Towards Using Visual Attributes to Infer Image Sentiment Of Social Events

Unaiza Ahsan
Georgia Institute of Technology
Atlanta, Georgia 30332–0250
Email: uahsan3@gatech.edu

Munmun De Choudhury
Georgia Institute of Technology
Atlanta, Georgia 30332–0250
Email: munmund@gatech.edu

Irfan Essa
Georgia Institute of Technology
Atlanta, Georgia 30332–0250
Email: irfan@gatech.edu

***Abstract*—Widespread and pervasive adoption of smartphones has led to instant sharing of photographs that capture events ranging from mundane to life-altering happenings. We propose to capture sentiment information of such social event images leveraging their visual content. Our method extracts an intermediate visual representation of social event images based on the visual attributes that occur in the images going beyond sentiment-specific attributes. We map the top predicted attributes to sentiments and extract the dominant emotion associated with a picture of a social event. Unlike recent approaches, our method generalizes to a variety of social events and even to unseen events, which are not available at training time. We demonstrate the effectiveness of our approach on a challenging social event image dataset and our method outperforms state-of-the-art approaches for classifying complex event images into sentiments.**

I. Introduction

Social media platforms such as Instagram, Flickr, Twitter and Facebook have emerged as rich sources of media, a large portion of which are images. Instagram reports that on average, more than 80 million photos are uploaded daily to its servers.[1] This includes images of personal major life events such as weddings, graduations, funerals, as well as of collective news events such as protests, presidential campaigns and social movements. While some images are usually accompanied with associated text in the form of tags, captions, tweets or posts, a large part of visual media does not contain meaningful captions describing the image content or labels describing visual affect.

Inference of psychological attributes such as sentiment from text is well-studied [26], however the extraction of sentiment via the visual content of images remains underexplored. Recent approaches that infer visual sentiment are limited to images containing an object, person or scene [2]. We address the problem of inferring the dominant affect of a photograph containing complex and often crowded scenes that characterize many social and news events. Our goal is to use only visual features of the given photograph and not rely on any metadata (See Figure 1).

Our motivation to use only visual data for sentiment prediction springs from three observations. (1) Automatically predicting sentiments on event images can help determine what users feel about the event and in what context they choose to share it online. This can help personalize social feeds of individuals, as well as improve recommendation algorithms. (2) News events are often shared in the form of collated articles with images. Accurately ascertaining the sentiment of the specific event images using text will lead to inherent biases that may be introduced by the text or caption of the image. (3) Text associated with an event image may not convey sufficient, accurate or reliable sentiment related information. For instance, some tags or captions may just describe the objects, actions or scenes occurring in the image without reflecting on the actual emotional state conveyed through the image.

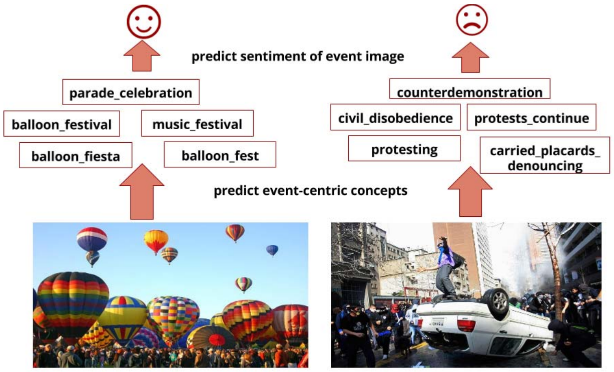

Fig. 1: Our major contribution is to map event concepts to sentiments for social event images.

Event images usually consist of objects (e.g. wedding gown, cake), scenes (e.g. church), people (e.g. bride), subevents (e.g. ring exchange), actions (e.g. dancing) and the like. We refer to these as *event concepts*. They are similar to the mid-level representations in sentiment prediction pipelines referred to as adjective noun pairs (ANPs) (e.g. cute baby, beautiful landscape) but there are no explicit adjectives or sentiments in our event concepts. In this paper we develop a sentiment detection framework that infers complex event image sentiment by exploiting visual concepts on event images. Our method discovers concepts for events and extracts intermediate representation of event images using probabilistic predictions from concept models [1].

