Journal of Open Humanities Data

Cutting the Frame: An In-Depth Look at the Hitchcock Computer Vision Dataset

RESEARCH PAPER

NABEEL SIDDIQUI

ubiquity press

ABSTRACT

This paper presents a comprehensive dataset comprising annotations generated by the Google Vision API for approximately 105,000 frames extracted from 15 Alfred Hitchcock films. These annotations include information about object detection, facial recognition, web-entity analysis, and explicit content filtering. With potential applications in the digital humanities and film studies, this dataset enables researchers to not only explore and evaluate cinematic content but also the ways that it resurfaces in various cultural contexts online.

The paper provides a detailed account of the dataset creation process, which involved the decryption of the DVD, frame extraction, and costs of annotations. Additionally, the paper outlines future research possibilities based on the dataset. These include statistical analysis of frame content and labels to identify patterns and trends, comparisons of different computer vision algorithms to assess their accuracy and effectiveness, and the utilization of bipartite networks to explore mise-en-scene in films.

CORRESPONDING AUTHOR:
Nabeel Siddiqui
Communications Department, Susquehanna University, Selinsgrove, PA, USA
siddiqui@susqu.edu

KEYWORDS:
film studies; digital humanities; visual turn; computer vision; network analysis

TO CITE THIS ARTICLE:
Siddiqui, N. (2024). Cutting the Frame: An In-Depth Look at the Hitchcock Computer Vision Dataset. *Journal of Open Humanities Data*, 10: 5, pp. 1–16. DOI: https://doi.org/10.5334/johd.163

DOI FOR DATASET AND LINK

10.5281/zenodo.10160597 (https://zenodo.org/records/10160597)

DATASET DESCRIPTION

Object name – Hitchcock Computer Vision Dataset

Format names and versions – CSV

Creation dates – September 1, 2023

Dataset creators – Nabeel Siddiqui

Language – English

License – Creative Commons Attribution 4.0 International

Repository name –Zenodo https://zenodo.org/records/10160597

Publication date – October 3, 2023

(1) INTRODUCTION

(1.1) THE VISUAL TURN IN THE DIGITAL HUMANITIES AND FILM STUDIES

In the digital humanities, text remains the dominant medium of expression and analysis (Manovich, 2020; McPherson, 2009; Meeks, 2013; Sayers, 2018). This has resulted in some scholars defining the field as "text heavy, visualization light, and simulation poor" (Champion, 2017). However, the rise of massively available visual data and the ability to computationally analyze it have opened new possibilities for scholarship. Wevers and Smits (2020) credit this "visual turn" to the proliferation and advancement of complex computer vision algorithms that utilize deep neural networks to make sense of images. These networks can recognize hierarchical patterns and apply filters to different parts of the image —like recognizing shapes, edges, and textures — gradually forming a more complex and nuanced understanding of the input data. The accuracy and impact of these models are only now being explored in the digital humanities, but they are posed to drastically alter the analysis, critique, and interpretation of cultural data (Arnold et al., 2022; Arnold & Tilton, 2019; di Lenardo et al., 2016; Gefen et al., 2021; Hu et al., 2017; Musik & Zeppelzauer, 2018; Pustu-Iren et al., 2020; Resig, 2014).

The shift away from logocentrism in computational analysis towards visuality carries significant implications for film studies. Yet, film scholars remain fraught with suspicion regarding computational approaches and question the ability of machines to appreciate the aesthetic and cultural nuances of cinematic content. As Nick Redfern (2023a) observes, while film studies has accepted that "film" may come in more formats than just celluloid—such as through VHS tapes, DVDs, and streaming platforms—its "studies" have remained the same. He calls for a new "post-disciplinary complex" that embraces methodological pluralism and the use of computational techniques to supplement the close reading of film. Likewise, drawing on notions of distant reading and listening, Tilton and Arnold (2019) postulate "distant viewing" as a new methodology that explicitly addresses the interpretive element of image semantics. They note, "To view images computationally, a representation of elements contained within the visual material—a code system in semiotics or, similarly, a metadata schema in informatics—must be constructed. Algorithms capable of automatically converting raw images into the established representation are then needed to apply the approach at scale" (Arnold & Tilton, 2019).

While the concept of distant viewing applies to both images and film, it is important to note their distinct characteristics. In the case of images, distant viewing involves the analysis of separate, static visual artifacts, such as paintings, photographs, or illustrations. Although these artifacts may be part of a series, in most cases, each individual object is meant to convey a singular visual message. Thus, the study and interpretation of such artifacts usually hinge upon metadata secondary to the objects themselves—information about their creators, contexts, or archive details, for instance. Conversely, film inherently consists of a temporally bound sequence of image frames and is consumed in a linear, chronological manner. The rich features and temporal dependencies of film necessitate a more complex modeling strategy for distant viewing. In other words, rather than focusing narrowly on static frame contents, it is essential to consider sequences, temporal transitions, and the unique grammar of film shots.

Tilton and Arnold's Distant Viewing Toolkit (DVT) is a prime example of their proposed framework for computer vision analysis of film in action. As they note, the analysis of an image or film can be broken down into two key steps (Arnold & Tilton, 2020). In the first step, the raw data is "viewed" through various annotation tools and algorithms, which create a metadata schema based on the visual elements in the images and frames. In the second step, these annotations are examined at a "distance" by aggregating their results across the datasets and comparing patterns, visual trends, or stylistic signatures. The distant viewing toolkit itself encapsulates these steps into an automated process that allows for a rapid, iterative, and exploratory framework.

The DVT is a free and open-source tool focused on ease of use. To avoid passing on costs related to the use of commercial APIs to users, it relies on open-source models for conducting its annotations. In the past few years, the reliability and capability of these open-source models have grown tremendously. For instance, models such as Yolo8 can not only detect a wide range of objects, but researchers can fine-tune it to specific tasks for film analysis (Jocher et al., 2023). Other models, such as OpenAI's CLIP can provide multimodal analysis capabilities that interpret both images and natural language together (Radford et al., 2021). Yet, while open-source models offer numerous advantages, they are limited in their ability to detect how images circulate in the broader cultural landscape. Consequently, researchers of social media have begun to turn to Google's commercial Vision API due to its ability to provide context-sensitive annotations that better capture trending interests and the dynamic nature of visual semantics in the digital age (Rogers & Giorgi, n.d.; Smith et al., 2022; Tommasini et al., 2023).

Digital humanities scholars are only now beginning to utilize Google's Vision API for the circulation of cultural objects, but exploration in film remains limited (Smits & Ros, 2021). As Richard Rogers notes, Google's ability to not only rank but index massive amounts of information gives it a unique edge in the realm of cultural analytics (Rogers, 2015). Although its API is proprietary, it offers exclusive features based on Google's Knowledge Graph, such as web entity analysis, which will likely never be implemented by open-source counterparts. According to Google, the knowledge graph is "a system that understands facts and information about entities from materials shared across the web, as well as from open source and licensed databases. It has amassed over 500 billion facts about five billion entities" (Singhal, 2012). In Google Search, when a user types in an entity—which Google Defines as a person, place, or thing—a "knowledge panel" to the side demonstrates a summary of that entity from the Knowledge Graph. Through the API, these extracted entities are reflected in the dataset as labels and enhance the interpretative potential of film frames. Rather than focusing purely on the visual content of the frame, this feature allows researchers to connect on-screen elements to a wider socio-cultural context exemplified by online trends.

In addition, the Vision API provides results about where film frames may circulate along with similar images found across the web in various cultural contexts, like fan art and memes. This opens a new way of understanding how certain images gain cultural significance, and how they're reinterpreted or repurposed to communicate new meanings in different settings. For example, in a popular film, a single frame featuring the protagonist in a dramatic climax may get circulated and become a symbol across fan communities or cause a viral trend on social media platforms. Likewise, online commentators and reviewers may use stills from the film to illustrate their critique or to clarify points of discussion, thereby increasing its visibility and potential for influencing public perceptions of the movie.

