Towards Understanding Crisis Events On Online Social Networks Through Pictures

Prateek Dewan[1], Anshuman Suri[1], Varun Bharadhwaj[2], Aditi Mithal[1], Ponnurangam Kumaraguru[1]
[1]Indraprastha Institute of Information Technology, Delhi (IIIT-D)
[2]National Institute of Technology, Tiruchirappalli, India
Email: [1]{prateekd,anshuman14021,aditim13122,pk}@iiitd.ac.in, [2]var6595@gmail.com

***Abstract*—Extensive research has been conducted to identify, analyze and measure popular topics and public sentiment on Online Social Networks (OSNs) through text, especially during crisis events. However, little work has been done to understand such events through pictures posted on these networks. Given the potential of visual content for influencing users' thoughts and emotions, we perform a large-scale analysis to study and compare popular themes and sentiment across images and textual content posted on Facebook during the terror attacks that took place in Paris in 2015. We propose a generalizable and highly automated 3-tier pipeline which utilizes state-of-the-art computer vision techniques to extract high-level human understandable image descriptors. We used these descriptors to associate themes and sentiment with images, and analyzed over 57,000 images related to the Paris Attacks. We discovered multiple visual themes which were popular in images, but were not identifiable through text. We also uncovered instances of misinformation and false flag (conspiracy) theories among popular image themes, which were not prominent in user-generated textual content. Further, our analysis revealed that while textual content posted after the attacks reflected negative sentiment, images inspired positive sentiment. These findings suggest that large-scale mining of images posted on OSNs during crisis, and other news-making events can significantly augment textual content to understand such events.**

I. Introduction

In the last few years, various researchers have analyzed textual content to identify and analyze sentiment, themes, and popular topics of discussion among the masses on OSNs during crisis events [1], [7], [10], [19], [24]. Sentiment and topic analysis of textual content on OSNs is widely used by researchers to understand the event, gauge the pulse of citizens, and draw inferences. However, all OSN content is not necessarily in textual format. Researchers have reported that high percentage of posts generated during real world events contain only images and no text at all [17]. This points to the fact that the methodology adopted for most aforementioned research misses out on such sections of content, which do not contain text. Moreover, even if a post contains text along with an image, the text may not be representative of the topic or sentiment depicted by the accompanying image. Consider the post in Figure 1 for instance. While sentiment analysis on text would reveal positive sentiment for this post, the sentiment associated with the image is contrasting.

Considering the human brain's affinity towards visual content [15], the pulse and sentiment of user-generated content as perceived by researchers through text, may differ from the true sentiment, since most past research does not consider images to draw inferences. In this paper, we attempt to answer two research questions, a) what are the popular themes and sentiment among images that are posted on OSNs during a crisis event? and b) how, if at all, are visual themes and sentiment different from their textual counterparts? To this end, we study a large dataset of over 57,000 images posted on Facebook during the terrorist attack in Paris in November 2015. We employ state-of-the-art computer vision techniques and construct a 3-tier pipeline for large-scale mining and measurement of the themes and sentiment of images posted on OSNs. Results of our measurement study reveal sizeable differences in prominent themes and sentiment drawn from images and text. We observed that textual content embedded in images, as well as text contained in posts, depicted negative sentiment. On the other hand, images, were found to inspire positive sentiment in general. Upon manual inspection, we observed that this contrasting behavior was largely due to the popularity of images offering support and solidarity to the victims of the attacks. We extracted visual themes from images and found that two of the top 10 themes among images were related to instances of misinformation and were not prominent in textual content. Further, textual content extracted from images revealed multiple (potentially sensitive) topics associated with "refugees", "passports", etc. which were popular in image text, but not in post text.

Fig. 1: Example of a Facebook post where sentiment associated with the post text is in contrast with the sentiment associated with the text embedded in the image.

These findings indicate the presence of useful information in the form of images posted on OSNs during crisis events, which haven't been widely explored in the literature. Such information can be of particular interest to researchers who currently resort to text for mining, analyzing, and understanding popular topics, sentiment, sensitive information, misinformation, etc. from OSNs during events. rurther, the resulting 3-tier pipeline we employed for our analysis scales to a generalizable model that can be applied to understand any similar dataset of images in a given context. For example, brands and organizations can utilize this pipeline to understand public response to products and marketing camapigns on a large scale.

II. Related Work

There exists literature in the space of studying images during crisis events on a small scale on OSNs, as well as studying crisis events on OSNs. Our work contributes towards enhancing the work done in both these areas.

A. Images on OSNs during crisis events

Multiple researchers have studied images posted on OSNs to analyze crisis events. Gupta et al. attempted to identify and characterize the spread of fake images on Twitter during Hurricane Sandy in 2013. Authors manually identified a set of fake images from news articles and blogs. This dataset was used to extract user and tweet level features to automatically identify tweets containing fake image URLs from tweets containing real image URLs [8]. Vis et al. conducted an exploratory analysis of images shared on Twitter during the 2011 UK riots. Authors manually classified images into 14 categories for characterization [26]. More similar work includes empirical analysis of Twitter images during the 2012 Israeli-Hamas conflict, where authors examined images shared by two Twitter accounts over a 2-month time frame. A total of 243 images were captured and studied manually to discover prominent themes and frames, human characters, etc. present in the images [21]. Kharroub et al. studied 581 Twitter images from the 2011 Egyptian revolution and found more images depicting crowds and protest activity as compared to images depicting violent content. In addition to most prominent visual themes, authors of this work tried to find whether user information helps in predicting image retweets, and whether image themes vary across different phases of the event [11].

