity. Finally, in section 4.3, the tests and models for non-linear relationships will be outlined before the final regression models are shown and explained in section 4.4.

4.1 Descriptive Statistics

Table 4 showcases the descriptive statistics of the data for the sample of Super Bowl videos, where n=132. The values in Table 4 are the cleaned and raw means, standard deviations, minima and maxima of the data, without scaling or normalisation techniques such as z-scaling having been performed yet.

Table 4: Descriptive statistics of the model

	Mean	Std dev	Max	Min
Views	8640998	24772450	216009800	656
Comments	906.09	2232.57	14209	0
Likes	11176.58	31790.66	261360	6
Colour Complexity	22.02	10.20	73.60	0.26
Edge Density	0.05	0.01	0.08	0.002
Luminance entropy	4.37	0.64	5.07	0.70
Unique objects	2.21	1.00	4.90	0
Asymmetry of OA	0.01	0.004	0.03	0.001
Irregularity of OA	0.04	0.01	0.08	0.03
Visual Variety	0.22	0.05	0.37	0.10
Duration (seconds)	43.73	20.58	147	7
Time (days)	532.03	519.25	1547	32
Subscribers	1684204	3795926	20400000	310

*Note: n=132 (all analysed videos)

When analysing the descriptive statistics, the variability of the dependent variables and the different scales of independent variables are the first things that stand out. All dependent variables are integer count data, whereas the independent variables, apart from unique objects, are all variables on a continuous complexity or variety scale. Due to the dependent variables being counts, the measures of engagement can only take on non-zero integer values. Considering the dependent variables are all non-negative integers, a Poisson-Regression would fit this type of analysis (Coxe et al., 2009).
Secondly, the standard deviation seems much higher than the mean of views, likes and comments. This could indicate high dispersion of the dependent variables in the data, frequently seen in social media data (Rietveld et al., 2020; Rooderkerk & Pauwels, 2016).

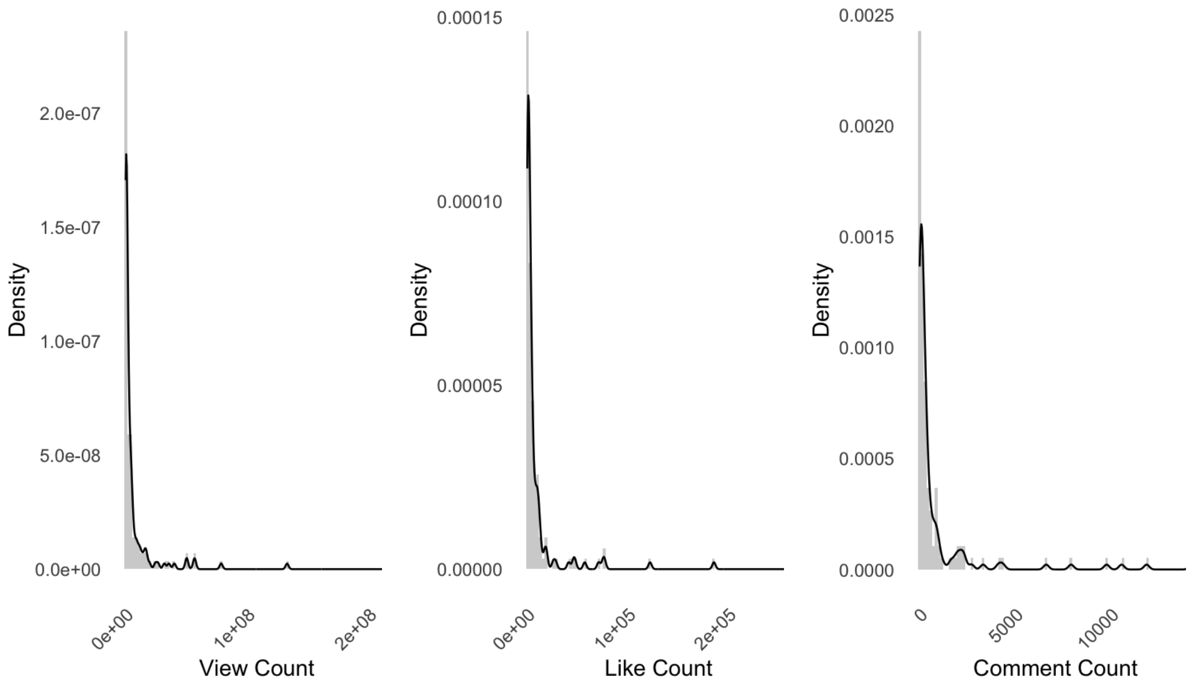
4.2 Pre-Analysis

Before a regression model can be fitted and interpreted, the most suitable model needs to be selected based on the data, and the related assumptions for this model have to be tested.

4.2.1 Regression Model Fitting

Due to all the dependent variables being non-negative discrete count data, a Poisson regression model is fitted to the data. A Poisson regression assumes that the mean of the dependent variable is equal to the variance. Figure 11 illustrates the density plot of the dependent variables views, likes and comments. The density plot uses kernel density to show the probability of the variable taking on certain values on the Y-axis and the value of the dependent variable on the X-axis. Overall, the plot shows the distribution of the data and, thus, the probability of the engagement metrics taking on certain values.

Figure 11: Density plots of the dependent variables views, likes and comments



As shown in Figure 11, the density plots of the dependent variables show right skewness and the long right tail. The descriptive statistics outlined in Table 4 in combination with the skewness in the density plot of dependent variables in Figure 11, could indicate overdispersion. A Negative Binomial regression deals with overdispersion well. Both a Poisson regression and a Negative Binomial regression are fitted to see if dispersion is high enough to an extent that a Negative Binomial model fits the data better.

The Negative Binomial regression model is a Generalised Linear Model (GLM) closely related to the Poisson family, where alpha is the dispersion parameter to deal with overdispersion. The difference between the two models is that a Poisson model estimates one parameter where the mean equals the variance. In contrast, the Negative Binomial distribution estimates the mean separate from the variance through the individual estimation of the dispersion parameter (=alpha). The negative binomial model significantly outperforms ordinary least-square regressions with log transformations for highly dispersed data (O'Hara & Kotze, 2010).

A nested model can be derived from a more complex model if some of the predictors of the complex model are 0. When the dispersion parameter alpha is 0 for a Negative Binomial regression, the variance is equal to the mean, which is the assumption of a Poisson Regression. Therefore, the Poisson model is a nested, simple version of the Negative Binomial model.

The Poisson regression model is first fitted to the data and compared to the model fit parameters of the Negative Binomial regression to confirm this model fits better with the data and overdispersion is actually present in the data. Before models can be equipped, independent variables must be standardised to deal with convergence and interpretation issues. Standardising and centring predictors that are not on the same scale is common to ensure the coefficients of variables with larger values are not overestimated by the model, especially when using Negative Binomial regression (Overgoor et al., 2022). The difference in variable scaling can be observed in Table 4 with the descriptive statistics, where variables like Colour Complexity have a mean of 22.02, and Irregularity of Object Arrangement has a mean value of 0.04. Variables are normalised with the use of z-scaling, where the variables are centred to a mean of 0 by subtracting the initial mean and scaled to have a standard deviation of 1 by dividing the scaled score by the standard deviation (Jayalakshmi & Santhakumaran, 2011). As an illustration, Colour Complexity has a mean of 22.02 and a standard deviation of 10.20. If the Colour Complexity score of a video is 30, the normalised value of colour complexity is (30-22.02)/10.20=0.78. The interpretation of the normalised coefficients will be discussed in section 5.2.

In the Poisson family, there are various ways of assessing how well the models fit the data. Two of the most used parameters to compare the fit of both nested and non-nested models are the Bayesian Information Criterion (BIC) and Alkaline Information Criterion (AIC). The AIC uses the log-likelihood of the models and considers the number of predictors to create a relative goodness of fit parameter with which Poisson models can be compared (Coxe et al., 2009). The BIC is calculated similarly to the AIC while accounting for the sample size. Log-likelihood measures the probability of observing the data given a specific model. It maps the natural logarithm of the likelihood that the estimated values match the observed data.

The AIC for the standard Poisson is much larger, as seen in Table 5, which means the Negative Binomial model better fits the overdispersed nature of the data.

The most used test to compare the goodness of fit (Log-likelihood) of two nested models is the Likelihood Ratio Test (LRT), which evaluates whether the inclusion of additional parameters in a more complex model significantly improves model fit compared to the simple model (Lewis et al., 2010; Neyman & Pearson, 1928). The test statistic is calculated as follows:

$$D=-2\times(\ln(\text{likelihood of simpler model})-\ln(\text{likelihood of more complex model}))$$

This D statistic follows the Chi-square distribution and the resulting p-value helps determine whether the additional complexity of a larger model is justified. A p-value smaller than 0.05 suggests that the added complexity is warranted due to a significantly higher log-likelihood.

Table 5 illustrates the model comparison between the Poisson model and the Negative Binomial model for all dependent variables based on the AIC value, log-likelihood and Likelihood Ratio Test to compare nested models. The LRT in R suggests that despite the addition of the dispersion parameter in the more complex negative binomial model, which was manually coded in R, the p-value is essentially equal to 0, as seen in Table 5. These low p-values for the LRT test corroborate the superior fit of the Negative Binomial model over the Poisson model through AIC and BIC assessment.

Table 5: Model fit assessment for a Poisson and Negative Binomial distribution

	Poisson AIC	Negative Binomial AIC	Poisson log-likelihood	Negative Binomial log-likelihood	LRT p-value
Views	2475434800	4309.4	-618858694	-2141.72	0.00
Likes	1538523	2537.2	-384624.65	-1255.61	0.00
Comments	202694	1874.7	-192312.33	-924.37	0.00

4.2.2 Multicollinearity Test

Finally, before model estimates can be interpreted, a check for multicollinearity is performed using the Variance Inflation Factor to check if independent variables are highly correlated (Nieuwenhuis, 2009). The Variance Inflation Factor indicates how much of a coefficient's variance is due to collinearity (Heij, 2004). A simple OLS model is fitted for the multicollinearity test using VIF, which is not interpreted but only used as input for the VIF function. The model shows a VIF score lower than 10 for all independent variables. All variables had a VIF of between 1 and 2.5, except for Asymmetry of Object Arrangement and Irregularity of Object Arrangement, which both had a score of 7.5 and 7.3, respectively, which is below the ultimate threshold of 10 but above 5, meaning it could be worrisome depending of the complexity of the model (Heij, 2004). A more appropriate regression model will be fitted to assess if these two multicollinearity scores are problematic so the impact of multicollinearity between the two Object Arrangement measures can be interpreted.

4.2.3 Dealing with Multicollinearity

The outputs of the full negative binomial regressions for all dependent variables confirm the signs of possible multicollinearity between Irregularity of Object Arrangement and Asymmetry of Object Arrangement. Considering both variables are calculated using the same feature congestion maps to outline object arrangement as described by Overgoor et al. (2022) based on the calculations introduced by Rosenholtz et al. (2007a) and Zhang et al. (2017), the multicollinearity likely stems from the omission of the feature congestion maps as a control variable in the current study. Due to the limited availability of Super Bowl content published on YouTube, the current paper cannot add the same complexity to the model as Overgoor et al. (2022) did. The feature congestion maps were left out of the current study to limit model complexity, which could lead to multicollinearity.

The standard error and p-value of the coefficient of Asymmetry of Object Arrangement are high when both object arrangement variables are in the Negative Binomial model. The inflated standard error disappears when one of the two Object Arrangement variables is omitted. To confirm multicollinearity, a correlation matrix (see Table 6) shows a 0.9 correlation between Asymmetry of OA and Irregularity of OA at $p<0.001$ significance level. The high VIF scores, inflated confidence intervals, and high correlation confirm multicollinearity.

Due to the high correlation between the two variables measuring object arrangement complexity, a decision is made to drop one of the variables to deal with multicollinearity instead. The study's objective is to explore the individual relationships between factors of visual complexity and the two object arrangement variables, which are so highly correlated that they do not warrant complex methods for dimensionality reduction. Asymmetry and Irregularity of Object Arrangement both seem to have explanatory power of engagement and are interpreted by humans with similar accuracy in the paper by Overgoor et al. (2022). Considering asymmetry is the only design complexity variable that did not follow a quadratic relationship in the study by Overgoor et al. (2022), and it has the highest VIF score, Asymmetry of OA is dropped.

