

Working with Color: How Color Quantization Can Aid Researchers of Problematic Information

Nina Lutz*
ninalutz@uw.edu
University of Washington
Seattle, Washington, USA

Jordyn W. Padzensky*
jzensky@uw.edu
University of Washington
Seattle, Washington, USA

Joseph S. Schafer
schaferj@uw.edu
University of Washington
Seattle, Washington, USA

ABSTRACT

Analyzing large sets of visual media remains a challenging task, particularly in mixed-method studies dealing with problematic information and human subjects. Using AI tools in such analyses risks reifying and exacerbating biases, as well as untenable computational and cost limitations. As such, we turn to adopting geometric computer graphics and vision methods towards analyzing a large set of images from a problematic information campaign, in conjunction with human-in-the-loop qualitative analysis. We illustrate an effective case of this approach with the implementation of color quantization towards analyzing online hate image at the US-Mexico border, along with a historicist trace of the history of color quantization and skin tone scales, to inform our usage and reclamation of these methodologies from their racist origins. To that end, we scaffold motivations and the need for more researchers to consider the advantages and risks of reclaiming such methodologies in their own work, situated in our case study.

CCS CONCEPTS

- Human-centered computing → Social media; • Computing methodologies → Image processing; Computer vision.

KEYWORDS

Problematic information, color quantization, visual analysis, social media, rumor, online hate

ACM Reference Format:

Nina Lutz, Jordyn W. Padzensky, and Joseph S. Schafer. 2024. Working with Color: How Color Quantization Can Aid Researchers of Problematic Information. In *Companion of the 2024 Computer-Supported Cooperative Work and Social Computing (CSCW Companion '24), November 9–13, 2024, San Jose, Costa Rica*. ACM, New York, NY, USA, 8 pages. <https://doi.org/10.1145/XXXXXX.XXXXXX>

1 INTRODUCTION

Scholars of problematic information, such as misinformation and online hate campaigns, often rely on large corpuses of social media data. Many of these datasets have not been largely visual, although

*Both authors contributed equally to this research.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

CSCW Companion '24, November 9–13, 2024, San Jose, Costa Rica.

© 2024 Copyright held by the owner/author(s). Publication rights licensed to ACM.
ACM ISBN 979-8-4007-1114-5/24/11
<https://doi.org/10.1145/XXXXXX.XXXXXX>

we know that visual media has a key role in such information campaigns. At scale, analysis of visual media sets requires new methodologies to be brought into CSCW and social computing spaces. Artificial intelligence (AI) and machine learning (ML) methods hold great potential to aid in such analyses, yet have presented biases and analysis gaps, particularly towards marginalized groups [12]. In the context of misinformation and online hate, using these methods can become particularly problematized, as we do not want to reify or exacerbate underlying biases, nor miss out on key data points for how marginalized individuals are being targeted by hateful and misleading rhetoric. Additionally, many of these technologies are costly and computationally prohibitive.

As such, we look towards lighter weight, non-AI and non-ML methods from computer vision and graphics, to enable better exploration and analysis of image datasets in problematic information campaigns. Such methods vary, and include approaches such as intraset cosine similarity to compare how images in datasets deviate from one another quantitatively, as used in the study of text-to-image generator outputs in auditing their bias [21] to averaging images together to understand biases in large datasets [12].

We extend these approaches in our ongoing, mixed-methods, and interpretivist case study of visual media regarding the US-Mexico border crisis. To analyze and debunk this misinformation, we utilize these quantitative image analysis methods with human-in-the-loop qualitative coding cooperative workflows to understand the role of visual media in this particular problematic information campaign.

This particular contribution focuses on one method from this pipeline: color quantization, a method by which summative color palettes and color information can be derived from a digital image. We examine the history of this non-AI and non-ML method as applied to human subjects, and how we may incorporate it into our pipeline for analyzing datasets and better sampling of images for human qualitative coding. To this end, we ask two research questions:

- (1) RQ1: What insights can be gained from non-machine learning methodologies to the study of visual misinformation about the U.S.-Mexico Border?
- (2) RQ2: How can we carefully and thoughtfully reclaim these methods for critical and empirical research at scale of crises events involving people?

To answer these questions, we contribute the following:

- (1) A critical, historical informed framework of using color quantization in mixed-methods, human-in-the-loop pipelines
- (2) A use case of our color quantization implementation with empirical results from our case study

In doing so, we position this work as a critical inquiry, reclaiming a non-ML image analysis method for large-scale image collections.

In the wake of the AI summer [10], we push back and say that we do not believe these tools are fully equipped to handle the problems that image-based misinformation presents, and we instead take a human-in-the-loop approach with thoughtful implementation and integration of quantitative methods.

2 RELATED WORK

2.1 Visual Misinformation

The misinformation field has traditionally relied heavily on textual analysis of social media posts [7, 45]. But given the increase of visual-first social networks and technologies of AI-generated imagery, misinformation research – particularly research that follows interpretive traditions to understand the role of misinformation in large social and political contexts [30, 44, 50] – stands to benefit significantly from incorporating visual analysis at scale into this work to address some of these platforms' novel challenges [69].

Misinformation is not a new area of study, building off decades of research [26, 30, 57] of studying large rumorizing campaigns, now largely coined mis- and disinformation, with misinformation as the umbrella terminology of such misleading, propagating information narratives. Such studies of these large-scale online rumors combine methodologies from various fields such as communications, political science, sociology, and computer science. These mixed-methods, interpretivist traditions have been utilized to study large amounts of misinformation that can do true harm on and offline [42], often through the study of problematic information which may contain hate, violence, and explicit content to fuel these narratives and conspiracy theories [15, 31].

In CSCW and related spaces that study online misinformation, common techniques involve building graphs of networks spreading information to understand the interactions between bad actors and unwitting spreaders [4, 61] and utilizing qualitative coding and content analysis to draw out campaign themes [50, 62]. This occurs through a process of sensemaking deeply rooted in organizational studies [70], where audiences, but also researchers, make sense of campaigns of information [18, 36].

As humans cannot possibly code all posts related to a campaign, quantitative techniques such as natural language processing [47], network analyses [4], and engagement data visualization [27] are applied to get larger-scale trends about these narratives and aid with exploratory data analysis in this mixed-methods work [50]. There has been work in visual methods for such exploratory data analysis of image collections, such as the work of Lev Manovich, who draws insights and new representations from large collections of images [37]. Additionally, there is increasing interest in sensemaking and qualitative coding in visual media as related to crises, political events, and misinformation [32, 39]. However, there is a need for further work in this space.