As evident, images play a crucial role in measuring public sentiment during crisis and mass emergency events like terror attacks, and in cases of detecting online radicalization. All aforementioned research however, is restricted to small scale, because of the manual effort involved in measurement and analysis. The use of images for analyzing events on a large-scale remains largely unexplored. We attempt to overcome this restriction by exploring automated methods to extract meaningful information from images.

B. Crisis event related studies on OSNs

Numerous researchers have looked at textual content to study crisis events on OSNs. Hughes et al. studied the use of the Twitter social network during four emergency events, and compared how this behavior was different from general Twitter use [10]. Gupta et al. presented a study to identify and characterize communities from a set of users who post messages on Twitter during three major crisis events that took place in 2011. Authors used textual content similarity in addition to link (network) and location similarity to identify clusters of users similar users [7]. Rudra et al. proposed a novel framework to assign tweets posted during mass emergency events into different situational classes, and then summarize those tweets. Similar to Hughes et al's approach of using textual features, authors extracted features like numerals, nouns, locations, verbs, etc. present in tweet text to identify and extract event summary [19]. Thelwall et al. studied sentiment of English tweets during a month long period and found that popular events were normally associated with increases in negative sentiment strength. Authors completely relied on tweet text to extract sentiment strength and draw inferences [24].

All aforementioned research used textual content to study events on OSNs and draw inferences, thereby missing out on a large section of content pertaining to images. As discussed previously, researchers have looked at images on OSNs using manual techniques, and reported interesting findings. Past research highlights the need for automated large-scale techniques to study and mine images to extract sentiment, themes, and other similar useful information that can be used by researchers to better understand the users' reactions with respect to crisis events on OSNs and draw more accurate inferences.

III. Methodology

A. Data collection

We used *#ParisAttacks* and *#PrayForParis* as queries and collected 131,548 public posts about the Paris attacks using Facebook's Graph API Search endpoint [1] between November 14, and November 25, 2015. Although the post search feature was deprecated in April 2015, we were able to access this feature through the OAuth token generated by Facebook's mobile app for iOS. This technique has been used in the past for gathering data from Facebook [9].

The Graph API returns posts in JSON format (JavaScript Object Notation), and each post has a *type* associated with it. However, posts returned as part of the search results do not include actual images. Therefore, we filtered out all posts of *type "photo"* and re-queried the Graph API in February 2016 to obtain a total of 57,748 images from these posts. This dataset of images has been anonymized, and is made publicly available for research purposes. [2] To keep the dataset size manageable, we identified duplicates using the difference hash (dHash) technique [3] and removed them from the publicly available version.

[1] https://developers.facebook.com/docs/graph-api/using-graph-api/v1.0#search

[2] https://goo.gl/jKgqJA

[3] http://www.hackerfactor.com/blog/?/archives/529-Kind-of-Like-That.html

Ethical considerations: We did not collect any private user data. At the time of writing this paper, we were not able to find any official documentation from Facebook about the feature we used for our data collection. To the best of our knowledge, this technique of collecting data does not violate any of Facebook's terms. [4] This data was strictly used for research purposes only. We respect the privacy of Facebook users, and our work does not disclose any personally identifiable information about any individual or group whose data was part of our dataset.

B. Image characterization

As previously discussed, past research on studying the role of images on OSNs has largely relied on manual methods to perform measurement studies on images [8], [18], [26]. This methodology is time-consuming and not scalable for bigger datasets containing more than a few hundred images. With millions of images generated on OSNs every day, manually looking at images is a futile way to understand visual content and draw any meaningful conclusions in a timely manner.

To overcome this drawback, we attempt to use automated methods to characterize and study images in our dataset. We utilize state-of-the-art object detection techniques coupled with minimal human effort, domain transfer deep learning, and optical character recognition techniques, and construct a 3-tier pipeline to extract human understandable descriptors from images quickly, and on a large-scale. Unlike previously used low-level image descriptors like SIFT [12] and SURF [2], our pipeline generates high level human understandable descriptors that associate abstract level concepts (themes) and sentiment with images.

This pipeline is almost entirely automated, and significantly reduces the amount of human involvement required for understanding news-making events through images on a large-scale. Further, this technique is the basis for a generalizable method that can be applied to any similar event. Figure 2 shows the architecture of our proposed pipeline.

