Combining Color-Based and Semantic-Based Approaches to Detect Cross-Cultural Differences in News Media

Marvin Limpijankit[1], John R. Kender[2]
[1]Undergraduate, School of Engineering, Columbia University, New York, USA
[2]Advisor, Computer Science Department, Columbia University, New York, USA
ml4431@columbia.edu, jrk3@columbia.edu

Abstract

While a fair amount of work has been done in previous semesters on examining cross-cultural differences through natural language expression, there has been little research focused on a visual medium. The deliberate use of images across news platforms can serve as a key indicator of cultural differences as it reflects the media's attempt to appeal to the interest of their respective audiences. One way in which this difference presents itself is in the content of the images themselves. However, subtle stylistic techniques are also often employed to evoke certain emotions, color grading being one such example. In this paper, we attempt to combine both color-based and semantic-based computer vision approaches to assess the differences in images from Chinese and Western news platforms, using the Ukraine conflict as a common, international news event. Utilizing color-based and semantic-based features, we are able to classify between photos emerging from Chinese and Western sources with ~70% accuracy. We find that Chinese news platforms tend to favor brighter images taken in an indoor setting whereas their Western counterparts are more likely to choose darker, outdoor images.

Index Terms: news media, cross-cultural differences, computer vision, color analysis, scene recognition, object recognition

1. Introduction

The adoption of technology into the everyday lives of many worldwide has caused the online space to become an increasingly accessible medium for self-expression. As such, the sheer amount of data available on these platforms makes the digital world an incredibly rich resource for assessing cultural differences across populations. Our project focuses around using computer vision and natural language processing among other techniques to detect cultural differences by contrasting online reactions to international news events from members of different affinity groups. This paper extends upon the larger investigation by applying new visual analysis methods to uncover the subtle differences in how culturally distinct media platforms choose to present global events.

Previously, other project members have leveraged digital resources by mining social media posts and employing text-based methods such as sentiment analysis. Using posts from social media platforms Twitter and Weibo, Zheng Hui and Zihang Xu[1] examined the difference in emotion from those in the UK versus those in China in response to the COVID-19 pandemic. They concluded that overall, Chinese posts demonstrated a slightly larger rate of negative sentiment (49.1% of posts) compared to the UK (45.8%). Additionally, by visualizing emotions over time, they were able to map changes in emotion to localized news events. For instance, events such as the Ürümqi fire that occurred in China or government lockdowns in the UK showed observable spikes in negative sentiment online. While this does provide valuable insight into how different groups experienced COVID-19, one potential question is how much of the analyzed public sentiment is attributable to differences across populations and how much is attributable to the local media coverage of country-specific COVID-19 events themselves.

There is also reason to believe that these cultural differences may present themselves through a visual medium, with online news platforms being one such example. One study conducted by Shahira Fahmy[2] found that when covering news related to terrorism, American, British, and Arab newspapers tended to emphasize slightly different narratives through image selection, with American sources highlighting government officials, British sources focusing on victims, and Arab newspapers concentrating on the protests. Other studies within this project have also experimented with utilizing vision-based techniques to analyze news videos from different cultures. Ruochen Liu[3] used object detection as an alternative to manual inspection for detecting similar video frames across Chinese and Western sources. One of the main findings was that only a few certain objects provided significant benefit in determining which events news videos were covering with people, and the frequency of people being a key feature across all the events studied. Other objects were important for event-specific identification, for instance the presence of a chair for AlphaGo videos or a TV/monitor for videos covering the Chang'e-5 space exploration. Another study, by Yu-Shih Chen[4], was able to use similar object detection principles to examine the differences in news video coverage of COVID-19 from sources based in China and the US. They concluded that despite reporting on the same event, the choice of content was vastly different, with US sources choosing to focus on scenes in a medical setting (e.g. hospitals and medical tents) whereas

Chinese sources tended to focus on authority figures (e.g. government press meetings and police). Another interesting observation was that Chinese videos tended to be more grand in setting compared to the US, containing many shots of large groups rather than individual people.