Finally, Google's Vision API provides the ability to determine if an image contains violence, adult content, parodies, or medical content. For film scholars, these annotations offer additional insights into portrayals and perceptions of certain themes or elements within films, whether overt or hidden among subtexts. Of course, it is important to understand the bias in these AI models as they tend to reproduce the ideologies and structures of their creators—an issue discussed later in the article. Yet at the same time, their use might unravel new aspects of our cultural heritage that have not yet been fully explored on a large scale.

Thus, while open-source models provide strong object detection, they cannot make sense of the socio-cultural implications of images. At the same time, the proprietary nature of commercial APIs requires a significant financial investment that makes them impractical for most film analyses. For instance, a 90-minute commercial film has 129,600 total frames, which—if analyzed using Google's Vision AI at $1.50 per 1000 units—would cost nearly $200 per film for just the basic annotation (*Pricing | Cloud Vision API*, n.d.). This cost, when scaled across many films for a comprehensive analysis is prohibitively expensive for most humanities researchers. Even if a frame was only taken every second or through more advanced frame selection techniques, the cost remains restrictive— this is particularly the case when accounting for more complex annotations, such as web entity analysis.

This paper and its accompanying dataset seek to bridge this accessibility gap by providing a comprehensive, free-of-charge resource that leverages Google's Vision API for film analysis. It introduces the Hitchcock Computer Vision Dataset which contains the results of the API's annotations of frames in fifteen Alfred Hitchcock films. These results include object detection, facial recognition, web-entity analysis, and explicit content filtering. Below, I detail the process of compiling this dataset, the choices made in selecting specific frames, issues of reproducibility, as well as the potential applications of the dataset for future researchers.

(1.2) OVERVIEW OF THE DATASET

The Hitchcock Computer Vision Dataset consists of a CSV showcasing the results of Google's Vision API on frames taken from 2017's *Alfred Hitchcock: Ultimate Collection* DVD box set. The collection itself contains fifteen of Hitchcock's canonical films directed between 1942 and 1976: *Saboteur* (1942), *Shadow of a Doubt* (1943), *Rope* (1948), *Rear Window* (1954), *The Trouble with Harry* (1955), *The Man Who Knew Too Much* (1956), *Vertigo* (1958), *North by Northwest* (1959), *Psycho* (1960), *The Birds* (1963), *Marnie* (1964), *Torn Curtain* (1966), *Topaz* (1969), *Frenzy* (1972), *Family Plot* (1976). Despite the absence of some noteworthy films, such as *Rebecca* (1940) and *To Catch a Thief* (1955), the collection offers a wide range of works that allow researchers to comprehensively examine the director's evolving cinematic techniques, narrative preferences, and recurring themes.

The choice of Hitchcock should not be seen as an aesthetic judgment. Instead, the selection was based on a desire to permit other scholars to also examine the discipline of film studies itself along with the works. According to John Belton (2003), "To some extent, the rise of Hitchcock studies mirrors the rise of film studies as an academic discipline." Likewise, David Colangelo (2018) shows that changes in Hitchcock studies can be directly traced to "perceptual shifts related to the technological conditions of film scholarship have shaped the analysis of film." Thus, by focusing on Hitchcock's films, one can gain valuable insights into how film studies as a discipline and film as an industry has evolved in response to the development of new technologies such as digital media.

(2) METHODS: EXTRACTING FRAMES AND UTILIZING COMMERCIAL COMPUTER VISION ANNOTATION APIS

After acquiring a copy of the DVD, there were three major steps taken to develop the dataset: decrypting the DVD, extracting the frames, and utilizing the Google Vision API for annotation. These steps brought several critical concerns to the fore about data integrity, the legal implications of decrypting DVDs in the United States, questions of which annotations would best serve a film studies audience, and the challenge of handling such a large amount of data. However, they provide a blueprint for those looking to undertake similar research in the future.

(2.1) STEP 1: DECRYPTING THE DVD

The first step in creating the dataset was to decrypt the DVD. However, this was not without challenges. A significant obstacle confronting researchers in the digital humanities is copyright. While other countries may have more lenient rules about research, this research was conducted in the United States where legislation can be notably restrictive. Because U.S. copyright law has crystalized around relatively narrow notions of literature that focus on "high art" textual production, film and other forms of visual media have often been overlooked in discussions of free use (Wharton, 2013). The 1998 Digital Millennium Copyright Act, which strengthened laws against the tampering of technological protection measures and devices specifically meant to bypass copyright restrictions, only exacerbated these issues (Calandrillo & Davison, 2008;

Cobia, 2009; Lipton, 2005; Lunney Jr, 2001). It was not until 2021 that the U.S. Copyright Office provided an exception for researchers seeking to bypass encryption for data mining books and films (Crump, 2022).

Numerous programs and scripts are available for DVD decryption, but due to the complexity of the law and the varying capabilities of the tools, it is key for researchers to understand both their technical and legal aspects. For this project, the application MakeMKV Inc. (GuinpinSoft inc, 2023) was used for its simplicity in accessing and unscrambling DVDs encrypted with Content Scramble Systems (Becker & Desoky, 2004; Eskicioglu & Delp, 2001).

Once the tool extracted the films, they were placed in an encrypted and password-protected hard drive to safeguard against unauthorized access. This ensured dataset integrity throughout the project while also adhering to relevant regulations. Once the annotation process described in step three was finished, all files were deleted, and the hard drive was wiped to maintain data security.

(2.2) STEP 2: EXTRACTING THE FRAMES

The second step in creating the dataset involved extracting individual frame shots from the movie files. As mentioned earlier, in a 90-minute film, there are approximately 129,600 frames, making both the storage and analysis of every frame a daunting task. Consequently, researchers have sought to reduce the number of frames to a manageable level while still providing a comprehensive representation of the film's narrative and cinematic scope (Gianluigi & Raimondo, 2006; Huang & Wang, 2019; Meng et al., 2016; Sebastian & Puthiyidam, 2015).

One approach to achieve frame reduction is to utilize keyframe extraction algorithms (Asha Paul et al., 2018; Sadiq et al., 2020; Sujatha & Mudenagudi, 2011). In a typical compressed video, the full-frame (I-frame) is only refreshed upon a significant change in the scene, while P-frames and B-frames merely update the pixels that have changed from the previous or future frame. This approach allows for efficient video storage and transmission while still capturing critical content.

An alternative method to keyframe extraction for frame reduction is the employment of scene detection algorithms. Scene detection algorithms segregate a film into scenes based on changes in the video frame sequences determined by color histogram differences and frame intensity/brightness. Popular tools, such as PySceneDetect, utilize both of these methods enabling adjustable sensitivity for scene change thresholds (Castellano, 2023; Gruzman & Kostenkova, 2014; Reddy & Jadhav, 2015). For film scholars, these tools can greatly ease data extraction needed for certain statistical analyses, such as finding the mean shot length (MSL) or the comparison of shot length frequency across different films (Redfern, 2023b; Salt, 1974).

Although scene detection algorithms and keyframe extraction tools minimize the number of frames, these methods are not infallible, and reliance on these techniques can inadvertently result in missing out on subtle, yet essential, aspects of a film. For instance, small changes in a character's expression might not register as significant to the frame but may be vital to the narrative or thematic interpretation. Likewise, a drawn-out scene with minimal visual variation but loaded with deep dialogue and character development would be under-represented through these methods. This is especially the case in Hitchcock's filmography, where films like *Rope* (1948) experimented with long unbroken takes.

Consequently, this project took a more straightforward time-based methodology for frame extraction. A film's frame was extracted each second using FFmpeg (Tomar, 2006). Thus, the extracted frames uniformly span the entire duration of the films, capturing both major and minor scenes, juxtapositions, transitions, and other visual narratives to ensure a comprehensive film summary. This made it relatively easy for each extracted frame to carry a timestamp that enables cross-referencing with the original film. At the end of the process, a total of 105,814 frames across the entire dataset were gathered. In the accompanying CSV file, each frame and its annotation are represented by a single row/observation.