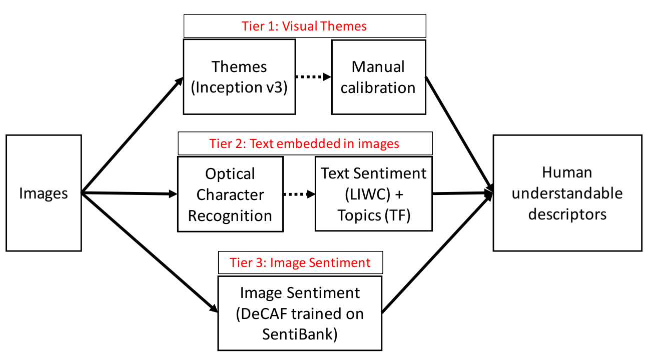


Fig. 2: Architecture of our 3-tier pipeline used to extract human understandable descriptors from images.

a) Tier 1 – Visual Themes: We use TensorFlow implementation of Google's Inception-v3 model [23] for image classification. Inception-v3 is a deep convolutional neural network (CNN) model trained for the ImageNet Large Visual Recognition Challenge using the data from 2012 and tries to classify images into 1,000 classes [20]. This model reports a top-1 error rate of 17.2%. However, the generated label comes from a fixed set of 1,000 labels, which may not be enough for characterizing the wide variety of images we come across on social networks. To establish the accuracy of this model on our dataset, we recruited human annotators through word-of-mouth publicity in an educational institute. All annotators were undergraduate computer science students and active Facebook users between the age of 18 and 21.

We got a random sample of 2,545 unique images annotated by two or more annotators. Each annotator was independently shown one image at a time. The job of each annotator was to mark whether they *agreed* with the label generated by the Inception-v3 model for the given image, or not. Using majority voting, we found that the model achieved an accuracy of 38.87% on our data sample. Given that there are 1,000 possible output labels, this accuracy value is much better than random guessing; assuming equal class sizes, the probability of a random guess being correct is 0.1%. However, through a small manual exercise, this accuracy improved significantly.

We hypothesized that images in our dataset with the same labels are highly likely to be similar, regardless of the labels associated with them being correct or not. This is because of the fact that CNNs model the vision system in animals, and are likely to group similar images together. We tested this hypothesis for the images associated with the top 30 most frequently occurring labels in our dataset, and found it to hold true. For example, the "Peace for Paris" symbol created by French graphic designer Jean Jullien, was labeled "bolo tie" by the model (Figure 3). Using this observation, we renamed 7 out of the top 30 labels to better suit the images under each of these labels. This exercise of renaming labels boosted the accuracy of the model to 51.3% on our random sample. We used these modified labels to associate abstract themes with images for our analysis. This step of manually calibrating the model output is the only human effort required in our pipeline.

Note that while retraining the Inception-v3 model on our dataset might seem an obvious option to achieve improved accuracy, retraining deep CNN based models like Inception is a time consuming and computationally expensive task, and requires large labeled datasets to perform well. Since the goal of our study is to only identify abstract level concepts for summarization, we opt to use the off-the-shelf version of this model coupled with minor human involvement over retraining from scratch.

b) Tier 2 – Text embedded in images: Past studies on analyzing topics, events, sentiment, etc. on OSNs have been largely limited to using textual content generated by users to draw inferences [7], [10], [19], [24]. This technique, however, misses out on a large section of textual information embedded in images, for example, Figure 4.

We used optical character recognition (OCR) to extract text from images in our dataset. We tested and evaluated two OCR

[4] Facebook Terms of Service, retrieved on December 22, 2016. https://www.facebook.com/terms.php

(a) What a "bolo tie" looks like

(b) Jean Jullien's "Peace for Paris" symbol

Fig. 3: Visual similarity between a bolo tie and the famous 'Peace for Paris' symbol. Such similarity is captured by the Inception-v3 model, and can be exploited to increase accuracy.

libraries – PyTesseract (Python wrapper for Tesseract OCR [5]), and OCRopy [6] – to extract text from the images in our dataset. To compare the performance of the two libraries, we first established a ground truth dataset by manually extracting text from a random sample of 1,000 images from our dataset. We then used five string similarity metrics (Jaro-Winkler distance, Jaccard index, Cosine similarity, Hamming distance, and Levenshtein distance) to compare the results produced by PyTesseract and OCRopy with the ground truth, separately. Scores from all five metrics indicated that PyTesseract performed better on our dataset than OCRopy (p-value < 0.01 for all five metrics), so we used PyTesseract for our final analysis. In all, we were able to extract text from a total of 31,869 images in our dataset. This text was used for sentiment analysis (using LIWC) and identifying popular topics (using term frequencies).

c) Tier 3 – Image sentiment: Sentiment derived from textual content generated by users on OSNs has been widely used by researchers in various contexts [3], [22], [24], [25]. However, few attempts have been made to understand the sentiment associated with images posted on OSNs [27], [28], [29]. Studies suggest that the human brain is hardwired to recognize and make sense of visual information more efficiently [15]. Thus, it is likely that sentiment extracted from textual content alone may not be representative of the overall sentiment associated with a theme or event. To this end, we attempt to extract sentiment from images using domain transfer deep learning.