This paper aims to expand on the existing body of work by addressing the areas of improvement of prior studies while also exploring and incorporating new methods for visual analysis. Instead of surveying public reaction to news events, the approach taken by Hui and Xu, this investigation focuses on the way media platforms themselves choose to present the same international events through a visual medium (still images), allowing for a more direct cultural analysis. Additionally, we adopt similar object-based visual approaches used by Liu and Chen to quantify these differences, once again using the count of people within frames as features but also generalizing the object-level approach to instead use scene-level attributes (ie. the location of the shot). Finally, we incorporate another mode, color, in our analysis through examination of the distributions of pixel values in RGB, HSV, and CIELAB spaces. By extracting these features and training a classifier to predict Chinese versus US images, we hope to use feature importance to uncover key differences between images, highlighting where differences in the presentation of news stories may exist.

2. Methods

2.1 Dataset Collection

For this investigation, we chose to use the Ukraine war conflict as the international news event of interest for a couple of reasons. Firstly, the length of the conflict, having lasted approximately a year at the time of this report, has led to news platforms worldwide publishing many images on the story. This is beneficial from a data quantity standpoint as it allows for larger training/testing dataset sizes, reducing the likelihood of overfitting our models while also allowing conclusions to be drawn from the data with more statistical confidence. Furthermore, the complex nature of the event allows for greater possibility in how news media might decide to cover the event (focusing on aspects like the war, the geopolitical implications, the humanitarian crisis, etc.) The hope is that similar to previous findings, there will be a noticeable distinction in the aspects different cultures choose to focus on. Finally, the event is extremely emotional, and as such platforms may choose to emphasize specific emotions or appeal to certain values through stylistic techniques, which will hopefully be observable by inspecting the properties of images related to color. Thus, the Ukraine conflict serves as an appropriate means of evaluating cultural difference in news media.

Our analysis focuses on 6 Chinese and Western news sources, China Daily, People's Daily, and Xinhua News Agency for Chinese media and CNN, NBC, and NYT as Western/US media. Using Google's search API, we automate

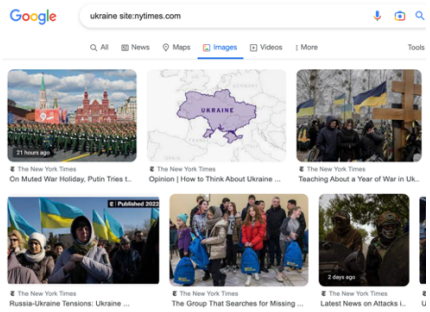

Figure 1. Example of a Domain Restricted Google Images Search

the process of downloading images by repeatedly calling the API with the appropriate parameters. We set the image format to *.jpg*, the query to '*Ukraine*', and the page number and the number of photos according to our current iteration. To restrict the website domain to the corresponding news outlet, we add *'site:[domain]'* to our query term (see figure 1). This process is repeated until 100 unique images per website are obtained, manually filtering throughout the process to remove duplicates. As an extra preprocessing step, we resize the images to 128x128 pixel square format to ensure all sources contain the same number of pixels. Finally, for efficient identification, the images are renamed in the convention 'img_[id].jpg' (ie. 'img_0.jpg' to 'img_100.jpg' for each source).

Figure 2. Sample Images from the Dataset (by source)

2.2 Color-Based Image Analysis

Colored pixels are traditionally represented in RGB format, where each pixel contains 3 channels each with an integer ranging from 0-255 corresponding to the amount of red, green, and blue in that pixel. However, one major limitation of RGB space is that the separation between features such as color and intensity is difficult to achieve because all channels need to be considered on a pixel-by-pixel basis (ie. channels do not represent specific features and the values of channels must be accounted for in relation to one another). Studies have demonstrated that for tasks such as content-based image retrieval, alternative color spaces such as HSV (hue, saturation, value) and LAB (lightness, a, b) significantly outperform RGB (Fadaei et al.) and are more aligned with human perception of color[5].

Accordingly, when considering how to represent our images on a pixel-by-pixel basis, we incorporate HSV and LAB as well. We parse through each image, storing the source and id for the image along with the pixel information into a .csv file.