Concretely, the contributions of our paper are:

- We propose a method to predict the sentiment of complex

[1] https://instagram.com/press, accessed April 2016

978-1-5090-6182-2/17/$31.00 ©2017 IEEE

event images using visual content and event concept detector scores without requiring any text analysis on test images.
- Our method outperforms state-of-the-art sentiment prediction approaches without extracting sentiment specific information from the images.
- We conduct comprehensive experiments on a challenging social event image dataset annotated with sentiment labels (*positive, negative, neutral*) from crowdworkers, and propose to share this dataset with the research community.
- To assess generalizability and validity, we employ our event sentiment detector on a large dataset of web images tagged with events *not considered* in model training, and characterize the nature of sentiments expressed in them.

II. Related Work

The increased use of social media by people in the last decade resulted in research opportunities to determine what people feel and emote about entities and events. Twitter emerged as a powerful platform to share opinions on daily events. Prior work includes developing frameworks to analyze sentiments on predidential debates [13, 8], SemEval Twitter sentiment classification task [11, 17] and brands [14]. De Choudhury *et al.* mapped moods into affective states [5] and also predicted depression from social media posts [6]. In attempts to make sense of large-scale community behavior, Kramer *et al.* utilized the text of posts made on Facebook to determine social contagion effects of emotion and affect [18]; whereas Golder and Macy [10] found that positive and negative affect expressed on Twitter can replicate known diurnal and seasonal behavioral patterns across cultures. All these approaches use text as a major source of sentiment discovery. We address the problem of identifying emotions conveyed by complex event images, without reliance on associated text.

Recent work on emotion prediction from images or videos leveraged low level visual features [15, 20, 28], user intention [12], attributes [2, 37], art theory-based descriptors [23] and face detection [31]. Our work is similar to the SentiBank [2] approach which extracts sentiment concepts-based representation of images and then predicts their sentiment using the concept representation as features but our method differs in one crucial way. We do not extract sentiment-related concepts on images such as 'cute baby' but event-related concepts such as 'birthday boy'. Hence our representation differs as it is *event specific* and not sentiment specific. Wang *et al.* [33] used web images and associated text to jointly learn image sentiment using a nonnegative matrix factorization approach. Our work differs from theirs in terms of image type. They predicted sentiment on images where objects and faces are clearly visible (hence dedicated object/scene/face detectors can be used). We focus on event sentiment detection from crowded event images where faces and objects may not be clearly visible.

Other similar work includes methods using deep networks for sentiment prediction but differ in that they either use sentiment specific features [4, 3], do not use intermediate concepts [35] or use probabilistic sampling to select training instances with discriminative features [36]. All of these methods do not address sentiment prediction of images containing complex and crowded scenes. A more recent line of work has started addressing emotion recognition in group images/videos [7, 25, 32, 30, 22, 34] however our problem domain is different as we do not require human beings or their faces to be visible in the image in order to predict the sentiment of the image.

III. Approach

In this section we present our sentiment classification framework starting from the proposed event concepts. Our method comprises three main steps: (1) Generating event concepts, (2) Computing event concept scores, and (3) Predicting sentiment labels from concept scores.

We first discover event concepts by mining an initial list of event categories from Wikipedia. Those categories are then used as search queries to mine Flickr tags. Thereafter, using a tweet segmentation algorithm [21] on these noisy tags, we generate generate relevant social event concepts. Finally, we combine these discovered concepts with nearest neighbors obtained by projecting event categories onto a semantic vector space (word2vec) [24]. For each discovered event concept, we crawl images shared on the web, compute convolutional neural network (CNN) features on them and train concept models. Once the models are trained, we predict concept scores on test images to compute our proposed features and finally use a linear Support Vector Machine (SVM) to predict the sentiment of the test images.

A. Generating Event Concepts

Using a concept-based intermediate representation as image features is an established technique for capturing high level semantic information from images [15, 20, 28]. Our main motivation behind generating event specific concepts is to formulate a discriminative representation for crowded event images using web-based results and social media tags. Off-the-shelf deep CNN features are useful for object and scene recognition from images but directly using these features for classifying sentiment of crowded event images is not sufficient due to the inherent ambiguity and complexity associated with visual manifestation of affect (as will also be illustrated in the results section).