The Inception-v3 model can be retrained to perform other visual recognition tasks using features extracted by the model during the training phase. This concept is known as domain transfer learning [14], and is available in the form of an open-source implementation, called Deep Convolutional Activation Feature (DeCAF). [7] DeCAF is a state-of-the-art deep CNN architecture for transfer learning based on a supervised pre-training phase [6]. We use this open-source implementation to retrain the Inception-v3 model on the SentiBank dataset to identify image sentiment. The SentiBank database comprises a total of half million Flickr images, extracted by querying the network using Adjective-Noun Pairs (ANPs) [4]. Since noun queries such as "dog", "baby", or "house" do not portray a well-defined emotion, these queries were prefixed with adjectives to form ANPs like "happy dog", "adorable baby", "abandoned house" etc., which associate these nouns with a strong emotion. We manually segregated these ANPs (and therefore, the images associated with them) into *positive* and *negative* classes for binary sentiment classification, and skipped the ANPs which did not fit clearly into a *positive* or *negative* sentiment. This exercise left us with a total of 305,100 positive sentiment images and 133,108 negative sentiment images. We performed a 10-fold random subsampling to balance the classes and obtain an unbiased model. For each fold, we split the dataset into three parts in an 80:10:10 ratio for training, validation, and testing respectively, and achieved a maximum accuracy of 69.8%.

Our 3-tier pipeline for image descriptor extraction is publicly available as a RESTful API.

IV. Analysis and Results

Themes and sentiment are two of the most widely studied aspects of OSN content during crisis events in literature [5], [13]. We therefore focus on these two aspects of the images in our dataset and present our findings.

A. Top visual themes featured misinformative images

Table I shows a list of the most commonly occurring image labels (obtained using Tier 1), and their description. We manually browsed through images corresponding to each of the top 20 labels and found that the most common types of images comprised of posters, banners, screenshots of Facebook posts, Twitter tweets, etc. Cartoons and animated posters resembling a comic book were also very popular. More examples include the Pray for Paris peace symbol by French artist Jean Jullien (label: Bola Tie), images of candles and lamps offering support to the victims of the attacks (label: Candle waxlight), and a variety of images of the Eiffel Tower, that became very popular (labels: Obelisk, Crane, etc.).

We identified some peculiar topics and themes which were popular on the network during the event. The "Malinois" label appearing in the top 20 (see Table I) corresponded to the breed of the police dog that died during the attacks, and became very popular. However, the cause of death of the dog was incorrectly quoted in multiple such images. Figure 4a shows one such picture of the police dog and states that the dog was killed when a suicide bomber detonated her explosive vest. However the real cause of the dog's death, as later clarified by French police, was multiple gunshot wounds caused by the French police forces' "Brenneke" bullets. [8] We collected all such images quoting misinformation in our dataset and found that these images had gathered over 1.1 million likes, 321,000 shares, and 38,000 comments.

[5]https://github.com/tesseract-ocr/tesseract

[6]https://pypi.python.org/pypi/ocropy

[7]https://www.tensorflow.org/versions/r0.8/how_tos/image_retraining/index.html

[8]http://www.dailymail.co.uk/news/article-3446511/Confirmed-Diesel-hero-police-dog-Paris-attacks-shot-dead-wounded-innocent-neighbours-reckless-shooting.html

TABLE I: Top 20 most common image labels in our dataset. Description of labels marked with * were recalibrated as discussed in Section III ("Tier 1 – Visual Themes").

Label	Count	Description
Website	12,416	Images of posts, tweets, banners, etc.
Book jacket*	5,383	Posters, banners, etc.
Comic book	3,803	Cartoons, animated posters
Fountain	1,264	Fountains at various locations
Envelope*	1,248	Posters, banners, etc.
Suit (clothing)	1,246	People wearing suit-like clothes
Stage	1,135	Stages during public speeches, mass gathering events, etc.
Candle waxlight	1,021	Lit candles and lamps offering support to victims
Malinois	995	Police dog who died during the attack
Scoreboard	971	Images of sports stadium
Microphone	906	Individuals, reporters, etc. using microphones
Menu	868	Images containing well formatted text
Bola Tie*	781	Peace for Paris symbol originally created by Jean Jullien
Bell cot	745	Various buildings
Jersey, T-shirt	743	People wearing t-shirts
Crane	677	Images of Eiffel Tower during twilight
Memorial Tablet	633	Variety of posters, hand written messages on boards, etc.
Church*	629	Grey scale images of Eiffel Tower
Palace	586	Large buildings, including Eiffel Tower from a distance
Obelisk*	547	Eiffel Tower

Similarly, one of the blasts during the attacks took place outside a football stadium, whose pictures quoted misinformation and became viral. These pictures were captured using the "Scoreboard" label; manual verification of images marked with this label revealed that most of these images captured the sports stadium. Figure 4b shows one such picture of the stadium and states that a Muslim security guard named Zouhier stopped a suicide bomber from entering the Stade de France stadium, and saved lives of hundreds of people inside the stadium. BBC later confirmed that it was not him who turned away the bomber. Instead, Zouheir was stationed elsewhere in the stadium, and related what he heard from colleagues who were closer to the bomb blast. [9] All instances of this misinformative image in our dataset garnered over 21,000 likes, 11,000 shares, and 450 comments.