Then, we visualize the difference qualitatively by observing the normalized frequency distributions of channels (ie. hue, saturation, and value distributions for HSV) as histograms. To quantify these differences, we first collect and compare standard metrics (mean, median, mode, standard deviation, skew, and kurtosis) for the resulting distributions. Then, we employ Jensen-Shannon Divergence (a symmetric, smoothed version of the Kullback–Leibler divergence, see equation 1) to calculate the pairwise distances between distributions from different sources. Finally, we pair these with a Kolmogorov-Smirnov (KS) test to obtain p-values and identify whether these differences are statistically significant or not.

$$JSD(P||Q) = \frac{1}{2}D(P||M) + \frac{1}{2}D(Q||M)$$
$$where\ M = \frac{1}{2}(P+Q)\ , D = KL\ Divergence \quad (1)$$

2.3 Semantic-Based Image Analysis

The semantic-based portion of our analysis contains two major sections, human detection, and scene-level classification. Following in similar fashion to Liu, we adopt a pre-trained, opensource YOLOv5 neural network from Ultralytics that has been trained on the COCO dataset for the human detection task[6] [7]. Operating on PyTorch, the model provides an efficient way to perform human detection on images with little startup cost. By iteratively feeding in our images as inputs, we are able to label images based on how many people are in the image. Additionally, we categorize images into 3 labels 0, 1, 2, indicating if there were no people present in the photo, there was one person present in the photo (ie. individual shot), or if there were multiple people in the photo (ie. group shot). Finally, we evaluate the distributions of labels across platforms using visual models (plots) as well as statistical tests when necessary.

Figure 3. Examples from Ultralytics YOLOv5 Model

To identify scene-level attributes, we use an 18-layer ResNet pre-trained model from MIT's Computer Science and Artificial Intelligence Laboratory, abbreviated CSAIL[8]. The model is one of many from MIT's Places project, an image database curated for the task of scene recognition containing 10 million+ images labeled with 400 scene categories. Figure 4 depicts the features the model provides for any given input image, including an indoor/outdoor score (0-1 metric), the corresponding probabilities for all scene labels, and some words describing the scene attributes. It also provides a saliency map using the class activation maximization (CAM) technique for each image, highlighting areas that were critical to the classification, providing some notion of interpretability. We limit our scope to extract only the indoor/outdoor score and the most-likely scene (scene with the largest probability), using these, along with the detection of people, as our semantic features.

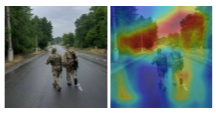

Figure 4. Examples Output from MIT CSAIL Places ResNet-18 Model

2.4 Classification Training

Using the feature sets gathered in both the color-based and semantic-based analyses, we train a Scikit learn decision tree model with varying max depths to classify between Chinese and Western images using only color-based metrics or only semantic-based metrics. To split the dataset into train/test, we employ stratify splitting with an 8:2 train/test ratio to ensure that all sources are equally represented in the split. We use the test accuracy from the top performing model for each feature set as the performance metric.

For color-based metrics, we use distribution metrics from RGB, HSV, and LAB space to train individual models and evaluate which results in the best accuracy. For semantic-based classifiers, we use all features (number of humans, indoor/outdoor score, and top-level scene prediction). Finally, we train a combined model by concatenating the features from the top color space with the semantic features as well. We investigate the node-splitting criteria at each step of the top performing tree to identify features that are most indicative of a Chinese versus Western photo, highlighting the potential cultural differences between news platforms.

2.5 Unsupervised Learning Methods

As an extension to using classification models trained on different feature sets, we present an alternative approach to discovering cultural differences using unsupervised learning. Instead of first splitting between news sources and then examining the differences between sources in the feature spaces, we first observe what natural clusters emerge from the images themselves, and then see if sources tend to produce images of one cluster over another. This approach may help

uncover natural topics/aspects of images that are present in news images of Ukraine, before evaluating each one in relation to specific news sources. The high-dimensional nature of the scene-level attributes outputted from the ResNet-18 model, with each image being mapped to a 400+ row vector corresponding to each scene, make it adequate for unsupervised learning, the hope being uncovering some natural lower-dimensional latent structures within the data. Using these vectors, we perform hierarchical agglomerative clustering using Scikit Learn and evaluate the emerging clusters qualitatively. Finally, we also propose an analogous method for pixel features, using K-means clustering in RGB space to extract color palettes from images.

