Cross-cultural Exploration of Visual Aesthetic in Magazine Covers

Marvin Limpijankit
M.S. Computer Science
Columbia University, NY 10027
ml4431@columbia.edu

John R. Kender
Professor Computer Science
Columbia University, NY 10027
jrk@cs.columbia.edu

Abstract

Cultural differences in visual preferences have been shown in a variety of applications including advertising, social media, and web design. We extend this investigation to magazines covers, collecting 2,733 images from a geographically diverse set of publications on the topic of fashion, food, and architecture. From this dataset, we extract 20 metrics that capture visual properties such as color, complexity, and lightness. Finally, we analyze the results for any cross-cultural differences and compare our findings to those previously discovered in the literature. Although magazines can have distinguishable style attributes, we did not find significant evidence for strong geographic patterns (similar stylistic behavior between covers from different publishers contained within the same region). However, we provide some support for previous claims about cultural preferences for colorfulness, visual complexity, and lightness.

Index terms: cross-cultural analysis, computational aesthetics, magazines

1 Introduction

Visual appeal is a key aspect in many areas, including news consumption, advertisement, and user experience design. Previous work has demonstrated that visual aesthetics can positively influence consumers' intentions on purchasing a certain product Bloch (1995). In a user experience setting, studies have shown that the initial impressions of a website's graphical interface will greatly determine whether a user will click off immediately or not Kim and Fesenmaier (2008). Therefore, it is important for companies, media corporations, and those designing products more generally, to be able to curate things in a way that is visually pleasing, which ultimately requires the ability to judge visual aesthetics. Such a task is further complicated by the fact that although some visual preferences appear to be universal, for instance the dislike of dark-yellow (olive) colors, others have shown to be culturally dependent Ou et al. (2004), Yokosawa et al. (2015), Palmer and Schloss (2010). Many attempts to study these differences involve survey-based experiments using participants from varying demographic groups. Another approach has been to gather data from online content-sharing platforms such as YouTube, Instagram, and Flickr, and use engagement metrics (e.g. views, likes, and comments) as implicit measures of visual preference.

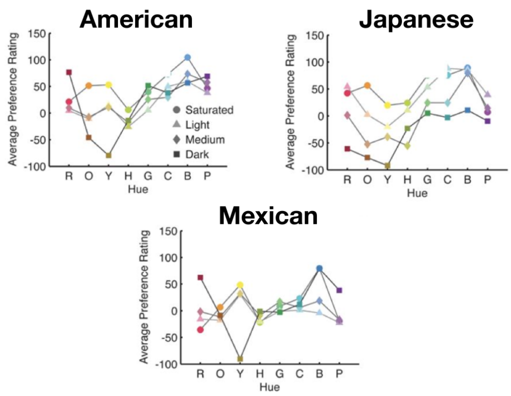

Figure 1: Color preferences across hues, lightness for American, Japanese, and Mexican individuals (Palmer et al., 2013)

In this paper, we adopt a similar methodology, collecting magazine covers from publications based in the United States and various Asian, European, and South American countries. We conduct our investigation under the pretense that in order to increase sales, publications will attempt to appeal specifically to their target region. Thus, we use these magazine covers as a proxy for the visual preference of a certain audience demographic. Inspired by past findings, which suggest that fashion Hansen (2004), food Cervellon and Dubé (2005), and architecture Koç et al. (2016) are strong indicators of culture, we limit our scope to magazines that are focused in those three domains. Then, on this set of magazine covers, we extract 20 predefined statistical image properties as numerical features. These features capture aspects of the image such as color and complexity. Finally, we analyze the dataset for cultural differences, cross referencing findings from other studies to validate whether our magazine covers reflect these same behavioral attributes or not. Also, we apply techniques such as principal component analysis (PCA), multidimensional scaling (MDS), hierarchical clustering, and local interpretable model-agnostic explanations (LIME) to aid this discussion.

Our contributions in this paper are:

1. A dataset of 2,733 magazine covers from 20+ fashion, food, and architecture publications based in the US, Asia, Europe, and South America.

2. An algorithmic extraction of 20 visual features from these covers capturing qualities such as color, complexity, luminance, and symmetry.

3. An analysis of the cultural differences in the prevalence of these aesthetics across magazine publications.

2 Related Work

2.1 Cultural differences in visual preference (using survey-based approaches)

Survey based approaches involve gathering individuals across certain demographic groups, presenting them with select images that differ in one or more visual aspect(s), and instructing them to rate the images based on aesthetics. Different studies may choose to employ slightly different methods, for instance having participants rate images versus ranking them relative to each other, but overall most adhere to this general framework.

Chiu et al. (2019) examined the effect of context in furniture product images. While both Asian and Western participants preferred matching contexts, Asians were more tolerant of product-context mismatches (when products were displayed in an unusual context). This is consistent with the view that Western cultures emphasize an analytic view, focusing on salient objects and using categorization, whereas East Asian cultures employ a more holistic view, focusing on the entire field and relationships between objects Masuda and Nisbett (2001). Another study found that Chinese participants rate traditional Chinese art as more aesthetically appealing than Western art and vice versa for Western participants, though both groups share a general preference for landscapes Bao et al. (2016). People from Hong Kong prefer less color coding and more detailed information on signage compared to those from Pakistan Iftikhar et al. (2021). Culture can also affect the associations between visual features (colors, shapes) and tastes Wan et al. (2014).

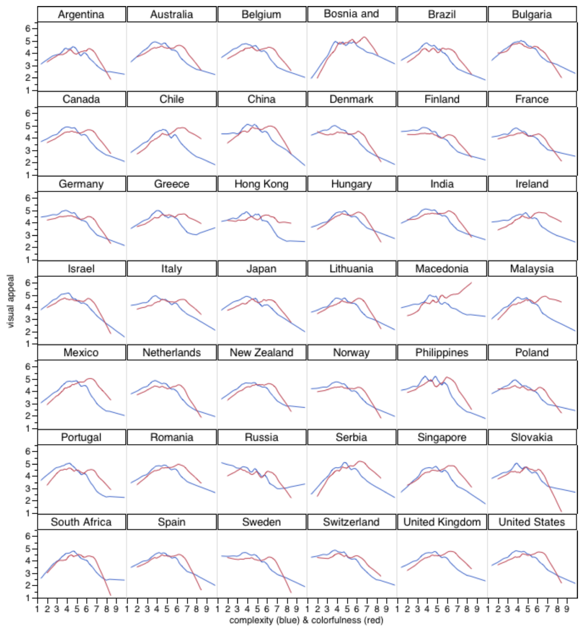

Figure 2: Color and complexity preferences across different countries (Reinecke and Gajos, 2014)

On visual preference generally, not in the context of product images, art, or signage, there are also a handful of universal and culture specific preferences that have been demonstrated. People generally prefer curved-contours as opposed to sharp angles Gómez-Puerto et al. (2018) and medium textual complexity Street et al. (2016). The preference for bluish colors and the dislike for dark-yellow (olive) colors (figure 1) also appears to be universal Palmer et al. (2013). Japan, Korea, and Taiwan show a stronger preference of whitish colors compared to other countries Saito (1996). This is likely because white symbolizes cleanliness and purity, which are valued more highly in these cultures. Similarly, Chinese people prefer red colors more than British people Ling and Hurlbert (2007). Japanese people prefer light colors (pastels) over dark colors more when compared to Americans Yokosawa et al. (2015).