Similar to the films themselves, the extracted frames were temporarily placed in an encrypted and password-protected storage to maintain data integrity and security. Upon completion of the frame extraction and subsequent annotation, as illustrated in the following step, all frames were subsequently deleted.

(2.3) STEP THREE: ANNOTATION OF FRAMES

The third step involved annotating the extracted frames using Google's Vision API. The API allows users to call it through their preferred programming/scripting language or through GUI applications that interface with it directly. For this project, Memespector-GUI was utilized (Chao, 2023). Although originally geared towards the study of social media content, the program's features are easily adaptable to film frames. Users need to simply specify their commercial API key and a folder containing the frame stills. Finally, they need to select which features/models they would like to invoke. For this project, safety, face, web, and label detection features were selected.

The result of the API call resulted in a CSV file containing the annotation of each frame. Only two small modifications were made to the output. First, those features not selected are returned as empty fields, and these were removed. Second, the original textual descriptions provided by the API for certain visual elements, specifically the likelihood of certain facial emotions, were modified to add a space for readability. For instance, by default, the API returns facial expressions as "VeryUnlikely," and this was changed to "Very Unlikely".

After the final compilation of the dataset, all video files and frames were deleted from the password-protected hard drive leaving only the CSV file. While this compromises reproducibility—an issue discussed at length later—it was necessary to ensure legal compliance with copyrighted material. That said, the steps outlined above, from the time-based frame extraction to the annotation, can be replicated by researchers who have access to the same video materials for verification or similar analysis.

(3) UNDERSTANDING THE ANNOTATIONS AND RESULTS

The final CSV file contains 25 columns that can broadly be categorized into the following five subcategories: film metadata, safety detection, face, and emotion detection, labels, and web entities/pages. Each has its own set of unique attributes that aid in the holistic interpretation of the frame. Below, I provide a detailed explanation of each category and elaborate on the specific details of each column belonging to it. A copy of the descriptions for each column is under the Zenodo Repository in the "Data Dictionary.md" file.

Film	The name of the film where the frame is from
Time	The time in seconds of the film where the frame was taken
Year	The year the film the frame is from was released

Table 1 Variables corresponding to information about film metadata for each frame in the dataset.

(3.1) FILM METADATA

The first three columns of the dataset have metadata about the frame including the film name, the timestamp of the frame in seconds (e.g., 300 represents that the frame is from the 5th minute of the film), and the release date of the film. These columns are likely uninteresting by themselves and will need to be combined with others to generate insights. While it is common to analyze genre in computational analysis of visual media, preexisting categories were purposefully omitted from this study. As film scholars have shown, film genres are not fixed categories and change over time, often influenced by societal and technological shifts (Altman, 1984; Deleyto, 2012; Gledhill, 2000; Jancovich, 2000; Klinger, 1984). Additionally, films frequently span multiple genres, making it challenging to categorize them accurately within the constraints of the data set. That said, future researchers can easily add these categories along with others as they deem fit.

GV_Safe_Adult	The Google Vision assessment of whether the frame contains adult content
GV_Safe_Spoof	The Google Vision assessment of whether the frame is likely to be a spoof or parody
GV_Safe_Medical	The Google Vision assessment of whether the frame contains medical content
GV_Safe_Violence	The Google Vision assessment of whether the frame contains violent content
GV_Safe_Racy	The Google Vision assessment of whether the frame contains racy content

Table 2 Variables corresponding to information about safety detection for each frame in the dataset.

(3.2) SAFETY DETECTION

The Google Vision API provides SafeSearch Detection to identify inappropriate content (Detect Explicit Content (SafeSearch), n.d.). Table 3 provides summary statistics of the likelihood of various categories across all the film frames. Through these results, researchers can filter and curate their dataset according to the context and requirements of their study. However, it must be noted that the API's judgment might not align with every researcher's definition of these terms, leading to potential discrepancies in interpretation. The implications of this and avenues for future research exploring this bias are detailed later in this article.

Table 3 Summary Statistics containing the likelihood of various Explicit Content classifications across all the film frames.

	GV_SAFE_ ADULT	GV_SAFE_ SPOOF	GV_SAFE_ MEDICAL	GV_SAFE_ VIOLENCE	GV_SAFE_ RACY
Very Unlikely	67000	77864	77976	39721	51368
Unlikely	34445	25180	26784	63402	39971
Possible	3082	2084	873	2336	11493
Likely	488	671	157	282	2338
Very Likely	98	14	23	72	643

Table 4 Variables corresponding to face and emotion detection for each frame in the dataset.

GV_Face_Joy	The Google Vision assessment of whether the facial expressions in the frame depict joy
GV_Face_Sorrow	The Google Vision assessment of whether the facial expressions in the frame depict sorrow
GV_Face_Anger	The Google Vision assessment of whether the facial expressions in the frame depict anger
GV_Face_ Surprise	The Google Vision assessment of whether the facial expressions in the frame depict surprise
GV_Face_ UnderExposed	The Google Vision assessment of whether the facial expressions in the frame are underexposed. This means the image could be too dark to accurately assess the emotions displayed on the face(s).
GV_Face_Blurred	The Google Vision assessment of whether the facial expressions in the frame are blurred. Blurriness can compromise the accuracy of emotion detection, so it's important to consider in an analysis
GV_Face_ Headwear	The Google Vision assessment of whether there is headwear present in the frame. Headwear can significantly impact the accuracy of emotion detection as it can obscure a portion of the face
GV_Face_Score	The Google Vision score of the overall clarity and quality of the facial expressions in the frame. A higher score indicates a clearer image, while a lower score may indicate possible issues like blurriness, underexposure, or obscured faces due to headwear.)

(3.3) FACE AND EMOTION DETECTION

The Google Vision API enables Face and Emotion Detection, which is particularly useful for film analysis. The tool can recognize faces within frames and label the detected facial expressions based on predefined categories like joy, sorrow, anger, and surprise. It additionally labels the likelihood of these emotions being present, equating to terms like "Very Likely," "Possible," "Unlikely," and "Very Unlikely." It is important to note that there can be multiple facial expressions in a single frame. In addition, the tool can detect how underexposed or overexposed an image is, the presence of headwear, and the blurriness of faces within frames. The confidence of the face detection is also provided under the "Face Score" column, giving researchers an estimation of the reliability of the detection results. Researchers must take a comprehensive account of the confidence and blurriness of the frame before making analytical claims about the emotional salience of a film. Table 5 below provides summary statistics of the likelihood of various facial emotions detected across all the film frames.

Table 5 Summary Statistics showcasing the likelihood of various facial emotions detected across all the film frames. Note that some film frames have multiple faces while others have none.

	GV_FACE_JOY	GV_FACE_SORROW	GV_FACE_ANGER	GV_SAFE_SURPRISE
Very Unlikely	169072	177071	182842	178276
Unlikely	6788	4731	817	2689
Possible	2982	1443	120	1035
Likely	2165	504	32	571
Very Likely	2806	64	2	242
NA	15882	15882	15882	15882

GV_Label_Descriptions	The Google Vision API's descriptions of the objects, activities, or concepts that are central to the frame. Multiple labels can be associated with a single fram.
GV_Label_Scores	The Google Vision score of the accuracy or relevance of the labels provided for the frame. Higher scores indicate that the label is likely highly relevant to the frame, while lower scores suggest less relevance or certainty.

Table 6 Variables corresponding to Google's Label detection for each frame in the dataset.

(3.4) GOOGLE LABELS

The API provides a feature named Label Detection which generates descriptive labels for the content in the image, and each label is returned with a confidence score indicating the degree of confidence for that label's accuracy. In some cases, the labels can represent high-level concepts or categories that the AI model has learned to identify, and analyzing these labels can provide valuable insights into the film's themes, settings, and visual styles. For instance, performing a frequency analysis of labels across a film or a dataset of films could reveal patterns and trends that would otherwise remain unnoticed. Figure 1 showcases the top 15 categories detected throughout all films.