3. Results and Discussion

3.1 Color-Based Image Analysis

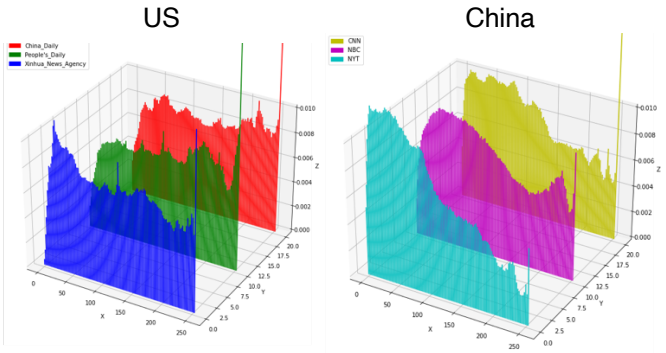

Figure 5. V-Distributions Histograms Across Sources

After exploring representations in the RGB, HSV, and LAB color spaces, we find that the channel distributions between sources are more distinct in HSV and LAB space compared to RGB space. The aggregated histograms for the R, G, and B channels (available in the supplementary materials) are extremely similar across sources. However, qualitatively observing the histograms in HSV and LAB space, differences in color between sources become evident, specifically within the V and L channels, both of which are associated with luminosity/lightness in their respective color spaces. While Chinese sources tend to demonstrate a more balanced distribution, Western sources across all 3 sources demonstrate strong positive skews. Thus, most pixels from images in Western sources tend to have smaller V and L values whereas Chinese sources are more balanced, suggesting that photos emerging from the former are generally darker in tone.

Additionally, these findings are also quantitatively evident in the pairwise Jensen-Shannon divergence distances, where the distances among Chinese sources and among Western sources are small, but large between Chinese and Western pairs. Furthermore, this separation seems to be more apparent in the V channel compared to the L channel, with the differences between regions being more exaggerated in the V-channel. For the V channel, sources from the same region tend to range in JSD distance from 0.0 - 0.1 whereas sources from different regions tend to range from 0.1 - 0.2. An interesting observation is that the New York Times demonstrates by far the most polarizing difference in V-distribution, with JSD distances of 0.2, 0.18, and 0.14 to each of the three Chinese websites, the largest 3 distances across all pairs. The New York Times also has the largest distances among sources within the same geographical category. These findings further suggest that Western sources tend to choose darker images when portraying the event whereas Chinese sources are slightly brighter, making them more balanced in terms of lightness. It also suggests that photos from the New York Times specifically tend to be especially dark, even in comparison to that of their US peers.

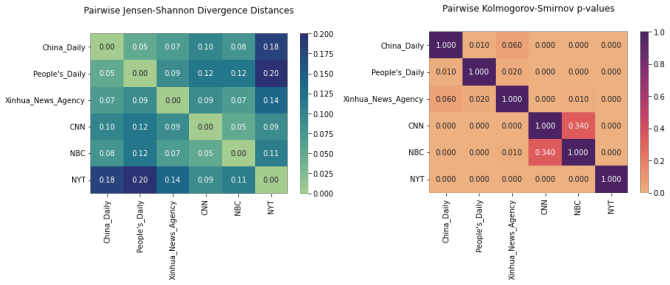

Figure 6. Pairwise JSD (left) and KS (right) Results for V-Distributions

The p-values obtained from the KS test for the V-distributions suggest that the difference between geographically distinct sources is statistically significant, meaning all p-values of a China-Western pair are < 0.05. Within regions, some pairs of sources tend to be significantly different and others not. Namely, CNN-NBC and China Daily-Xinhua News Agency pairs seem to show no significant difference between their distributions (p-values of 0.34 and 0.06 respectively). The New York Times appears to be significantly different from all other sources (both Chinese and Western), validating that their photos are noticeably darker than all other sources.