We generate relevant social event concepts using the following steps:

1) We use Wikipedia to mine a list of 150 social event categories from its category 'Social Events'. This list is generic in order to cover all possible types and categories of events. Some sample event categories are: basketball match, art festivals, beauty pageants, black friday etc..
2) We use the event categories as exact queries to Fickr and retrieve top 200 tags for public images.
3) We preprocess the tags and employ them to a tweet segmentation algorithm proposed by [21] to generate coherent segments (phrases). This algorithm uses a dynamic programming approach to select only those combination of words that have high probability of

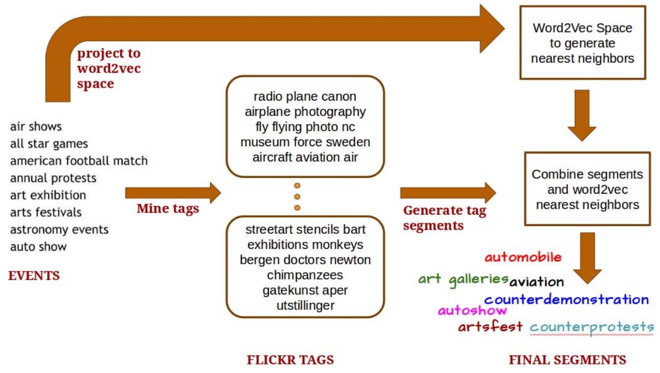
project to word2vec space
Word2Vec Space to generate nearest neighbors
air shows
all star games
american football match
annual protests
art exhibition
arts festivals
astronomy events
auto show
EVENTS
Mine tags
radio plane canon airplane photography fly flying photo nc museum force sweden aircraft aviation air
streetart stencils bart exhibitions monkeys bergen doctors newton chimpanzees gatekunst aper utstillinger
FLICKR TAGS
Generate tag segments
Combine segments and word2vec nearest neighbors
automobile
art galleries
aviation
counterdemonstration
autoshow
artsfest
counterprotests
FINAL SEGMENTS

Fig. 2: Generating event concepts for social events [1]

occurence in large text corpuses and words that are named entities. We also make sure the extracted segments are visually representative [29]. We inspect the highest scoring segments after computing the final scores and remove ambiguous or slang words.

4) Finally, we project each event category (mined from Wikipedia) on to a word embedding using the popular word2vec [24] approach. The word embedding is pretrained on the Google News Dataset—a large corpus of text from Google News articles comprising around 100 billion words. We extract 20 nearest neighbors to each event category and add them to the pool of segmented phrases. We use the word vectors pretrained on Google News Dataset because as it is a collection of words from news articles, the word vectors refer to those words and phrases which involve news events and are hence relevant to our work. After pruning irrelevant concepts, we finally end up with 856 social event concepts. Figure 2 shows the event concept discovery pipeline. For further details, please see [1].

B. *Computing Event Concept Scores*

Each generated event concept is used as a search query on the Microsoft Bing search engine to extract the top 100 public images. MS Bing is a convenient platform for scraping highly discriminative images for a wide variety of search queries. The images are used to train linear classifiers to predict concept scores on our test images. The image features used are the activations of the last layer (fc7) in a Convolutional Neural Network (CNN) pretrained on ImageNet [27] and Places Databases [38] and the CNN architecture used is AlexNet [19][1]. We compute fc7 features on each image and use event concept classifiers to predict the concept probabilistic scores. For each image I, the feature vector f_I is a concatenation of all concept classifier scores predicted on the image. Thus $f_I = \{x_i\}_{i=1}^{m}$ where m is the total number of concepts and x_i is the score predicted for ith concept classifier. In our proposed method, $m = 856$.

C. *Predicting Sentiment Labels*

Given that event concepts generated from similar images are likely to be semantically similar, our hypothesis is that these concepts would capture the sentiment conveyed in the image. For example, a birthday event image may contain top predicted concepts such as ‘celebrations’, ‘party’ *etc.* These are all positive concepts and thus, the overall image is predicted to be a positive image, as opposed to neutral or negative. Event concepts can thus predict the emotion conveyed by the image without any explicit sentiment-related feature computation. Figure 3 shows the complete event image sentiment classification pipeline.

IV. Experiments

In this section we describe our event image dataset, the user study conducted to generate sentiment labels for the dataset and our experimental setup to predict event image sentiments on the test set.

A. *Dataset*

We retrieve public images from Microsoft Bing using 24 event categories as search queries. Our event categories include **accidents**, **airplane crash**, **baby shower**, **birthday**, **carnivals**, **concerts**, **refugee crises**, **funerals**, **wedding**, **protests**, **wildfires**, **marathons etc.** These events are diverse, capture both planned and unplanned events and include personal as well as community-based events. We obtain around 10,500 images. We pass these images to the crowdsourcing platform Amazon Mechanical Turk and request crowdworkers to rate the sentiment of each image. We ask them to mark images with **one** of the following five options: (1) Positive, (2) Negative, (3) Neutral, (4) Not an event image or (5) Image does not load. Each image is labeled by three crowdworkers. We accept responses only from those workers who are located in the US and who have an approval rating of more than 95%.