Although a majority of visual preferences seem to be universally shared, subtle culture-specific differences seem to exist. One reason for this, mentioned previously in the case of Japan, Korea, and

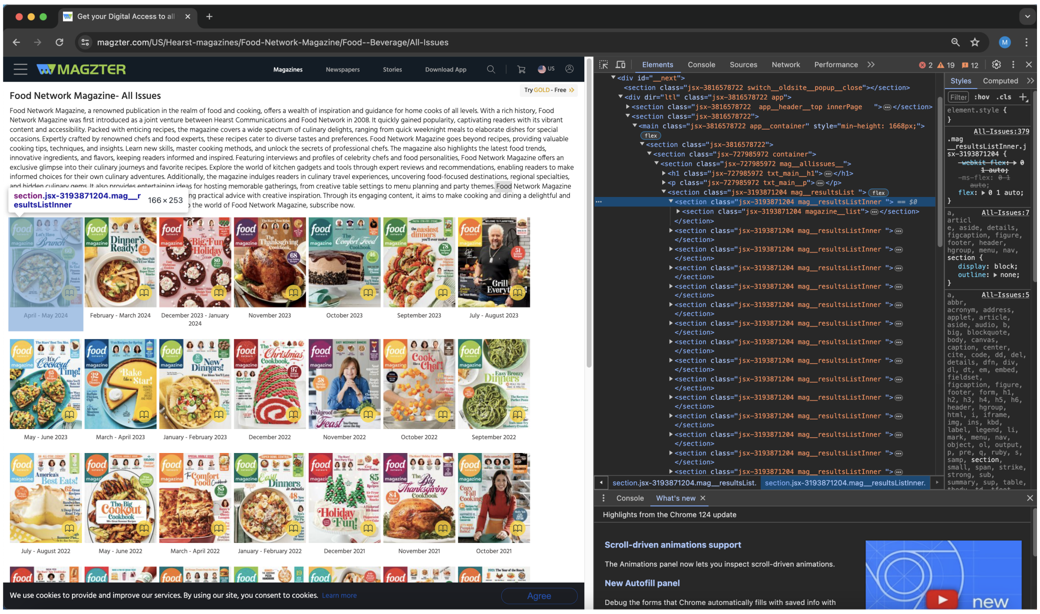

Figure 3: Magzter webpage for 'Food Network Magazine'

Taiwan preferring white/whitish colors, is because of cultural values, of which colors can be somewhat associated with. Colors can also have ingrained meaning in some cultures for instance the color red and Chinese culture. Ecological valence theory of human color preference suggests that cultural differences in color preference can be attributed to "people's average affective responses to color-associated objects" Palmer and Schloss (2010). The authors propose that a culture group's affective rating of objects, weighted by how similar they associated that object with a given color, can partially explain the variance in color preference across cultures.

2.2 Cultural differences in visual preference (using online data)

With the growing amount of data being made available online, an alternative approach to discovery culture-specific visual preferences involves analyzing metrics such as views, likes, comments, and ratings on images or other visual content. Combined with the geographic information linked to the accounts, cultural rules can be studied.

Analyzing 2.4 million ratings of the visual appeal of websites from nearly 40,000 participants, it was found that the level of colorfulness and visual complexity at which appeal peaks (figure 2) varies significantly across countries Reinecke and Gajos (2014). For instance, Russians preferred much lower complexity and Macedonians preferred designs with lots of colorful elements. In Facebook profile pictures, East Asian users were more likely to de-emphasize their face compared to American users who tended to prioritize their face over the background Huang and Park (2013).

Cultural differences in the preference of the visual content itself have also been found. The same study concluded that East Asians also expressed less emotion than Americans in their profile pictures. A YouTube analysis, showed that specific objects occurred more frequently in the thumbnails of trending videos in certain countries compared to others Zhang et al. (2021). Indian users displayed a greater interest in things such as 'train' and 'vegetable' whereas US and Canadian users preferred 'car' and foods with higher calorie content. Advertisements from 'high context' nations (Japan, Korea, China) used symbolic visuals, celebrity ambassadors, and indirect portrayals of product whereas 'low context' nations (US, UK, Germany) demonstrated the opposite behavior An (2007).

2.3 Computational Aesthetics

The task of automatic image aesthetic assessment aims to assign a score to an image based on its aesthetic appeal (i.e. quality or beauty). The hand-crafted approach involves extracting a set of pre-defined image metrics (measures of aspects such as color, texture, and light) and fit a machine learning model Anwar et al. (2021). Alternatively, neural networks trained on large datasets can, without (or with relatively little) human instruction, learn low and high-level image features and use these to predict an aesthetic score Doshi et al. (2020) Kao et al. (2017). In this work, we focus on extracting hand-crafted features since it allows us to interpret cultural differences in terms of a clearly defined set of visual aspects.

These image features, also known as statistical image properties (SIPs), have been shown to useful in a variety of downstream tasks. For instance, the SIPs from ’favorited’ images, can used to infer a Flickr users’ personality traits Segalin et al. (2016). For aesthetic assessment, they have been shown to be good predictors of aesthetic for abstract paintings Redies and Bartho (2023) and website/app interfaces Miniukovich and De Angeli (2015). Images taken from different domains (print advertisements, visual artworks, natural scenes, and more) contain patterns in their SIPs that can distinguish them from one another Braun et al. (2013).

3 Methods

3.1 Data Collection

Based on previous findings suggesting that fashion, food, and architecture are strong expressions of culture, we focus on magazines from those domains. For fashion, we select Vogue Magazine as our data source since 1) images of the magazine covers are readily available via. the archive page `voguescovers.blogspot.com` and 2) these pages are already separated by country, meaning we can group by the local/regional Vogue publisher. This also allows us to control the source since the variance in SIPs may be somewhat attributed to the different design styles publications have rather cultural preferences.

For food and architecture as magazine topics, we use `www.magzter.com`, an online repository of magazines, to source publications from multiple countries by filtering for topic and language. Once a publisher has been identified, we then scrape the sites HTML structure, parse the page (see figure 3), and download the image of each cover one by one. The resulting distribution of magazine covers across sources and countries for food publications is given in figure 4.

3.2 Feature Extraction

We extract a total of 20 visual features for each image. These features have been extensively investigated in the literature and roughly capture aspects

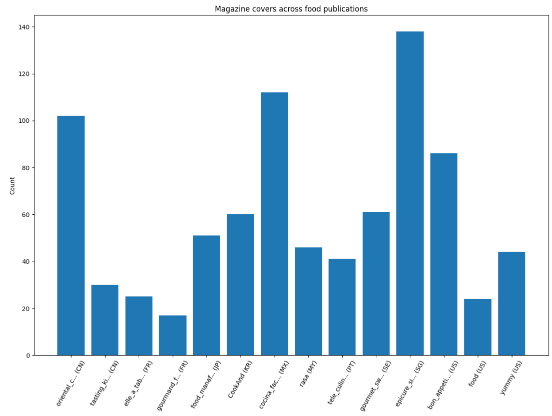

Figure 4: Distribution of food magazine covers

such as color, texture (edges), light, complexity, and symmetry. An exhaustive coverage of these SIPs, detailing the visual quality they relate to and how they are calculated, is given in appendix A.

3.3 Data Analysis

PCA. Dimensionality reduction is used to reduce the 20-dimensional feature space while preserving the maximum variance within the data. This emphasizes the visual aspects in which magazine covers in our dataset differ most, while also allowing us to visualize the data in 2D space. Principal component analysis (PCA) is one method to reduce dimensionality. In PCA, for n data points $X = X_1, X_2, ...X_n$ each with d features, p weight vectors w (where each w is d dimensional) are chosen such that when the data is projected onto the span of these w vectors the variance is maximized. An example of the optimization to find first weight vector $w_{(1)}$ is given in equation 1. The following w vectors are then chosen by iteratively following the same process but on the reduced data (the data after the previous principal components have been subtracted out). The raw features differ in scale so we first standardize our features via Z-score normalization before applying PCA.

$$w_{(1)} = \underset{\|w\|=1}{\operatorname{argmax}} \left\{ \sum_i \left(x_{(i)} \cdot w \right)^2 \right\} \tag{1}$$

MDS. We visualize the similarity between sources using multidimensional scaling (MDS). Once each data point has been reduced to a 2D representation, for each distinct pair of sources we calculate a silhouette score. This metric measures the quality of a clustering assignment and ranges from -1 to 1.