3.2 Semantic-Based Image Analysis

3.2.1 People

The aggregated results from labelling images as containing 'no people', 'individual' (containing one person), and 'group' (containing more than 1 person) across all sources are displayed in figure 7. One dominating theme is that the entire

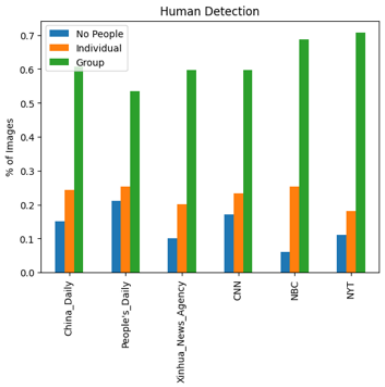

Figure 7. Distribution of # of Humans by Source

dataset is composed mainly of photos of multiple individuals, making up around 64% of the images, then individual photos at 23%, and lastly no people shots. These results are likely due to the fact that the conflict is an event with a strong narrative driven around people (e.g. the soldiers, the civilians, the political figures).

Although it is evident that the rate of individual shots is relatively similar, Western sources tend to choose images with large groups of people slightly more than Chinese sources (0.66% vs. 0.61%), whereas Chinese sources seem to use photos with no people at a slightly higher rate (0.15% vs. 0.11%). Thus, one potential differentiating cultural factor might be the use of group images versus scenic images, with Western photos focusing more on people whereas Chinese sources may focus more on photos of places. From inspecting the dataset, another possible explanation is that graphics using a clipart style tend to be very popular in Chinese sources whereas these are less common in Western sources which, when they do use graphics, tend to be informative rather than entertaining.

Figure 8. Graphics from Chinese (left) and US (right) Sources

3.2.2 Setting

Scoring images using the indoor/outdoor metric reveals a strong pattern that Chinese platforms tend to focus on indoor settings, with an average score of 0.41 slightly favoring indoor scenes, whereas Western sources more commonly make use of outdoor shots with an average score of 0.67. This can be further explained by examining the distribution of scenes for each region, which have been visualized in figure 9. Since the network consists of hundreds of scenes, a lot of which are extremely specific, it may be susceptible to error/noise. This might explain the large head, with a few scenes, such as 'army base' having large frequencies and many scenes, such as 'hockey', containing only 1 or 2 images. Also, class labels aren't necessarily exclusive, for instance 'road' and 'highway' are both potential classes according to the model but it can be argued that 'highway' is a subset of 'road'. One potential improvement may be to use a model with fewer, more generalized scenes.

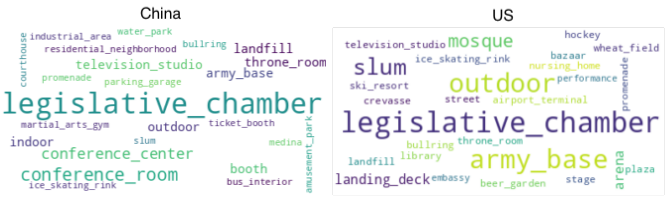

Figure 9. Word Clouds for Image Scenes

Despite this, one notable takeaway is that although both Western and Chinese sources have 'legislative chamber' as one of the most popular settings, Western sources also tend to contain a significant portion of photos corresponding to 'army base' whereas Chinese sources do not. Rather, Chinese sources focus more on scenes such as 'conference_center' and 'conference_room'. Principal component analysis on this scene matrix confirms that the scenes 'legislative chamber' and 'army base' tend to be the most heavily weighted features among the top components, meaning a lot of images can be described by how much they are classified as one or the other. One potential explanation is that Western sources tend to focus their narrative on the war aspect, whereas Chinese sources are more concerned the geopolitical context of the event, which may also explain the difference in indoor/outdoor scores. These interests may be attributable to their respective countries stake in the conflict, with China being a close political ally of Russia whereas the US is less involved but still siding with Ukraine, and thus trying to raise awareness through the emotional, human aspects of the war.

3.3 Classification Training

The previous sections provide some indication into the potential differences between Western and US images. We test these hypotheses by fitting a classifier using the extracted feature sets and evaluate the top performing models to identify which ones are indicative of a Chinese versus Western photo and highlight which factors seem to be more important than others.