We build our event sentiment database based on the following rules:

- We only keep images if at least 2 out of 3 crowdworkers agree on its sentiment label, whether positive, negative or neutral.
- We discard all images on which fewer than 2 crowdworkers agree on the sentiment label of the event image. We also discard those images crowdworkers mark as ‘Not an event image’ and ‘Image does not load’.

We discard images on which crowdworkers disagree because of the subjective nature of the task. The final number of images retained is 8,748. Hence we find that crowdworkers agree on the sentiment labels of 83.3% of the initial images.

The distribution of sentiments in our final dataset is shown in Figure 4. As the pie chart shows, the positive and neutral images are more than six times as many as the negative images. This is because social media platforms are generally perceived as places that promote the sharing and dissemination of positive thoughts and behaviors. Further, the recent Facebook emotional contagion study [18], pointed to the fact that people engage more with positive posts, while negative posts

[1] Hybrid-CNN model is publicly available at https://github.com/BVLC/caffe/wiki/Model-Zoo

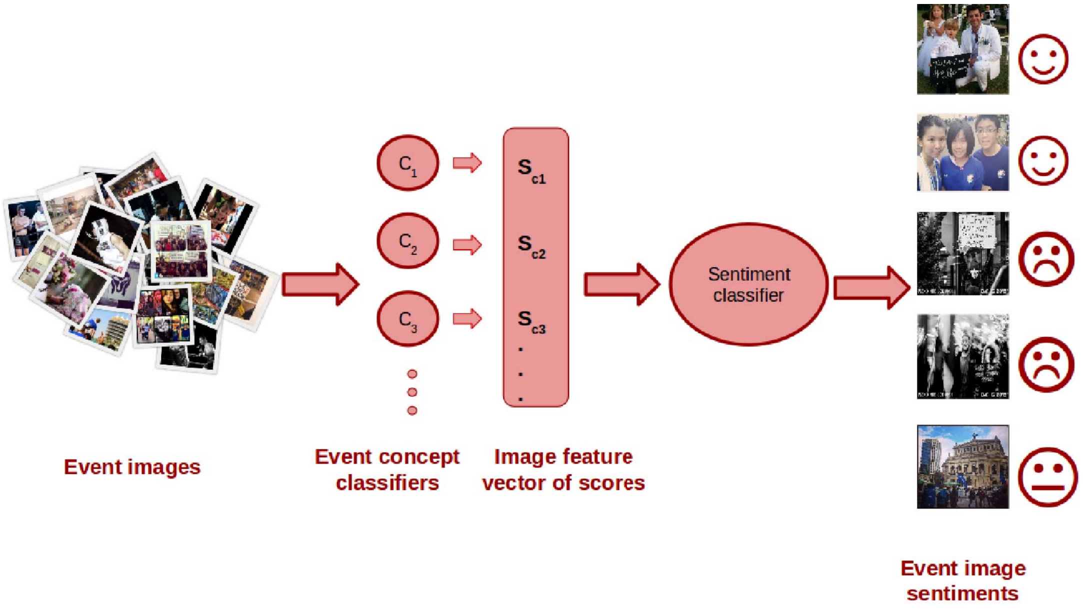

Fig. 3: Sentiment classification pipeline.

decrease user engagement. Hence, even for events that are negative in general (such as earthquakes, societal upheavals and crises), images related to rehabilitation efforts, political liberty or community solidarity may be perceived as positive.

Figure 5 shows a few examples of positive, negative and neutral images as annotated and agreed upon by crowdworkers. The top row shows positive images and it can be seen that many different events can convey positive emotions. Similarly, negative images show clear cases of violence and attacks. The bottom row shows neutral events and this is what the bulk of the images are annotated as; as no clear positive or negative emotion is conveyed by these images.

B. Experimental Setup

We set up our experiments with the annotated event image dataset. For training, we randomly sample 70% of the images from each sentiment class as positive training data and an equal number of training images from the rest of the sentiment classes as negative training data. We test on the remaining (30%) of images per class. Our test set also consists of an equal number of negative test data sampled from the other sentiment classes than the one being tested. Hence our sentiment prediction baseline accuracy is always 50%. We use this one-vs-all strategy, repeat this procedure 5 times and average the sentiment prediction accuracies per class to obtain the final accuracy.