4.3 Non-Linearity and Quadratic Testing

Three steps are followed to test for nonlinearity and (inverted) U-relationships, as outlined by Lind and Mehlum (2010) and Haans et al. (2015). From the research done by Haans et al. (2015), it seems most methodologies testing for the presence of quadratic relationships are based on many assumptions, causing false positives that are too easily confirmed to check for real (inverted) U-relationships. The papers outline three steps which will be followed to assess the presence of quadratic relationships.

1a. The quadratic model should fit the data better than the linear model while accounting for model complexity. Model AIC can be used to determine the model that best balances model complexity and fit.

1b. The linear predictor's coefficient should be positive, and the quadratic predictor's coefficient should be significant and negative, or vice versa.

2. To confirm that the (inverted) U-relationship is present, the slope should be positive and negative for the smallest and largest independent variable values or contrariwise. To check the slope at extreme values, the minimum (X_L) and maximum value (X_H) of the independent variable are used to calculate the slope for both the low and high end of the independent variable according to the formulas:

$$Slope\ at\ X_L = \beta_1 + 2 \times \beta_2 X_L$$

$$Slope\ at\ X_H = \beta_1 + 2 \times \beta_2 X_H$$

When the slope has been calculated, a variance-covariance matrix is used to calculate the standard error of the slope, which is then used to perform a t-test to see if the slope is significantly different from 0 when $p<0.05$. For an inverted U-relationship, the slope at X_L should be significantly positive, while the slope at X_H should be significantly negative. For a U relationship, the slope at X_L should be negative with $p< 0.05$ and the slope at X_H should be positive with $p<0.05$.

3. The turning point of the quadratic should be within the bounds of the data to ensure the observed curve is a natural relationship and not due to data limits. The turning point is calculated as follows, as described by Haans et al. (2015):

$$\chi_{turn} = \frac{-\beta_1}{(2-\beta_2)}$$

Whether the (inverted) U-relationship occurs naturally in the data can be tested by estimating the turning point's 95-percent confidence interval (CI). If the CI is within the data range, one can be reasonably sure a natural U-relationship exists. If the lower (CI_{lower}) or upper bound (CI_{upper}) is outside the data range, one-half of the curve may only be revealed by the data. The 95% confidence interval is estimated using the delta method, which is intuitive and intuitive for confidence interval estimations (Hole, 2007; Xu & Long, 2005). The delta method uses the standard error of the estimators β_1 and β_2 to estimate the confidence interval using linear methods. The only issue is that the delta method could cause a bias in small samples. Still, this risk is a mere sidenote considering the extensiveness of the additional steps taken to test for (inverted) U-relationships and the additional complexity of using a bootstrapping mechanism.

4.4 Final Regression Models

Finally, the Negative Binomial model, without Asymmetry of OA, will be fitted using the normalised predictors, creating the following models:

$$log(Views) = \beta_{1,0} + \beta_{1,1} Colour\ Complexity + \beta_{1,2} Edge\ Density + \beta_{1,3} Luminance\ Entropy +$$
$$+ \beta_{1,4} Object\ count + \beta_{1,5}\ Object\ Arrangement\ Irregularity + \beta_{1,6} Visual\ Variety +$$
$$\beta_{1,7} Subscribers + \beta_{1,8}\ Duration + \beta_{1,9} Time + \beta_{1,10} Status$$

$$log(Likes) = \beta_{2,0} + \beta_{2,1}X_1 + \dots + \beta_{j,i}X_i$$

$$log(Comments) = \beta_{3,0} + \beta_{3,1}X_1 + \dots + \beta_{j,i}X_i$$

The formulas in section 3.5 will be used for non-linearity testing with a log-link transformation that the Negative Binomial model uses, and without the error term in OLS models. Therefore, all predicted variables in the non-linear models are changed to "log(DV) = …"

This testing method for non-linearity means 36 models will be fitted just for non-linearity and (inverted) U-relationship testing. The exclusion of Asymmetry of OA due to multicollinearity means this variable will also be excluded from non-linearity testing.

5. Results

In this section, the individual effects or correlations between variables in the study are presented and examined in section 5.1 before the full linear and non-linear regression models are presented and interpreted in section 5.2. Section 5.3 goes deeper into the limitations of the results and the study.

5.1 Individual Effects

Table 6 shows the independent correlation coefficients between all variables in the study, calculated using the Pearson Correlation coefficient in R Studio to determine the underlying relationship of variables in the study. The dependent variables show a strong positive correlation between likes and comments at a 0.01 significance level, while the correlation between comments or likes, and views is not strong nor significant.

Table 6: Correlations between variables in the study

	1	2	3	4	5	6	7	8	9	10	11	12	13
1 Views													
2 Likes	.07												
3 Comments	.20*	.74***											
4 Colour Complexity	.09	-.12	-.16										
5 Edge Density	.33***	-.09	-.05	.29***									
6 Luminance Entropy	.13	-.18*	-.13	.61***	.31***								
7 Asymmetry of OA	.11	-.18*	-.17*	.19*	.52***	.44***							
8 Irregularity of OA	.08	-.15	-.14	-.03	.50***	.30***	.90***						
9 Unique Objects	.15	-.21*	-.15	.55***	.13	.49***	.13	-.01					
10 Visual Variety	.12	.07	-.02	.44***	.20*	.28**	.25**	.14	.17				
11 Subscribers	-.06	.72***	.44***	-.21*	-.12	-.27**	-.24**	-.18*	-.34***	-.03			
12 Time	-.07	.19*	.04	-.06	-.12	.03	-.10	-.04	-.11	-.05	.21*		
13 Length	-.01	.06	.25**	.09	-.14	-.04	-.25**	-.27**	.06	.04	.08	-.16	
14 Status	-0.06	.02	.03	-.15	.04	-.16	.04	.07	-.33***	.15	.14	.33***	-.00

*Note: significance levels are * for p< 0.1, ** for p<0.05, *** for p<0.01*

This finding aligns with the division of levels of engagement by Khan (2017), where views are described as passive engagement while liking and commenting show a more active form of engagement. Some significantly low, moderate and strong correlations are observed between measures of Visual Complexity. Colour Complexity has a moderately positive correlation at a 0.01 significance level with Luminance Entropy, the Number of Unique Objects and Visual Variety.

Edge Density has a moderate positive correlation with Luminance Entropy, Asymmetry of OA and Irregularity of OA at a 0.01 significance level. A high level of Edge Density means a frame has complex textures and patterns, often caused by a considerable variation in brightness. The correlation between Edge Density and the object arrangement variables is easily explained by the number of textures and general outlines throughout the frame, which directly correlates with the number of object outlines and feature congestion maps used to calculate object arrangement. Luminance Entropy has a moderate to low positive correlation with the Object Arrangement variables and the Number of Unique Objects at a 0.01 significance level. When Object Arrangement is complex, scene structure, surfaces and orientations widely vary, which goes hand in hand with a high entropy of light. Many objects in a frame are accompanied by many contours and shadows, which is related to the variability and entropy of light throughout a frame.

Irregularity of OA and Asymmetry of OA are highly positively correlated at a 0.01 significance level, indicating a problematic multicollinearity issue, as discussed in section 4.2. As for the correlations of Visual Variety and other variables in the study, the correlations with Colour Complexity ($p<0.01$), Luminance Entropy ($p<0.05$) and Asymmetry of OA ($p<0.05$) are the only significant correlations which range from low to moderately positive. Videos high in Visual Variety are highly dynamic, which indicates a dynamic arrangement of objects within frames due to continuous changes in scene composition. The correlations also align with the complex use of colours and a varying level of light throughout frames. As for the relevant correlations related to control variables, subscribers are moderately to highly correlated to likes and comments at a 0.01 significance level but not to views. This correlation between subscribers and engagement metrics indicates that channels high in subscribers tend to get more likes and comments but not more views. Subscribers are more significantly related to more active forms of engagement, which could be due to the brand having a tight-knit brand community on YouTube (Brodie et al., 2013).

The significant moderate relationships between visual variables indicate an interconnectedness of the visual information studied in this paper. Visual Variety is less interconnected than other Visual Complexity variables, highlighting its independent role in measuring the dynamic nature of video content.

5.2 Explaining Engagement Through Regression Models

5.2.1 Main Linear Regression Models

Table 7 shows the full regression results for visual video features per online engagement variable with only linear predictors. The coefficient β indicates the relationship between the predictor and the online engagement metric. The value within brackets indicates the standard error of the coefficients, and the asterisks (*) indicate the significance level of the coefficient. The AIC, BIC, log-likelihood and Pseudo R^2 indicate model fit, while α indicates the dispersion parameter of the engagement metric, as discussed in section 4.2.1.

The interpretation of coefficients is not straightforward because a Negative Binomial regression with z-scaled predictors is used. The effect of a predictor on a measure of online engagement should be exponentiated before it can be interpreted since the Negative Binomial model uses a log transformation to deal with overdispersion. The goodness of fit measures log-likelihood, AIC, BIC and Nagelkerke Pseudo R^2 measure how well the data fits the model. The log-likelihood, AIC and BIC are frequently used goodness of fit measures for Negative Binomial models (Coxe et al., 2009; Lewis et al., 2010; Neyman & Pearson, 1928). The Nagelkerke Pseudo R^2 measures how much of the variance in the dependent variable is explained by the predictors in the model (Nagelkerke, 1991). The most impactful significant visual variables and their effects on online engagement will be discussed, as they are most compelling for marketers.

Table 7: Negative Binomial Regression Results with all predictors

	Log(Views)		Log(Likes)		Log(Comments)	
	No Controls	Full Model	No Controls	Full Model	No Controls	Full Model
	β (S.E.)	β (S.E.)	β (S.E.)	β (S.E.)	β (S.E.)	β (S.E.)
Colour Complexity	0.51*** (0.18)	0.53*** (0.18)	0.07 (0.15)	0.11 (0.13)	0.22 (0.18)	0.39** (0.17)
Edge Density	0.16 (0.21)	0.36* (0.20)	-0.19 (0.17)	-0.02 (0.15)	0.01 (0.21)	0.42** (0.20)
Luminance Entropy	-0.14 (0.23)	-0.29 (0.22)	-0.38* (0.20)	-0.13 (0.17)	-0.71*** (0.23)	-1.01*** (0.22)
Unique Objects	0.31* (0.18)	0.31 (0.19)	-0.50*** (0.15)	-0.12 (0.14)	0.17 (0.18)	0.35* (0.19)
Irregularity of OA	-0.54*** (0.19)	-0.44** (0.19)	-0.34** (0.16)	0.08 (0.14)	-0.45** (0.19)	0.07 (0.19)
Visual Variety	0.42*** (0.16)	0.48*** (0.16)	0.51*** (0.14)	0.43*** (0.12)	0.14 (0.17)	0.14 (0.16)
log(Subscribers)		0.28 (0.29)		1.43*** (0.22)		1.19*** (0.29)
Time		-0.34** (0.16)		0.21* (0.12)		0.14 (0.16)
Length		0.19 (0.15)		0.52*** (0.11)		0.91*** (0.15)
Status		-1.18** (0.47)		-0.77** (0.35)		-0.09 (0.46)
(Intercept)	15.75*** (0.14)	16.75*** (0.43)	8.95*** (0.12)	9.56*** (0.33)	6.58*** (0.15)	6.68*** (0.43)
Log-likelihood	-2143.83	-2140.02	-1284.40	-1256.21	-933.42	-923.07
AIC	4306.4	4304	2584.8	2536.4	1882.8	1870.1
BIC	4341.27	4338.64	2607.87	2571.01	1905.89	1904.73
Observations	N = 132	N = 132	N = 132	N = 132	N = 132	N = 132
Nagelkerke Pseudo R^2	0.14	0.19	0.28	0.53	0.14	0.27
α (Dispersion Parameter)	2.69	2.59	2.01	1.49	2.85	2.54

*Note: significance levels are * for p< 0.1, ** for p<0.05, *** for p<0.01*

Coefficient interpretation

A coefficient of 0.53 for the normalised scale of Colour Complexity means that an increase of one standard deviation in Colour Complexity, which is 10.20, as shown in Table 4, means that views are multiplied by (exp(0.53) = 1.7. This means that overall, a video with a Colour Complexity score of 10.20, or one standard deviation, higher is expected to have 70% more views.