To encode the data, we use the mean, median, mode, standard deviation, skew, and kurtosis for each channel in color space and one-hot encode the top scene prediction for the semantic features. We then employ stratified sampling with a train/test ratio of 8:2 to ensure that our test set is representative of all sources equally, and choose to simple decision tree model, experimenting with max_depths of 3, 5, 10, and 20 as a way of navigating the issue of overfitting/underfitting the dataset. For individual feature sets (color/semantics), it was found that a max depth of 3 produced the best test results whereas when expanding the feature set by combining both color and semantics, the optimal depth was 5. Finally, we visualize the top splits of the decision tree to see the properties that optimally separate photos into Chinese or Western.

Feature Set	Accuracy
RGB	0.66
HSV	0.70
LAB	0.68
Semantic	0.67
HSV + Semantic	0.66

Table 1. Decision. Tree Performance Across Feature Sets

Comparing color spaces, we validate the hypothesis that HSV and LAB, when compared to RGB, allow for a better representation of the images. However, this difference is marginal, with only a 4% accuracy gain when transitioning

from RGB to HSV representation and a 2% gain from RGB to LAB. Only using color features, we are able to achieve an accuracy of approximately 70% (HSV). Visualizing the corresponding decision tree reveals that a key indicator of Chinese images is having a mean V value of =>164, with the node fitting that condition containing 53 samples, a Gini index of 0.28, and 83% Chinese images. We interpret this result as having a large V value (ie. brighter images) is more indicative of a Chinese news image rather than a Western one. The skew of the H channel also seems to be a strong feature to split on, with images having a large positive skew in H values (more H values concentrated around the upper end around blue/purple/red) tending to be US images. However, one potential limitation is that in HSV space, the H values are circular, meaning 255 and 0 are contiguous, which is not accounted for in this model.

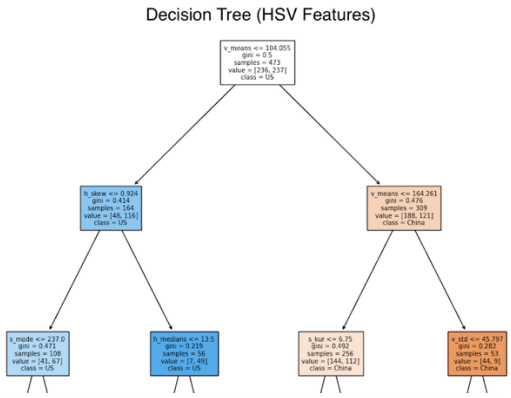

Figure 10. Top Level Decision Tree Splits for HSV Features

Classifying only on sematic features yielded an accuracy of 67%, comparable to the color features only models. The first split occurred on indoor/outdoor score, where most sources with low scores tended to be Chinese images and vice versa. The second most indicative feature was the one hot encoded column for army base, where being classified as an army base was the optimal second split on both paths following the first. With these two splits, images that were classified as an army base (whether they were classified as indoor or outdoor) were already strongly indicative of a US image, with resulting nodes having a makeup of 89% (indoor, army base) and 84% (outdoor, army base) US. Images that didn't fall into either category remained somewhat ambiguous, with the corresponding nodes having Gini values of 0.40 and 0.49. Then, only on the third level was the number of people in a scene considered. Overall, these results suggest that the relative order of importance of semantic features seems to be indoor/outdoor, whether the image was an army base or not, and finally the number of people in the scene. Despite being a popular feature, legislative chamber was not used as a splitting feature at any point in the tree, likely because it seems to be a common feature that both Chinese and Western images share.

Finally, combining both feature sets resulted in a classifier that performed slightly worse (~66% accuracy) than using each feature set individually, likely due to overfitting as a consequence of expanding the number of possible features to split on during training. Despite these shortcomings,