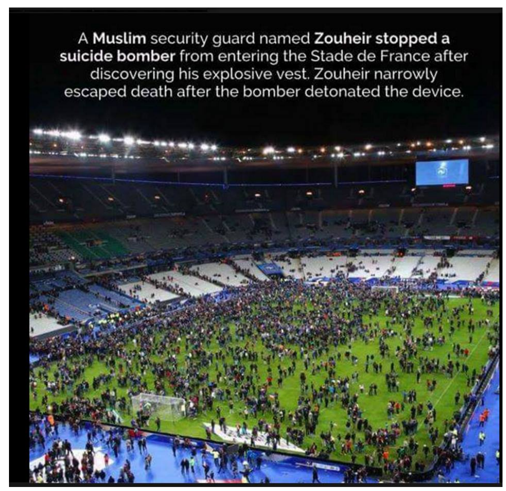
This technique of automatic identification of themes and topics from images on a large-scale can be especially helpful to identify popular instances of misinformation spread through images. Slight modifications to convolutional neural network based labeling models like Inception-v3, can aid in identifying potentially harmful and sensitive content such as guns, blood, etc. in images, and help monitor the flow of such images, and react in a timely manner, if needed.

B. Text embedded in images featured sensitive topics and reflected negative sentiment

Applying optical character recognition (OCR) on images in our dataset revealed 31,869 images (55% of all images in our dataset) which contained text embedded in them.

[9] http://www.bbc.com/news/blogs-trending-34845882

(a) Diesel the dog who was allegedly killed by terrorists



(b) Zouheir, the security guard who was claimed to have stopped a terrorist from entering the stadium

Fig. 4: Rumors spread on Facebook in the form of images during the Paris Attacks in 2015. We used CNN based computer vision techniques to identify image themes and discovered that some of the most popular image themes were associated with rumors.

Prominent topics: Table II shows a mutually exclusive set of the 10 most frequently occurring relevant words in the text we extracted from images and posts. We picked 500 most commonly occurring words in images that were not present in post text, and vice versa, to identify prominent themes among image and post text independently. We noticed that the most commonly occurring words among image and post text had less than 45% overlap, highlighting that popular words among image text were considerably different from those in post text.

Text extracted from posts was dominated by words like *"prayers", "prayfortheworld", "life", "support", "god"* etc., depicting support and solidarity for the victims. Text extracted from images however, revealed some potentially sensitive top-

TABLE II: Mutually exclusive set of 10 most frequently occurring relevant keywords in post and image text, with their normalized frequency. We identified some potentially sensitive topics among image text, which were not present in post text. Word frequencies are normalized independently by the total sum of frequencies of the top 500 words in each class.

	Top words in posts		Top words in images	
	Word	Norm. freq.	Word	Norm. freq.
1.	retweeted	0.0055	house	0.0045
2.	time	0.0052	safety	0.0044
3.	prayers	0.0050	washington	0.0042
4.	news	0.0047	sisters	0.0039
5.	prayfortheworld	0.0044	learned	0.0038
6.	life	0.0043	mouth	0.0038
7.	let	0.0042	stacy	0.0037
8.	support	0.0042	passport	0.0037
9.	god	0.0040	americans	0.0036
10.	war	0.0039	refugee	0.0035

ics like *"refugees", "passports"*, etc. which were not amongst the most talked about topics in post text. We also uncovered a popular conspiracy theory surrounding the Syrian "passports" that were found by French police near the bodies of terrorists who carried out the attacks, and were allegedly used to establish the identity of the attackers as Syrian citizens. [10] Text embedded in images depicting this theme questioned how the passports could have survived the heat of the blasts and fire. This conspiracy theory was then used by miscreants to label the attacks as a *false flag* operation, influencing citizens to question the policies and motives of their own government. The popularity of such memes on OSN platforms can have undesirable outcomes in the real world, like protests and mass unrest. It is therefore vital to be able to identify such content and counter / control its flow to avoid repercussions in the real world. Interestingly, 8,273 of these 31,869 images (25.95%) did not contain any user-generated textual content, indicating that most prior work on event analysis using text on OSNs would have entirely missed this set of data during their analysis, as discussed previously [10], [7], [24], [19].

Text sentiment: Past research has studied text sentiment to draw inferences about the overall sentiment and emotion of users on OSNs. Since most modern content monitoring techniques also focus on textual content, obfuscating sensitive textual content like hate speech and propaganda by embedding it in images is a lucrative way for malicious entities to avoid detection. Thus, we hypothesize that the sentiment of text embedded in images would be different from the sentiment of textual content posted by users in the conventional form. To confirm our hypothesis, we employed Linguistic Inquiry and Word Count (LIWC) [16] to determine and compare the emotion of the image text and post text in our dataset. We found that text embedded in images was negative on average, and twice in magnitude as positive emotion (Mann-Whitney U test: $p-value < 0.01$). Although we found more negative emotion in both images and posts as compared to positive emotion, the magnitude of negative emotion as compared to positive emotion was much higher in images as compared to text. We also noticed that positive emotion in posts was 2.6 times higher in magnitude than positive emotion in images. For negative emotion, this magnitude dropped down to 1.25.