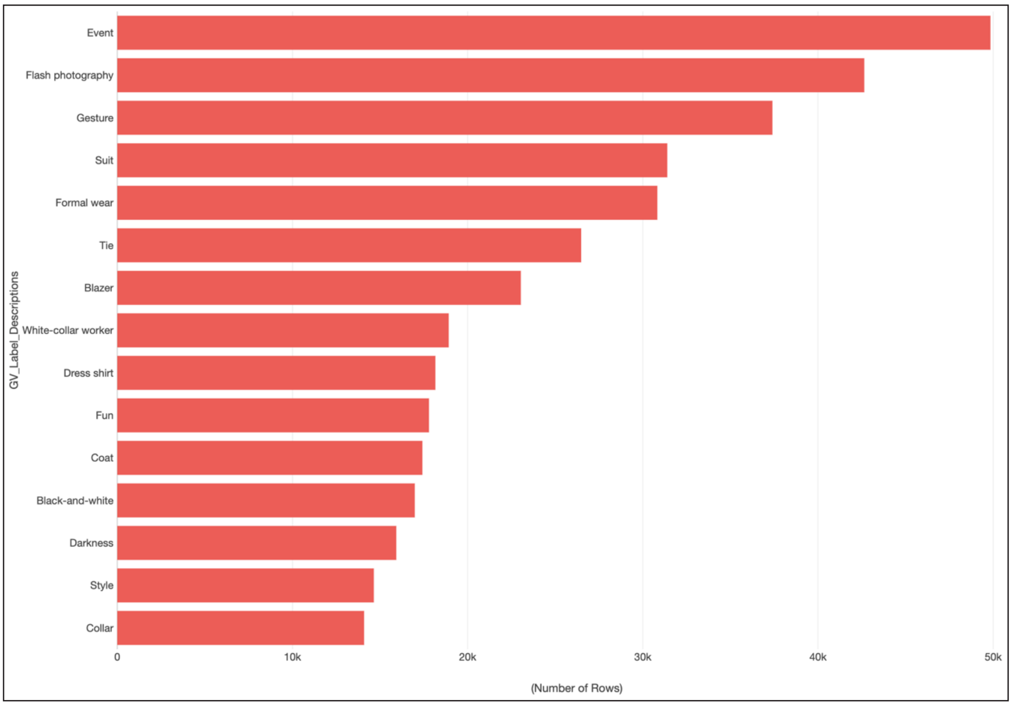

Figure 1 The top 15 labels detected by Google's Vision API throughout all fifteen films.

GV_Web_Entity_Descriptions	The Google Vision API's descriptions of any entities associated with the frame that can be found on the web. These might include names of actors, film titles, locations, or anything else that might be identified and linked to web-based information.
GV_Web_Entity_Scores	The Google Vision score of the accuracy or relevance of the web entities associated with the frame. Higher scores indicate that the entity is likely highly relevant to the frame, while lower scores suggest less relevance or certainty.
GV_Web_BestGuessLabels	The Google Vision API's best guess labels for the frame based on its content. These labels represent the API's highest confidence associations for the frame, based on all analyzed elements and data gathered from the web
GV_Web_FullMatchingImages	The Google Vision's output of any full matching images that can be found on the web
GV_Web_PagesWithFullMatchingImages	The Google Vision's output of any web pages that contain full matching images for the frame.
GV_Web_PartialMatchingImages	The Google Vision's output of any partially matching images that can be found on the web.
GV_Web_VisuallySimilarImages	The Google Vision's output of visually similar images that can be found on the web. This could include images that share similar colors, composition, subject matter, or other visual elements with the frame

Table 7 Variables corresponding to Google Web Entities and Web Pages for each frame in the dataset.

(3.5) GOOGLE WEB ENTITIES AND WEB PAGES

Perhaps the key advantage of Google's Vision API is its Web Entities and Web Pages detection, which cross-references identified objects and scenes with Google's Knowledge Graph. Through the API, these extracted entities are reflected in the dataset as labels and enhance the interpretative potential of the film frames. Rather than focusing purely on the visual content of the frame, this feature allows researchers to connect the on-screen elements to wider socio-cultural contexts. Figure 2 below showcases the top 15 web entities detected throughout all the film frames.

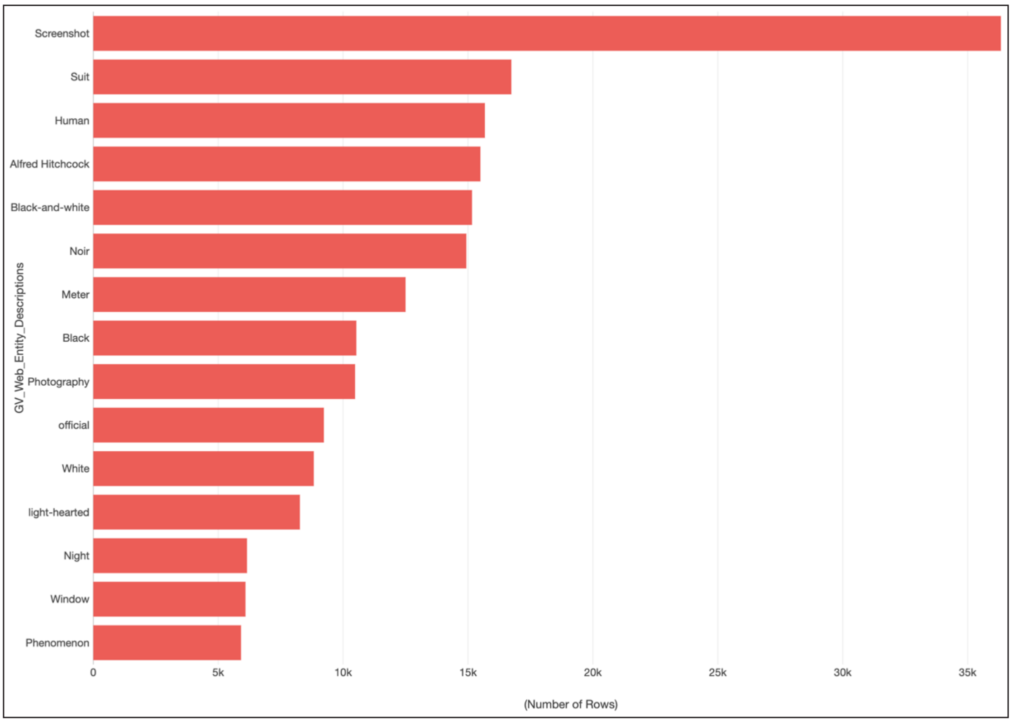

Figure 2 An analysis of the top 15 web entity labels based on Google's Knowledge Graph throughout all 15 films.

In addition to the Web Entities, the API provides the URL of web pages where the images or frames of the movie have appeared along with a direct link to the image itself. This holds immense value for copyright infringement cases and studying the dissemination of film discourse across the web by tracking how a film's visual material is interpreted and repurposed in different contexts.

(4) ISSUES OF REPRODUCIBILITY AND REPLICABILITY

One of the chief challenges with utilizing proprietary computer vision algorithms is issues of replicability and reproducibility. This is exacerbated by the fact that the films and film shots are under copyright and cannot easily be made available to researchers who do not have access to the original materials. Thus, it is worth exploring these matters in more depth and how they intersect with the Hitchcock Computer Vision Dataset before utilizing it for future research.

Issues of reproducibility and replicability are longstanding concerns in scientific research. In 2015, the Open Science Collaboration published a paper that attempted to reproduce 100 psychology studies and found that only 36% could be successfully redone with the same results (Open Science Collaboration, 2015). In 2016, an article in *Nature* surveyed fifteen hundred scientists and found that more than 70% of participants had failed to reproduce experiments conducted by other scientists while more than 50% had failed to reproduce their experiments (Baker, 2016). The results showed that 52% of those surveyed believed there was a significant reproducibility crisis. However, less than 31% reported that the failure to reproduce research meant that the original results were incorrect.