Along with how visual media contributes to misinformation campaigns and trends within these collections, there is also a spectrum of “cheap fakes” to “deep fakes” [48] and the need to investigate human and AI manipulation and creation of media. The field of media forensics utilizes quantitative, computational methodologies to understand the tampering or synthesizing of digital visual media (photographs, videos, GIFs, etc) [5, 13, 20], doing just this. There is ample opportunity to bring more computational methods into

misinformation campaign studies to perform baseline media forensics, such as leveraging color quantization to see if an image has been edited by color space inconsistencies. We demonstrate this case later in this work.

2.2 Historicism in CSCW and Color Quantization

CSCW focuses on technologies and systems of cooperative and collaborative work that pre-date modern computers. Sensemaking and rumorizing have long histories before internet mis- and disinformation, and the idea color schemas and relationships to human skin long predate digital mixing and color spaces like RGB. We ground our work in historicism within CSCW, an area that has been identified as under-utilized [60], to understand how looking back at not just theories and underpinnings but uses of technology and methods can influence how we study and view phenomena today – and add a sensibility towards countering what we may see as present, progress, and novelty.

For many scholars in CSCW space, skin tone and imagery have come to a major head within the last decade, with advances such as scholars Joy Buolamwini and Timnit Gebru's pivotal “Gender Shades” paper exposing how computer vision algorithms incorporated into high stakes technical systems often ignore darker skin tones [12]. This utilized the Fitzpatrick scale to measure these discrepancies, but acknowledged its limitation but that it was the best and most reasonable measure available. In subsequent years, various studies have examined the biases against and problematization of skin tone categories and representations [24, 66].

However, such skin tone scales have a lineage tracing back to 18th-century anthropology, where Johann Blumenbach is often credited with being amongst the first to formally and academically introduce phenotypical skin tone categories [6, 53]. Blumenbach posited 5 racial groups and went on to become the father of scientific racism, proclaiming an evolutionary superiority of white people based on cranial geometry and color theory [6, 53]. This paved the way for the work of Von Luschan, who in the early 20th century developed the Hautfarbe Tafel [65] a tool for systematically identifying 36 skin tones by region, based on their deviation from whiteness, and went on to be used by the Nazis in WWII eugenics movements [16].

Long after scientific racism was less accepted, there was still a drive for categorization and summarization of human skin tones for medical and scientific purposes. The Fitzpatrick scale directly adopted the Von Luschan scale to make a scale of human skin tones with respect to skin cancer risks in 1975 [22]. This was later found to be flawed, leading to the under diagnosis of cancer in people with darker skin tones [22]. Regardless, scale has become ubiquitous, becoming the standard of emojis across many platforms [40]. Many CSCW and social computing scholars have studied its role in self-presentation and digital identity via emojis [51, 52, 69] and data visualization scholars have considered its repercussions on how human data is represented outside of emojis [19].

In the last few years, there have been advancements toward improving the representation and analysis of human skin tone shades, both within and outside of academic pursuits. Fenty's 2017 makeup launch with a record number of 40 foundation shades, particularly

servicing a variety of customers with darker skin, increased and has forever reset the standard for diversity of skin tone coverage in makeup lines [71]. Building off the flaws of the Fitzpatrick scale and answering the calls of bias against darker skin tones in machine learning and camera systems, the Monk Skin Tone Scale, in collaboration with Google, was developed as a new and improved standard of skintone annotation for machine systems, informed by social science and historical methods [41].

Such quantified skin tone classification schemas will never fully capture the range of human skin, but methodologies of quantification to begin such measurements may benefit investigations such as ours in identifying the potential presence of human subjects across large datasets, particularly towards capturing a larger variety of skin tones in such datasets to expose and understand the narratives against people of color.

2.3 Digital Color Quantization

Color quantization is the digital process by which distinct colors of image pixels are quantified into color spaces by representing them in 3D geometric vectors [11, 46]. As such, this process can find averages and trends in the colors of an image, and reduce the colors of an image for image compression and analysis.

This process occurs via projecting these vectors into geometric color spaces, such as RGB (red, green, and blue). The RGB space, derived from the trichromatic theory of human color perception, represents all colors as numerical functions of red, green, and blue and has become a technological standard in many digital and physical displays [49, 64]. However, this space can obscure the visibility of key colors, such as representing browns and pinks as mathematically similar, although they are dissimilar to human perception. Color spaces such as HSV utilize units of hue, saturation, and value to solve this, making all hues (i.e. yellow vs blue) as geometrically having their own planes, making analysis of pixel distributions easier at first glance, and more intuitive to human vision, optimizing its use for image editing and analysis [58].

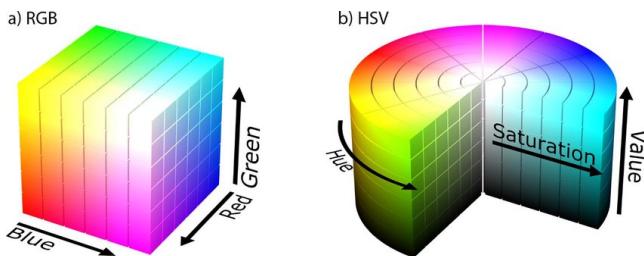


Figure 1: Showing color spaces in RGB and HSV geometries.
Figure credit: Michael Horvath [17].