Visual Features that have the largest positive influence on each of the engagement metrics

Colour Complexity has a highly significant positive effect on two online engagement metrics: views and comments. Colour Complexity has a normalised β coefficient of 0.53 related to log(Views), where $p<0.01$. Each coefficient's exponent is calculated, resulting in a β coefficient of 1.70 to make it more interpretable. The exponentiated coefficient indicates a significant positive relationship between Views and Colour Complexity, where for every standard deviation increase in Colour Complexity, which is 10.20, as shown in Table 4, a 70% increase in views is observed. The normalised β coefficient of Colour Complexity is 0.39 when it comes to its relationship to Comments at a significance level of $p<0.01$. After exponentiating the coefficient, we can conclude that for every standard deviation increase in Colour Complexity, Comments increase by 48% ($\exp(\beta) = 1.48$). Visual Variety also has significant positive effects on two of the three measures of online engagement. Visual Variety seems to have a significant positive relationship to both log(views) and log(likes), where the normalised β coefficients are equal to 0.48 and 0.43, respectively, both at $p<0.01$ significance level. When exponentiating the coefficients of Visual Variety, a standard deviation increase of 0.05, as shown in Table 4, implies a 61% increase ($\exp(\beta)=1.61$) in Views and a 54% increase ($\exp(\beta)=1.54$) in Likes.

Edge Density only has a positive relationship with one of the three measures of online engagement: Comments. The relationship between Views and Edge Density, however, is very close to significant, where $p<0.1$ but $p>0.05$. The normalised coefficient of Edge Density regarding its relationship to Comments is 0.42, with the respective p-value being smaller than 0.05 and thus positively significant. The exponentiated coefficient is 1.52, which means that for every standard deviation increase in Edge Density, which is 0.01, as shown in Table 4, the number of Comments on a video is 52% percent higher.

Visual Features that have the largest negative influence on each of the engagement metrics

All visual variables with a significant negative coefficient for their relation to a measure of online engagement only do so with one of the three measures. Luminance Entropy has a significant negative relationship with Comments, while Irregularity of OA has a significant negative relationship with Views. The normalised coefficient of Luminance Entropy at $p<0.01$ significance level regarding its relationship to log(Views) is -1.01, indicating a very negative relationship. The exponentiated coefficient is 0.36, meaning a one standard deviation increase in Luminance Entropy, which is 0.64, as shown in Table 4, implies a 64% decrease in Comments.

Finally, the last significant negative relationship between a visual feature and one of the online engagement metrics is the relationship between the Irregularity of OA and log(Views). The normalised coefficient of this relationship is -0.44, where $p<0.05$. When exponentiated, this significant negative relationship between the Irregularity of OA and Views warrants a normalised β coefficient of 0.64. This normalised exponentiated relationship implies that for every standard deviation increase in Irregularity of Object Arrangement, which is 0.01, as shown in Table 4, the number of Views on a video is estimated to be 36% lower.

Non-significant results

Some of the coefficients estimated in Table 7 are non-significant for the relationships between visual features and online engagement. Some are marginally statistically insignificant, like the relationship between Edge Density and log(views) or the Number of Unique Objects and log(comments), as the p values are smaller than 0.1 but larger than the cutoff of 0.05. These relationships are strong, but there is insufficient statistical evidence for the observed effects to be considered statistically significant. Other estimates, like most estimates for the online engagement metric Likes, have standard errors larger than the coefficients, indicating a large variability of effects. One reason could be that the dispersion parameter is much lower for Likes than for the models estimating Views and Comments. While a lower dispersion parameter indicates more stable estimates, this could suggest that the effects of visual variables on Likes are more subtle and harder to detect, increasing the need for a larger sample size. Considering the complexity and sample size of this study (n=132), it is likely that more subtle effects are not estimated accurately, leading to higher S.E. and lower p-values.

Important Control Variables

As shown in Table 7, the control variables "Length of the video in seconds" and "Subscribers" significantly influence Likes and Comments. Especially the estimates of Subscribers related to log(Likes) ($\beta = 1.43$, $p<0.01$) and log(Comments) ($\beta = 1.19$, $p<0.01$) indicate that the number of Subscribers on an account strongly influences the active engagement on a video from this channel (Brodie et al., 2011). Length has a significant positive relationship with both log(Likes) ($\beta=0.52$, $p<0.01$) and log(Comments) ($\beta = 0.91$, $p<0.01$). Considering how short and expensive Super Bowl advertisements are, companies that can afford more air time might be larger and more relevant, in general, or can spend more on things like celebrity features or production. Status, on the other hand, indicating whether a video is private or not, has significant negative coefficients for both log(Views) ($\beta=-1.18$, $p<0.01$) and log(Likes) ($\beta=-0.77$, $p<0.01$). These coefficients are easily explained by either the fact that the videos are no longer in the YouTube algorithm and can only be visited by link or that these videos were deleted due to them performing poorly on the platform. A strong case for either of these explanations can be made because causality cannot be determined from a regression analysis.

Model evaluations

As explained in section 4.2, the AIC and BIC are goodness of fit measures for non-nested models based on the maximalisation of the log-likelihood. The log-likelihood is the basis of other goodness of fit estimators for likelihood maximalisation models. Like AIC and BIC, log-likelihood should be as close to 0 as possible to warrant the best model fit. The log-likelihood estimates the chance the data would occur under the current model. The Nagelkerke Pseudo R^2 is similar to the regular R^2 used in ordinary least squares (OLS) regression models but for models using maximum likelihood estimations, like a Negative Binomial model. The Pseudo R^2 is a measure of explanatory power rather than model fit. The Pseudo R^2 shows how much of the variance in the dependent variables is explained by the model, whereas log-likelihood, AIC and BIC measure how well the model fits the data.

The robustness checks in the models showcase that the control variables significantly improve model fit. While the visual features already account for a large portion of the model fit, the control variables do warrant more complexity in the model since the model fits significantly better, considering the lower AIC and BIC scores. The models with control variables also have significantly higher explanatory power, although Pseudo R^2 does not control for the number of estimators in the model like the Adjusted R^2 does in OLS regression.

The model for comments fits the data best. This model has the lowest AIC and BIC values, indicating it balances model fit and complexity most effectively. While this model fits best with the data, the model for Likes explains a significant portion of the variance in likes with a Nagelkerke Pseudo R^2 of 0.53. The model for Views fits the data worst, as it has the highest AIC and BIC values. The predictors are less effective at explaining the variability in views, possibly due to additional, unmeasured factors influencing views. When assessing the log-likelihood, BIC and AIC estimations, there is a clear division between the models predicting Views versus those estimating Likes and Comments. The reason for this estimation of AIC, BIC and log-liklihood parameters could be due to the dispersion of data. As seen in Figure 11, the model estimating views has the highest dispersion, while the number of likes and comments is much lower. The lower dispersion in comments and likes could simply explain the lower model fit parameters, while the Pseudo R^2 does clearly show the degree in which the model explains the variance in the data. This measure of explanatory power is high for Likes, medium for comments and low for Views.

Following the Visual Stimulation theories by Berlyne (1970) and Hebb (1955), Likes and Comments more accurately reflect the enjoyment of the content and active participation. In contrast, Views are influenced much more by the YouTube algorithm and other extraneous factors. The results from the regression do, however, align well with the findings in existing literature despite the small sample size and, therefore, limited complexity of the model. Despite the sample size, various significant large effects indicate strong relationships between some of the visual features and online engagement. Inferences from the results should be made cautiously, which will be discussed in section 5.3. As described in Section 2.3, there is strong evidence for non-linearity concerning the relationships between Visual Complexity, Visual Variety and engagement. The following section will explore these non-linear relationships and utilise rigorous testing methodologies to find (inverted) U-relationships.

5.2.2 Exploring Non-Linearity

This section uses the three steps outlined in section 4.3 to test for non-linear and (inverted) U-relationships. As outlined in sections 3.5, 4.3 and 4.4, many small models will be used to explore non-linear relationships within the data based on the evidence found in existing literature outlined in section 2.3. Table 8 is based on a linear and a quadratic model for each combination of visual features and online engagement metrics. Considering the study has six independent variables, given that Asymmetry of OA has been eliminated from the analysis due to multicollinearity, and three measures of online engagement, the model below comprises 36 small models. Even though control variables are used in the models of Table 8, their coefficients are irrelevant to the outcome of non-linearity testing; therefore, they are not presented. This section solely tests for the presence of quadratic relationships but does not try to estimate coefficients and relate those to their absolute estimated effect on each online engagement metric.

Table 8 illustrates the regression results from the small models for non-linearity testing. The coefficients, standard errors and significance levels are presented as in Table 7. Only the AIC is presented in Table 8, as this is the only model fit parameter needed to determine if the data fits the linear or non-linear model better. Additionally, it is important to note that each combination of the dependent and independent variable is based on a linear and non-linear model, meaning this table comprises 18 linear models and 18 quadratic models.

Step 1: Non-linearity assessment, Model coefficients and significance

Step 1a is taken for each relationship by comparing the AIC values of each independent variable's linear and quadratic models using the (inverted) U assessment method from section 4.3. Only the quadratic model of Irregularity of OA has a significantly better fit to the data than the linear counterpart for the dependent variable views. The AIC of the quadratic model is 4309.8, and the AIC of the linear model is 4312.3, meaning the quadratic model fits the data better while accounting for the extra quadratic term. Colour Complexity and Visual Variety also have more of a quadratic relationship to the dependent variable likes than a linear relationship.

The quadratic model indicating the relationship between Likes and Colour Complexity has an AIC value of 2535.4 for the quadratic model, while the linear model has an AIC of 2535.8. The model estimating the relationship between Likes and Visual Variety has an AIC of 2528.1 for the quadratic variant and an AIC of 2530.3 for the linear model. For the models with comments as dependent variable, all variables except Edge Density and Irregularity of OA have lower AIC values for the quadratic models than their linear counterparts. The model fit for Colour Complexity (Linear AIC: 1869.4, Quadratic AIC: 1851.3), Luminance Entropy (Linear AIC: 1860.2, Quadratic AIC: 1859.3), Unique Objects (Linear AIC: 1863.4, Quadratic AIC: 1861.7) and Visual Variety (Linear AIC: 1863.3, Quadratic AIC: 1859.5) all favour more of a quadratic relationship.

Now that the independent-dependent variable pairs for which the non-linear model warrants a better fit are clear, the coefficients, p-values and type of relationship can be interpreted for step 1b, outlined in section 4.3. The only non-linear relationship with views as a dependent variable is the relationship between views and the Irregularity of Object Arrangement. According to Table 8, Irregularity of OA has a positive linear coefficient ($\beta = 0.68$) and a negative quadratic coefficient ($\beta = -0.14$) with a p-value smaller than 0.05. This means that the signs of the coefficients and the significance level indicate a possible inverted U-relationship.

As for the models with likes as a dependent variable, Colour Complexity and Visual Variety were the only non-linear models that fit the data well. The linear coefficient of Colour Complexity is slightly positive ($\beta = 0.03$). However, the quadratic coefficient is both non-significant and non-negative ($\beta = 0.08$), which means that although the relationship is non-linear, no significant quadratic relationship is present as $p > .05$. The linear coefficient of Visual Variety is positive ($\beta = 0.35$), and the quadratic coefficient is negative ($\beta = -0.19$). While the p-value of the negative quadratic coefficient is smaller than 0.1, it is larger than 0.05. Hence, there is no significant inverted U-relationship.