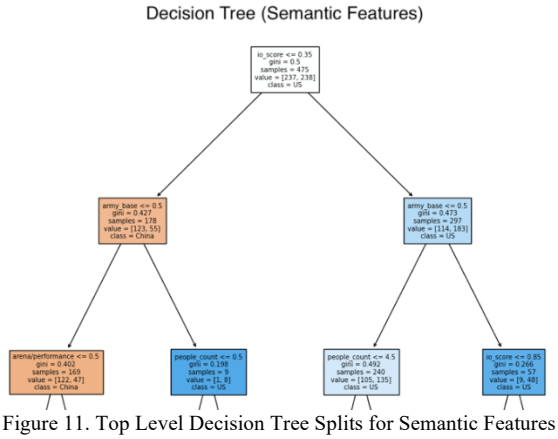

Figure 11. Top Level Decision Tree Splits for Semantic Features

inspecting the decision tree did reveal some insight on the relative importance of color versus semantic features, with indoor/outdoor score being the strongest, most important feature followed by metrics about the V-channel. Overall, these findings suggest that using the properties of the lightness of the image (the V channel in HSV) as well as the high-level features about the scene (indoor/outdoor score) are generally enough to classify images into Chinese or Western with approximately 65-70% accuracy. More granular features such as the specific location predicted, the number of people in the photo, and the saturation distribution seemed to be not as indicative.

3.4 Unsupervised Learning Methods

Hierarchical agglomerative clustering using the Euclidean distance between the 400+ length scene vector for each image revealed 3 main clusters (see figure 12). The first two clusters, depicted in orange and green, contained 47 and 31 images respectively whereas the red contained the remaining images.

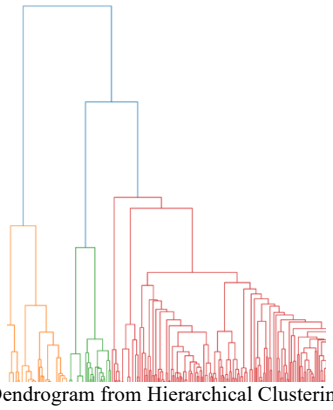

Figure 12. Dendrogram from Hierarchical Clustering on Scenes

Upon further inspecting each cluster (see figure 13), it appears that the orange cluster represents images taken in the political context, often containing photos of individuals speaking at public events or press meetings. These are likely the same images corresponding to scenes such as legislative chamber and conference center. On the other hand, the second group, depicted on the right, is strongly associated with a militaristic context. These dominant clusters reiterate what

was found previous in semantic analysis and suggest that two natural narratives/aspects to the conflict present within the images seem to be the geopolitical context as well as the physical war effort. As expected, Chinese images make up an overwhelming majority of images in the first cluster (the geopolitical narrative) at approximately 66%, whereas Western sources make up the majority of the second cluster, with around 74% of photos. There appears to be no underlying theme among images from the third cluster (red), though there remains potential in exploring the various subgroups as depicted in the dendrogram that exists within the third cluster. Overall, these results once again suggest that two main narratives apparent in the photos about the Ukraine war are the geopolitical conflict as well as the military effort, and that Chinese sources favor covering the former whereas Western sources demonstrate a tendency towards the latter.

Figure 13. Images Sampled from Each Cluster

Similarly, we present another method of uncovering natural patterns within photos using the color properties of images. Using K-means clustering within the RGB space of images, we are able to uncover dominating color palettes within photos by observing the resulting cluster centroids. With color being a potential stylistic property used to enhance certain emotions, by extracting dominating color palettes, we might be able to identify which sources may be attempting to appeal to which emotions, though such a pattern is not guaranteed to exist. It should be noted that the results from this section are preliminary rather than definitive, and more work will be required before any insights can be drawn.