In digital images, each pixel has an encoded color. In many programming libraries, like OpenCV, this color can be easily extracted in either RGB or HSV values [9]. Alternatively, these values can be converted between one another mathematically in code, by projecting RGB vectors to HSV space and vice versa [14]. In order to implement color quantization, these values are stored in one data structure – quantizing them from abstract pixels to an ordered collection of vectors. Analysis may be performed on this set,

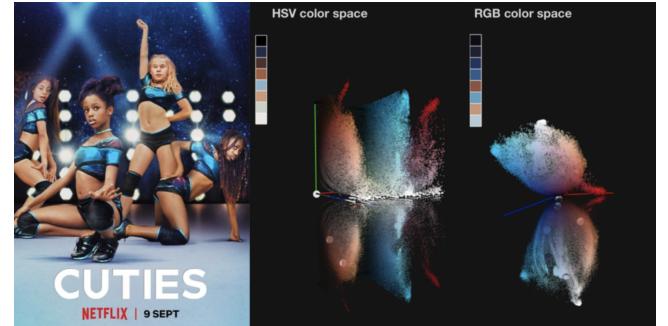


Figure 2: Differences between HSV and RGB spaces in image color quantization. While the summative palette remains the same, the left HSV space distinguishes the hues for more intuitive human summarization. Source; Lutz, 2020 [35].

commonly toward understanding a smaller representative subset of colors to describe this image. This can be done in many ways, such as finding global maximums in color histograms or utilizing k-means clustering and reporting out the cluster centroids [46].

3 METHODS

3.1 Color Quantization Implementation

We have implemented a color quantization protocol in Python to take in image(s) to aid analysis in our case study. This code offers three functionalities:

- (1) Constructs the color quantization image distributions and summative color palette with k-means clustering
- (2) Flags images if it is suspected that humans are within them based on the Monk skin tone scale
- (3) If a particular object or symbol of interest is likely to exist in the image

This functionality supports the analysis of large-scale image collections in projects like ours, which have humans in the loop and seek thematic analysis and descriptive statistics regarding the image collection for further analysis. The overall color quantization is a fairly standard implementation [11, 46]. It takes the following steps, the results of which are captured in Figure 3:

- (1) Image is uploaded in Python with the OpenCV library [9].
- (2) Converts each pixel to HSV color space, storing them in an ordered list, based on the pixel ordering from the default OpenCV structure [9].
- (3) K-means clustering, with a user identified k^1 , is then performed to cluster the colors, with the k centroids of these clusters saved to another list.
- (4) A distribution of pixels are then plotted in 3D space with the centroids visually denoted for quick debugging.

It is worth noting that the centroids themselves can also be returned alone, representing a summative color palette. Additionally, this code can either be batched or done on an image of choice for media forensics settings, explained in 4.3.

¹We recommend k in the range of 4-6 based on our trials and palette results.

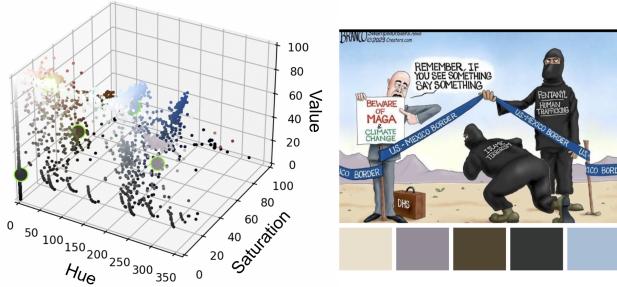


Figure 3: On the left, the pixel color distribution of the image on the right in HSV color space, with centroids plotted as larger points. The summative color palette for the image is below the image itself. This image was selected from our sample set of imagery on misinformation and online hate about the US-Mexico border rhetoric.

3.2 Human Subject or Chosen Object Implementation

To signify if a human or other object of interest may be in the image, we can search for pixel recurrences and ratios of certain colors, through a correlational search of particular color ratios in relation to one another [25, 28, 59]. This allows us to search extensive collections of images with a low weight, rapid query for ones that we suspect contain either human subjects or objects of interest. This method is lighter weight and removes potential biases by not relying on existing recognition approaches. However, this methodology will still have errors, which are acceptable in our implementation given its goal of being leveraged to aid our human qualitative coders to explore image collections.

To detect humans, we utilize the Monk Skin Tone scale color, bounding it by 15% on each dimensional side of the color in HSV space to adjust for shadows and lightning differences and capture the orb distribution from their scale for annotation [41]. These are meant to serve as tonalities, not exact skin shades as the Fenty 40 point scale would be [24], and thus more appropriate for more lighting conditions for our particular study, as per Monk et al.'s design of this scale as a guardrail and bucketing [24, 41]. Such approximations can help flag images that potentially incorporate human subjects for review. Given that our qualitative coding or analysis does not look for skin tones, these are not units of measure but rather a methodology to capture as many potential human subjects as possible.

For the detection of objects of key interest, we focus on symbols of hate groups for our particular research, which we sourced from the Global Extremist Symbols Database maintained by the nonprofit Global Project Against Hate and Extremism [1]. For any given image or symbol, either from the database or another object of interest, we run color quantization on it, soliciting summative palettes to search for within an image, as seen in Figure 6. We then utilize the same search approach as human subjects to flag if an image may contain an object of interest based on clusters of pixels being similar to the palette. We suspect that this approach could be used

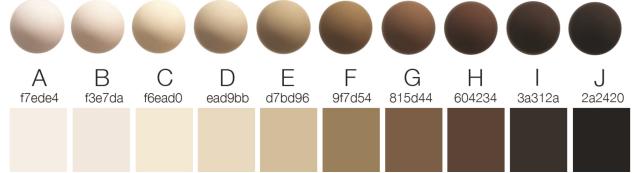


Figure 4: Monk scale and orb palettes. We use starting HEX colors and projected them into HSV value ranges with 15% bound on all axes permutations, to capture colors across the orb distribution and within our images. Figure credit: Dr. Ellis Monk [41].

for other kinds of key objects as well, but focus on hate symbols in our case study.

3.3 Mixed-Methods Human-in-the-Loop Content Analysis

This implementation of color quantization for summarization and exploration is meant for integration into a human-in-the-loop pipeline. In cases of problematic information campaigns, humans often qualitatively code samples of media as part of content analysis [38]. In our current study of harmful rhetoric at the US-Mexico border, a codebook has been inductively developed from the trending media we collected in this space. We utilize Label Studio [67], an open-source software similar to AtlasTI, to code our images with this codebook.

This color quantization allows us to better sample from our overall dataset, to get a more representative sample for human qualitative coders to identify the most salient points in this narrative based on our research questions, prioritizing how migrants are depicted in these narratives at the US-Mexico border. We can also sample for hate groups or particular objects of interest such that qualitative coders can explore if an image is misappropriated or has additional context.

Along with the qualitative coding samples, this method also allows us to analyze collections of images that are too large for coders to code, by outputting descriptive statistics and summary trends for the image collection.