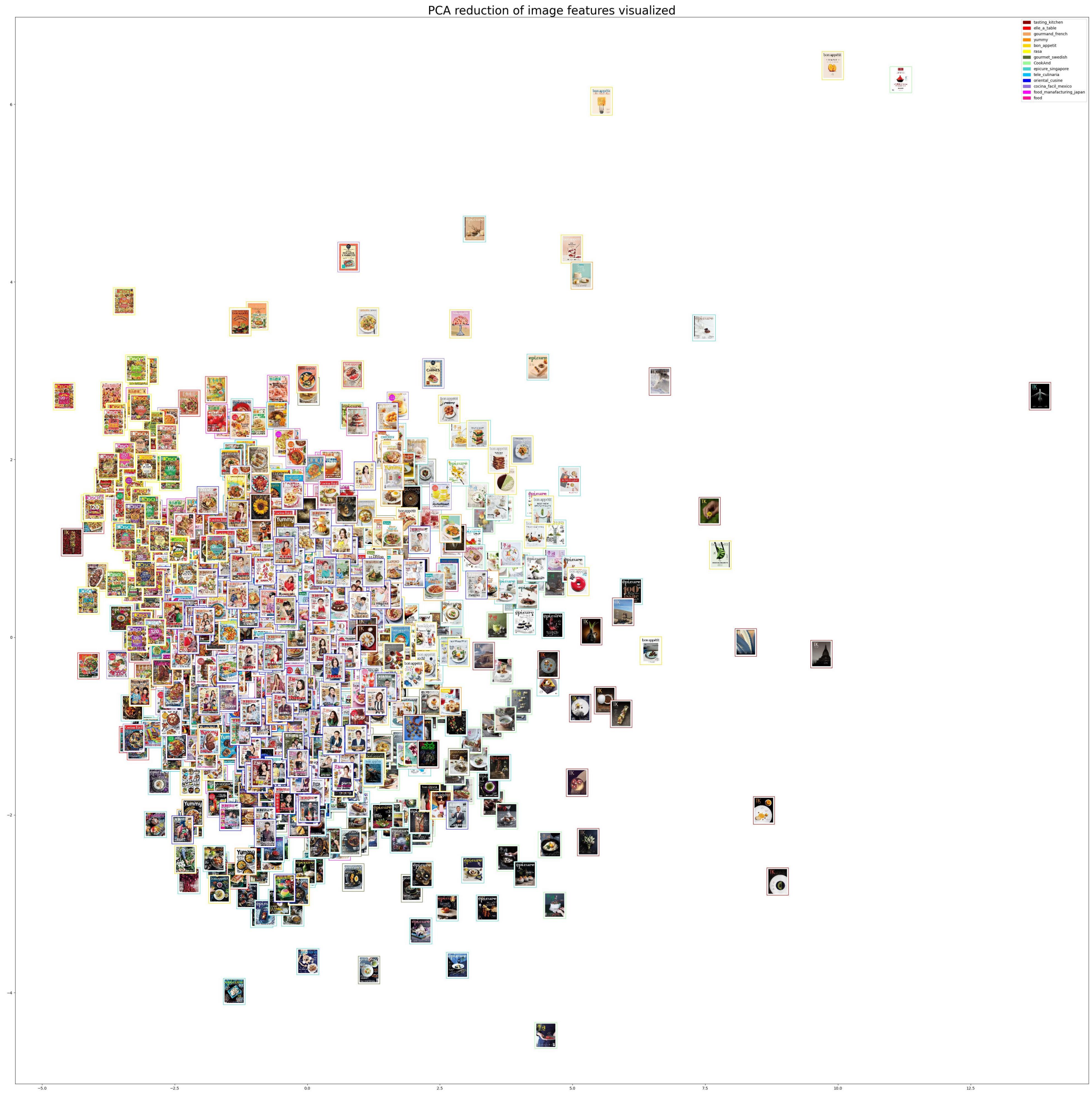

Figure 5: Visualization of food magazine covers in PCA-reduced 2D space

A score of 1 indicates a good clustering where data points assigned to the same cluster are close to each other and those belonging to different clusters are far. A score of -1 indicates the opposite, meaning that the clusters are not separable and there is a lot of overlap. By taking the source labels as clustering assignments, we use this as an approximation for the similarity between sources in terms of their magazine cover visuals.

Since a large silhouette score means sources are less similar (more separable clusters), we use these scores as similarity 'distances'. These similarities are then visualized in a 2D plot using MDS, where we use the computed scores as ground truth distances between sources and represent each source as a point in 2D space such that their distances to other points are reflective of their similarity to other sources. Formally, for n sources and similarity scores $d_{i,j}$ (the silhouette score between source i, j), we want to find 2D representations $x_1, ..., x_n$ such that the discrepancy between the Euclidean distance between points x_i, x_j and the similarity $d_{i,j}$ is minimized. This is done by solving the optimization in equation 2 by randomly initializing points $x_1, ..., x_n$ and using gradient descent.

$$\operatorname{argmin} x_1, ..., x_n \sum_{i<j} (\|x_i - xj| - d_{i,j})^2 \quad (2)$$

Other. To compliment the analysis of the magazines in our reduced space, we also validate key features and patterns that distinguish one magazine's visual style from other magazines. To do so, we construct a random forest classifier (using the non-reduced dataset) for each source in a one vs. rest manner. Then, on a balanced, held out test set of data, we run the model and use local interpretable model-agnostic explanations (LIME). LIME is a method for model explainability and returns the features that were most important for a model's prediction (i.e. why the reasons model predicted one outcome vs. another). These are used to verbalize the features that best distinguish one magazine's style from the others.

Finally, we use hierarchical agglomerative clustering to discover natural groups of visual styles among magazines. This does not rely on having the source labels and also operates on the entire, non-reduced feature space. From these, we identify style categories and analyze the distribution of sources within each style.

Note that the following results section is currently limited to the data gathered for food magazines.

4 Results

4.1 'Visual space' of magazine covers

The visualization of magazine covers in 2D space is shown in figure 5. The resulting feature weights for the two components is given in appendix B. The diagonal direction from the top-left of the space to the bottom-right seems to correspond to the usage of color. Magazine covers in the top-left are very colorful, often using reds, yellows, and greens as the dominant color palette. Progressing towards the bottom-right, the usage of color decreases as magazine covers begin to feature a plain white background instead of vibrant colors. Finally, in the bottom-right of the space, the dominating background color becomes black. Similarly, the opposing direction (a diagonal from the bottom-left to the top-right) seems to correspond to visual complexity, namely in the form of sharpness and edges. Images that fall toward the bottom-left seem to be very visually complex. These are often images with many sharp details, lots of text, complex textual patterns as backgrounds, and occasionally sub-images within the cover itself. In the top-right of the space, magazines tend to have fewer objects, a more homogeneous color palette over all (leading to less pronounced edges), and a more minimalist style with fewer text.

Color usage. Of the magazine publications analyzed, Asian ones tend to occupy the bottom-right space. Asian magazines, with the exception of "Rasa" magazine based in Malaysian, are less vibrant, use a more limited color palette, and have a darker background overall. The publications that dominate this space include Epicure (Singapore), Tasting Kitchen (China), and CookAnd (Korea). These three magazines mostly feature images of dishes prepared by acclaimed, professional fine dining venues. This up-scale feel is further accentuated by the use of black. In contrast, magazine covers in the top right seem to come from Western magazines, Cocina Facil (Mexico), Tele Culinara (Portugal), and Bon Appetit (US). Among the most frequently expressed colors are warm hues such as reds, yellows, and greens. As mentioned previously, the Malaysian magazine "Rasa" also accounts for many of the vibrant covers in this region. Chinese magazines, along with the rest of the publications (France, Japan, US, and Sweden), fall in the middle of this spectrum, neither super vibrant nor dark. These magazines use a lot of white or single-color pastels as the background.