To address the issue of reproducibility and diverging perceptions of its significance, the US Congress, through the National Science Foundation, commissioned a team of academics "to define what it means to reproduce or replicate a study, explore issues related to reproducibility and replicability across science and engineering, and assess any impact of these issues on the public's trust in science" (Committee on Reproducibility and Replicability in Science, 2019) In their report, the commission noted the need to better define both terms. As they make apparent, reproducibility is tied directly with "computational reproducibility," and requires that code, data, and computational steps are provided to the greatest extent possible. Meanwhile, for research to be replicable, other researchers should be able to repeat the experiment and reach similar conclusions/results.

Yet, while the committee stressed computational reproducibility, it also noted that non-reproducibility and replicability can occur for a variety of reasons and do not necessarily reflect that data analysis, or the composition of a dataset is inconclusive. In particular, they contend that "some fields of scientific inquiry...involve complex data gathering from multiple sensors, modeling, and algorithms that cannot all be readily captured and made available for other investigators to reproduce" (Committee on Reproducibility and Replicability in Science, 2019). They also argue that proprietary information cannot always be made fully available for reasons regarding privacy, copyright, or legal constraints. Thus, reproducibility should be seen as a range and understood in the context of research or dataset creation. While every effort should be taken to make the creation of a dataset available, full reproducibility is not always achievable.

Regarding the Hitchcock Computer Vision Dataset, there are two key constraints to reproducibility that one should bear in mind before utilizing it for additional research. One, the Google Vision API used for generating tags and annotations is proprietary and constantly being updated. Hence, a different analysis could be generated depending on when a researcher calls the API. Additionally, since the API also retrieves information about where still shots of the images are located online, the results can change dynamically based on how the images move around the web. In the Zenodo repository, a "Readme.md" file contains the date of the API call, the Memespector-GUI version number, and system information to provide as close a replica of the data generation process as possible. Yet, one should keep in mind that this does not guarantee a perfect duplication.

Two, the images that the Vision API assesses are drawn from well-known Hitchcock films that remain under copyright. This limited open access complicates the communication of the base data, further hampering reproducibility. Legal requirements set by copyright holders prevent researchers from possessing or disseminating actual film frames, despite their relevance to the study involved. In the Zenodo dataset, the "Readme.md" file contains the MakeMKV version number, ffmpeg version number, system information, and the command line prompt for ffmpeg to get the still frames listed.

In short, when working with the Hitchcock Computer Vision Dataset, future researchers should bear in mind these nuances. They should recognize the specificity of the data, as well as the legal and proprietary constraints associated with its use. A detailed understanding of non-replicability and its justifications can facilitate improved error analysis when using this dataset. They should also note that non-replicability doesn't invalidate the entire dataset but calls for a thorough understanding of the variables at play in the study they hope to conduct.

(5) AVENUES FOR FUTURE RESEARCH

(5.1) STATISTICAL ANALYSIS OF FRAME CONTENT AND LABELS

One of the most valuable and relatively direct ways to gain insights from the data available through the Hitchcock Dataset is by performing statistical analyses on the content and labels generated for the film frames. This approach is akin to a detailed 'census' of the cinematographic elements, where every frame is methodically explored, and its visual features noted and categorized according to the labels the API provides. The goal is to extract patterns and discernible trends embedded within the dataset which may otherwise go unnoticed in qualitative analysis. For instance, an examination of recurring labels could reveal the filmmaker's use of symbolism and visual cues, offering insights into recurring visual patterns and motifs that contribute to the storytelling.

Comparing frequencies across a selection of films could likewise reveal larger trends and patterns. Perhaps films from a particular era show a distinctive bias toward certain themes, objects, or stylistic choices. The possibility of integrating temporal data into the analysis adds another dimension to the investigation. By noting when these visual elements occur in the timeline of the film, researchers can potentially trace the development of narratives, the evolution of themes, or the transformation of certain motifs throughout the film. These analyses could then play a vital role in comparative film studies, enabling researchers to discern broader patterns across the cinematic landscape.

Given its capability to identify facial expressions and emotions within frames, the Google Vision API opens the door for a "macroanalysis" of sentiment-derived plots. In literary analysis, packages, such as the Syuzhet, package have been used to plot the emotional trajectory of narratives over time (Jockers, 2023; Kim, 2022; Naldi, 2019; Rinker, 2021). Applying a similar approach to film studies, researchers can identify moments of joy, anger, sorrow, and surprise expressed by the characters. For instance, in Figure 3, the top seven films containing faces showing joy the Vision API determines is at least "Possible" are shown.

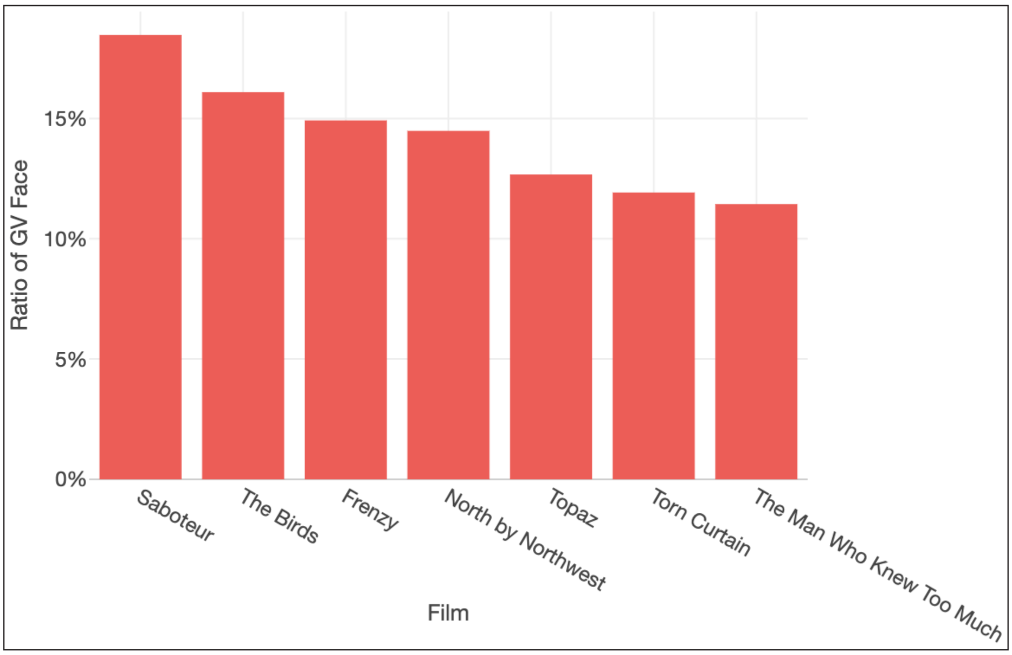

Figure 3 Top Seven Films Containing Ratio of Faces Showing Joy Detected as at least "Possible" by the Vision API.

(5.2) COMPARISON OF DIFFERENT COMPUTER VISION ALGORITHMS TO UNDERSTAND BIAS

Another potential avenue of research involves comparing performance and results from different computer vision algorithms in the context of film studies by joining the Google Vision results with those obtained from other commercial APIs, such as Microsoft's Azure Computer Vision, Amazon Rekognition, or Clarifai. In addition, comparisons can be made between open-source models, such as ResNet50 and VGG19 (He et al., 2016; Keras Team, n.d.; Simonyan & Zisserman, 2015). In film studies and digital humanities more broadly, these juxtapositions between models remain relatively unexplored and in need of further inquiry.

One particularly important area of research that needs further investigation is cultural bias in computer vision. Since computer vision models are trained on datasets that reflect the biases of their creators, these shape the perception and interpretation of the film frames they're applied to. These issues, if overlooked, could skew research outcomes and interpretations. For instance, the Vision APIs' categorization of emotion may not accurately reflect a film's emotional content. Emotional analysis detection has also been criticized for over-relying on norms constructed around Western, adult, able-bodied individuals, potentially sidelining non-standard expressions or outlying demographics (Rhue, 2018). Others have argued that emotional analysis is a new form of physiognomy due to its practice of using outer appearances to infer inner characteristics (Arcas, 2017).