Table 8: The independent linear and quadratic models for each of the six independent variables per dependent variable

	Views				Likes				Comments			
	Linear Model		Quadratic Model		Linear Model		Quadratic Model		Linear Model		Quadratic Model	
	β (S.E.)	AIC	β (S.E.)	AIC	β (S.E.)	AIC	β (S.E.)	AIC	β (S.E.)	AIC	β (S.E.)	AIC
Colour Complexity	0.45** (0.15)	4300.4	0.37* (0.18)	4302.9	0.16 (0.11)	2535.8	0.03 (0.14)	2535.4	-0.04 (0.15)	1869.4	-0.07 (0.17)	1851.3
Colour Complexity^2			0.05 (0.07)	4302.9			0.08 (0.05)	2535.4			0.18** (0.07)	1851.3
Edge Density	0.43** (0.15)	4308	0.52** (0.16)	4308.9	0.05 (0.12)	2538.5	0.01 (0.12)	2539.4	0.05 (0.15)	1863.4	0.08 (0.15)	1865.2
Edge Density^2			0.09 (0.08)	4308.9			-0.06 (0.06)	2539.4			0.04 (0.07)	1865.2
Luminance Entropy	0.35* (0.15)	4312	0.35 (0.23)	4314	0.11 (0.12)	2537.9	0.05 (0.17)	2539.7	-0.31* (0.14)	1860.2	-0.54** (0.21)	1859.3
Luminance Entropy^2			0.00 (0.07)	4314			-0.02 (0.05)	2539.7			-0.11. (0.06)	1859.3
Unique Objects	0.54** (0.17)	4306.9	0.53** (0.18)	4308.8	-0.06 (0.13)	2538.5	-0.03 (0.14)	2539.8	-0.05 (0.16)	1863.4	-0.12 (0.17)	1861.7
Unique Objects^2			0.03 (0.12)	4308.8			-0.08 (0.14)	2539.8			0.22* (0.11)	1861.7
Irregularity of OA	0.48** (0.16)	4312.3	0.68** (0.24)	4309.8	0.18 (0.12)	2537.1	0.31. (0.18)	2538	0.36** (0.15)	1859.6	0.44. (0.23)	1861.4
Irregularity of OA^2			-0.14*** (0.04)	4309.8			-0.04 (0.03)	2538			-0.04 (0.04)	1861.4
Visual Variety	0.51*** (0.15)	4307.3	0.47** (0.15)	4308.2	0.37*** (0.11)	2530.3	0.35** (0.11)	2528.1	-0.08 (0.14)	1863.3	-0.05 (0.14)	1859.5
Visual Variety^2			-0.13 (0.12)	4308.2			-0.19* (0.09)	2528.1			-0.29* (0.11)	1859.5

*Note: Significance levels are * for p< 0.1, ** for p<0.05, *** for p<0.01. There are six independent variables. For each of the independent variables, a linear and quadratic model for each of the three dependent variables is shown with the respective AIC. Control variables are present in the models, but intercepts and control coefficients are not relevant for non-linearity analysis

Finally, the coefficients and significance levels of the quadratic models predicting the number of comments using Colour Complexity, Luminance Entropy, Unique Objects and Visual Variety as individual predictors are interpreted. Colour Complexity has a negative linear predictor (β= -0.07) and a significant positive quadratic coefficient (β= 0.18) with a p-value smaller than 0.05, which indicates a U-relationship instead of an inverted U-relationship where the coefficient is negative for low values, comes to a turning point and starts rising at higher values of Colour Complexity. Luminance Entropy has a negative linear predictor (β = -0.54) and a less negative quadratic predictor (β = -0.11) with a p-value larger than 0.1, meaning it is insignificant. Although the quadratic relationship between comments and the Number of Unique Objects has a negative linear coefficient (β = -0.12) and a positive quadratic coefficient (β = 0.22), the p-value is slightly higher than the 0.05 threshold, meaning there is a non-significant U-relationship present. Visual Variety does not have a quadratic relationship, considering both the linear coefficient (β = -0.05) and the quadratic coefficient (β = -0.29) are negative and insignificant.

Step 2: Slope test at extreme values

As for step 2, the minimum and maximum values of the independent variable need to be added into the slope formulas. The formulas are as follows:

$$\beta_1 + 2 \times \beta_2 X_L \text{ and } \beta_1 + 2 \times \beta_2 X_H$$

Using the minimum and maximum normalised values of Irregularity of Object Arrangement and the model estimating the relationship between Views and Irregularity of OA, the slope can be assessed at extreme values of the independent variable. The slope is 0.98 (p=0.00) at X_L and -1.53 (p=0.00) at X_H. These slopes and significance levels indicate strong evidence that an inverted U-relationship is present due to the initially significantly positive and the later significantly negative slope.

The model uses the same formula to estimate the relationship between Colour Complexity and comments. Due to the suspected U-relationship, the slope should be significantly negative at low and significantly positive at high values. At X_L the slope is -0.83 ($p<0.05$), and the slope is 1.71 (p=0.00) at X_H. These slopes and significance levels indicate a U-relationship due to the initial significantly negative and positive slope for higher values of Colour Complexity, indicating a negative trend before coming to a stop, or minimum, before increasing again.

Step 3: Turning Point and 95% Turning Point Confidence Interval

After steps 1 and 2, only two possible quadratic relationships remain between the dependent and independent variables. The 95% z-scaled confidence interval estimates are now used to test whether the whole turning point CI fits into the data range for this independent variable.

Table 9 illustrates the turning point and the normalised estimated 95% confidence intervals of the turning point for each engagement metric per column. The rows represent the relevant visual features, so the cells contain the turning point and its CI per relevant relationship. The turning points are calculated using the formula outlined in section 4.3, step 3, while the confidence interval is calculated using the Delta method.

Table 9: Relevant normalised turning points and confidence interval estimations

	Views		Likes		Comments	
	Turning point	95% CI (Lower: Upper)	Turning point	95% CI (Lower: Upper)	Turning point	95% CI (Lower: Upper)
Colour Complexity					0.21	-0.67: 1.09
Irregularity of OA	2.50	1.38: 3.61				

**Note: These are the Z-transformed values and still need to be re-calculated into absolute values*

The z-transformed confidence intervals are recalculated to the original values and compared to the minimum and maximum observed values in the descriptive statistics in Table 4.

The relationship between OA irregularity and Views has demonstrated a significant quadratic relationship. The z-transformed turning point of OA irregularity is 2.5, as shown in Table 9, and the absolute mean and standard deviation in the descriptive statistics (Table 4) are 0.035 and 0.005, respectively.

The following formula can be used to re-calculate absolute values:

$$Z - transformed\ score\ = \frac{(absolute\ score - absolute\ mean\ score)}{absolute\ standard\ deviation}$$

$$2.5\ = \frac{(absolute\ score - 0.035)}{0.05}$$

$$2.5\ \times\ 0.05\ =\ 0.0125$$

$$0.0125\ =\ (absolute\ score\ -\ 0.035)$$

$$absolute\ score\ =\ 0.0475$$

The confidence intervals of the turning point when plotting the relationship between Views and Irregularity of OA leads to an absolute turning point of 0.0475. Using the same formula for recalculation, the $95\% - CI_{lower}$ is 0.042 and $95\% - CI_{upper}$ is 0.053. When these values are compared to the descriptive statistics of Irregularity of OA in Table 4, we can see that the minimum and maximum values are 0.033 and 0.082. The confidence interval of the turning point fits well within the possible data points, and the inverted U-relationship occurs naturally and is not due to data boundaries.

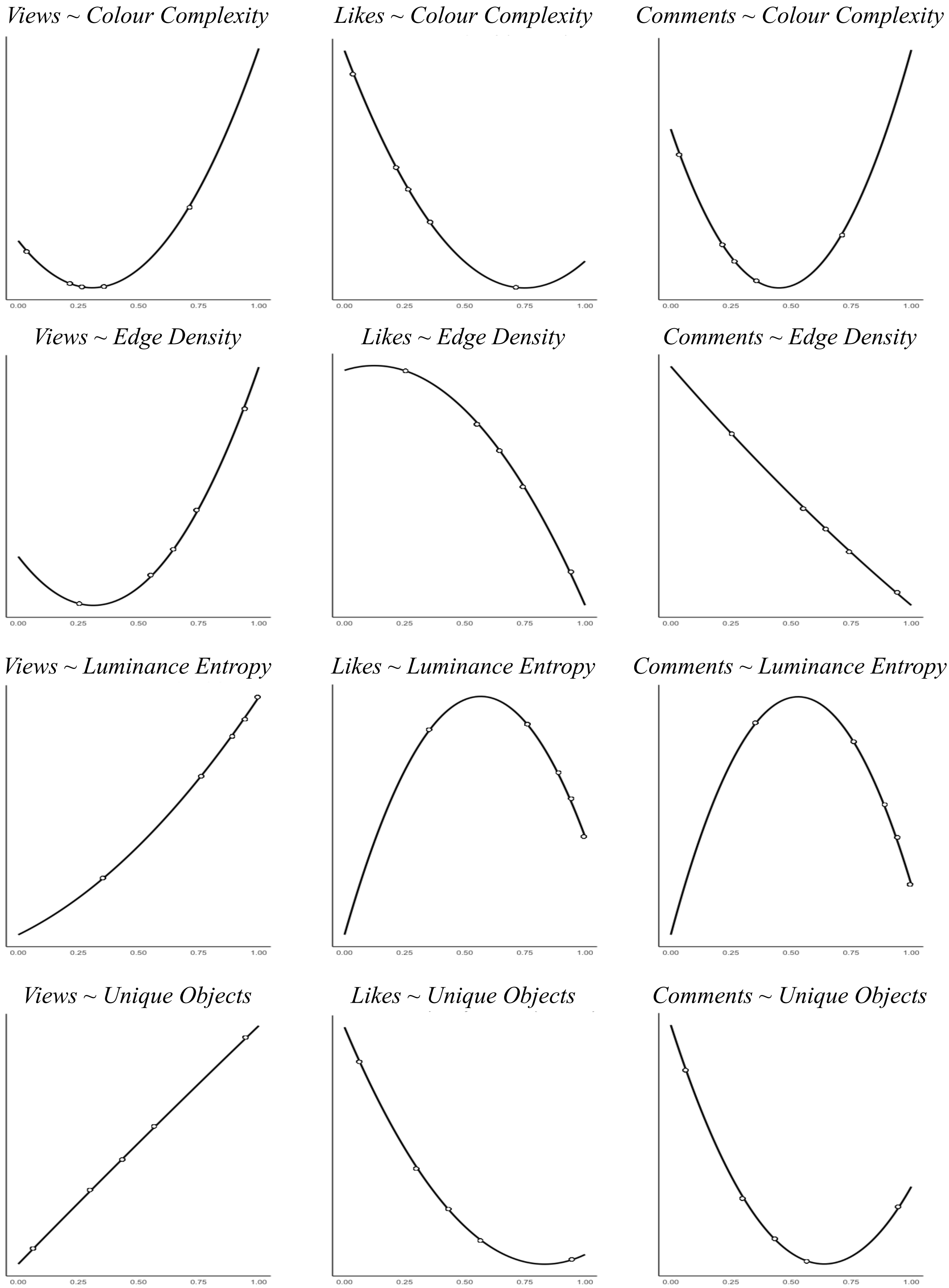
When assessing the other quadratic relationship between Comments and Colour Complexity, the same method is used as for Irregularity of OA. The normalised turning point 0.21 is 24.16 when re-calculated to an absolute value. The turning point's absolute 95% confidence interval ranges from CI_{lower}=15.19 to CI_{upper}=31.14. Colour Complexity's absolute minimum and maximum values are 0.26 and 73.60, respectively, as shown in the descriptive statistics in Table 4. Therefore, the U-relationship between Colour Complexity and Comments occurs naturally and is not created by the data range.