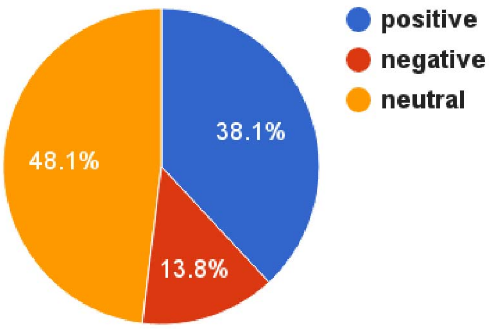

Fig. 4: Distribution of sentiments in our crowd-annotated social event image dataset.

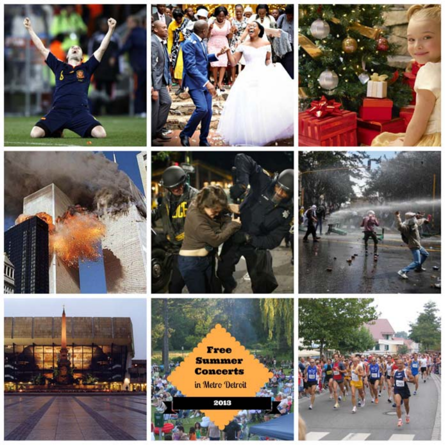

Fig. 5: Event images with sentiments agreed upon by majority vote: The top row shows positive event images, middle row shows negative images and bottom row shows neutral images.

We compute our event concept scores on the images by using the Caffe [16] deep learning framework. This tool extracts CNN layer 7 activations ('fc7') as features for all the images using AlexNet [19] architecture pre-trained on HybridCNN. Each feature is 4096-dimensional. HybridCNN is a CNN model pretrained on 978 object categories from ImageNet database [27] and 205 scene categories from Places dataset [38].

Then we use our trained event concept classifiers to predict the concept score for each image. We concatenate the concept scores to form the final feature vector for each image. These scores are then input to a linear SVM (We use the publicly available LIBLINEAR library [9]) that trains a sentiment detection model for each sentiment class and predicts the sentiment of the 30% test samples per class. We evaluate the effectiveness of our algorithm by computing the sentiment prediction accuracy for each class and the overall average accuracy.

V. Results and Discussion

Table I shows the sentiment prediction accuracies for several powerful state-of-the-art baselines and our proposed event concept features on our event sentiment dataset. We use the SentiBank [2] and Deep SentiBank [4] implementations provided by the authors. We also compare against the baselines of directly using fc7 features from AlexNet [19] and Hybrid-CNN and training a sentiment classifier on top of the fc7 features. For all the sentiment classes as well as overall average sentiment prediction, our proposed approach outperforms the state-of-the-art. This is achieved given that our method does not use sentiment-specific concepts such as 'smiling baby'. Our method also shows superior performance to deep CNN features (AlexNet and HybridCNN), demonstrating that off-the-shelf deep CNN features are insufficient to recognize sentiments in event images containing crowded and complex scenes.

The reason why sentiment-specific mid-level representation (adjective noun-pairs) does not work well with social event images is that concepts such as 'magical sunset' or 'amazing sky' may be relevant for general images shared on the web but social event images comprise complex interplay of objects, people and scenes. Our event concepts such as 'shouting slogans' or 'birthday girl' are event specific and generalize to many different events.

Sample positive and negative images correctly classified by our proposed method are shown in Figure 6. The positive images (first row) have the following event concepts predicted on them: 'crowd parade', 'troupe performs', 'party students', 'streets' etc. The second row depicts negative sentiment images that are correctly identified. It is apparent that the colors in the image also affect the sentiment annotation and thus we see dark black and gray tones in some of the negative images. Sample negative images with their top predicted concepts are shown in Figure 7.

However, there are some event images where our sentiment classifier does not predict the correct sentiment. This is due to the subjectivity in deciding which image evokes a neutral or negative emotion as can be seen in Figure 8. Since there are images in these color tones in the dataset which are labeled as negative, the classifier predicted negative sentiment on these images.

TABLE I: Per-class and average accuracy (in %) of event image sentiment prediction.