Along with the emotional detection algorithm, explicit content filtering may also raise critical questions. Designed with certain societal standards often aligned to the technology's region of creation, explicit content filters can have a particular biased perspective on what is deemed inappropriate or explicit. Google's Vision API, for example, has been criticized for its conservative standards of indecency, resulting in over-censorship of non-heteronormative materials (Monea, 2023). For researchers using the Hitchcock Dataset or other similar datasets in film analysis, attention must be paid to the potential for both unintentional censoring of content and the narrowing of interpretive possibilities brought by these automated content moderators.

An exhaustive investigation into such biases is beyond the scope of this paper, but future researchers must conduct resources such as the Critical Dataset Studies Reading List to think critically about these inherent biases, incorporate mitigation strategies in their research design, and act with informed skepticism about these models' analytical outputs (Knowing Machines Team, n.d.).

(5.3) IMAGE NETWORKS OF MISE-EN-SCENE

One potentially fruitful approach for further investigation is to examine the co-occurrence of labels with frames using a bipartite network, which can shed light on the exploration of mise-en-scene. A bipartite network is a graph containing two types of nodes where nodes from the same set do not connect. In the case of interpreting films, one of these nodes would represent the elements detected, such as labels or facial emotions, while the other would represent the frame itself. The connection of two nodes of one type to the same node of another type represents a co-occurrence. These co-occurrences can be "projected" into two distinct monopartite networks – one for each node type – containing information about which elements frequently appear together within the same frame or which frames usually exhibit similar elements. For instance, Figure 4 showcases the co-occurrence of different web-entity labels with node size representing higher co-occurance counts. Centrality measures like degree, closeness, and betweenness can then be applied to these projected networks to rank the importance of certain elements or frames based on different characteristics. The resulting network visualization not only provides an overview of the relationships among elements within a film but also identifies key nodes that are central to the narrative structure or contribute significantly to the visual composition of the film.

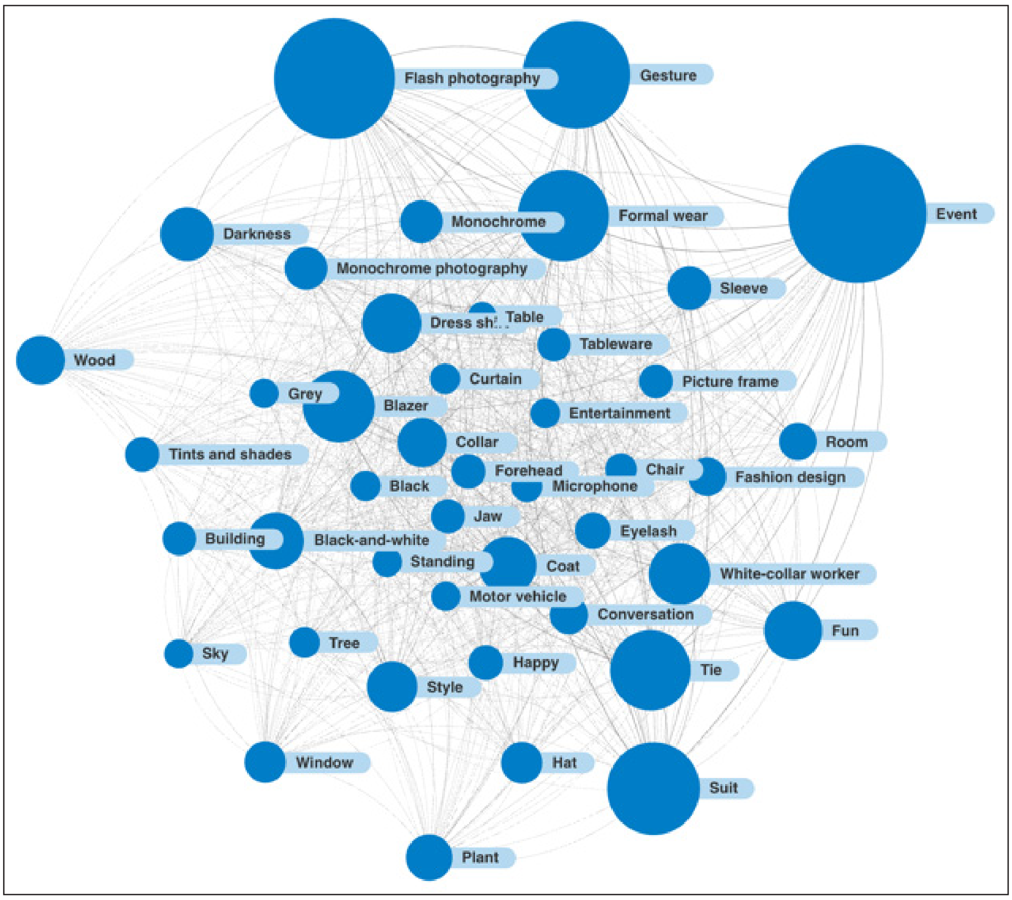

Figure 4 Monopartite Projection showcasing approximately the top 40 co-occurrences of web entities throughout all film frames.

Bipartite "computer vision networks" can be particularly beneficial even if not projected. Omena, et al. (2023) demarcates three different types of computer vision networks that utilize Google's Vision API in the study of digital content:

> *Image-Label Network*-A bipartite network where images and labels are represented as nodes while the connections represent the labels associated with images.
>
> *Image-Web Entities Network*-A bipartite network where images and web entities – concepts and categories detected in an image – are represented as nodes. Links between nodes indicate a connection between a specific image and a web entity. Figure 3 demonstrates the top.
>
> *Image-Domain Network*-A bipartite network where images and domains – the websites where images are found- are represented as nodes. Connections between nodes indicate that a specific image is associated with a particular domain.

All three of these networks can be used in the study of film with the only major change being the substitution of 'image' with 'frame' (i.e. frame-label networks, frame-web entities networks, and frame-domain networks.

One thing to note is that Omena et al. use "label" to refer specifically to the output of the Google Vision API's label detection feature, which identifies general objects, locations, and activities in a frame. However, for film scholars, labels should take a broader meaning referring to all mise-en-scene elements. Considering the unique context of film analysis, it may also be helpful to propose three additional computer vision networks that could offer valuable insight:

> *Frame-Emotion Network* – A bipartite network where individual frames and detected emotions are represented as nodes. The connections between nodes represent the emotions identified in each frame.
>
> *Frame-Explicit Content Network* – A bipartite network where individual frames and detected explicit content are represented as nodes. The connections between nodes symbolize the explicit content identified within each frame.
>
> *Frame-Landmark Network* – A bipartite network where individual frames and detected landmarks are represented as nodes. The connections between nodes denote the presence of specific landmarks within each frame.

In short, the application of computer vision bipartite networks in the analysis of mise-en-scène presents a potent instrument for deciphering the intricate interconnections between various elements contained within a film frame. This methodology enriches the study of mise-en-scène within the sphere of film analysis, facilitating the examination of the film's architectural framework and dynamic interplay of visual components. Furthermore, these networks provide a new avenue for a network science approach to film analysis that complements traditional qualitative methodologies.

(6) CONCLUSION

The Hitchcock Computer Vision Dataset presented in this article serves as a valuable resource for researchers in the fields of film studies and digital humanities. By extracting screen frames from a comprehensive collection of Alfred Hitchcock films, this dataset allows for an in-depth analysis of Hitchcock's stylistic techniques, narrative choices, and thematic elements. Moreover, it invites a reflexive approach towards the evolution of film studies as a discipline in the digital age.

Through the utilization of computer vision technology, this data-driven approach complements qualitative methods of film analysis. The rich metadata obtained from Google's Vision API, including object detection, facial recognition, web-entity analysis, and explicit content filtering, enhances researchers' understanding of the visual elements present within each frame. Furthermore, the dataset opens doors for further research in various areas of film studies and digital media concerning emotional arcs, the analysis of recurring visual motifs across different films, the identification of explicit content patterns, and the exploration of cultural symbols and landmarks in film narratives.

The Hitchcock Computer Vision dataset pushes the boundaries of film studies, showing that the merger of technology and art can lead to new, unexplored horizons that could further enrich our understanding and appreciation of cinema. Perhaps most importantly, it provides a blueprint for future research on other directors or film corpora, demonstrating the vast potential of using machine learning and computer vision in combination with traditional film analysis techniques systematically and comprehensively.