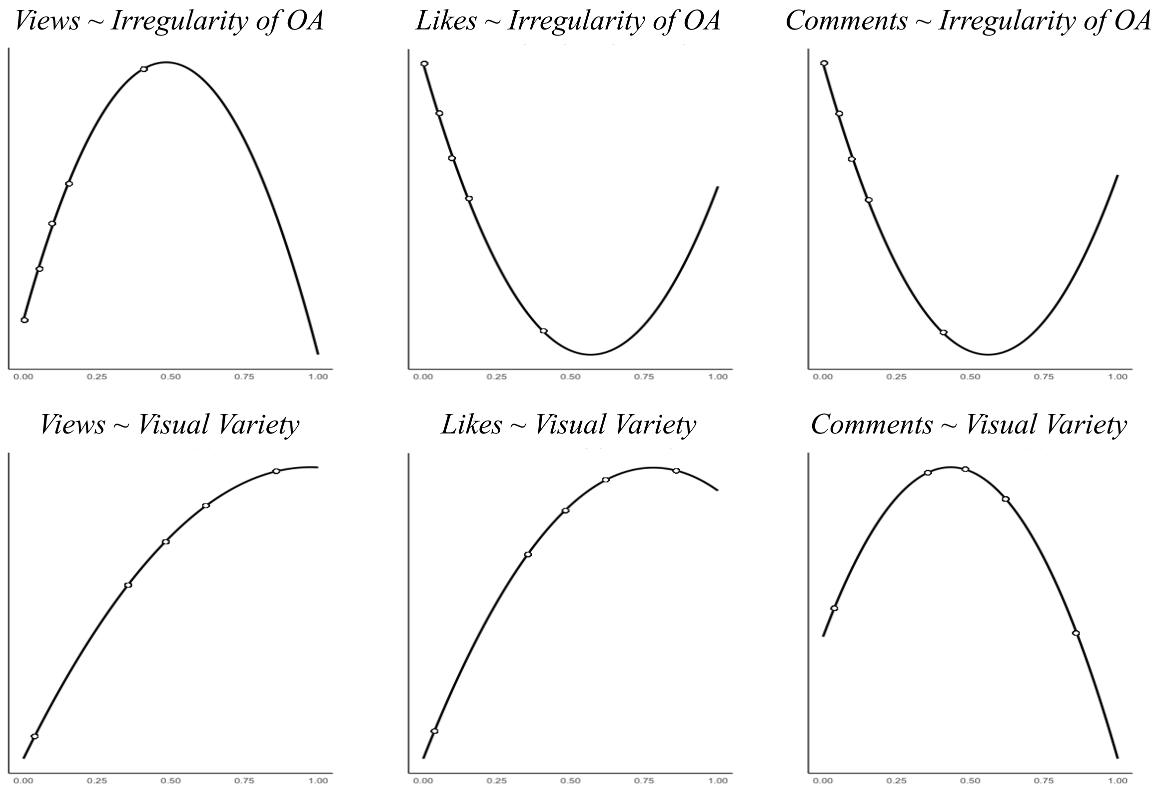
The results from steps 1a, 1b, 2 and 3 indicate that 2 U-type relationships are found. One is between Views and Irregularity of OA, where an inverted U-relationship was observed, and the other is between Comments and Colour Complexity, where a U-relationship is observed.

Figure 12 illustrates the non-linear smoothed relationships of each visual variable with the engagement metrics based on the small model estimates in Table 8. Each column represents the relationships for one dependent variable, and each row represents the relationships of an independent variable with each of the three dependent variables. The Y-axes of all graphs represent the value of the dependent variable, which can be views, likes, or comments, depending on the column in the table. The X-axes showcase the min-max scaled values of the independent variables from (0-1). This scaling is only used for the purpose of visualisation. The white dots on the plot indicate the 1st, 25th, 50th, 75th and 99th percentile values of the independent variable to show the full range of the relationship between all levels of the visual variable and the online engagement metric.

As outlined in Figure 12, if significance levels had been lower, many more full and half (inverted) U-relationships could have been found in the data, as indicated by the shape of the smoothed plots. This means there is a significant likelihood that more non-linear relationships could be present in the data if the sample size had been larger.

Figure 12: Non-linear relationships of all Visual Variables and Engagement





*Non-linear smoothed relationship between each independent variable on the x-axis and views, likes and comments on the y-axis, respectively. The independent variables are on a 0 to 1 min-max scale purely for visualisation. The dots on the line indicate the 1st, 25th, 50th, 75th and 99th percentile, respectively.

5.3 Limitations and Further Research

This paper aims to use the operationalised Visual Complexity measures by Overgoor et al. (2022) and the Visual Variety measures by Li et al. (2019) to showcase the use of objective static and dynamic visual measures to perform video analytics and measure the effects on online engagement.

5.3.1 Omitted Variables

As mentioned in the literature review, factors such as the popularity of the content creator, the traction of a brand and the video title may have influenced the engagement metrics. Additionally, while this study primarily examines visual complexity and variety, other elements like audio quality, video sentiment and the overall production quality of the video may also play significant roles in driving engagement but were not the primary focus (Hautz et al., 2014). Other potentially relevant factors like humour and celebrity appearances are also not accounted for due to the already complex nature of video analytics, the limitation of time and resources and the added complexity.

Furthermore, the personal branding of the creators of the advertisements can significantly influence user engagement, as outlined by Overgoor et al. (2022). Well-known or highly regarded brands are more likely to attract views, likes, and comments due to their established audience. For example, some automotive brands might use visual features that align with their luxurious brand image. At the same time, other companies might use visuals to highlight a sense of safety and reliability. Despite this variable's importance, the sample's nature and size do not allow controlling for individual brand characteristics, as it would cause significant model overfitting.

The industry or content category of the video also affects engagement levels. Different types of content generate varying levels of interest and interaction from viewers. For example, entertainment videos typically garner more engagement than educational or technical content. The nature of the content influences how viewers interact with the video, with some categories naturally encouraging more discussion and interaction. The tech sector uses more minimalistic advertising, while movies and consumer goods brands use more expressive colours, details and irregularity.

Another critical aspect to consider is that online engagement with a video changes over time, as discussed in the literature review in section 2.1.2. This study captures engagement metrics at a single point in time, which may only partially represent the dynamic nature of how a video accumulates views, likes, and comments over its lifecycle. Over time, videos can gain more engagement as they are shared across different platforms and reach broader audiences. This limitation was minimised by adding the variable Time into the study, which measures the time since a video was posted; however, a time-series analysis would show more factors related to the initial buzz created by the ad and the lifespan of the content on YouTube. The collected data spans videos posted over several years, meaning that the timing of data collection could affect the results. However, research suggests that a significant portion of engagement (views, likes, comments) occurs shortly after the video is posted, especially when the buzz is high, like for Super Bowl advertisements, as shown in Figure 3. Given the rapid rise in exposure, the timing should not critically impact the engagement metrics, which can be seen in the regression outputs in Table 7, as Super Bowl videos mature quickly due to the initially created buzz. Nevertheless, longer-term exposure could result in more views, likes and comments.

5.3.2 Sample Size

One of the main limitations of the current study is the number of videos analysed. The choice to analyse brand-generated content makes creating one or multiple search queries to collect data through the YouTube API challenging. Despite YouTube being the most prominent online video-sharing platform and one of the websites with the most traffic on the interwebs, this limitation makes it so a specific set of videos needs to be chosen. In this case, analysing Super Bowl videos limits availability since every year, about 50-80 advertisements are published online, with a large set of them being deleted from the platform within a year or two. Even the additional data gathered from the four previous renditions of the Super Bowl does not garner a large dataset. Especially when fitting models with many predictors like quadratic or cubic terms to overdispersed data, the general rule of thumb to have 10-20 observations per predictor might not hold up. The sample size limitation was especially detrimental to the significance levels of the inverted U-relationships. As shown in Table 8 and Figure 12, there is a possibility of uncovering more quadratic relationships within the data. If unlimited resources were available, a larger sample should be analysed for which more processing power is necessary to perform Object Detection and Object Arrangement calculations. In the case of a much larger dataset, a custom object detection algorithm could be trained on a sample of the data models, the full models can be used for non-linearity testing, more control variables can be added, and prediction models can be construed. Estimating the confidence interval of the turning points for (inverted) U testing is based on the delta methodology described in section 5.2.2. The delta methodology is the most widely used method to estimate confidence intervals for location quotients of the generalised linear model framework (Beyene & Moineddin, 2005). The confidence interval estimates using the delta method are prone to be inaccurate for small samples, and considering the sample of this study is relatively small (n=132), this could lead to inaccurate estimates (Xu & Long, 2005). Considering the strict testing methodology for (inverted) U-relationships and the CI of the turning point being well within data bounds, this limitation should not influence outcomes significantly.

5.3.3 Limitations of Visual Feature Measurement

Despite capturing a large set of low-level video features, the current features could capture more information or some information more accurately. With a larger sample size would come the possibility of training a custom Object Detection algorithm, where a pre-trained version is used now as done in previous studies by Overgoor et al. (2022) and Shin et al. (2019), while Li et al. (2019) uses custom models to perform object and facial detection. A custom-trained model is more accurate and more fitting to the specifics of the dataset. While the number of objects does not rely heavily on object labelling accuracy, object outline accuracy would also be improved. Additional analyses could be performed regarding the impact of specific objects or the presence of celebrities on engagement, which would help create a deeper understanding of online engagement and online marketing videos.
Finally, while Visual Variety does capture a broad spectrum of dynamic video features, movement within frames is not analysed using object tracking or movement analysis, like optical flow analysis. Optical flow or movement analysis has not been excessively studied yet in the field of marketing related to engagement, apart from some interesting papers, like the study by Zhou et al. (2021), which leverages optical flow to measure movement propensity in video courseware videos and its effect on engagement. Adding specific measures of dynamic flow and object tracking would help gain a deeper understanding of the specific effects of visual variety levels on engagement.

5.3.4 Limitations of Regression Models

Two significant changes could have been made to make the regression models more complete. Firstly, non-linearity testing could have been done with the full quadratic models if the sample size had allowed it. Secondly, interaction terms could be added between visual variables to study the effect of more clearly categorised visual variables on online engagement without the interaction of one variable on the other affecting coefficients. Further research could investigate the impact of visual features of advertisements and the effect on shares or other more active forms of engagement since these variables fit the data much better and seemed to be predicted more accurately by the models in Table 7 in section 5.2.1. Despite these facts, the regression outputs still offer some interesting results, mainly because significant effects can still be observed despite various limitations. Despite data limitations, the significance of various complex relationships makes an interesting case for the strength of the relationship between some of the visual features and online engagement.

5.3.5 Generalisability of the Study

Despite YouTube being the largest video-sharing platform, the sample only reflects some of the consumer population. YouTube generally has a younger audience since an older audience might not be up to speed on internet usage. The sampled advertisements are also of very high quality, with very high advertising spending, making the sample somewhat biased toward advertisements of higher quality. While higher-quality advertisements ensure that visual features and bad video quality are not confused or miscalculated, most advertisements have a different quality than those studied. The Super Bowl advertisements also deal with a significant initial buzz, creating more complex effects and being influenced by more extraneous variables. Most advertisements are not aired with such an underlying buzz and exposure, which makes it harder to make inferences based on this study alone.

6. Discussion

This study aimed to explore the relationship between low-level visual video features and online engagement on YouTube. The videos were sampled using an online Super Bowl archive and manual search to ensure the Super Bowl spots on YouTube were original. Frames were sampled using even frame sampling to lower computational overhead and achieve accurate, representative and non-biased snippets of each video. Computer Vision and Object Detection Algorithms were leveraged to find objective Visual Complexity and Visual Variety scores for each Super Bowl video.
As a second part of the analysis, the visual video features were regressed over three online engagement measures: views, likes and comments. The full Negative Binomial model showed that likes and comments fit the data better than the model predicting views. The model for Comments showed relatively low log-likelihood, AIC, and BIC values, which indicates that the most active form of engagement, writing a comment versus just watching the video or clicking one button, is most accurately predicted by visual video features. The Pseudo R^2 of all models is relatively low, which is expected for videos with such cultural significance that can lead to the relationships being impacted by many variables such as history, pre-perceived opinions about a brand, industry, message congruence of the brand perception and the message, emotional factors in the advertisement, audio and many more cultural factors. The model for Likes seemed to explain most of the variance in the data, which is probably due to the lower dispersion of the model (α=1.49) compared to the other models, which is a sign of less deviation in the data, as it also explains the relatively low number of significant predictors. Colour Complexity and Visual Variety seemed to be the most accurate positive predictors of online engagement, with two measures showing significant positive relationships. On the other hand, Luminance Entropy had a significant negative relationship with Comments, while Irregularity of Object Arrangement had a significant negative relationship with Views.