4 RESULTS

Our research team has had success with this lightweight and flexible methodology towards supporting the collaborative work of qualitative coding of US-Mexico border rhetoric. It has allowed our qualitative coders to gather more relevant and representative samples of images for analysis, gather descriptive statistics and summaries, and incorporate media forensics into our work.

4.1 Human and Select Subjects of Interest

Color quantization, like any method, is not foolproof, and will have margins of error in over or under-selecting images. Our current case study found an error rate of approximately 18.4% for human subjects² in photographic images, with a tendency to over-select

²For our purposes, human subjects do not include far range crowd photos where individuals are not discernible, a common image that was excluded from this measurement.

images (false positives) and an error rate of 9.6% for relevant hate symbols in illustrative and photographic images. We arrive at this error bound from a test run on 2000 images in our dataset for 5 hate symbols.

Regardless, we have found this methodology helpful in a larger human-in-the-loop pipeline to study these large-scale campaigns towards identifying human subjects in images, as seen in Figure 5, or objects of interest, such as Figure 6. Such images are flagged by the code for their pixel relationships to either the Monk skin tones or an object of interest from our database.

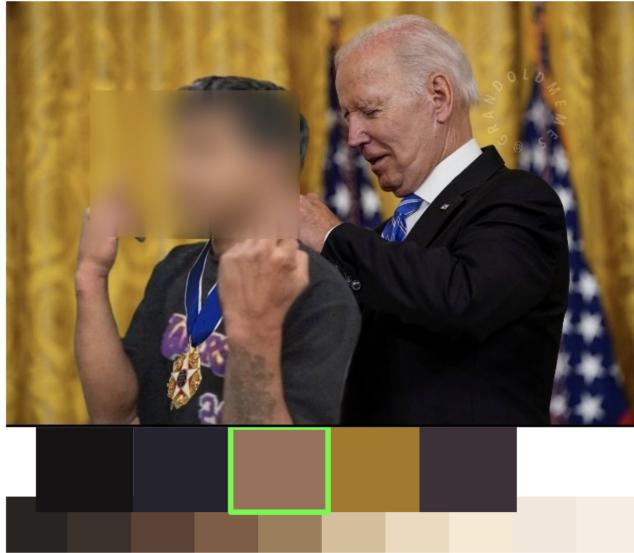


Figure 5: Summative palette extracted from Biden image, with the green highlighted color falling into the Monk Skin Tone Spectrum (below palette). Face blurred for privacy [54]. This image was Photoshopped and is not an accurate image, and has been used as part of an online hate campaign against Latin American migrants.

This method is not robust enough for a fully automated analysis. In our pipeline, to ensure nuance is added to these quantifiers, human coders verify and contextualize these images.

4.2 Human-in-the-loop Interaction

This code is lightweight and efficient with very few dependencies, making it easy to integrate into a larger pipeline. This setup is particularly useful in aiding a human-in-the-loop approach, with methods such as exploratory data analysis, descriptive statistics, and improving qualitative coding. Our team uses this mixed-methods pipeline, where exploratory data analysis is paramount. Using this method allows us to attain high-level summaries, such as the frequency by which certain subjects may appear, and allows us to leverage color quantization data to cluster images into visually similar groupings for future study and sampling. All this aids the qualitative coding experience by giving human coders a better bird's-eye view of their data than what exploratory text methods such as word clouds may otherwise give.



Figure 6: Image on right flagged for hate imagery due to clusters of yellow pixels around proportionately more black and black adjacent pixels, as per the color distribution from the logo taken from the Global Extremist Symbols Database [1]. Eyes blurred for privacy [54].

4.3 Media Forensics

There are times in such human-in-the loop projects where team members may need to prove if an image was altered, such as to explicitly call out misleading and harmful actors. There are a plethora of media forensics methods for detecting image manipulation, aforementioned in 2.1. Color quantization can also be a media forensics tool, by seeing if regions of an image match the color distribution to see if they are from the same image or if the image has been tampered with. Figure 7 features an image that qualitative coders marked as tampered. However, analysis of the color quantization and color channels show that the last line of text was added to this image and is not from this same image, given that its color space is drastically different than the other lines. This means this line of text could not have been physically made with the same marker, lighting, and surface, supporting the qualitative observation.

5 DISCUSSION

There is an increased and important need for critical theory and methods regarding the treatment of people of color in extensive and complex sociotechnical systems powered by AI and emerging computational technologies [43]. Many of these models and methods reify biases and exacerbate harms in working and environmental conditions [3, 12, 23, 56]. Thus, in our investigation, we argue for CSCW researchers seeking methods to parse large collections of data, particularly of human subjects, to reflect and apply criticality and historicism [60] and ask if they need to use blackbox tools and models over lighter-weight, physics-informed approaches such as

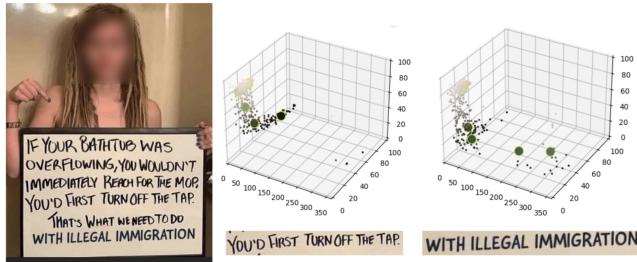


Figure 7: Different color spaces of the lines in this image show quantitative evidence that the last line could not be physically from the photo, supporting the qualitative observation. Face blurred for privacy [54]

color quantization. We posit such methods as a beneficial alternative in research regarding sensitive and marginalized populations, such as problematic information and hate content. However, these solutions are not without their own flaws, and studies using these lighter-weight techniques must still proceed with rigorous care for studied populations.

While these visual methods may not rely on current issues around AI models and their climate costs, they are derived from dark histories which have still become ubiquitous, such as adoption in emojis [40]. What does it mean to reclaim and discuss methodologies used to historically harm and marginalize people of color? What does it mean to reclaim them when in the backdrop, they have become so baked into technology and visual culture? When we use this method to study xenophobic misinformation at the US-Mexico border, we are truly utilizing a point of a lineage that once aided, justified, and made scientific racism de facto and de jour, and allowed the colonization and socioeconomic disenfranchisement of individuals that are now migrating largely from the Global South to the Global North. Therefore, reclaiming the method of color quantization is not simple. As Audre Lorde cautions against utilizing methods of domination to dissemble it, which is stated in her oft-cited warning “The master’s tools will never dismantle the master’s house.” [33].