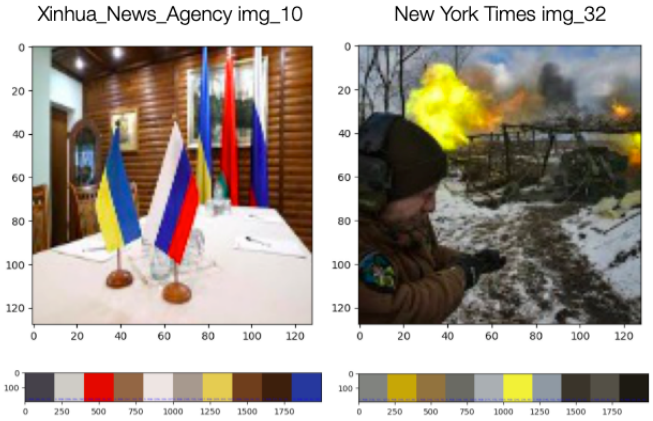

Figure 14. K-Means Clustering in RGB Space to Extract Color Palettes

4. Conclusion and Future Work

Representing images using color-based and semantic-based features, we are able to classify news photos of the Ukraine war as emerging from Chinese and US online media platforms with an accuracy of 70%. In doing so, we manage to uncover several key properties that are indicative of a Chinese vs. Western image. Chinese sources demonstrate a strong tendency to publish images of indoor settings, framing the Ukraine conflict as a geopolitical event whereas Western sources are more likely to post images of the outdoor scenes, highlighting the war effort. Additionally, Western photographs favor darker tones in their images compared to Chinese ones. One possible explanation for this disparity is that because the different countries have different stakes in the conflict they may choose to appeal to different narratives as their respective audiences are likely interested in differing aspects. With China being a political ally of Russia, Chinese readers may be more interested in the political implications of the conflicts. On the other hand, US platforms, being more aligned with the views of Ukraine, may be attempting to raise awareness by appealing to emotion, highlighting the human impact of the war. These are further supported by the differences in lightness, as the darker colors may reflect the publishers attempt at creating a grim feeling and drive emotional impact.

The direct relationship between geographic region and images remains somewhat ambiguous. For one, it may be that the lighting differences are a direct consequence of the choice of scene sources portray, with conference rooms being lit fairly brightly whereas the colors of the battlefront, soldiers' uniforms, and more are darker. With all images having been encoded with scene features, it may be worthwhile to compare images that are semantically the same (of the same content) and then investigate potential differences in lighting. Furthermore, although these findings might not generalize to other news events, the hope is that this paper has outlined a possible framework for approaching the problem of detecting cultural differences. Applying such a framework to other events may provide insightful results. There is also room to improve in the construction of the feature space, perhaps experimenting with more general (or potentially hierarchically structured) scene features as well as encoding HSV in a way that reflects the circular property of the hue channel. Finally, the unsupervised methods presented may also be worthwhile to explore, whether that be finding 'topics' or 'narratives' through images, extracting color palettes, or more.

References

[1] Hui, Zheng, and Zihang Xu. "Cross-Cultural Differences of Sentiment in Social Media Posts Respond to Major Event over Time." TAGGING AND BROWSING VIDEOS ACCORDING TO THE PREFERENCES OF DIFFERING AFFINITY GROUPS.

[2] Fahmy, Shahira. "Contrasting Visual Frames of Our Times: A Framing Analysis of English- and Arabic-Language Press Coverage of War and Terrorism." International Communication Gazette, vol. 72, no. 8, 2010, pp. 695–717, https://doi.org/10.1177/1748048510380801.

[3] Liu, Ruochen, and John R Kender. “Determining Video Similarity With Object Detection.” THE PREFERENCES OF DIFFERING AFFINITY GROUPS, 18 May 2021.

[4] Chen, Yu-Shih, and John R Kender. “Differences in Visual Context with Near-Identical Textual Taggings in COVID-19 Videos from China and the US.” TAGGING AND BROWSING VIDEOS ACCORDING TO THE PREFERENCES OF DIFFERING AFFINITY GROUPS, 4 Jan. 2022.

[5] Fadaei, Sadegh. “Comparision of Color Spaces in DCD-Based Content-Based Image Retrieval Systems.” 2021 7th International Conference on Signal Processing and Intelligent Systems (ICSPIS), 2021, https://doi.org/10.1109/icspis54653.2021.9729360.

[6] Jocher, G. YOLOv5 by Ultralytics (Version 7.0) [Computer software], 2020, https://doi.org/10.5281/zenodo.3908559

[7] Lin, Tsung-Yi, et al. “Microsoft Coco: Common Objects in Context.” Computer Vision – ECCV 2014, 2014, pp. 740–755, https://doi.org/10.1007/978-3-319-10602-1_48.

[8] Zhou, Bolei, et al. “Places: A 10 Million Image Database for Scene Recognition.” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 40, no. 6, 2018, pp. 1452–1464, https://doi.org/10.1109/tpami.2017.2723009.