The findings of the non-linearity analysis, which is the final part of the study, indicate strong evidence for at least two significant quadratic relationships in the data when assessing the relation between visual video features and online engagement for the sample of Super Bowl videos. The relationship between Colour Complexity and Comments followed a U-relationship based on the testing logic imposed by Haans et al. (2015) and Lind and Melum (2010). Views and Irregularity of OA followed an inverted U-relationship. While the findings are interesting and offer some insights into using Visual Complexity and Visual Variety measures to determine effective online advertising methods, the findings need to be taken with a grain of salt. The relatively small sample and overdispersed data could lead to inaccurate estimates, even though the relationships must be strong to be statistically significant. Especially the model for Views, which has a high log-likelihood value, could fit the data poorly enough that even significant results might be practically irrelevant. This issue would need to be tested across different contexts and samples. When clustering this research with prior studies on Visual Complexity (Overgoor et al., 2022; Pieters et al., 2010), Visual Variety (Li et al., 2019), and relevant video analyses (Song et al., 2021), the non-linear Irregularity of OA finding related to views contrasts with the results of Overgoor et al. (2022). They identified a U-shaped relationship between likes—another measure of online engagement—and irregularity of OA for images. Despite Overgoor et al. (2022) applying the theory of Visual Complexity to images rather than videos, they found a similar relationship. The U-shaped relationship found for Colour Complexity and views is more exploratory in nature. The finding related to the relationship between Colour Complexity is not in line with prior research on other more active engagement metrics, like likes (Overgoor et al., 2022) and click behaviour (Song et al., 2021). Most studies find an inverted U-relationship for all variables related to feature complexity (Colour, Edges and Lighting) for image data. This adverse finding could be due to there being a difference between image and video data and engagement or the Super Bowl advertisements being of such high quality that the noise due to video quality issues leads to a different finding. The results outline the strength of the non-linear relationships between visual features and engagement, outlining the relevance of the Stimulation theories by Hebb (1955) and Berlyne (1970).

7. Appendix

7.1 GitHub Repository

All code used during data collection, video analysis and data analysis can be found on this GitHub repository:

https://github.com/RoelofBlommaert/Thesis-2024

7.2 Coding Packages Used

Table 10: Noteworthy packages used in this paper:

Language	Package	Description
Python	OpenCV	An open-source library for computer vision and image processing tasks
Python	TorchVision	A library that provides datasets, model architectures, and image transformations for computer vision tasks
Python	PyTorch	An open-source deep learning framework for building and training neural networks
Python	Detectron2	A PyTorch-based library for object detection and segmentation tasks
R	MASS	A package for statistical functions and datasets
R	car	A package that provides tools for regression analysis and diagnostics
R	pscl	A package for Bayesian analysis and simulation-based inference, particularly for political science data

*Note that packages used through MatLab are not added here as MatLab has fewer open-source and modular packages or Toolboxes. For MatLab toolboxes, refer to the Github page used for this paper and the referenced Github pages by Walthers and Koch (2007), Overgoor (2022) and Rosenholtz et al. (2007b).

7.3 Visual Variety Calculation in Python

```python
import cv2 as cv
import numpy as np
import pandas as pd
import os


#Since the calculation of visual variety demands a different manner of looping and comparing frames,
instead of measuring
#independent scores based on one frame at a time, the visual variety calculations are separate from the
functions and main script


def convert_and_normalise(frame):
    g_scale = cv.cvtColor(frame, cv.COLOR_BGR2GRAY)
    x_min, x_max = np.min(g_scale), np.max(g_scale)
    normalised = (g_scale - x_min) / (x_max - x_min) if x_max - x_min > 0 else
np.zeros(g_scale.shape)
    return normalised


def calculate_visual_variety(video_path):
    cap = cv.VideoCapture(video_path)
    if not cap.isOpened():
        print(f"Error: Could not open video {video_path}.")
        return None

    total_frames = int(cap.get(cv.CAP_PROP_FRAME_COUNT))
    interval = total_frames // 10
    distances = []
    norm_old_frame = None

    for i in range(10):
        frame_pos = i * interval
        cap.set(cv.CAP_PROP_POS_FRAMES, frame_pos)
        ret, frame = cap.read()
        if not ret:
            continue
        norm_frame = convert_and_normalise(frame)
        if norm_old_frame is not None:
            distance = np.sum(np.abs(norm_frame - norm_old_frame)) / norm_frame.size
            distances.append(distance)
        norm_old_frame = norm_frame

    cap.release()
    return np.mean(distances) if distances else 0
```

```
# Directory containing video files
video_dir = 'Data/downloaded_videos'
video_files = [os.path.join(video_dir, f) for f in os.listdir(video_dir) if f.endswith('.mp4')]


# Dictionary to store video names and their visual variety scores
visual_variety_scores = {}


# Calculate visual variety scores for each video
for video_path in video_files:
   video_name = os.path.basename(video_path)
   score = calculate_visual_variety(video_path)
   if score is not None:
       visual_variety_scores[video_name] = score
       print(f"Visual Variety Score for {video_name}: {score}")


print(visual_variety_scores)
# Read existing CSV file with visual complexity scores
df_complexity = pd.read_csv('Data/video_analysis_complexity_results.csv')


# Convert the visual variety dictionary to a DataFrame
df_variety = pd.DataFrame(list(visual_variety_scores.items()), columns=['Video Name', 'Visual
Variety'])
df_variety['Video Name'] = df_variety['Video Name'].str.replace('.mp4', '', regex=False)
# Merge the DataFrames on video name
merged_df = df_complexity.merge(df_variety, on='Video Name', how='outer')


# # Save the updated DataFrame to CSV
merged_file_path = 'Data/complexity_and_variety_scores.csv'
merged_df.to_csv(merged_file_path, index=False)
```

7.4 Number of Objects Calculation in Python

```
import cv2 as cv
import detectron2
from detectron2.utils.logger import setup_logger
setup_logger()
# Import some common detectron2 utilities
from detectron2 import model_zoo
from detectron2.engine import DefaultPredictor
```

```
from detectron2.config import get_cfg
from detectron2.utils.visualizer import Visualizer
from detectron2.data import MetadataCatalog, DatasetCatalog
import os
import numpy as np
import pandas as pd


# Setup detectron2 logger and configuration
cfg = get_cfg()
cfg.merge_from_file(model_zoo.get_config_file("COCO-InstanceSegmentation/mask_rcnn_R_50_FP
N_3x.yaml"))
cfg.MODEL.ROI_HEADS.SCORE_THRESH_TEST = 0.5  # set the testing threshold for this model
cfg.MODEL.DEVICE = 'cpu'  # Run on CPU
cfg.MODEL.WEIGHTS =
model_zoo.get_checkpoint_url("COCO-InstanceSegmentation/mask_rcnn_R_50_FPN_3x.yaml")
predictor = DefaultPredictor(cfg)


def detect_and_return_unique_objects(frame):
  outputs = predictor(frame)
  instances = outputs["instances"].to("cpu")
  classes = instances.pred_classes
  scores = instances.scores
  labels = [MetadataCatalog.get(cfg.DATASETS.TRAIN[0]).thing_classes[i] for i in classes]
  unique_labels = {label for label, score in zip(labels, scores) if score > 0.5}
  return len(unique_labels)


def calculate_unique_objects(video_path):
  cap = cv.VideoCapture(video_path)
  if not cap.isOpened():
    print(f"Error: Could not open video {video_path}.")
    return None

  total_frames = int(cap.get(cv.CAP_PROP_FRAME_COUNT))
  interval = total_frames // 30  # Process 30 frames evenly distributed throughout the video
  cum_unique_objects = 0

  for i in range(30):
    frame_pos = i * interval
    cap.set(cv.CAP_PROP_POS_FRAMES, frame_pos)
    ret, frame = cap.read()
    if not ret:
      continue
    unique_objects = detect_and_return_unique_objects(frame)
    cum_unique_objects += unique_objects
```

```
    cap.release()
    return cum_unique_objects


# Directory containing video files
video_dir = 'Data/downloaded_videos'
video_files = [os.path.join(video_dir, f) for f in os.listdir(video_dir) if f.endswith('.mp4')]


# Dictionary to store video names and their sum of unique objects over 30 frames
unique_objects_counts = {}


# Calculate unique objects for each video
for video_path in video_files:
    video_name = os.path.basename(video_path)
    count = calculate_unique_objects(video_path)
    if count is not None:
        unique_objects_counts[video_name] = count
        print(f"Unique objects count for {video_name}: {count}")


# Convert the unique objects dictionary to a DataFrame
df_unique_objects = pd.DataFrame(list(unique_objects_counts.items()), columns=['Video Name',
'Unique Objects Count'])
df_unique_objects['Video Name'] = df_unique_objects['Video Name'].str.replace('.mp4', '',
regex=False)
df_unique_objects['Unique Objects Count'] = df_unique_objects['Unique Objects Count']/30
df_complexity_variety = pd.read_csv('Data/complexity_and_variety_scores.csv')


# Merge the DataFrames on video name
merged_df = df_complexity_variety.merge(df_unique_objects, on='Video Name', how='outer')


# # Save the updated DataFrame to CSV
merged_file_path = 'Data/visual_feature_matrix.csv'
merged_df.to_csv(merged_file_path, index=False)


# Save the DataFrame to CSV
df_unique_objects.to_csv('Data/unique_objects_counts.csv', index=False)
```

7.5 Other Visual Complexity Calculation in MatLab

The MatLab code and paths added below are based on the code from Walthers and Koch (2007), Overgoor (2022) and Rosenholtz et al. (2007b).

```
% Add paths to necessary directories
addpath 'Data/downloaded_videos/'
addpath 'Helper/June01/'
addpath 'Helper/matlabPyrTools-master/'
addpath 'IJRM_visual_complexity/Complexity/'
addpath 'IJRM_visual_complexity/Alternative_Complexity/'
addpath 'IJRM_visual_complexity/Helper/SFFCMCode/SFFCMCode/'

% Define the directory containing videos
videoDir = 'Data/downloaded_videos/';
videoFiles = dir(fullfile(videoDir, '*.mp4'));  % Ensure files are being found

if isempty(videoFiles)
    error('No video files found. Check the directory path and file extensions.');
end

% Preallocate a cell array for storing results
numVideos = length(videoFiles);
resultsData = cell(numVideos, 6);  % Ensure this matches the number of metrics

% Loop through each video file
for k = 1:numVideos
    videoPath = fullfile(videoFiles(k).folder, videoFiles(k).name);
    v = VideoReader(videoPath);

    fprintf('Processing %s\n', videoPath);  % Debug: output which file is being processed

    if ~hasFrame(v)
        warning('No frames in %s. Skipping this video.', videoPath);
        continue;
    end

    % Calculate the interval for sampling 30 frames evenly across the video
    totalFrames = floor(v.Duration * v.FrameRate);
    interval = floor(totalFrames / 30);  % Adjust if not correct

    % Initialize sum variables
    sumLc = 0; sumCc = 0; sumEd = 0; sumAh = 0; sumAv = 0; sumIrv = 0;
    frameProcessed = 0; frameCount = 0;

    % Loop through frames, sampling at calculated intervals
    while hasFrame(v)
```

```
        frame = readFrame(v);
        frameCount = frameCount + 1;

        if mod(frameCount, interval) == 1 || frameCount == 1
            % Debug: Check if this part of the loop is executing
            fprintf('Analyzing frame %d of %s\n', frameCount, videoPath);

            % Simulate the function calls and processing
            sumLc = sumLc + luminance_complexity(frame);
            sumCc = sumCc + colorfulness(frame);
            sumEd = sumEd + edge_density(frame);
            [ah, av, irv] = arrangement(frame);
            sumAh = sumAh + ah;
            sumAv = sumAv + av;
            sumIrv = sumIrv + irv;
            frameProcessed = frameProcessed + 1;
            if frameProcessed >= 30
                break;
            end
        end
    end

    % Store results if any frames were processed
    if frameProcessed > 0
        [~, name, ~] = fileparts(videoPath);
        resultsData{k, 1} = name;
        resultsData{k, 2} = sumLc / frameProcessed;
        resultsData{k, 3} = sumCc / frameProcessed;
        resultsData{k, 4} = sumEd / frameProcessed;
        resultsData{k, 5} = ((sumAh + sumAv) / 2) / frameProcessed;
        resultsData{k, 6} = sumIrv / frameProcessed;
    else
        fprintf('No frames were processed for %s. Check interval calculation and frame conditions.\n',
videoPath);
    end
end

% Convert the cell array to a table
resultsHeader = {'Video Name', 'Luminance Complexity', 'Color Complexity', 'Edge Density',
'Asymmetry of Object Arrangement', 'Irregularity of Object Arrangement'};
finalResultsTable = cell2table(resultsData, 'VariableNames', resultsHeader);

% Display the final table
disp(finalResultsTable);

% Optionally save the table to a CSV file
writetable(finalResultsTable, 'Data/video_analysis_complexity_results.csv');
```

8. References

Arthurs, J., Drakopoulou, S., & Gandini, A. (2018). Researching YouTube. *Convergence*, *24*(1), 3–15. https://doi.org/10.1177/1354856517737222

Babyak, M. A. (2004). What you see may not be what you get: a brief, nontechnical introduction to overfitting in Regression-Type models. *Psychosomatic Medicine, 66(3), 411-421.* Retrieved May 18, 2024, from https://journals.lww.com/psychosomaticmedicine/fulltext/2004/05000/What_You_See_May_Not_Be_What_You_Get__A_Brief.21.aspx

Bergen, J. R., & Landy, M. S. (1991). Computational modeling of visual texture segregation. *Computational models of visual processing*, *17*, 253-271.