Such challenges remain open for further scholarship and development, alongside the assumption that there is a need to perform color quantization and summation of human skin tones themselves. However, we hope this application and intellectual contribution will pave forward a path to challenge CSCW scholars to step back and interrogate alternatives to AI-enabled analysis of visual images, along with the complexities of these sociotechnical alternatives.

Along with the criticality of such approaches to datasets, there comes a point where the value of humans in the loop becomes paramount. For decades, qualitative, human coding has been used in conjunction with computational methods to bring rich insights around how people interact at scale in digital spaces [8, 62, 68]. In our study, this approach allows images to be flagged for qualitative coders as potential images with human subjects or objects of interest. This allows for better sampling but also for better analysis, where researchers may be able to consider new questions in such datasets. Furthermore, in cases of extreme hate content, this

methodology may assist in protecting qualitative researchers from the worst of the worst in their data [29, 55].

6 LIMITATIONS

Color quantization, and by extension geometric pixel matching, is a convenient method for these types of tasks in many ways – it runs fast and it is fully dependent on geometry and color spaces as opposed to models of object detection or AI tools. However, for this particular use case, this method could benefit from some select model implementations, such as technologies like MediaPipe [34] which can identify humans with greater accuracy or other object recognition approaches [2, 63, 72]. Color quantization is additionally limited in identifying human subjects where no skin is visible, where other geometric methods may thrive. There are also limitations in objects of interest, such as “alternate color” logos, which may have different color spaces. There are other object detection methods that can be implemented with machine learning or other graphics methodologies which may serve better for identifying objects in such cases where our system may struggle. Regardless, this method provides an extremely durable, lightweight, and useful way for researchers to rapidly explore large amounts of images in studying problematic information campaigns.

7 CONCLUSION AND FUTURE WORK

We have presented our implementation of color quantization for use in a mixed-methods human-in-the-loop pipeline to study the visual rhetoric of problematic information campaigns, using empirical evidence and test results from our case study of the visual media surrounding the U.S.-Mexico Border on social media. We have demonstrated the functionality of this methodology and how it enhances our collaborative workflow within our research team for insights and analysis regarding a large visual dataset. Furthermore, we have shown the critical and theoretical merit of incorporating historic visual methods, which have typically been previously used to invoke harm or have been radically exclusionary.

Towards our own work, we will continue iterating and enhancing such methodologies into our pipelines, by incorporating methods with color quantization but also beyond it. Going forward, we hope that more researchers not only explore visual information campaigns online in mixed methods ways, but also consider the histories and careful implementations and limits of their methods. It is our sincere ambition that such inquiries will develop more rich visual research methodologies and theories at the intersection of CSCW and visual culture.

8 ACKNOWLEDGEMENTS

We are grateful to Professor Kate Starbird for supporting this work. We are also grateful to our colleagues involved in this larger research project from whom we received feedback: Katie Arriaga, Mert Bayar, Wilson Chen, Dominic Monteperto, Logan Tuttle, and Tiffany Yan. Nina Lutz has been supported in conducting this by an NSF SaTC Grant (No. 2120496). Joseph S. Schafer has been supported by an NSF Graduate Research Fellowship (No. DGE-2140004). Any opinions, findings, conclusions, or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation or other funders.