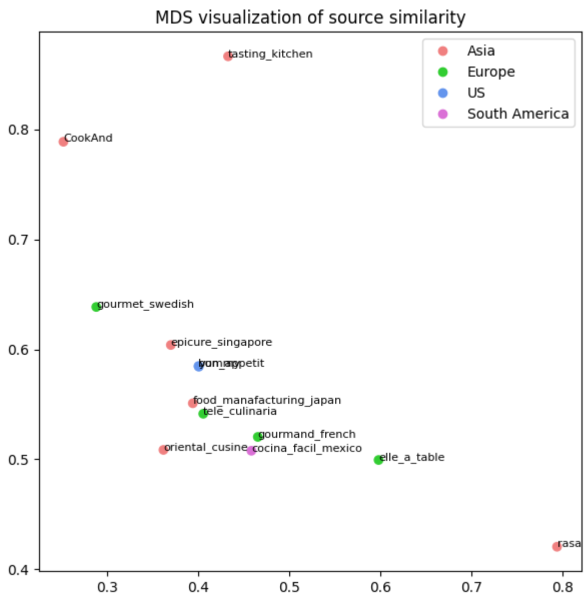

Figure 6: Style similarity visualization via. MDS

Complexity. Images from the same publication are less consistent in terms of complexity. In other words, the visual style of sources appears to be reflected more in their use of color rather than complexity. Still, some patterns can be noticed. Images with high complexity typically can have many smaller images embedded within the larger image (especially Rasa) and lots of text overlaid, for instance the US-based Yummy magazine and Mexican-based Cocina Facil. Additionally they also may depict multiple dishes/objects and contain lots of text which contributes to sharp edges (especially Chinese characters). In the case of one Chinese magazine, Oriental Cuisine, the complexity arises from the fact that they are more indirect, often showcasing images of celebrities which can contain many visually complex elements in their clothing and are often complimented with sub-images of food. This is consistent with previous findings that 'high context' nations, such as China, prefer indirect portrayals of product An (2007). Lower complexity images, which seem to be less prevalent, often use a more homogeneous color palette, depict one or few objects, and contain very little text. These are often covers from ones expressing a more minimalist style. It is important to note that no single source tends to dominate this lower complexity space, rather, on occasion some images from magazines will appeal to this style.

4.2 Visualizing similarities between sources

Computed pairwise similarity scores using the silhouette score metric on the reduced 2D data are shown in appendix B. These similarity scores are visualized in 2D space via. MDS in figure 6. Tasting Kitchen (China), CookAnd (Korea), and Rasa (Malaysia) appear to be separated from the other sources, which are grouped much tighter in the central of the space. This is likely due to their distinct styles, Tasting Kitchen with a much darker and upscale style, CookAnd with a more minimalist style, and Rasa with vibrant, visually complex covers with multiple sub-images. The rest of the sources from various countries in Asia, Europe, South America, and the US seem to be much more similar. The 2 US sources are extremely similar in style (on top of one another in the plot). The remaining 3 Asian sources are also relatively grouped together. Out of the four European sources, three of them (two French magazines and one Portuguese) appear to be similar with the exception of Gourmet, a Swedish publication. These results suggest that the difference in style cannot be solely attributed to Asia vs. US vs. Europe etc. but rather a property of styles of the individual publications within these regions. However, it is noteworthy that the styles that deviate the most from the majority in this reduced space of color and complexity are Asian publications. This may suggest that Asian publications are more likely to pursue unique styles that target specific feelings.

4.3 Validating style observations

Figure 7: Example LIME explanation for CookAnd vs. Rasa

An example of the LIME technique is shown in figure 7, where we perform a sanity check by training a random forest model to learn the difference between CookAnd and Rasa. In the 'visual space', CookAnd appears in the region corresponding to low complexity whereas Rasa is the opposite. Correctly, LIME identifies complexity to be the most important feature causing the model to predict CookAnd on this test sample. Using this technique but training a model to identify one source vs. the rest for each magazine, we collect the most indicative stylistic features for each magazine. The results confirm our previous findings suggested in

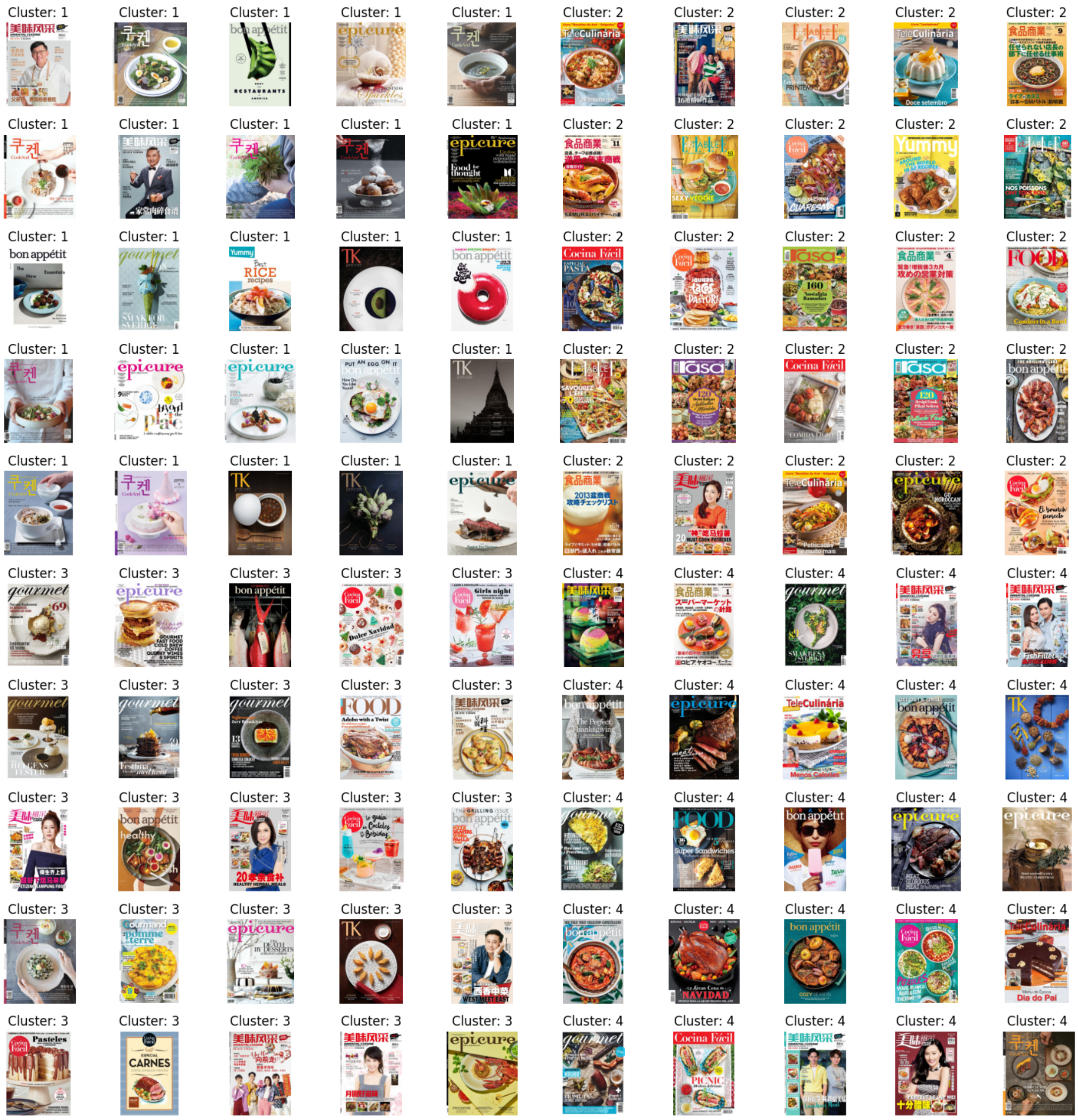

Figure 8: Example covers from different clusters

section 4.1, for instance 'high complexity' is a key feature for Rasa, 'low luminance mean' is a key feature for Epicure, and 'low complexity' is a key feature for Tasting Kitchen.