Berlyne, D. E. (1970). Novelty, complexity, and hedonic value. *Attention Perception & Psychophysics*, *8*(5), 279–286. https://doi.org/10.3758/bf03212593

Beyene, J., & Moineddin, R. (2005). Methods for confidence interval estimation of a ratio parameter with application to location quotients. *BMC Medical Research Methodology*, *5*(1). https://doi.org/10.1186/1471-2288-5-32

Bhandari, U., Chang, K., & Neben, T. (2019). Understanding the impact of perceived visual aesthetics on ser evaluations: an emotional perspective. *Information & Management*, *56*(1), 85–93. https://doi.org/10.1016/j.im.2018.07.003

Book, A. C., & Schick, C. D. (1997). *Fundamentals of copy & layout*. McGraw Hill Professional.

Brodie, R. J., Hollebeek, L. D., Jurić, B., & Ilić, A. (2011). Customer engagement. *Journal of Service Research*, *14*(3), 252–271. https://doi.org/10.1177/1094670511411703

Brodie, R. J., Ilić, A., Jurić, B., & Hollebeek, L. D. (2013). Consumer engagement in a virtual brand community: An exploratory analysis. *Journal of Business Research*, *66*(1), 105–114. https://doi.org/10.1016/j.jbusres.2011.07.029

Calvo, M. G., & Lang, P. (2004). Gaze patterns when looking at emotional pictures: motivationally biased attention. *Motivation and Emotion*, *28*(3), 221–243. https://doi.org/10.1023/b:moem.0000040153.26156.ed

Campbell, C., Thompson, F. M., Grimm, P. E., & Robson, K. (2017). Understanding why consumers don’t skip Pre-Roll video ads. *Journal of Advertising*, *46*(3), 411–423. https://doi.org/10.1080/00913367.2017.1334249

Chen, Y., Wang, S., Zhang, W., & Huang, Q. (2018). Less is more: picking informative frames for video captioning. In *Lecture notes in computer science* (pp. 367–384). https://doi.org/10.1007/978-3-030-01261-8_22

Coxe, S., West, S. G., & Aiken, L. S. (2009). The Analysis of count Data: A gentle introduction to poisson regression and its alternatives. *Journal of Personality Assessment*, *91*(2), 121–136. https://doi.org/10.1080/00223890802634175

Dessart, L. (2017). Social media engagement: a model of antecedents and relational outcomes. *Journal of Marketing Management*, 1–25. https://doi.org/10.1080/0267257x.2017.1302975

Donderi, D. C. (2006). Visual complexity: A review. *Psychological Bulletin*, *132*(1), 73–97. https://doi.org/10.1037/0033-2909.132.1.73

Farace, S., Roggeveen, A. L., Ordenes, F. J. V., De Ruyter, K., Wetzels, M., & Grewal, D. (2019). Patterns in motion: How visual patterns in ads affect product evaluations. *Journal of Advertising*, *49*(1), 3–17. https://doi.org/10.1080/00913367.2019.1652120

Ferguson, R. (2008). Word of mouth and viral marketing: taking the temperature of the hottest trends in marketing. ˜*the* œ*Journal of Consumer Marketing/Journal of Consumer Marketing*, *25*(3), 179–182. https://doi.org/10.1108/07363760810870671

Fornalczyk, K., Bortko, K., Disterheft, A., & Jankowski, J. (2023). Modeling the Impact of Video Dynamics on User Engagement and Eye Tracking Patterns. *Procedia Computer Science*, *225*, 4740–4749. https://doi.org/10.1016/j.procs.2023.10.473

Fredricks, J. A., Blumenfeld, P. C., & Paris, A. H. (2004). School engagement: potential of the concept, state of the evidence. *Review of Educational Research*, *74*(1), 59–109. https://doi.org/10.3102/00346543074001059

Germeys, F., & D'Ydewalle, G. (2005). The psychology of film: perceiving beyond the cut. *Psychological Research*, *71*(4), 458–466. https://doi.org/10.1007/s00426-005-0025-3

Girshick, R., Donahue, J., Darrell, T., & Malik, J. (2014). *Rich feature hierarchies for accurate object detection and semantic segmentation*. Retrieved February 20, 2024, from https://openaccess.thecvf.com/content_cvpr_2014/html/Girshick_Rich_Feature_Hierarchies_2014_CVPR_paper.html

Google. (n.d.). Interest in 'Super Bowl' over time via YouTube search [Graph]. Retrieved May 26, 2024, from https://trends.google.com/trends/explore?date=2024-01-01%202024-12-31,2023-01-01%202023-12-31,2022-01-01%202022-12-31,2021-01-01%202021-12-31,2020-01-01%202020-12-31&geo=,,,,&gprop=youtube&q=Super%20Bowl,Super%20Bowl,Super%20Bowl,Super%20Bowl,Super%20Bowl&hl=en-US

Gross, J., & Von Wangenheim, F. (2022). Influencer Marketing on Instagram: Empirical Research on Social Media Engagement with Sponsored Posts. *Journal of Interactive Advertising*, *22*(3), 289–310. https://doi.org/10.1080/15252019.2022.2123724

Haans, R. F. J., Pieters, C., & He, Z. (2015). Thinking about U: Theorizing and testing U‐ and inverted U‐shaped relationships in strategy research. *Strategic Management Journal*, *37*(7), 1177–1195. https://doi.org/10.1002/smj.2399

Hajli, N. (2014). A study of the impact of social media on consumers. *International Journal of Market Research*, *56*(3), 387–404. https://doi.org/10.2501/ijmr-2014-025

Hartmann, W. R., & Klapper, D. (2018). Super Bowl ads. *Marketing Science (Providence, R.I.)*, *37*(1), 78–96. https://doi.org/10.1287/mksc.2017.1055

Hasler, D., & Süsstrunk, S. (2003). Measuring colorfulness in natural images. *Proceedings of SPIE*. https://doi.org/10.1117/12.477378

Hautz, J., Füller, J., Hutter, K., & Thürridl, C. (2014). Let users generate your video ads? The impact of video source and quality on consumers' perceptions and intended behaviors. *Journal of Interactive Marketing*, *28*(1), 1–15. https://doi.org/10.1016/j.intmar.2013.06.003

He, K., Gkioxari, G., Dollar, P., & Girshick, R. (2017). *Mask R-CNN*. Retrieved March 13, 2024, from https://openaccess.thecvf.com/content_iccv_2017/html/He_Mask_R-CNN_ICCV_2017_paper.html

Heij, C., de Boer, P., Franses, P. H., Kloek, T., & van Dijk, H. (2004). Econometric Methods with Applications in Business and Economics. Oxford University Press.

Hole, A. R. (2007). A comparison of approaches to estimating confidence intervals for willingness to pay measures. *Health Economics, 16, 827–840.* https://doi.org/10.1002/hec.1197

Hollebeek, L. D., Glynn, M. S., & Brodie, R. J. (2014). Consumer brand engagement in social Media: conceptualization, scale development and validation. *Journal of Interactive Marketing*, *28*(2), 149–165. https://doi.org/10.1016/j.intmar.2013.12.002

Hollebeek, L. D. (2011). Exploring customer brand engagement: definition and themes. *Journal of Strategic Marketing*, *19*(7), 555–573. https://doi.org/10.1080/0965254x.2011.599493

Hudson, S., Roth, M. S., Madden, T. J., & Hudson, R. (2015). The effects of social media on emotions, brand relationship quality, and word of mouth: An empirical study of music festival attendees. *Tourism Management*, *47*, 68–76. https://doi.org/10.1016/j.tourman.2014.09.001

Hung, D. L., & Tzeng, O. J. (1981). Orthographic variations and visual information processing. *Psychological Bulletin*, *90*(3), 377–414. https://doi.org/10.1037/0033-2909.90.3.377

Jahn, B., & Kunz, W. H. (2012). How to transform consumers into fans of your brand. *Journal of Service Management*, *23*(3), 344–361. https://doi.org/10.1108/09564231211248444

Jayalakshmi, T., & Santhakumaran, A. (2011). Statistical normalization and backpropagation for classification. *International Journal of Computer Theory and Engineering*, 3(1), 89–93. https://doi.org/10.7763/ijcte.2011.v3.288

Kahn, W. A. (1990). PSYCHOLOGICAL CONDITIONS OF PERSONAL ENGAGEMENT AND DISENGAGEMENT AT WORK. *Academy of Management Journal*, *33*(4), 692–724. https://doi.org/10.2307/256287

Khan, M. L. (2017). Social media engagement: What motivates user participation and consumption on YouTube? *Computers in Human Behavior*, *66*, 236–247. https://doi.org/10.1016/j.chb.2016.09.024

Kusumasondjaja, S., & Tjiptono, F. (2019). Endorsement and visual complexity in food advertising on Instagram. *Internet Research*, *29*(4), 659–687. https://doi.org/10.1108/intr-11-2017-0459

Lavie, T., & Tractinsky, N. (2004). Assessing dimensions of perceived visual aesthetics of web sites. *International Journal of Human-computer Studies*, *60*(3), 269–298. https://doi.org/10.1016/j.ijhcs.2003.09.002

Lee, D., Hosanagar, K., & Nair, H. S. (2018). Advertising Content and Consumer Engagement on Social Media: Evidence from Facebook. *Management Science*, *64*(11), 5105–5131. https://doi.org/10.1287/mnsc.2017.2902

Lee, J., & Ahn, J. (2012). Attention to banner ads and their effectiveness: An Eye-Tracking approach. *International Journal of Electronic Commerce*, *17*(1), 119–137. https://doi.org/10.2753/jec1086-4415170105

Lewis, F., Butler, A., & Gilbert, L. (2010). A unified approach to model selection using the likelihood ratio test. *Methods in Ecology and Evolution*, *2*(2), 155–162. https://doi.org/10.1111/j.2041-210x.2010.00063.x

Li, F. F., VanRullen, R., Koch, C., & Perona, P. (2002). Rapid natural scene categorization in the near absence of attention. *Proceedings of the National Academy of Sciences of the United States of America*, *99*(14), 9596–9601. https://doi.org/10.1073/pnas.092277599

Li, X., Shi, M., & Wang, X. (2019). Video mining: Measuring visual information using automatic methods. *International Journal of Research in Marketing*, *36*(2), 216–231. https://doi.org/10.1016/j.ijresmar.2019.02.004

Lind, J. T., & Mehlum, H. (2010). With or without U? the appropriate test for a U-Shaped relationship*. *Oxford Bulletin of Economics and Statistics*, *72*(1), 109–118. https://doi.org/10.1111/j.1468-0084.2009.00569.x

Liu, X., Shi, S. W., Teixeira, T. S., & Wedel, M. (2018). Video Content Marketing: The making of Clips. *Journal of Marketing*, *82*(4), 86–101. https://doi.org/10.1509/jm.16.0048

Marks, H. M. (2000). Student engagement in instructional activity: Patterns in the elementary, middle, and high school years. *American Educational Research Journal*, *37*(1), 153. https://doi.org/10.2307/1163475

Matz, S., Segalin, C., Stillwell, D., Müller, S. R., & Bos, M. W. (2019). Predicting the personal appeal of marketing images using computational methods. *Journal of Consumer Psychology*, *29*(3), 370–390. https://doi.org/10.1002/jcpy.1092

McParlane, P. J., Moshfeghi, Y., & Jose, J. M. (2014). “Nobody comes here anymore, it’s too crowded”; Predicting Image Popularity on Flickr. *In Proceedings of International Conference on Multimedia*, *14 (pp. 385–391).*. https://doi.org/10.1145/2578726.2578776

Miniukovich, A., Sulpizio, S., & De Angeli, A. (2018). Visual complexity of graphical user interfaces. *Proceedings of the 2018 International Conference on Advanced Visual Interfaces (pp. 1-9)*. https://doi.org/10.1145/3206505.3206549

Munaro, A. C., Barcelos, R. H., Maffezzolli, E. C. F., Rodrigues, J. P. S., & Paraíso, E. C. (2021). To engage or not engage? The features of video content on YouTube affecting digital consumer engagement. *Journal of Consumer Behaviour*, *20*(5), 1336–1352. https://doi.org/10.1002/cb.1939

Nagelkerke, N. J. D. (1991). A note on a general definition of the coefficient of determination. *Biometrika*, *78*(3), 691–692. https://doi.org/10.1093/biomet/78.3.691

Nagle, F., & Lavie, N. (2020). Predicting human complexity perception of real-world scenes. *Royal Society Open Science*, *7*(5), 191487. https://doi.org/10.1098/rsos.191487

Nanne, A., Antheunis, M. L., Van Der Lee, C. G., Postma, E., Wubben, S., & Van Noort, G. (2020). The use of Computer Vision to analyze Brand-Related user generated image content. *Journal of Interactive Marketing*, *50*(1), 156–167. https://doi.org/10.1016/j.intmar.2019.09.003

Neyman, J., & Pearson, E. S. (1928). On the Use and Interpretation of Certain Test Criteria for Purposes of Statistical Inference: Part II. *Biometrika*, *20A*(3/4), 263. https://doi.org/10.2307/2332112

Nieuwenhuis, G. (2009). Statistical methods for business and economics. McGraw Hill.