These results indicate that textual content flowing on the network in the form of images is a critical source of information that can provide intriguing and detailed insights into crisis event related content posted on social networks. Such insights beg to be taken into consideration and analyzed thoroughly while making a judgment on the pulse and sentiment of the audience about the event.

C. Images inspired positive sentiment

Inferring image sentiment through the sentiment of text embedded in it (as discussed previously), is a small part of understanding the sentiment associated with an image. Text is only a part of the overall sentiment that an image may reflect. Moreover, there may be no text present in an image at all. Researchers have acknowledged the problem of understanding image sentiment, and come up with some solutions recently [27], [28], [29]. Using some of the most advanced techniques in the domain of image sentiment extraction, we performed sentiment analysis of images in our dataset.

Contrary to text sentiment, we found that images, on average, portrayed a positive sentiment. We observed that close to 60% of the 57 thousand images in our dataset depicted a positive sentiment. However, as already discussed, the accuracy of our image sentiment model was not too high (approximately 70%). Therefore, to verify the validity of our observations, we recruited human annotators to manually mark a small random sample of 2,545 images from our dataset as positive, negative, or neutral. Participants were also given an option to skip. Each image was annotated by at least 2 (and at most 3) participants. After removing the skipped images and using majority voting, we found that 50.95% images were marked as positive, whereas only 16.21% images were marked as negative. The remaining images were marked as neutral. This exercise confirmed our findings and affirmed the dominance of positive sentiment images in our dataset.

This observation can be attributed to the large number of pictures depicting support and solidarity for the victims of the attacks, which included posters, banners, people holding lit candles and lamps, the famous Peace for Paris symbol, etc. Such images inspire a positive sentiment on the viewer, as confirmed by our human annotators as well as the pre-trained sentiment prediction model.

Interestingly, we came across a substantial number of instances where image sentiment conflicted with the sentiment of the text present in the post. Consider the post shown in Figure 5 for example. While the text in the post reads, *"Horrible news.. No words :(:("* reflecting highly negative sentiment, the image depicts the Eiffel tower lit up in French colors, signifying support for the victims and reflecting a positive sentiment. We observed that, out of the 19,954 posts

[10] http://www.aljazeera.com/news/2015/11/paris-attacks-give-rise-conspiracy-theories-151118093352559.html

in our dataset which contained textual content as well as an image, 25.33% of the posts (5,056 posts) had conflicting image and text sentiments. Out of these, 10.98% of the posts (2,192 posts) contained an image depicting a negative sentiment, whereas the textual content present in the post reflected positive sentiment. Similarly, 14.35% of the posts (2,864 posts) contained an image depicting a positive sentiment, whereas the textual content present in the post reflected negative sentiment.

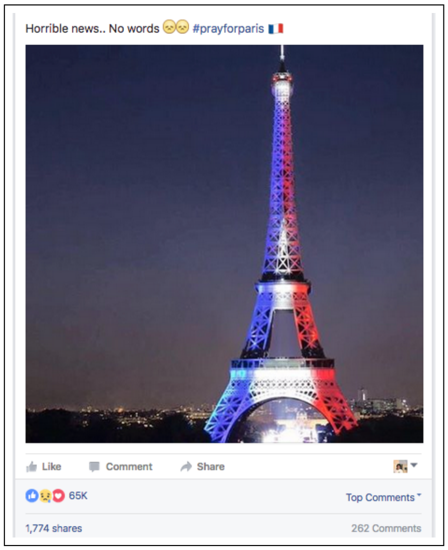


Fig. 5: Example of a post published during the Paris attacks, showing conflicting sentiments across image and text. This post was published by a verified page, garnering over 65 thousand likes and was shared 1,774 times.

We studied sentiment across post text, image text, and images temporally in order to understand how public sentiment changed across these three categories over time (Figure 6). Images were found to depict positive sentiment throughout the 12 days of our observation period after the attacks. We observed that post text sentiment was negative during the first few hours, but gradually moved to the positive side over time. Contrarily, image text reflected positive sentiment initially but moved towards negative after the first few hours. This trend persisted for the rest of our observation period.

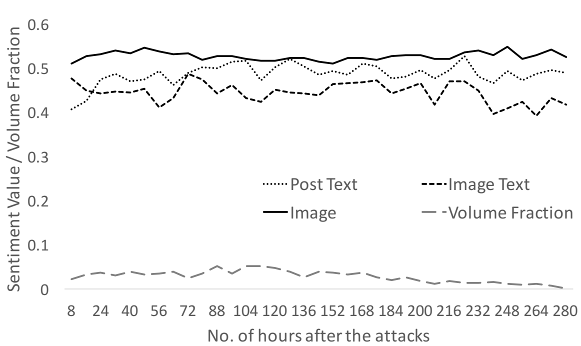


Fig. 6: Sentiment values across post text, image text, and images over time. A value of 0.5 depicts neutral sentiment. Volume fraction represents the fraction of data from our dataset posted over time.