REFERENCES

- [1] [n. d.]. Hate Symbol Directory. <https://globalextremism.org/global-extremist-symbols-database/>
- [2] Rakhsith L. A, Anusha K. S, Karthik B. E, Arun Nithish D, and Kishore Kumar V. 2021. A Survey on Object Detection Methods in Deep Learning. In *2021 Second International Conference on Electronics and Sustainable Communication Systems (ICESC)*. 1619–1626. <https://doi.org/10.1109/ICESC51422.2021.9532809>
- [3] Lasse F Wolff Anthony, Benjamin Kanding, and Raghavendra Selvan. 2020. Carbontracker: Tracking and predicting the carbon footprint of training deep learning models. *arXiv preprint arXiv:2007.03051* (2020).
- [4] Andrew Beers, Joseph S. Schafer, Ian Kennedy, Morgan Wack, Emma S. Spiro, and Kate Starbird. 2023. Followback Clusters, Satellite Audiences, and Bridge Nodes: Coengagement Networks for the 2020 US Election. *Proceedings of the International AAAI Conference on Web and Social Media* 17 (June 2023), 59–71. <https://doi.org/10.1609/icwsm.v17i1.22126>
- [5] Kratika Bhagtnani, Amit Kumar Singh Yadav, Emily R. Bartusiak, Ziyue Xiang, Ruiting Shao, Sriram Baireddy, and Edward J. Delp. 2022. An Overview of Recent Work in Media Forensics: Methods and Threats. <https://doi.org/10.48550/arXiv.2204.12067> arXiv:2204.12067 [cs].
- [6] Raj Bhupal. 2007. The beautiful skull and Blumenbach's errors: the birth of the scientific concept of race. *BMJ : British Medical Journal* 335, 7633 (Dec. 2007), 1308–1309. <https://doi.org/10.1136/bmj.39413.463958.80>
- [7] Nadya Bliss, Elizabeth Bradley, Joshua Garland, Filippo Menczer, Scott W. Ruston, Kate Starbird, and Chris Wiggins. 2020. An Agenda for Disinformation Research. <https://doi.org/10.48550/arXiv.2012.08572> arXiv:2012.08572 [cs].
- [8] Prashant Bordia and Nicholas Difonzo. 2004. Problem Solving in Social Interactions on the Internet: Rumor As Social Cognition. *Social Psychology Quarterly* 67, 1 (March 2004), 33–49. <https://doi.org/10.1177/01902750406700105> Publisher: SAGE Publications Inc.
- [9] G. Bradski. 2000. The OpenCV Library. *Dr. Dobb's Journal of Software Tools* (2000).
- [10] Amy Bruckman. 2024. Surviving the AI Summer. <https://asbruckman.medium.com/surviving-the-ai-summer-64626e5547e3>
- [11] Luc Brun and Alain Tréneau. 2017. Color quantization. In *Digital color imaging handbook*. CRC press, 589–637.
- [12] Joy Buolamwini and Timnit Gebru. 2018. Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification. In *Proceedings of the 1st Conference on Fairness, Accountability and Transparency*. PMLR, 77–91. <https://proceedings.mlr.press/v81/buolamwini18a.html> ISSN: 2640-3498.
- [13] Rainer Böhme, Felix C. Freiling, Thomas Gloe, and Matthias Kirchner. 2009. Multimedia Forensics Is Not Computer Forensics. In *Computational Forensics (Lecture Notes in Computer Science)*. Zeno J. M. H. Gerards, Katrin Y. Franke, and Cor J. Veenman (Eds.). Springer, Berlin, Heidelberg, 90–103. https://doi.org/10.1007/978-3-642-03521-0_9
- [14] Vladimir Chernov, Jarmo Alander, and Vladimir Bochko. 2015. Integer-based accurate conversion between RGB and HSV color spaces. *Computers & Electrical Engineering* 46 (2015), 328–337.
- [15] Matteo Cinelli, Andraž Pelicon, Igor Mozetič, Walter Quattrociocchi, Petra Kralj Novak, and Fabiana Zollo. 2021. Dynamics of online hate and misinformation. *Scientific Reports* 11, 1 (Nov. 2021), 22083. <https://doi.org/10.1038/s41598-021-01487-w>
- [16] United States Holocaust Memorial Museum Collection. [n. d.]. Tool used to classify skin color in racial studies conducted in Nazi Germany. <https://collections.ushmm.org/search/catalog/irn564926>
- [17] Wikimedia Commons. 2020. File:RGBCube_b.svg – Wikimedia Commons, the free media repository. https://commons.wikimedia.org/w/index.php?title=File:RGBCube_b.svg&oldid=452251057 [Online; accessed 23-May-2024].
- [18] Dharmा Dailey and Kate Starbird. 2015. “It’s Raining Dispersants”: Collective Sensemaking of Complex Information in Crisis Contexts. In *Proceedings of the 18th ACM Conference Companion on Computer Supported Cooperative Work & Social Computing*. ACM, Vancouver BC Canada, 155–158. <https://doi.org/10.1145/2685553.2698995>
- [19] Priya Dhawka, Helen Ai He, and Wesley Willett. 2023. We are the Data: Challenges and Opportunities for Creating Demographically Diverse Anthropographies. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. ACM, Hamburg Germany, 1–14. <https://doi.org/10.1145/3544548.3581086>
- [20] Hany Farid. 2020. Image Forensics. In *Computer Vision*. Springer International Publishing, Cham, 1–10. https://doi.org/10.1007/978-3-030-03243-2_877-1
- [21] Sourojit Ghosh and Aylin Caliskan. 2023. ‘Person’ == Light-skinned, Western Man, and Sexualization of Women of Color: Stereotypes in Stable Diffusion. In *Findings of the Association for Computational Linguistics: EMNLP 2023*. Houda Bouamor, Juan Pino, and Kalika Bali (Eds.). Association for Computational Linguistics, Singapore, 6971–6985. <https://doi.org/10.18653/v1/2023.findings-emnlp.465>
- [22] P Goon, C Banfield, O Bello, and NJ Levell. 2021. Skin cancers in skin types IV–VI: does the Fitzpatrick scale give a false sense of security? *Skin health and disease* 1, 3 (2021), e40.