Nikolinakou, A., & King, K. W. (2018). Viral video ads: Emotional triggers and social media virality. *Psychology & Marketing*, *35*(10), 715–726. https://doi.org/10.1002/mar.21129

O'Brien, H. L. (2017). Antecedents and learning outcomes of online news engagement. *Journal of the Association for Information Science and Technology (Print)*, *68*(12), 2809–2820. https://doi.org/10.1002/asi.23854

Oğuz, E. A., Košir, A., Strle, G., & Burnik, U. (2023). Low-Level video features as predictors of consumer engagement in multimedia advertisement. *Applied Sciences (Basel)*, *13*(4), 2426. https://doi.org/10.3390/app13042426

O'Hara, R. B., & Kotze, D. J. (2010). Do not log‐transform count data. *Methods in Ecology and Evolution*, *1*(2), 118–122. https://doi.org/10.1111/j.2041-210x.2010.00021.x

Overgoor, G., Rand, W., Van Dolen, W., & Mazloom, M. (2022). Simplicity is not key: understanding firm-generated social media images and consumer liking. *International Journal of Research in Marketing*, *39*(3), 639–655. https://doi.org/10.1016/j.ijresmar.2021.12.005

Overgoor. (2022). *IJRM_Visual_complexity*. GitHub. Retrieved March 12, 2024, from https://github.com/Gyys1992/IJRM_visual_complexity

Picchi, A. (2024, February 11). How much do Super Bowl commercials cost for the 2024 broadcast? *CBS News*. Retrieved February 15, 2024, from https://www.cbsnews.com/news/how-much-super-bowl-commercial-cost-2024/

Pieters, R., Wedel, M., & Batra, R. (2010). The Stopping Power of Advertising: Measures and Effects of Visual Complexity. *Journal of Marketing*, *74*(5), 48–60. https://doi.org/10.1509/jmkg.74.5.048

Pozharliev, R., Verbeke, W., Van Strien, J. W., & Bagozzi, R. P. (2015). Merely Being with you Increases My Attention to Luxury Products: Using EEG to Understand Consumers' Emotional Experience with Luxury Branded Products. *Journal of Marketing Research*, *52*(4), 546–558. https://doi.org/10.1509/jmr.13.0560

Rietveld, R., Van Dolen, W., Mazloom, M., & Worring, M. (2020). What you Feel, Is what you like Influence of Message Appeals on Customer Engagement on Instagram. *Journal of Interactive Marketing*, *49*(1), 20–53. https://doi.org/10.1016/j.intmar.2019.06.003

Roma, P., & Aloini, D. (2019). How does brand-related user-generated content differ across social media? Evidence reloaded. *Journal of Business Research*, *96*, 322–339. https://doi.org/10.1016/j.jbusres.2018.11.055

Rooderkerk, R. P., & Pauwels, K. H. (2016). No comment?! The drivers of reactions to online posts in professional groups. *Journal of Interactive Marketing*, *35*, 1–15. https://doi.org/10.1016/j.intmar.2015.12.003

Rosenholtz, R., Li, Y., & Nakano, L. (2007a). Measuring visual clutter. *Journal of Vision*, *7*(2), 17. https://doi.org/10.1167/7.2.17

Rosenholtz, R., Li, Y., & Nakano, L. (2007b). *Feature Congestion and Subband Entropy measures of visual clutter*. *MIT Libraries DSpace*. Retrieved March 26, 2024, from http://hdl.handle.net/1721.1/37593

Schaufeli, W. B., Salanova, M., González-Romá, V., & Bakker, A. B. (2002). The Measurement of Engagement and Burnout: A two sample confirmatory Factor analytic approach. *Journal of Happiness Studies (Print)*, *3*(1), 71–92. https://doi.org/10.1023/a:1015630930326

Shin, D., He, S., Lee, G. M., Whinston, A. B., Cetintas, S., & Lee, K. (2020). Enhancing social media analysis with visual data Analytics: a deep learning approach. *Management Information Systems Quarterly*, *44*(4), 1459–1492. https://doi.org/10.25300/misq/2020/14870

Shukla, A., Katti, H., Kankanhalli, M., & Subramanian, R. (2018). Looking Beyond a Clever Narrative. *Proceedings of the 20th ACM International Conference on Multimodal Interaction (ICMI '18)*, 210–219. https://doi.org/10.1145/3242969.3242988

Simons, R. F., Detenber, B. H., Cuthbert, B. N., Schwartz, D. D., & Reiss, J. E. (2003). Attention to television: alpha power and its relationship to image motion and emotional content. *Media Psychology*, *5*(3), 283–301. https://doi.org/10.1207/s1532785xmep0503_03

Smith, K. A., Sheppard, S., Johnson, D. W., & Johnson, R. T. (2005). Pedagogies of Engagement: Classroom-Based Practices. *Journal of Engineering Education (Washington, D.C.)*, *94*(1), 87–101. https://doi.org/10.1002/j.2168-9830.2005.tb00831.x

Song, D. Y., Wang, S. C., Ou, C., Chen, X., Liu, R., & H.H., T. (2021). How do video features matter in visual advertising? An elaboration likelihood model perspective. ICIS 2021 Proceedings. 8. Retrieved April 13, 2024, from https://research.tilburguniversity.edu/en/publications/how-do-video-features-matter-in-visual-advertising-an-elaboration-2

Statista. (2024a, February 7). *Countries with the highest monthly traffic volume to YouTube.com 2022*. Retrieved April 24, 2024, from https://www.statista.com/statistics/1357163/youtube-global-monthly-visits-by-country/

Statista. (2024b, March 28). *Global top websites by monthly visits 2023 | Statista*. Retrieved April 24, 2024, from https://www.statista.com/statistics/1201880/most-visited-websites-worldwide/

Stuppy, A., Landwehr, J. R., & McGraw, A. P. (2023). The art of slowness: Slow motion enhances consumer evaluations by increasing processing fluency. *Journal of Marketing Research, 61(2),* 185-203. https://doi.org/10.1177/00222437231179187

Teixeira, T. S., Wedel, M., & Pieters, R. (2010). Moment-to-Moment optimal branding in TV commercials: preventing avoidance by pulsing. *Marketing Science*, *29*(5), 783–804. https://doi.org/10.1287/mksc.1100.0567

Tellis, G. J., MacInnis, D. J., Tirunillai, S., & Zhang, Y. (2019). What drives virality (Sharing) of online digital content? The critical role of information, emotion, and brand prominence. *Journal of Marketing*, *83*(4), 1–20. https://doi.org/10.1177/0022242919841034

Tuch, A. N., Bargas-Avila, J. A., Opwis, K., & Wilhelm, F. H. (2009). Visual complexity of websites: Effects on users' experience, physiology, performance, and memory. *International Journal of Human-computer Studies*, *67*(9), 703–715. https://doi.org/10.1016/j.ijhcs.2009.04.002

Vraga, E. K., Bode, L., Wells, C. L., Driscoll, K., & Thorson, K. (2014). The Rules of Engagement: Comparing two social protest movements on YouTube. *Cyberpsychology, Behavior and Social Networking*, *17*(3), 133–140. https://doi.org/10.1089/cyber.2013.0117

Wagemans, J., Elder, J. H., Kubovy, M., Palmer, S. E., Peterson, M. A., Singh, M., & Von Der Heydt, R. (2012). A century of Gestalt psychology in visual perception: I. Perceptual grouping and figure–ground organization. *Psychological Bulletin*, *138*(6), 1172–1217. https://doi.org/10.1037/a0029333

Walthers, D., & Koch, C. (2006). Modeling attention to salient proto-objects. *Neural Networks*, *19*(9), 1395–1407. https://doi.org/10.1016/j.neunet.2006.10.001

Walthers, D., & Koch, C. (2007). SaliencyToolbox. GitHub. Retrieved March 26, 2024, from https://github.com/DirkBWalther/SaliencyToolbox

Wedel, M., & Pieters, R. (2006). Eye tracking for visual marketing. *Foundations and Trends in Marketing*, *1*(4), 231–320. https://doi.org/10.1561/1700000011

Wu, Y., Kirillov, A., Massa, F., Lo, W.-Y., & Girshick, R. (2019). Detectron2. Retrieved April 2, 2024, from https://github.com/facebookresearch/detectron2

Xiao, L., Li, X., & Zhang, Y. (2023). Exploring the factors influencing consumer engagement behavior regarding short-form video advertising: A big data perspective. *Journal of Retailing and Consumer Services*, *70*, 103170. https://doi.org/10.1016/j.jretconser.2022.103170

Xu, J., Fedorowicz, J., & Williams, C. B. (2019). Effects of symbol sets and needs gratifications on audience engagement: Contextualizing Police social media communication. *Journal of the Association for Information Systems*, 536–569. https://doi.org/10.17705/1jais.00543

Xu, J., & Long, J. S. (2005). Confidence intervals for predicted outcomes in regression models for categorical outcomes. *The Stata Journal*, *5*(4), 537–559. https://doi.org/10.1177/1536867x0500500405

Yin, Y., Jia, J. S., & Zheng, W. (2021). The effect of slow motion video on consumer inference. *Journal of Marketing Research*, *58*(5), 1007–1024. https://doi.org/10.1177/00222437211025054

YOLOV8: a new State-of-the-Art Computer Vision model. (n.d.). Retrieved March 13, 2024, from https://yolov8.com/

Youtube [Team Youtube]. (2021, November 10). An update to dislikes on YouTube. *blog.youtube*. Retrieved February 19, 2024, from https://blog.youtube/news-and-events/update-to-youtube/

YouTube Culture & Trends - data and cultural analysis for you. (n.d.). YouTube Culture & Trends. Retrieved February 19, 2024, from https://www.youtube.com/trends/ads-leaderboard/

YouTube for press. (n.d.). Blog.Youtube. February 18, 2024, from https://blog.youtube/press

Youtube Team. (2021, November 10). An update to dislikes on YouTube. *blog.youtube*. March 2, 2024, from https://blog.youtube/news-and-events/update-to-youtube/

Zhang, S., Lee, D., Singh, P. V., & Srinivasan, K. (2017). How much is an image worth? Airbnb property demand estimation leveraging large scale image analytics. *Social Science Research Network*. https://doi.org/10.2139/ssrn.2976021

Zhou, M., Chen, G. H., Ferreira, P., & Smith, M. D. (2021). Consumer behavior in the online classroom: Using video analytics and machine learning to understand the consumption of video courseware. *Journal of Marketing Research*, *58*(6), 1079–1100. https://doi.org/10.1177/00222437211042013

Zhu, J., Cheng, M., & Wang, Y. (2024). Viewer In-Consumption Engagement in Pro-Environmental Tourism Videos: A Video Analytics approach. *Journal of Travel Research*. https://doi.org/10.1177/00472875231219634