Upon manual inspection, we observed that for the first few hours after the attacks, news and updates mentioning words like "*killed, blast, attack, shooting, explosion*", etc. in post text were prominent, resulting in negative sentiment. This trend changed towards positive sentiment after a few hours, when prayers, support, condolences, solidarity, etc. for the victims started pouring in. On the other hand, image text started off with positive sentiment, dominated by a variety of images with text offering support (like "*Pray for Paris*") posted initially. However, after a few hours, backlash against a terrorist organization, Syrian refugees, conspiracy theories, etc. skewed image text sentiment towards the negative side.

Through this analysis, we uncovered a new dimension for mining sentiment from user-generated content on OSNs, which has been largely unexplored in prior OSN related literature. Our results shed light on the varying sentiment depicted by images and text during the Paris attacks. While textual sentiment analysis revealed negative sentiment, we found that images shared on Facebook during the event depicted positive sentiment. We also found a considerable proportion of posts where textual sentiment and image sentiment depicted opposite polarity (8.75% of all images in our dataset). It is important to note that while text has been widely accepted in literature as a means to infer user sentiment on OSNs, the sentiment perceived by users is not restricted to text only. Instead, given the affinity of the human mind towards visual content, images are likely to contribute much more to the perceived sentiment of users as compared to text.

V. Limitations

The Graph API documentation does not specify any details regarding the data sample it returns. Hence, we do not claim that our dataset is representative of the entire Facebook population. However, the strength of this work lies in the proposed methodology that scales to a generalizable model for studying any context-specific dataset of images on a large scale, and a perspective from the visual angle that has largely been missing in social media research.

The accuracy of our image labeling and sentiment detection models is limited. However, validation through manual inspection and human annotations revealed that our models sufficed for producing high-level summaries for images, which was the primary objective of this work. Moreover, these models are trained using CNNs, which form the basis for state-of-the-art techniques for computer vision. These models can be further improved by feeding them true positive datasets of images. Generating a big enough dataset for such models is however, a challenging task, and out of the scope of this work. We also compared the distribution of sentiment values for text and images across our dataset, and did not find any statistically significant difference, thereby reducing the chance of bias due to skewed distributions.

Text extracted using Tesseract is limited by the performance of modern OCR techniques. We came across instances in our dataset where we were manually able to recognize text, but the OCR failed to identify this text. Most such instances involved the presence of calligraphic text, and noisy background. The percentage of such instances was low as compared to the number of images for which we were able to extract text.

VI. Conclusion

Images are an integral part of OSN content and are naturally more appealing to the human mind than text. In this paper, we highlighted the vast amount of information that can be mined from images, by making use of modern computer vision techniques. We also highlight how this information may differ from textual content that has been previously used in literature as the primary source to infer users' pulse and sentiment.

Various researchers have used text to measure the sentiment and mood of users in diverse contexts like natural calamities, politics, sports, etc. With the large volume and popularity of images on OSNs in recent times, it is imminent that results drawn from text alone fail to capture the pulse of the audience accurately. We believe that the results drawn from past studies can be improved by taking visual content into consideration.

Brands and organizations invest heavily in social media marketing and rely on textual responses generated by users to gauge their reactions, and in turn, the performance of their products. Being able to understand the users' pulse through images is likely to help such organizations measure the response to their products much better, and cover a larger section of the audience. Moreover, while analyzing sentiment and emotion through text is largely limited by language, such a barrier does not exist for images.

Pictures posted on OSNs can be a critical source of information for law enforcement organizations to understand public sentiment, especially during crisis events. With the enormous volume and velocity of data being generated on OSNs, it is extremely tough to monitor visual media at present because of lack of automated methods for measurement and analysis of such content. Our proposed pipeline can be used in such scenarios for mining knowledge from visual content and identifying popular themes and citizens' pulse during crisis events. Although this methodology has its limitations, it can be very effective for producing high-level summaries and reducing the search space for organizations in terms of content that may need attention. We also described how our methodology can be used for identifying popular instances of misinformation spread through images, which may lead to major implications in the real world.

References

[1] A. Acar and Y. Muraki. Twitter for crisis communication: lessons learned from japan's tsunami disaster. *International Journal of Web Based Communities*, 7(3):392–402, 2011.

[2] H. Bay, T. Tuytelaars, and L. Van Gool. Surf: Speeded up robust features. In *ECCV*, pages 404–417. Springer, 2006.

[3] J. Bollen, H. Mao, and X. Zeng. Twitter mood predicts the stock market. *Journal of Computational Science*, 2(1):1–8, 2011.

[4] D. Borth, T. Chen, R. Ji, and S.-F. Chang. Sentibank: large-scale ontology and classifiers for detecting sentiment and emotions in visual content. In *ACM MM*, pages 459–460. ACM, 2013.