Features	positive	negative	neutral	avg. accuracy
AlexNet CNN	64.67	35.25	63.96	54.63
Hybrid CNN	72.15	67.08	61.27	66.83
SentiBank	62.31	60.79	59.09	60.73
Deep SentiBank	74.52	71.74	65.83	70.69
Event concepts (ours)	**77.11**	**74.13**	**67.94**	**73.06**

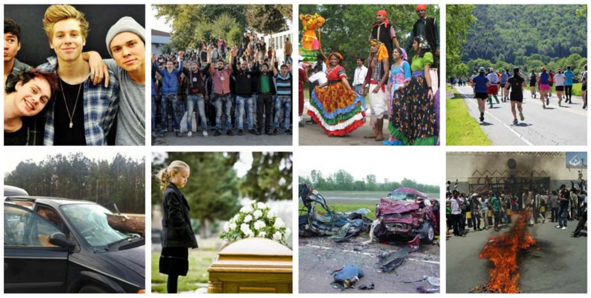

Fig. 6: Correct positive (top row) and negative (bottom row) sentiment predictions by our proposed method on the social event dataset

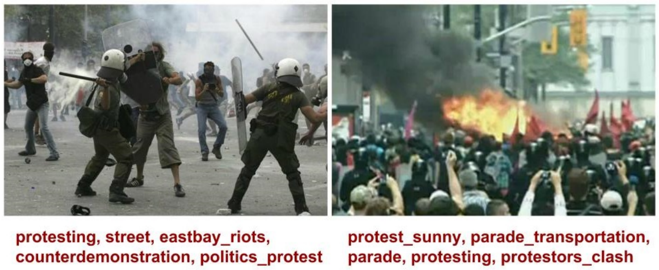
protesting, street, eastbay_riots, counterdemonstration, politics_protest
protest_sunny, parade_transportation, parade, protesting, protestors_clash

Fig. 7: Top predicted concepts for sample negative images in our dataset

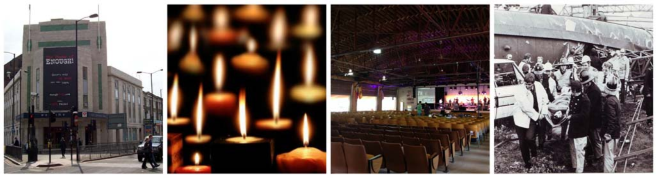

Fig. 8: Neutral sentiment images but classifier predicts them as negative images

Similarly there are images annotated as 'neutral' but the classifier predicts them as positive due to the stronger positive cues present in these images as depicted in Figure 9. A possible solution to this is to add more training data explicitly drawing the line between positive and neutral sentiment and

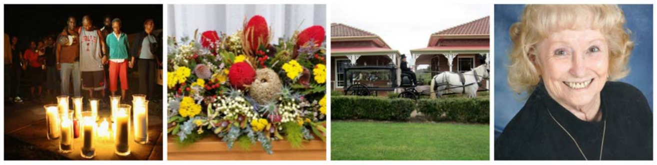

Fig. 9: Neutral sentiment images but classifier predicts them as positive images

negative and neutral sentiment in complex event images. It constitutes a promising direction for future extensions of this work.

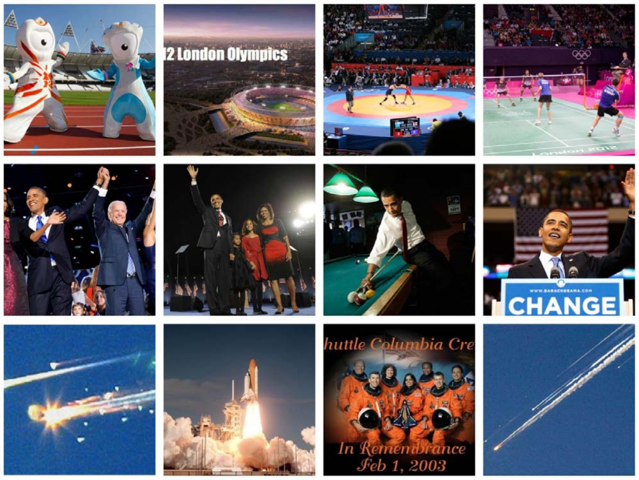

Fig. 10: Sample images from the characterization dataset used for qualitative analysis. From top to bottom, the events are: *Summer Olympics 2012, Obama wins elections 2008 and Columbia Space Shuttle Disaster*

A. *Generalizability & Validity*

We augment our experiments with a sentiment characterization study on a dataset of specific news event images crawled from the web. Our purpose is to qualitatively analyze our algorithm's performance on *unknown event images* (events not present in the training set) and to generalize and validate the use of event concept scores as features to classify sentiment in social event images. We mine 8,000 images from Microsoft Bing for 24 specific events such as *royal wedding, election campaign Trump, Summer Olympics 2012, Obama wins elections 2008, Columbia Space Shuttle Disaster, Arab Spring, Hurricane Katrina, Boston Bombing etc.* Sample images from this dataset are shown in Figure 10. This dataset is different from the previous one in that these events are specific (happened in a particular place and time). These events are chosen such that they should contain images conveying a balanced range of emotions. We do not use these images for training any model. We compute event concept scores on all the images and input them to the trained SVM model to predict the underlying sentiment. This model predicts whether the images are positive, negative or neutral. The model predictions are then qualitatively analyzed to see which images result in what kind of sentiment predictions.