4.4 Exploring clusters of magazine covers

The resulting cluster assignment was computed using hierarchical clustering and the number of clusters was chosen by optimizing for the Dunn index, a scoring metric for clustering. Random examples from each of the clusters are shown in figure 8. The distribution of sources within the cluster is given in figure 9. From the samples, it seems that clusters correspond to, 1) low visual complexity, low colorfulness 2) high visual complexity, high colorfulness 3) and 4) medium visual complexity, medium colorfulness. Assessing the source distribution confirms what was previously found in section 4.1. For category 1, covers with low complexity and colorfulness, the Asian-based magazines CookAnd (Korea), Epicure (Singapore), and Tasting Kitchen (China) dominate, accounting for over 82.8% of the images. For category 2, visually complex and colorful magazines, Rasa (Malaysia) along with the European-based magazines make up a large proportion. Categories 3 and 4 are seem to be equally made up of a variety of sources.

4.5 Comparing to previous findings

In this section, we use the results from our data set to cross-validate cultural visual preferences found in the literature. For each claim, we first find the subset of features relevant to the claim, then

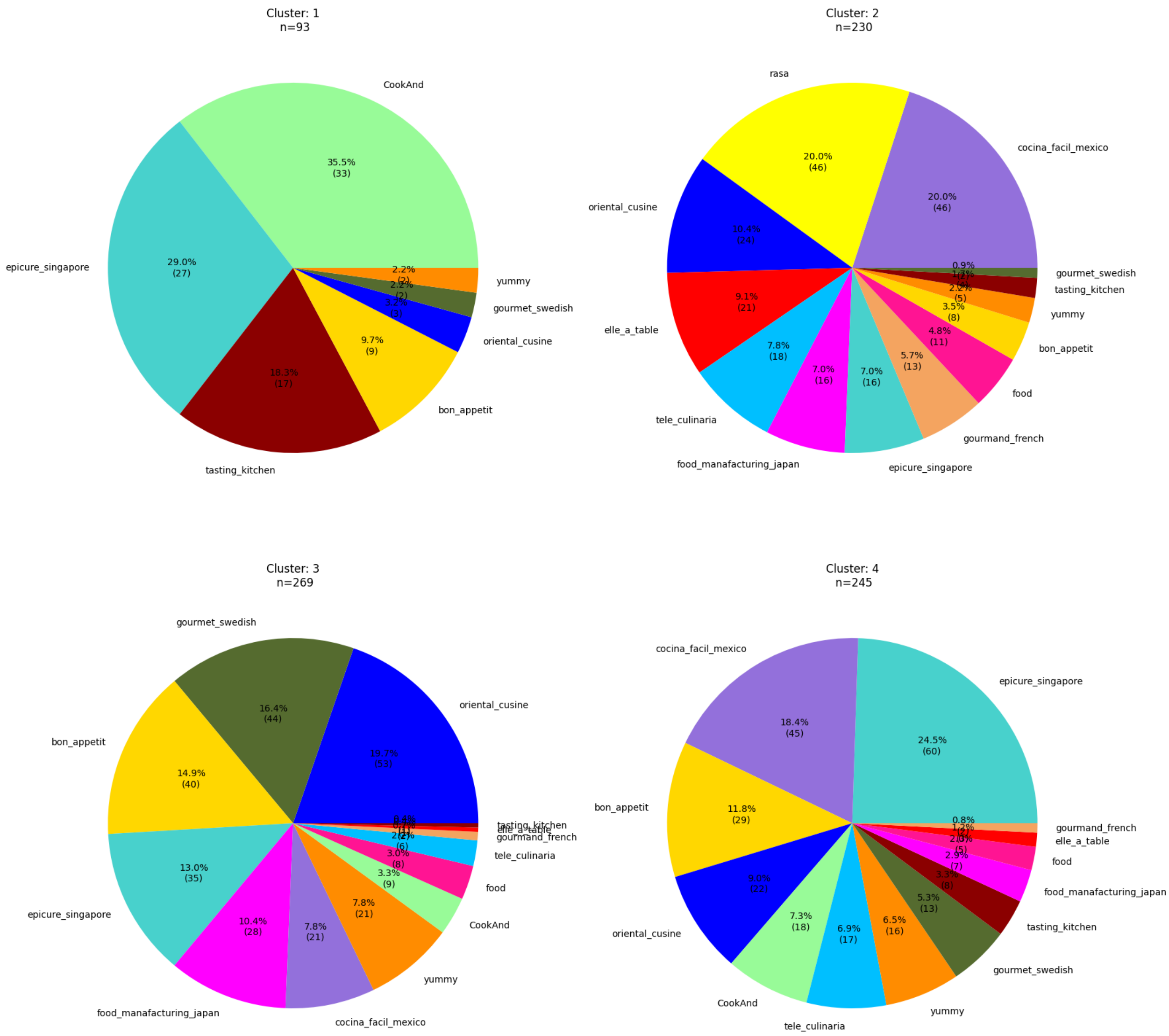

Figure 9: Source distribution within each cluster

identify which sources to compare against, and finally perform a two sample t-test.

People from Mexico preferred websites with substantially higher complexity scores. Reinecke and Gajos (2014) Complexity, fractal dimension, color entropy, luminance entropy, and variability were identified as features relating to 'complexity'. These capture the complexity of edges, colors, and light. Comparing the features for Cocina Facil (Mexico) vs. all other magazines, it was found that Cocina Facil covers, on average, demonstrated larger values for all 5 features. With the exception of variability, it was found that all of these differences were statistically significant ($p < 0.05$). Thus, our data supports the claim that Mexicans prefer more complex visuals.

For France, appeal peaks at a lower colorfulness than most other countries. Reinecke and Gajos (2014) The features identified for colorfulness were hue mean , saturation mean, lab_a mean, lab_b mean, and color entropy. Although French magazines (Elle a Table and Gourmand French) had a lower average value for hue mean, their covers had larger values for saturation, color entropy, and lab_b. The only non-statistically significant feature was lab_a mean, in which there was no discernible difference between French magazine covers and others. Thus, we did not find that French people preferred lower colorfulness. One thing to note is that colorfulness

is difficult to measure in terms of metrics such as hue mean, due to how they are calculated (see appendix A for more details).
Malaysia has one of the highest preferences for colorful websites. Reinecke and Gajos (2014) Using the same features from the previous comparison, we identify that the Malaysian magazine Rasa did have larger values for all except hue mean, and out of these they were all statistically significant. Therefore, our results support the claim that Malaysians prefer colorful visuals.

Japanese like light colors (pastels) more and dark colors less than Americans. Palmer et al. (2013) For this claim we used luminance mean as a measure of light colors. Covers from the Japanese magazine Food Manufacturing Japan had greater values for luminance than those from US magazines (Bon Appetit, Food, Yummy), and this difference was significant. Thus, this was supported by our findings.

There is a stronger preference for white and whitish colors in Japan and Korea than other countries. Saito (1996) Luminance was examined and we compared Japanese and Korean magazine covers against the rest of the data. It was found that on average luminance was lower for Japan and Korea versus other countries, likely due to the fact that the only Korean magazine was CookAnd, which, to capture the up-scale feeling, employed lots of dark colors in the background. So, our findings did not reflect a preference for white for these countries.

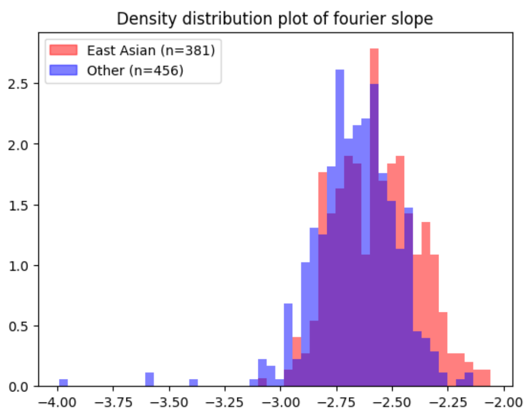

Figure 10: Fourier slope comparison for Asian vs. Other magazine covers

High-context nations (Asian countries) focus more on the entire image whereas Western countries focus more on the salient features. An (2007) We separated sources into Asian vs. Non-Asian and used Fourier slope for this claim. Fourier slope measures the presence of fine, detailed, high frequency image features. The assumption is that a larger focus on salient features will lead to more focus on an object in the foreground and blur in the background, whereas the opposite will include more details which translates to a larger (less negative) Fourier slope. We found that indeed Asian covers did have a greater presence of finer details (figure 10), supporting this claim.