- [23] Mary L Gray and Siddharth Suri. 2019. *Ghost work: How to stop Silicon Valley from building a new global underclass*. Eamon Dolan Books.
- [24] Courtney M. Heldreth, Ellis P. Monk, Alan T. Clark, Candice Schumann, Xango Eye, and Susanna Ricco. 2024. Which Skin Tone Measures Are the Most Inclusive? An Investigation of Skin Tone Measures for Artificial Intelligence. *ACM Journal on Responsible Computing* 1, 1 (March 2024), 7:1–7:21. <https://doi.org/10.1145/3632120>
- [25] Jing Huang, S. Ravi Kumar, Mandar Mitra, Wei-Jing Zhu, and Ramin Zabih. 1999. Spatial Color Indexing and Applications. *International Journal of Computer Vision* 35, 3 (Dec. 1999), 245–268. <https://doi.org/10.1023/A:1008108327226>
- [26] Caroline Jack. 2017. Lexicon of lies: Terms for problematic information. (2017).
- [27] Jasser Jasser. 2019. Dynamics of Misinformation Cascades. In *Companion Proceedings of The 2019 World Wide Web Conference (WWW '19)*. Association for Computing Machinery, New York, NY, USA, 33–36. <https://doi.org/10.1145/3308560.3314194>
- [28] TERENCE J Keating, PR Wolf, and FL Scarpace. 1975. An improved method of digital image correlation. *Photogrammetric Engineering and Remote Sensing* 41, 8 (1975), 993–1002.
- [29] Brit Kelley and Stephanie Weaver. 2020. Researching People who (Probably) Hate You: When Practicing “Good” Ethics Means Protecting Yourself. *Computers and Composition* 56 (June 2020), 102567. <https://doi.org/10.1016/j.compcom.2020.102567>
- [30] Zarine Kharazian, Madeline Jalbert, Saloni Dash, Shahan Ali Memon, Kate Starbird, Emma Spiro, and Jevin West. 2024. Our field was built on decades-old bodies of research across a range of disciplines. It wasn’t invented by a ‘class of misinformation experts’ in 2016. <https://www.cip.uw.edu/2024/01/24/misinformation-field-research/>
- [31] Jae Yeon Kim and Aniket Kesari. 2021. Misinformation and Hate Speech: The Case of Anti-Asian Hate Speech During the COVID-19 Pandemic. *Journal of Online Trust and Safety* 1, 1 (Oct. 2021). <https://doi.org/10.54501/jots.vii.13> Section: Peer-reviewed Articles.
- [32] Sukwon Lee, Sung-Hee Kim, Ya-Hsin Hung, Heidi Lam, Youn-Ah Kang, and Ji Soo Yi. 2016. How do People Make Sense of Unfamiliar Visualizations?: A Grounded Model of Novice’s Information Visualization Sensemaking. *IEEE Transactions on Visualization and Computer Graphics* 22, 1 (2016), 499–508. <https://doi.org/10.1109/TVCG.2015.2467195>
- [33] Audre Lorde. 2018. *The master's tools will never dismantle the master's house*. Penguin UK.
- [34] Camillo Lugaresi, Jiuqiang Tang, Hadon Nash, Chris McClanahan, Esha Ubowea, Michael Hays, Fan Zhang, Chuo-Ling Chang, Ming Yong, Juhyun Lee, Wan-Teh Chang, Wei Hua, Manfred Georg, and Matthias Grundmann. 2019. MediaPipe: A Framework for Perceiving and Processing Reality. In *Third Workshop on Computer Vision for AR/VR at IEEE Computer Vision and Pattern Recognition (CVPR) 2019*. https://mixedreality.cs.cornell.edu/s/NewTitle_May1_MediaPipe_CVPR_CV4ARVR_Workshop_2019.pdf
- [35] Nina Lutz. 2020. ‘Cuties’, Netflix, and Computation. <https://ninalutz.medium.com/cuties-netflix-and-computation-colors-b89a3ec302b1>
- [36] Lena Mamýkina, Drashko Nakikj, and Noémie Elhadad. 2015. Collective Sensemaking in Online Health Forums. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems (Seoul, Republic of Korea) (CHI '15)*. Association for Computing Machinery, New York, NY, USA, 3217–3226. <https://doi.org/10.1145/2702123.2702566>
- [37] Lev Manovich. 2020. *Chapter 10 - media visualization*. The MIT Press. <https://doi.org/10.7551/mitpress/11214.001.0001>
- [38] Philipp Mayring. 2015. Qualitative content analysis: Theoretical background and procedures. *Approaches to qualitative research in mathematics education: Examples of methodology and methods* (2015), 365–380.
- [39] Juan Carlos Medina Serrano, Orestis Papakyriakopoulos, and Simon Hegelich. 2020. Dancing to the Partisan Beat: A First Analysis of Political Communication on TikTok. In *12th ACM Conference on Web Science*. ACM, Southampton United Kingdom, 257–266. <https://doi.org/10.1145/3394231.3397916>
- [40] Kate M. Milner. 2021. “One part politics, one part technology, one part history”: Racial representation in the Unicode 7.0 emoji set. *New Media & Society* 23, 3 (March 2021), 515–534. <https://doi.org/10.1177/1461444819899623> Publisher: SAGE Publications.
- [41] Ellis Monk. 2023. The Monk Skin Tone Scale. (May 2023). <https://doi.org/10.31235/osf.io/pdf4c> Publisher: OSF.
- [42] Sadiq Muhammed T and Saji K. Mathew. 2022. The disaster of misinformation: a review of research in social media. *International Journal of Data Science and Analytics* 13, 4 (2022), 271–285. <https://doi.org/10.1007/s41060-022-00311-6>
- [43] Ihudiya Finda Ogbonnaya-Ogburu, Angela D.R. Smith, Alexandra To, and Kentaro Toyama. 2020. Critical Race Theory for HCI. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems (CHI '20)*. Association for Computing Machinery, New York, NY, USA, 1–16. <https://doi.org/10.1145/3313831.3376392>
- [44] Onook Oh, Manish Agrawal, and H. Raghav Rao. 2013. Anxiety and rumor: exploratory analysis of Twitter posts during the Mumbai terrorist attack. In *Rumor and Communication in Asia in the Internet Age*. Routledge, 143–155.