Figure 11 shows images predicted as positive in this dataset. Since there is no ground truth, we qualitatively inspect the results. As the figure shows, the positive prediction makes intuitive sense on most of the images. Recall we do not use any of these images in the training set. We also show images that are predicted as negative in this database. Figure 12 shows such images. These images belong to events such as *Russian airstrikes, Arab Spring, Humanity washed ashore, US war Afghanistan, Nepal earthquake* etc. These predictions also make sense; visually as well as cognitively. However there are also cases where images from almost all events are classified into sentiment categories that do not make cognitive sense (for example, classifying a Hurricane Sandy image as positive as shown in Figure 11). The explanation behind such misclassification is that these images contain very little visual cues to direct our sentiment classifier to recognize the underlying event. Another scenario where our algorithm can give random predictions (or just classify everything as neutral since this is the largest class in our data) is when the images are ambiguous. Subjectivity remains an open challenge, but we believe we have addressed this issue and taken steps towards the right direction.

B. *Limitations and Future Work*

We recognize limitations in our approach. The learnt model can recognize positive images with great accuracy where strong visual cues are present in the image but makes errors when differentiating between positive/negative and neutral sentiments.

To elaborate on this, consider Table II. It shows the top most frequent event concepts for all positive, negative and neutral images respectively in our social event dataset. We can qualitatively validate that our event concepts computed on images marked as positive are associated with positive sentiments (e.g. festivities, party, birthday celebrations etc.). Similarly, there are many predicted concepts associated with negative sentiments but a few of these remain ambiguous e.g. parading. This shows us some limitations with our event

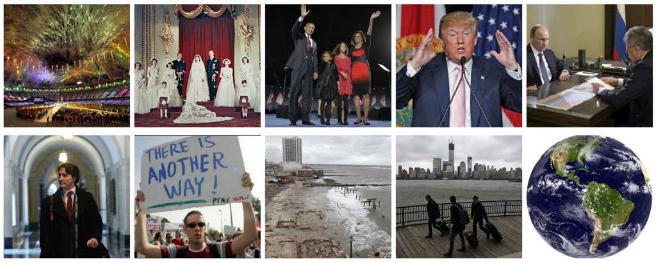

Fig. 11: Images in the characterization dataset which are predicted as positive

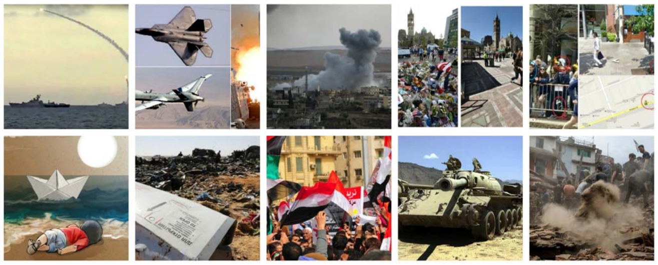

Fig. 12: Images in the characterization dataset which are predicted as negative

concept modeling approach where some predicted concepts on images may not correspond to the actual image content thus rendering their sentiment different to what the images should convey. Our top predicted concepts for neutral images in the dataset contain a variety of event concepts, ranging from protest-related concepts to birthdays and holidays. This can result in neutral predictions by the sentiment classifier which is biased towards the largest class present in our dataset (neutral).

Summarily, we find that there is a gap between human perception of an event (e.g. 'all images of Nepal earthquake must be negative') and actual images obtained from the web which contain a variety of emotions associated with the events. However, we believe that our approach generally captures the nuanced nature of affect around an event on the image level satisfactorily.

Future work includes extending the richness of social event data by adding more training data and richer labels to the sentiment recognition pipeline and potentially improving the classifier confusion between the three sentiments.

VI. Conclusion

Our work introduces a framework to predict complex image sentiment using visual content alone. We introduce an annotated social event dataset and demonstrate that our proposed event concept features can be mapped effectively to sentiments. We evaluate our algorithm against state-of-the-art approaches and our method outperforms them by a significant margin. We also examine the performance of our event sentiment detector on an unseen dataset of images spanning events not considered in model training, and thus assess our proposed method's broader generalizabilty and validity.