5 Discussion

In this paper, we provide a cultural analysis into the visual aspects of magazine covers. We find that the majority of the variance in style can be attributed to the usage of color and image complexity. Three groups of styles emerge from our data, low colorfulness low complexity, high colorfulness high complexity, and medium colorfulness medium complexity. The first group seemed to contain many images from Asian magazines, which expressed a minimalist style, used less text, used darker backgrounds, and contained fewer objects. On the other hand, magazines with greater usage of color and more complexity tended to come from Western magazines from the US, Europe, and South America. These contained elements such as sub-images, lots of text overlaid, and warm colors such as reds and yellows. The third category seemed to contain a mix of sources.

There were multiple magazines that did not follow these general observations, for instance Rasa (Malaysia) had one of the most colorful and complex cover profiles out of the entire dataset. One Chinese publication also used brighter colors and had relatively complex visual elements (text, objects, etc.). This magazine also made noticeably frequent use of celebrities and personalities unlike the other magazines that focused on images of food. On occasion, covers from the US magazine Bon Appetit used a minimalist approach, with little text and low color variety. Therefore, these properties seem to be publication specific rather than something that can be generalized across an entire region. We also provide some support for findings previously discussed in the literature regarding cultural preferences for color, complexity, and lightness.

Future investigation may include increasing the size of the data, extending the dataset to other topics such as sports or automobile magazines, or expanding the feature set to include more visual aspects. Studying the visual aesthetics based on not only culture but also genre may be worthwhile. Another interesting question is whether the content of these magazines differ culturally, which may reflect in the objects portrayed or the words displayed on the cover, for instance.

References

Daechun An. 2007. Advertising visuals in global brands' local websites: a six-country comparison. *International Journal of Advertising*, 26(3):303–332.

Abbas Anwar, Saira Kanwal, Muhammad Tahir, Muhammad Saqib, Muhammad Uzair, Mohammad Khalid Imam Rahmani, and Habib Ullah. 2021. A survey on image aesthetic assessment. *arXiv preprint arXiv:2103.11616*.

Yan Bao, Taoxi Yang, Xiaoxiong Lin, Yuan Fang, Yi Wang, Ernst Pöppel, and Quan Lei. 2016. Aesthetic preferences for eastern and western traditional visual art: Identity matters. *Frontiers in psychology*, 7:1596.

Peter H Bloch. 1995. Seeking the ideal form: Product design and consumer response. *Journal of marketing*, 59(3):16–29.

Julia Braun, Seyed A Amirshahi, Joachim Denzler, and Christoph Redies. 2013. Statistical image properties of print advertisements, visual artworks and images of architecture. *Frontiers in Psychology*, 4:808.

Marie-Cécile Cervellon and Laurette Dubé. 2005. Cultural influences in the origins of food likings and dislikes. *Food quality and preference*, 16(5):455–460.

Tseng-Ping Chiu, Carolyn Yoon, Shinobu Kitayama, and Colleen Seifert. 2019. Cultural differences in aesthetic preferences: Does product-to-context match matter? *Design Discourse on Culture and Society: Re: Research, Volume 5*, 5:35.

Nishi Doshi, Gitam Shikkenawis, and Suman K Mitra. 2020. Image aesthetics assessment using multi channel convolutional neural networks. In *Computer Vision and Image Processing: 4th International Conference, CVIP 2019, Jaipur, India, September 27–29, 2019, Revised Selected Papers, Part II 4*, pages 15–24. Springer.

Gerardo Gómez-Puerto, Jaume Rosselló, Guido Corradi, Cristina Acedo-Carmona, Enric Munar, and Marcos Nadal. 2018. Preference for curved contours across cultures. *Psychology of Aesthetics, Creativity, and the Arts*, 12(4):432.

Karen Tranberg Hansen. 2004. The world in dress: Anthropological perspectives on clothing, fashion, and culture. *Annu. Rev. Anthropol.*, 33:369–392.

Chih-Mao Huang and Denise Park. 2013. Cultural influences on facebook photographs. *International Journal of Psychology*, 48(3):334–343.

Hassan Iftikhar, Salman Asghar, and Yan Luximon. 2021. A cross-cultural investigation of design and visual preference of signage information from hong kong and pakistan. *The Journal of Navigation*, 74(2):360–378.

Yueying Kao, Ran He, and Kaiqi Huang. 2017. Deep aesthetic quality assessment with semantic information. *IEEE Transactions on Image Processing*, 26(3):1482–1495.

Heejun Kim and Daniel R Fesenmaier. 2008. Persuasive design of destination web sites: An analysis of first impression. *Journal of Travel research*, 47(1):3–13.

Gül Koç, Marie-Thérèse Claes, Bryan Christiansen, et al. 2016. *Cultural Influences on Architecture*. IGI Global.

Yazhu Ling and Anya C Hurlbert. 2007. A new model for color preference: Universality and individuality. In *Color and Imaging Conference*, volume 15, pages 8–11. Society of Imaging Science and Technology.

Takahiko Masuda and Richard E Nisbett. 2001. Attending holistically versus analytically: comparing the context sensitivity of japanese and americans. *Journal of personality and social psychology*, 81(5):922.

Aliaksei Miniukovich and Antonella De Angeli. 2015. Computation of interface aesthetics. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, pages 1163–1172.

Li-Chen Ou, M Ronnier Luo, Andrée Woodcock, and Angela Wright. 2004. A study of colour emotion and colour preference. part i: Colour emotions for single colours. *Color Research & Application*, 29(3):232–240.

Stephen E Palmer and Karen B Schloss. 2010. An ecological valence theory of human color preference. *Proceedings of the National Academy of Sciences*, 107(19):8877–8882.

Stephen E Palmer, Karen B Schloss, and Jonathan Sammartino. 2013. Visual aesthetics and human preference. *Annual review of psychology*, 64:77–107.

Christoph Redies and Ralf Bartho. 2023. Statistical image properties and aesthetic judgments on abstract paintings by robert pepperell. *Journal of Vision*, 23(6):1–1.

Katharina Reinecke and Krzysztof Z Gajos. 2014. Quantifying visual preferences around the world. In *Proceedings of the SIGCHI conference on human factors in computing systems*, pages 11–20.

Miho Saito. 1996. Comparative studies on color preference in japan and other asian regions, with special emphasis on the preference for white. *Color Research & Application*, 21(1):35–49.

Crisitina Segalin, Alessandro Perina, Marco Cristani, and Alessandro Vinciarelli. 2016. The pictures we like are our image: continuous mapping of favorite pictures into self-assessed and attributed personality traits. *IEEE Transactions on Affective Computing*, 8(2):268–285.

Nichola Street, Alexandra M Forsythe, Ronan Reilly, Richard Taylor, and Mai S Helmy. 2016. A complex story: Universal preference vs. individual differences shaping aesthetic response to fractals patterns. *Frontiers in human neuroscience*, 10:213.

Xiaoang Wan, Andy T Woods, Jasper JF van den Bosch, Kirsten J McKenzie, Carlos Velasco, and Charles Spence. 2014. Cross-cultural differences in cross-modal correspondences between basic tastes and visual features. *Frontiers in psychology*, 5:122267.

K Yokosawa, KB Schloss, M Asano, and SE Palmer. 2015. Cross-cultural studies of color preferences: Us and japan. *Cognitive Science*.