COMPETING INTERESTS

The author has no competing interests to declare.

AUTHOR AFFILIATIONS

Nabeel Siddiqui orcid.org/0000-0002-6126-5833
Communications Department, Susquehanna University, Selinsgrove, PA, USA

REFERENCES

Altman, R. (1984). A Semantic/Syntactic Approach To Film Genre. *Cinema Journal*, *23*(3), 7. DOI: https://doi.org/10.2307/1225093

Arcas, B. A. y. (2017, May 20). Physiognomy's New Clothes. *Medium*. https://medium.com/@blaisea/physiognomys-new-clothes-f2d4b59fdd6a

Arnold, T., & **Tilton, L.** (2019). Distant Viewing: Analyzing Large Visual Corpora. *Digital Scholarship in the Humanities*, *34*(Supplement 1), 3–16. DOI: https://doi.org/10.1093/llc/fqz013

Arnold, T., & **Tilton, L.** (2020). Distant Viewing Toolkit: A Python Package for the Analysis of Visual Culture. *Journal of Open Source Software*, *5*(45), 1800. DOI: https://doi.org/10.21105/joss.01800

Arnold, T., Tilton, L., & **Wigard, J.** (2022). Automatic Identification and Classification of Portraits in a Corpus of Historical Photographs. *Proceedings Ceur*, 25–35.

Asha Paul, M. K., Kavitha, J., & **Jansi Rani, P. A.** (2018). Key-frame extraction techniques: A review. *Recent Patents on Computer Science*, *11*(1), 3–16. DOI: https://doi.org/10.2174/2213275911666180719111118

Baker, M. (2016). 1,500 scientists lift the lid on reproducibility. *Nature*, *533*(7604), Article 7604. DOI: https://doi.org/10.1038/533452a

Becker, M., & **Desoky, A.** (2004). A study of the DVD content scrambling system (CSS) algorithm. *Proceedings of the Fourth IEEE International Symposium on Signal Processing and Information Technology*, 353–356. DOI: https://doi.org/10.1109/ISSPIT.2004.1433792

Belton, J. (2003). Can Hitchcock Be Saved from Hitchcock Studies? *Cineaste*, *28*(4), 16–21.

Calandrillo, S. P., & **Davison, E. M.** (2008). The dangers of the digital millennium copyright act: Much ado about nothing. *William and Mary Law Review*, *50*, 349.

Castellano, B. (2023). *PySceneDetect* (0.6.2) [Computer software]. https://www.scenedetect.com/copyright/

Champion, E. M. (2017). Digital humanities is text heavy, visualization light, and simulation poor. *Digital Scholarship in the Humanities*, *32*(Supplement 1), 25–32. DOI: https://doi.org/10.1093/llc/fqw053

Chao, J. (2023). *Memespector-GUI: Graphical User Interface Client for Computer Vision APIs* (0.2.5) [Computer software]. DOI: https://doi.org/10.5281/zenodo.7704877

Cobia, J. (2009). The digital millennium copyright act takedown notice procedure: Misuses, abuses, and shortcomings of the process. *Minnesota Journal of Law Science and Technology*, *10*, 387.

Colangelo, D. (2018). Hitchcock, Film Studies, and New Media: The Impact of Technology on the Analysis of Film. *Film Theory in Media History*, *127*. DOI: https://doi.org/10.2307/j.ctt1zqrmrh.10

Committee on Reproducibility and Replicability in Science. (2019). *Reproducibility and Replicability in Science*. National Academies Press. https://www.nap.edu/catalog/25303

Crump, C. (2022, August 16). *Publishers are blocking digital humanities research*. Berkeley News. https://news.berkeley.edu/2022/08/16/publishers-are-blocking-digital-humanities-research

Deleyto, C. (2012). Film Genres at the Crossroads: What Genres and Films Do to Each Other. In *Film Genre Reader Iv* (pp. 218–236). University of Texas Press. DOI: https://doi.org/10.7560/742055-019

Detect explicit content (SafeSearch). (n.d.). Google Cloud. Retrieved from https://cloud.google.com/vision/docs/detecting-safe-search (Last accessed: October 3, 2023).

di Lenardo, I., Seguin, B. L. A., & **Kaplan, F.** (2016). *Visual patterns discovery in large databases of paintings*. https://infoscience.epfl.ch/record/220638

Eskicioglu, A. M., & **Delp, E. J.** (2001). An overview of multimedia content protection in consumer electronics devices. *Signal Processing: Image Communication*, *16*(7), 681–699. DOI: https://doi.org/10.1016/S0923-5965(00)00050-3

Gefen, A., Saint-Raymond, L., & **Venturini, T.** (2021). AI for Digital Humanities and Computational Social Sciences. In B. Braunschweig & M. Ghallab (Eds.), *Reflections on Artificial Intelligence for Humanity* (pp. 191–202). Springer International Publishing. DOI: https://doi.org/10.1007/978-3-030-69128-8_12

Gianluigi, C., & **Raimondo, S.** (2006). An innovative algorithm for key frame extraction in video summarization. *Journal of Real-Time Image Processing, 1*(1), 69–88. DOI: https://doi.org/10.1007/s11554-006-0001-1

Gledhill, C. (2000). Rethinking Genre. In L. Williams & C. Gledhill (Eds.), *Reinventing film studies* (pp. 221–243). Bloomsbury.

Gruzman, I. S., & **Kostenkova, A. S.** (2014). Algorithm of scene change detection in a video sequence based on the three dimensional histogram of color images. *2014 12th International Conference on Actual Problems of Electronics Instrument Engineering (APEIE)*, 1–1. DOI: https://doi.org/10.1109/APEIE.2014.7040826

GuinpinSoft inc. (2023). *MakeMKV* (1.17.5) [Computer software]. GuinpinSoft inc. https://www.makemkv.com/download/

He, K., Zhang, X., Ren, S., & **Sun, J.** (2016). Deep Residual Learning for Image Recognition. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 770–778. DOI: https://doi.org/10.1109/CVPR.2016.90

Hu, R., Gayol, C. P., Odobez, J.-M., & **Gatica-Perez, D.** (2017). Analyzing and visualizing ancient Maya hieroglyphics using shape: From computer vision to Digital Humanities. *Digital Scholarship in the Humanities*, *32*(Supplement 2), 179–194. DOI: https://doi.org/10.1093/llc/fqx028

Huang, C., & **Wang, H.** (2019). A novel key-frames selection framework for comprehensive video summarization. *IEEE Transactions on Circuits and Systems for Video Technology*, *30*(2), 577–589. DOI: https://doi.org/10.1109/TCSVT.2019.2890899

Jancovich, M. (2000). "A Real Shocker": Authenticity, Genre and the Struggle for Distinction. *Continuum, 14*(1), 23–35. DOI: https://doi.org/10.1080/713657675

Jocher, G., Chaurasia, A., & **Qiu, J.** (2023). *YOLO by Ultralytics* (8.0.0) [Computer software]. https://github.com/ultralytics/ultralytics

Jockers, M. (2023). *syuzhet: Extracts Sentiment and Sentiment-Derived Plot Arcs from Text* (1.0.7) [Computer software]. https://cran.r-project.org/web/packages/syuzhet/index.html

Keras Team. (n.d.). *Keras documentation: Keras Applications*. Keras. Retrieved from https://keras.io/api/applications/ (Last accessed: October 3, 2023).

Kim, H. (2022). Sentiment Analysis: Limits and Progress of the Syuzhet Package and Its Lexicons. *Digital Humanities Quarterly*, *16*(2).

Klinger, B. (1984). 'Cinema/Ideology/Crititicism' Revisited: The Progressive Text. *Screen, 25*(1), 30–44. DOI: https://doi.org/10.1093/screen/25.1.30

Knowing Machines Team. (n.d.). *Critical Dataset Studies Reading List*. Retrieved from https://knowingmachines.org/reading-list (Last accessed: November 19, 2023).