- <https://api.taylorfrancis.com/content/chapters/edit/download?identifierName=doi&identifierValue=10.4324/9780203403327-9&type=chapterpdf>
- [45] Onook Oh, Manish Agrawal, and H. Raghu Rao. 2013. Community Intelligence and Social Media Services: A Rumor Theoretic Analysis of Tweets During Social Crises. *MIS Quarterly* 37, 2 (2013), 407–426. <http://www.jstor.org/stable/43825916> Publisher: Management Information Systems Research Center, University of Minnesota.
- [46] M.T. Orchard and C.A. Bouman. 1991. Color quantization of images. *IEEE Transactions on Signal Processing* 39, 12 (Dec. 1991), 2677–2690. <https://doi.org/10.1109/78.107417>
- [47] Ray Oshikawa, Jing Qian, and William Yang Wang. 2020. A Survey on Natural Language Processing for Fake News Detection. In *Proceedings of the Twelfth Language Resources and Evaluation Conference*, Nicoletta Calzolari, Frédéric Béchet, Philippe Blache, Khalid Choukri, Christopher Cieri, Thierry Declerck, Sara Goggi, Hitoshi Isahara, Bente Maegaard, Joseph Mariani, Hélène Mazo, Asuncion Moreno, Jan Odijk, and Stelios Piperidis (Eds.). European Language Resources Association, Marseille, France, 6086–6093. <https://aclanthology.org/2020.lrec-1.747>
- [48] Britt Paris and Joan Donovan. 2019. Deepfakes and Cheap Fakes. <https://datasociety.net/library/deepfakes-and-cheap-fakes/> Publisher: Data & Society Research Institute.
- [49] Ralph W. Pridmore. 2011. Complementary colors theory of color vision: Physiology, color mixture, color constancy and color perception. *Color Research & Application* 36, 6 (2011), 394–412. <https://doi.org/10.1002/col.20611> arXiv:<https://onlinelibrary.wiley.com/doi/pdf/10.1002/col.20611>
- [50] Stephen Prochaska, Kayla Duskin, Zarine Kharazian, Carly Minow, Stephanie Blucker, Sylvie Venuto, Jevin D. West, and Kate Starbird. 2023. Mobilizing Manufactured Reality: How Participatory Disinformation Shaped Deep Stories to Catalyze Action during the 2020 U.S. Presidential Election. *Proceedings of the ACM on Human-Computer Interaction* 7, CSCW1 (April 2023), 1–39. <https://doi.org/10.1145/3579616>
- [51] Alexander Robertson, Walid Magdy, and Sharon Goldwater. 2018. Self-Representation on Twitter Using Emoji Skin Color Modifiers. *Proceedings of the International AAAI Conference on Web and Social Media* 12, 1 (June 2018). <https://doi.org/10.1609/icwsm.v12i1.15055> Number: 1.
- [52] Alexander Robertson, Walid Magdy, and Sharon Goldwater. 2021. Black or White but Never Neutral: How Readers Perceive Identity from Yellow or Skin-toned Emoji. *Proceedings of the ACM on Human-Computer Interaction* 5, CSCW2 (Oct. 2021), 350:1–350:23. <https://doi.org/10.1145/3476091>
- [53] Nicolaas Rupke and Gerhard Lauer. 2018. *Johann Friedrich Blumenbach: race and natural history, 1750–1850*. Routledge.
- [54] Joseph S. Schafer, Brett A. Halperin, Sourojit Ghosh, and Julie Vera. 2024. TO SCREENTSHOT OR NOT TO SCREENTSHOT? TENSIONS IN REPRESENTING VISUAL SOCIAL MEDIA PLATFORM POSTS. In *AoIR Selected Papers of Internet Research*. Sheffield, United Kingdom.
- [55] Joseph S. Schafer, Kate Starbird, and Daniela K. Rosner. 2023. Participatory Design and Power in Misinformation, Disinformation, and Online Hate Research. In *Proceedings of the 2023 ACM Designing Interactive Systems Conference (DIS '23)*. Association for Computing Machinery, New York, NY, USA, 1724–1739. <https://doi.org/10.1145/3563657.3596119>
- [56] Morgan Klaus Scheuerman, Kandrea Wade, Caitlin Lustig, and Jed R. Brubaker. 2020. How We've Taught Algorithms to See Identity: Constructing Race and Gender in Image Databases for Facial Analysis. *Proc. ACM Hum.-Comput. Interact.* 4, CSCW1, Article 58 (may 2020), 35 pages. <https://doi.org/10.1145/3392866>
- [57] Tomotsu Shibutani. 1966. *Improvised news: A sociological study of rumor*. Bobbs-Merrill, Oxford, England. Pages: x, 262.
- [58] Li Shuhui and Guo Gaizhi. 2010. The application of improved HSV color space model in image processing. In *2010 2nd International Conference on Future Computer and Communication*, Vol. 2. Ieee, V2–10.
- [59] John R Smith and Shih-Fu Chang. 1996. Tools and techniques for color image retrieval. In *Storage and retrieval for still image and video databases iv*, Vol. 2670. SPIE, 426–437.
- [60] Robert Soden, David Ribes, Seyram Avle, and Will Sutherland. 2021. Time for Historicism in CSCW: An Invitation. *Proceedings of the ACM on Human-Computer Interaction* 5, CSCW2 (Oct. 2021), 459:1–459:18. <https://doi.org/10.1145/3479603>
- [61] Kate Starbird, Renée DiResta, and Matt DeButts. 2023. Influence and Improvisation: Participatory Disinformation during the 2020 US Election. *Social Media + Society* 9, 2 (2023), 20563051231177943. <https://doi.org/10.1177/20563051231177943> arXiv:<https://doi.org/10.1177/20563051231177943>
- [62] Kate Starbird, Emma Spiro, Isabelle Edwards, Kaitlyn Zhou, Jim Maddock, and Sindhuja Narasimhan. 2016. Could This Be True?: I Think So! Expressed Uncertainty in Online Rumoring. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. ACM, San Jose California USA, 360–371. <https://doi.org/10.1145/2858036.2858551>
- [63] CM Sukanya, Roopa Gokul, and Vince Paul. 2016. A survey on object recognition methods. *International Journal of Science, Engineering and Computer Technology* 6, 1 (2016), 48.
- [64] Sabine Süsstrunk, Robert Buckley, and Steve Swen. 1999. Standard RGB color spaces. In *Color and imaging conference*, Vol. 7. Society of Imaging Science and Technology, 127–134.
- [65] Anna K Swiatoniowski, Ellen E Quillen, Mark D Shriver, and Nina G Jablonski. 2013. Comparing von Luschans skin color tiles and modern spectrophotometry for measuring human skin pigmentation. *American journal of physical anthropology* 151, 2 (2013), 325–330. Publisher: Wiley Online Library.
- [66] William Thong, Przemyslaw Joniak, and Alice Xiang. 2023. Beyond Skin Tone: A Multidimensional Measure of Apparent Skin Color. [arXiv:2309.05148 \[cs.CV\]](https://arxiv.org/abs/2309.05148)
- [67] Maxim Tkachenko, Mikhail Malyuk, Andrey Holmanyuk, and Nikolai Libubimov. 2020–2022. Label Studio: Data labeling software. <https://github.com/heartexlabs/label-studio> Open source software available from <https://github.com/heartexlabs/label-studio>.
- [68] Sara Vieweg, Leyisa Palen, Sophia B. Liu, Amanda L. Hughes, and Jeannette N. Sutton. 2008. Collective intelligence in disaster: Examination of the phenomenon in the aftermath of the 2007 Virginia Tech Shooting. In *Proceedings of ISCRAM 2008 – 5th International Conference on Information Systems for Crisis Response and Management*, F. Friedrich and B Van de Walle (Eds.). Information Systems for Crisis Response and Management, ISCRAM, Washington, DC, 44–54. http://idl.iscram.org/files/vieweg/2008/1051_Vieweg_etal2008.pdf ISSN: 2411-3387 Journal Abbreviation: ISCRAM 2008.
- [69] Yuan Wang, Yukun Li, Xinning Gui, Yubo Kou, and Fenglian Liu. 2019. Culturally-Embedded Visual Literacy: A Study of Impression Management via Emoticon, Emoji, Sticker, and Meme on Social Media in China. *Proceedings of the ACM on Human-Computer Interaction* 3, CSCW (Nov. 2019), 68:1–68:24. <https://doi.org/10.1145/3359170>
- [70] Karl E. Weick, Kathleen M. Sutcliffe, and David Obstfeld. 2005. Organizing and the Process of Sensemaking. *Organization Science* 16, 4 (2005), 409–421. <https://doi.org/10.1287/orsc.1050.0133> arXiv:<https://doi.org/10.1287/orsc.1050.0133>
- [71] Adrienne Werle. 2019. Beyond Light, Medium, and Dark: Diversity and Inclusivity in the Makeup and Beauty Industries. (2019).
- [72] Zhengxia Zou, Keyan Chen, Zhenwei Shi, Yuhong Guo, and Jieping Ye. 2023. Object Detection in 20 Years: A Survey. *Proc. IEEE* 111, 3 (2023), 257–276. <https://doi.org/10.1109/JPROC.2023.3238524>

Received 23 May 2024