Songyang Zhang, Tolga Aktas, and Jiebo Luo. 2021. Mi youtube es su youtube? analyzing the cultures using youtube thumbnails of popular videos. In *2021 IEEE International Conference on Big Data (Big Data)*, pages 4999–5006. IEEE.

A Feature Extraction

Color Features. *Hue, Saturation, Lab(a), Lab(b), Color Entropy*. Color is an important part of images especially when it comes to aesthetics. The use of colors can affect the mood of an image, certain colors carry certain connotations, and more, making it a commonly assessed aspect in image-based tasks. There are multiple ways to represent colors numerically, RGB, where the color of each pixel is described using three numbers corresponding to red, green, and blue, being the most popular. However, alternative color spaces exist, such as HSV (hue, saturation, value) and Lab (lightness, a, b), both of which have been shown to be more aligned with the human perception of color [34]. *Hue* and *saturation* are measured as the mean value across all pixels of the H, and S channels of the image represented in HSV space. Similarly, *lab(a)* and *lab(b)* are the mean values of the A, and B channels respectively in LAB space. Finally, *color-entropy* is introduced to capture the 'colorfulness' of an image, and is measured as the Shannon entropy of the hue channel for a given image. Images with many different values of hue across its pixels will have high entropy whereas ones with predominantly one or few colors will have low entropy.

Image Dimension Features. *Aspect Ratio, Image Size*. Although not as important in this study, considering thumbnails are reshaped into a pre-defined dimension by YouTube, for completeness we measure the *aspect ratio* and *image size* of the thumbnail. *Aspect ratio* is defined as the ratio between the height and width of the image, calculated by dividing the image width by the image height, capturing how horizontally dominated the image dimensions are. *Image size* is calculated as the sum of image width and image height. It is necessary to include both as images can have the same aspect ratio while differing in image size, where one is a proportional scale up/down of the other. Image width and image height are measured in pixel units. While thumbnails are a pre-defined dimension, it is important to note that both of these features can still vary within our dataset for two reasons. First, YouTube videos can be one of two content types, videos or shorts, which are presented slightly differently. Second, since we automatically trim any black bars as part of the preprocessing, thumbnails with black bars that are added by the user, in addition to YouTube's automatic padding, will cause slight variation.

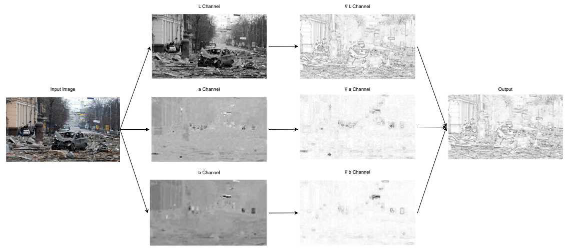

Figure 11: Process of calculating the gradient

Light Features. *Contrast, Luminance, Luminance Entropy*. To complement color, several features for lightness (i.e. brightness) are also considered since color alone doesn't capture everything. For instance, the effect of viewing a very bright vibrant red, such as Salmon, may not be equivalent to that of a darker red such as Burgundy. *Luminance*, which describes how bright an image is overall, is measured as the mean value of the L channel in LAB space. In this paper, *contrast*, a concept that often arises in the context of Photography, is calculated as the standard deviation of the L channel. Images with many varying light and dark spots will have high contrast whereas images where the brightness is relatively unchanged throughout will have low contrast. *Luminance entropy* is the Shannon entropy of the L channel and describes roughly the same concept as contrast but with some small differences.

$$g = \sqrt{g_x^2 + g_y^2} \tag{3}$$

$$\theta = \arctan \frac{g_x}{g_y} \tag{4}$$

Histogram of Oriented Gradients Features. *Self-Similarity, Complexity, Anisotropy*. A histogram of orientated gradients (HOG) is a feature descriptor that was originally introduced for object recognition [45]. The method describes the orientation of gradients within localized regions of an image, storing them as a probability density histogram where the bins are defined by the angle and the weight is determined by the magnitude of the gradient. To calculate the gradient, x and y Sobel filters, which capture the first derivatives, are applied to each channel (L, a, b) of the image, producing g_x, g_y respectively. Then, the total gradient g and angle θ can be determined using the equations above. Finally, the three channels are merged together by taking the maximum gradient and corresponding angle for each pixel. The process up until this point is depicted in figure 11.

Self-similarity measures the similarity of HOG features in a smaller, localized patch of an image with larger regions. There are several ways to calculate self-similarity which differ in the granularity at which these patches are considered and how the larger patch that each smaller patch is compared to is selected. In this study, we calculate self-similarity using neighboring patches on a level of 3 (other variations are described in full in Braun et al. [46]). First, an image is divided evenly into four quadrants, which are each divided again into even quadrants, and so on. In our case, this is repeated 3 times, i.e. the level we are performing self-similarity at. The result is an even 8x8 partition of the image with 64 patches in total. Then, for each patch, a similarity score between its HOG feature (represented as a d-dimensional vector where d is the number of bins) and each adjacent patch is calculated using the Histogram Intersection Kernel outlined in equation 5, where $h, h^{'}$ are two HOG vectors, $h_i, h_i^{'}$ are the corresponding i-th entries, and n is the number of bins used in the histogram. The overall *self-similarity* metric is the median similarity score between all patch, neighboring patch pairs in the image. *Complexity* is calculated as the mean gradient strength throughout an image. An image with lots of sharp edges and therefore very strong gradients would have high complexity and vice versa. Finally, *anisotropy* measures the relative magnitude of gradients with respect to different orientations, calculated as the standard deviation of HOG features. Low values of *anisotropy*, and thus lower standard deviations, indicate a more uniform strength across all orientations (bins) whereas a high value would suggest strongly differing strengths at different orientations. It describes whether there is a balance of edges with respect to the angle of orientation or not throughout an image.

$$\text{HIK}\left(h, h^{'}\right) = \sum_{i=1}^{n} \min(h_i, h_i^{'}) \tag{5}$$

Fourier Transformation Features. *Fourier Slope, Fourier Sigma*. The Fourier transformation is a commonly used technique in image processing that maps an image to the frequency domain, decomposing the image into a sum of periodic sine and cosine components. An example of the transformed image is supplied in figure 12, where the coordinates represent the frequency of the components (lower frequency components are closer to the center) and the color represents the magnitude of that component to the image. Low-frequency components are larger, less variant parts of the image whereas high-frequency components capture sharp details or rapid changes in intensity. From here, we calculate the power (magnitude squared) of the Fourier transformation as a function of the radial average from the center taken at 1-pixel intervals, which describes the relative decrease in presence from low-frequency components to high-frequency components. The log plot is then computed and *Fourier slope* is calculated as the slope of the line of best fit whereas the *Fourier sigma* is calculated as the root mean squared error of the line of best fit. The former describes how steep the drop-off is, ie. the relative strength of low-frequency components compared to high-frequency components, whereas the *Fourier sigma* describes how much this decrease deviates from a linear course. An image with larger *Fourier slope* (less negative) suggests a stronger presence of details whereas smaller values suggest an overall smoother image.