[5] M. Cheong and V. C. Lee. A microblogging-based approach to terrorism informatics: Exploration and chronicling civilian sentiment and response to terrorism events via twitter. *Information Systems Frontiers*, 13(1):45–59, 2011.

[6] J. Donahue, Y. Jia, O. Vinyals, J. Hoffman, N. Zhang, E. Tzeng, and T. Darrell. Decaf: A deep convolutional activation feature for generic visual recognition. In *ICML*, pages 647–655, 2014.

[7] A. Gupta, A. Joshi, and P. Kumaraguru. Identifying and characterizing user communities on twitter during crisis events. In *Workshop on Data-driven User Behavioral Modelling and Mining from Social Media*, DUBMMSM, pages 23–26, New York, USA, 2012. ACM.

[8] A. Gupta, H. Lamba, P. Kumaraguru, and A. Joshi. Faking sandy: characterizing and identifying fake images on twitter during hurricane sandy. In *Proceedings of WWW companion*, pages 729–736, 2013.

[9] S. Gupta, P. Gupta, M. Ahamad, and P. Kumaraguru. Exploiting phone numbers and cross-application features in targeted mobile attacks. In *Proceedings of the 6th Workshop on Security and Privacy in Smartphones and Mobile Devices*, pages 73–82. ACM, 2016.

[10] A. L. Hughes and L. Palen. Twitter adoption and use in mass convergence and emergency events. *International Journal of Emergency Management*, 6(3):248–260, 2009.

[11] T. Kharroub and O. Bas. Social media and protests: An examination of twitter images of the 2011 egyptian revolution. *new media & society*, 18(9):1973–1992, 2016.

[12] D. G. Lowe. Object recognition from local scale-invariant features. In *ICCV*, volume 2, pages 1150–1157. IEEE, 1999.

[13] M. Mendoza, B. Poblete, and C. Castillo. Twitter under crisis: Can we trust what we rt? In *Proceedings of the first workshop on social media analytics*, pages 71–79. ACM, 2010.

[14] S. J. Pan and Q. Yang. A survey on transfer learning. *IEEE Transactions on knowledge and data engineering*, 22(10):1345–1359, 2010.

[15] M. Parkinson. The power of visual communication. *Billion Dollar Graphics*, 2012.

[16] J. W. Pennebaker, C. K. Chung, M. Ireland, A. Gonzales, and R. J. Booth. The development and psychometric properties of LIWC2007, 2007.

[17] Precog. A picture is worth 32.33 words: Importance of analyzing images on online social media. *http://precog.iiitd.edu.in/blog/2016/08/01/imagesononlinesocialmedia/*, 2016.

[18] R. Q. Redondo. The image use on twitter during the 2015 municipal election campaign in spain. *Revista Latina de Comunicación Social*, 71:85–107, 2016.

[19] K. Rudra, S. Banerjee, N. Ganguly, P. Goyal, M. Imran, and P. Mitra. Summarizing situational tweets in crisis scenario. In *ACM Conference on Hypertext and Social Media*, pages 137–147. ACM, 2016.

[20] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A. C. Berg, and L. Fei-Fei. ImageNet Large Scale Visual Recognition Challenge. *International Journal of Computer Vision (IJCV)*, 115(3):211–252, 2015.

[21] H. Seo. Visual propaganda in the age of social media: An empirical analysis of twitter images during the 2012 israeli–hamas conflict. *Visual Communication Quarterly*, 21(3):150–161, 2014.

[22] S. Stieglitz and L. Dang-Xuan. Political communication and influence through microblogging–an empirical analysis of sentiment in twitter messages and retweet behavior. In *System Science (HICSS), 2012 45th Hawaii International Conference on*, pages 3500–3509. IEEE, 2012.

[23] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna. Rethinking the inception architecture for computer vision. *CoRR*, abs/1512.00567, 2015.

[24] M. Thelwall, K. Buckley, and G. Paltoglou. Sentiment in twitter events. *Journal of the American Society for Information Science and Technology*, 62(2):406–418, 2011.

[25] A. Tumasjan, T. O. Sprenger, P. G. Sandner, and I. M. Welpe. Predicting elections with twitter: What 140 characters reveal about political sentiment. In *ICWSM*, pages 178–185, 2010.

[26] F. Vis, S. Faulkner, K. Parry, Y. Manyukhina, and L. Evans. Twitpic-ing the riots: analysing images shared on twitter during the 2011 uk riots. *Twitter and Society*, pages 385–398, 2013.

[27] C. Xu, S. Cetintas, K.-C. Lee, and L.-J. Li. Visual sentiment prediction with deep convolutional neural networks. *arXiv preprint arXiv:1411.5731*, 2014.

[28] Q. You, J. Luo, H. Jin, and J. Yang. Robust image sentiment analysis using progressively trained and domain transferred deep networks. *arXiv preprint arXiv:1509.06041*, 2015.

[29] J. Yuan, S. Mcdonough, Q. You, and J. Luo. Sentribute: image sentiment analysis from a mid-level perspective. In *Proceedings of the Second International Workshop on Issues of Sentiment Discovery and Opinion Mining*, page 10. ACM, 2013.