Lipton, J. (2005). The law of unintended consequences: The Digital Millennium Copyright Act and interoperability. *Washington and Lee Law Review*, *62*, 487.

Lunney Jr, G. S. (2001). The death of copyright: Digital technology, private copying, and the digital millennium copyright act. *Virginia Law Review*, 813–920. DOI: https://doi.org/10.2307/1073857

Manovich, L. (2020). *Cultural Analytics*. MIT Press. DOI: https://doi.org/10.7551/mitpress/11214.001.0001

McPherson, T. (2009). Introduction: Media Studies and the Digital Humanities. *Cinema Journal*, *48*(2), 119–123. DOI: https://doi.org/10.1353/cj.0.0077

Meeks, E. (2013, July 26). Is Digital Humanities Too Text-Heavy? | Digital Humanities Specialist. *Stanford University Libraries*. https://dhs.stanford.edu/spatial-humanities/is-digital-humanities-too-text-heavy/

Meng, J., Wang, H., Yuan, J., & **Tan, Y.-P.** (2016). From keyframes to key objects: Video summarization by representative object proposal selection. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 1039–1048. DOI: https://doi.org/10.1109/CVPR.2016.118

Monea, A. (2023). I Know It When I See It: The Heteronormativity of Google's Safesearch. *Porn Studies*, *10*(2), 135–153. DOI: https://doi.org/10.1080/23268743.2022.2086163

Musik, C., & **Zeppelzauer, M.** (2018). Computer vision and the digital humanities: Adapting image processing algorithms and ground truth through active learning. *VIEW: Journal of European Television History and Culture*, *7*(14), 59–72. DOI: https://doi.org/10.18146/2213-0969.2018.jethc153

Naldi, M. (2019). *A review of sentiment computation methods with R packages* (arXiv:1901.08319). arXiv. http://arxiv.org/abs/1901.08319

Omena, J. J., Bitencourt, E., Chao, J., M. Flores, A. M., Sepúlveda, R., Draisci, L., Berg, A., Ngo, M., Leite, E., Molin, L., Brennan, M., Svoboda, J., Cruz, D., Qin, L., & **Du, Q.** (2023, February 18). Cross Vision-APIs Studies. Digital methodologies for understanding computer vision. *Digital Methods Initiative*. DOI: https://doi.org/10.13140/RG.2.2.33287.27040

Open Science Collaboration. (2015). PSYCHOLOGY. Estimating the reproducibility of psychological science. *Science (New York, N.Y.)*, *349*(6251). DOI: https://doi.org/10.1126/science.aac4716

Pricing | Cloud Vision API. (n.d.). *Google Cloud*. Retrieved September 29, 2023, from https://cloud.google.com/vision/pricing

Pustu-Iren, K., Sittel, J., Mauer, R., Bulgakowa, O., & **Ewerth, R.** (2020). Automated Visual Content Analysis for Film Studies: Current Status and Challenges. *Digital Humanities Quarterly*, *14*(4).

Radford, A., Kim, J. W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., Sastry, G., Askell, A., Mishkin, P., Clark, J., Krueger, G., & **Sutskever, I.** (2021). *Learning Transferable Visual Models From Natural Language Supervision* (arXiv:2103.00020). arXiv. DOI: https://doi.org/10.48550/arXiv.2103.00020

Reddy, B., & **Jadhav, A.** (2015). Comparison of Scene Change Detection Algorithms for Videos. *2015 Fifth International Conference on Advanced Computing & Communication Technologies*, 84–89. DOI: https://doi.org/10.1109/ACCT.2015.44

Redfern, N. (2023a). *Computational Film Analysis with R*. https://cfa-with-r.netlify.app/cfa

Redfern, N. (2023b). The average shot length and the ecological fallacy in Film Studies. *Humanities Bulletin*, *6*(1).

Resig, J. (2014). Using computer vision to increase the research potential of photo archives. *Journal of Digital Humanities*, *3*(2), 3–2.

Rhue, L. (2018). *Racial Influence on Automated Perceptions of Emotions* (SSRN Scholarly Paper 3281765). DOI: https://doi.org/10.2139/ssrn.3281765

Rinker, T. (2021). *sentimentr: Calculate Text Polarity Sentiment* (2.9.0) [Computer software]. https://cran.r-project.org/web/packages/sentimentr/index.html

Rogers, R. (2015). Digital Methods for Web Research. In *Emerging Trends in the Social and Behavioral Sciences* (pp. 1–22). John Wiley & Sons, Ltd. DOI: https://doi.org/10.1002/9781118900772.etrds0076

Sadiq, B. O., Muhammad, B., Abdullahi, M. N., Onuh, G., Muhammed, A. A., & **Babatunde, A. E.** (2020). Keyframe extraction techniques: A review. *ELEKTRIKA-Journal of Electrical Engineering*, *19*(3), 54–60.

Salt, B. (1974). Statistical style analysis of motion pictures. *Film Quarterly*, *28*(1), 13–22. DOI: https://doi.org/10.2307/1211438

Sayers, J. (2018). Introduction: Studying Media through New Media. In *The Routledge Companion to Media Studies and Digital Humanities* (pp. 1–6). Routledge. DOI: https://doi.org/10.4324/9781315730479-1

Sebastian, T., & **Puthiyidam, J. J.** (2015). A survey on video summarization techniques. *International Journal of Computer Applications*, *132*(13), 30–32. DOI: https://doi.org/10.5120/ijca2015907592

Simonyan, K., & **Zisserman, A.** (2015). *Very Deep Convolutional Networks for Large-Scale Image Recognition* (arXiv:1409.1556). arXiv. DOI: https://doi.org/10.48550/arXiv.1409.1556

Singhal, A. (2012, May 16). Introducing the Knowledge Graph: Things, Not Strings. *Google*. https://blog.google/products/search/introducing-knowledge-graph-things-not/

Smith, A. O., Tacheva, J., & **Hemsley, J.** (2022). Visual Semantics of Memes: (Re)Interpreting Memetic Content and Form for Information Studies. *Proceedings of the Association for Information Science and Technology*, *59*(1), 800–802. DOI: https://doi.org/10.1002/pra2.731

Smits, T., & **Ros, R.** (2021). Distant Reading 940,000 Online Circulations of 26 Iconic Photographs. *New Media & Society*, *25*(12), 3543–3572. DOI: https://doi.org/10.1177/14614448211049459

Sujatha, C., & **Mudenagudi, U.** (2011). A study on keyframe extraction methods for video summary. *2011 International Conference on Computational Intelligence and Communication Networks*, 73–77. DOI: https://doi.org/10.1109/CICN.2011.15

Tomar, S. (2006). Converting Video Formats with Ffmpeg. *Linux Journal*, *2006*(146), 10.

Tommasini, R., Ilievski, F., & **Wijesiriwardene, T.** (2023). IMKG: The Internet Meme Knowledge Graph. In C. Pesquita, E. Jimenez-Ruiz, J. McCusker, D. Faria, M. Dragoni, A. Dimou, R. Troncy, & S. Hertling (Eds.), *The Semantic Web* (Vol. 13870, pp. 354–371). Springer Nature Switzerland. DOI: https://doi.org/10.1007/978-3-031-33455-9_21

Wevers, M., & **Smits, T.** (2020). The Visual Digital Turn: Using Neural Networks to Study Historical Images. *Digital Scholarship in the Humanities*, *35*(1), 194–207. DOI: https://doi.org/10.1093/llc/fqy085

Wharton, R. (2013). Digital Humanities, Copyright Law, and the Literary. *Digital Humanities Quarterly*, *7*(1).

TO CITE THIS ARTICLE:
Siddiqui, N. (2024). Cutting the Frame: An In-Depth Look at the Hitchcock Computer Vision Dataset. *Journal of Open Humanities Data*, 10: 5, pp. 1–16. DOI: https://doi.org/10.5334/johd.163

Submitted: 04 October 2023
Accepted: 30 November 2023
Published: 11 January 2024

Journal of Open Humanities Data is a peer-reviewed open access journal published by Ubiquity Press.