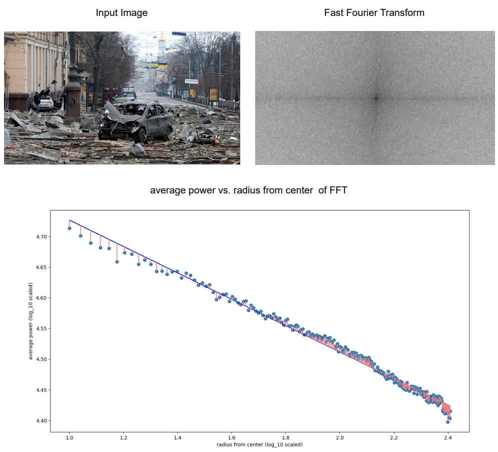

Figure 12: Process of calculating the Fourier features

Symmetry Features. *Symmetry-lr, Symmetry-ud*. Pre-trained models, such as AlexNet, that have been trained on large datasets often learn certain low-level features such as color and texture in their initial layers before advancing to more abstract, complicated features in later layers. Previous work has demonstrated that using activation maps from these initial layers can produce a symmetry metric that is more closely aligned with human perception [48]. To calculate symmetry, these first layer AlexNet activations for an image I_l and a flipped copy F_l are used as shown in equation 6. The term on the right sums over all first layer filters f and all pixel coordinates x, y and calculates the difference between those from the original vs. flipped activations. This number describes asymmetry, and so the final score is calculated by taking $1-$ this value. A horizontal flip is used to calculate the *lr* (left-right) symmetry whereas a vertical flip is used for *ud* (up-down) symmetry.

$$S(I_l, F_l) = 1 - \frac{\sum_{x,y,f} |I_l(x,y,f) - F_l(x,y,f)|}{\sum_{x,y,f} \max\left(I_l(x,y,f), F_l(x,y,f)\right)} \tag{6}$$

CNN Features. *Sparseness, Variability*. First layer AlexNet activations are also used to calculate *sparseness* and *variability*. *Sparseness* is defined as the median of the variances for each resulting response map (max-pooled). Images with low sparseness indicate less variance in response maps, meaning many filters respond to a similar degree in different max-pooled patches of the image, and thus the image is richer. Conversely, images with high sparseness often correspond to images with homogeneous patches, hence the term sparse. *Variability* is measured as the variance over all response maps, and captures the inverse of *self-similarity*. If an image has high variability, then low-level features across different subregions are very diverse, whereas an image with low variability suggests the presence of very similar features across subregions.

B Supporting Figures

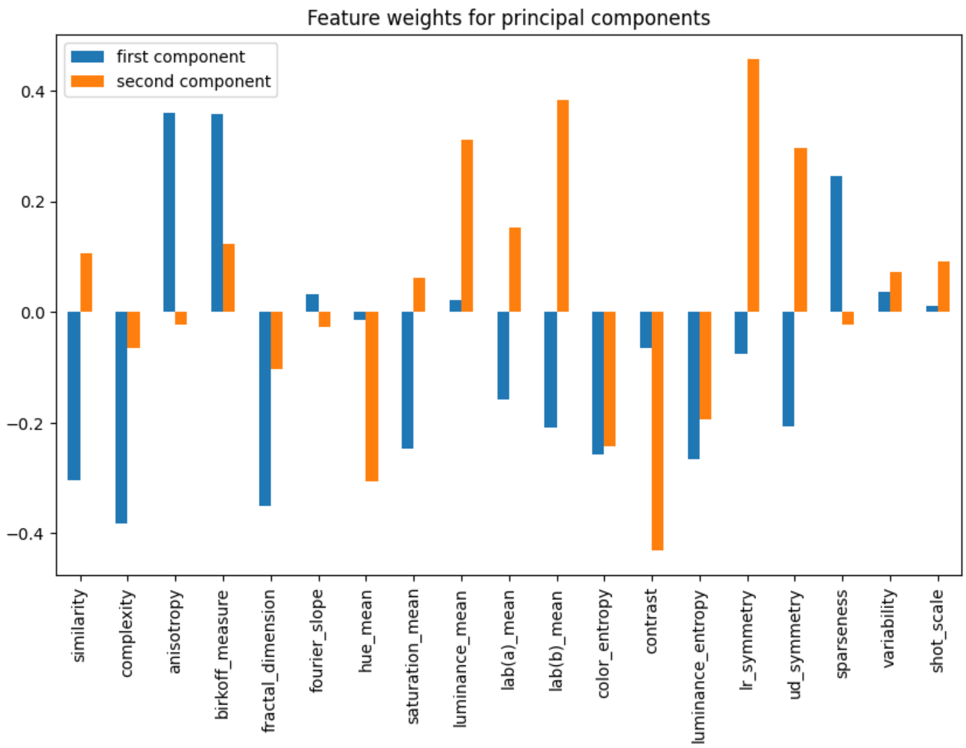

Figure 13: Feature weights from 2D PCA reduction

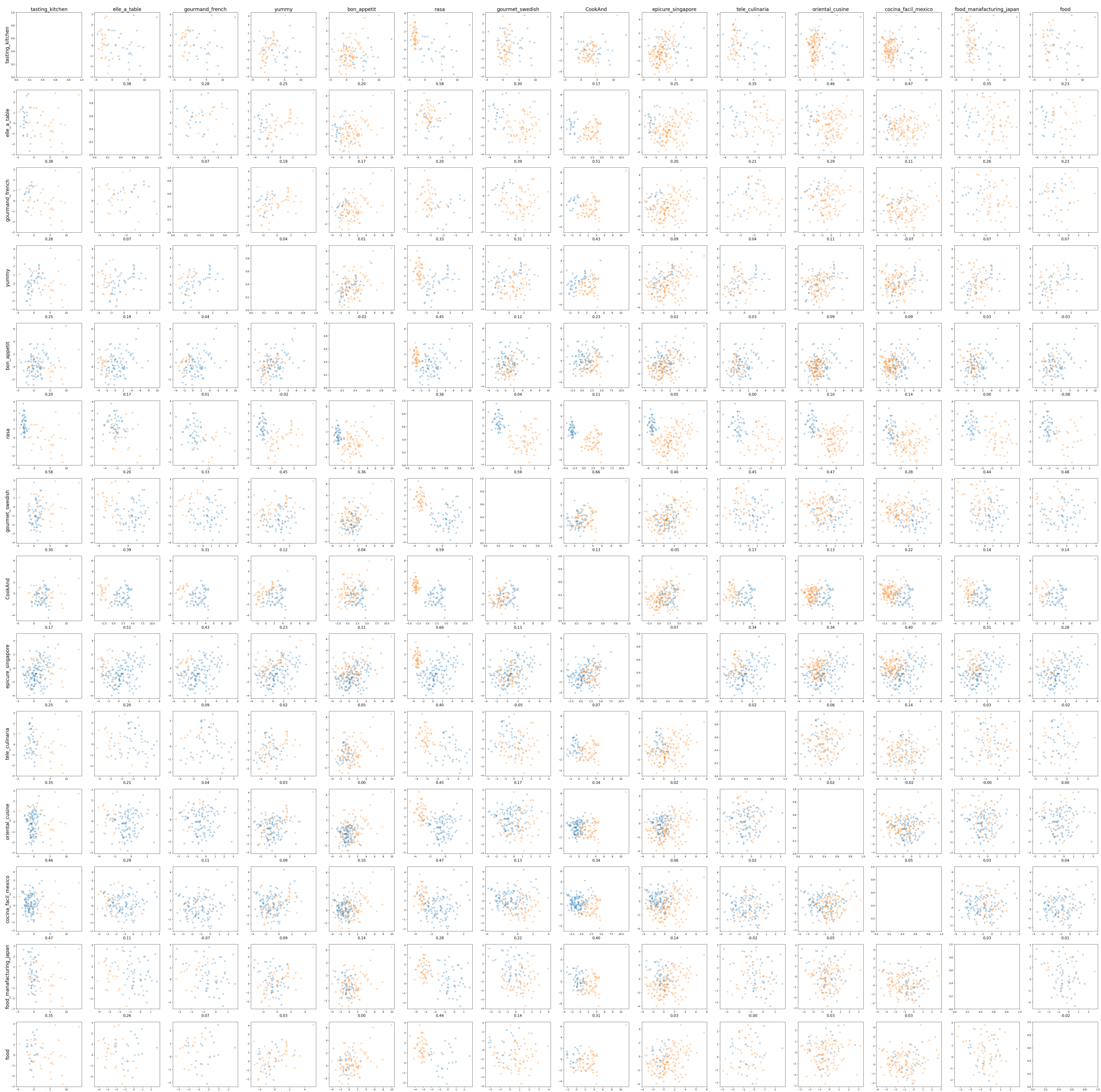

Figure 14: Pairwise similarity computation for magazines