References

[1] U. Ahsan, C. Sun, J. Hays, and I. Essa. Complex event recognition from images with few training examples. *arXiv preprint arXiv:1701.04769*, 2017.

[2] D. Borth, R. Ji, T. Chen, T. Breuel, and S.-F. Chang. Large-scale visual sentiment ontology and detectors using adjective noun pairs. In *Proceedings of the 21st ACM international conference on Multimedia*, pages 223–232. ACM, 2013.

[3] G. Cai and B. Xia. Convolutional neural networks for multimedia sentiment analysis. In *Natural Language Processing and Chinese Computing*, pages 159–167. Springer, 2015.

[4] T. Chen, D. Borth, T. Darrell, and S.-F. Chang. Deepsentibank: Visual sentiment concept classification with deep convolutional neural networks. *arXiv preprint arXiv:1410.8586*, 2014.

[5] M. De Choudhury, M. Gamon, and S. Counts. Happy, nervous or surprised? classification of human affective states in social media. In *Sixth International AAAI Conference on Weblogs and Social Media*, 2012.

[6] M. De Choudhury, M. Gamon, S. Counts, and E. Horvitz. Predicting depression via social media. In *ICWSM*, 2013.

[7] A. Dhall, R. Goecke, and T. Gedeon. Automatic group happiness intensity analysis. *IEEE Transactions on Affective Computing*, 6(1):13–26, 2015.

[8] N. A. Diakopoulos and D. A. Shamma. Characterizing debate performance via aggregated twitter sentiment. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 1195–1198. ACM, 2010.

[9] R.-E. Fan, K.-W. Chang, C.-J. Hsieh, X.-R. Wang, and C.-J. Lin. Liblinear: A library for large linear classification. *The Journal of Machine Learning Research*, 9:1871–1874, 2008.

[10] S. A. Golder and M. W. Macy. Diurnal and seasonal mood vary with work, sleep, and daylength across diverse cultures. *Science*, 333(6051):1878–1881, 2011.

[11] M. Hagen, M. Potthast, M. Büchner, and B. Stein. Twitter sentiment detection via ensemble classification using averaged confidence scores. In *Advances in Information Retrieval*, pages 741–754. Springer, 2015.

[12] A. Hanjalic, C. Kofler, and M. Larson. Intent and its discontents: the user at the wheel of the online video search engine. In *Proceedings of the 20th ACM international conference on Multimedia*, pages 1239–1248. ACM, 2012.

[13] Y. Hu, F. Wang, and S. Kambhampati. Listening to the crowd: automated analysis of events via aggregated twitter sentiment. In *Proceedings of the Twenty-Third international joint conference on Artificial Intelligence*, pages 2640–2646. AAAI Press, 2013.

TABLE II: Top predicted concepts for positive, negative and neutral images on characterization dataset.

Sentiment	**Top predicted concepts**
Positive	concert, festivities, party, birthday celebrations, food, wedding church, bride heart, homecoming parade
Negative	protesting, politics protest, police parade, riots, parading, marchers protest, antiwar demonstrations
Neutral	horribles parade, diploma, rally, activism, house concert, street, paint balling, party students, fall graduates

[14] B. J. Jansen, M. Zhang, K. Sobel, and A. Chowdury. Twitter power: Tweets as electronic word of mouth. *Journal of the American society for information science and technology*, 60(11):2169–2188, 2009.

[15] J. Jia, S. Wu, X. Wang, P. Hu, L. Cai, and J. Tang. Can we understand van gogh's mood?: learning to infer affects from images in social networks. In *Proceedings of the 20th ACM international conference on Multimedia*, pages 857–860. ACM, 2012.

[16] Y. Jia, E. Shelhamer, J. Donahue, S. Karayev, J. Long, R. Girshick, S. Guadarrama, and T. Darrell. Caffe: Convolutional architecture for fast feature embedding. *arXiv preprint arXiv:1408.5093*, 2014.

[17] S. Kiritchenko, X. Zhu, and S. M. Mohammad. Sentiment analysis of short informal texts. *Journal of Artificial Intelligence Research*, pages 723–762, 2014.

[18] A. D. Kramer, J. E. Guillory, and J. T. Hancock. Experimental evidence of massive-scale emotional contagion through social networks. *Proceedings of the National Academy of Sciences*, 111(24):8788–8790, 2014.

[19] A. Krizhevsky, I. Sutskever, and G. E. Hinton. Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems*, pages 1097–1105, 2012.

[20] B. Li, S. Feng, W. Xiong, and W. Hu. Scaring or pleasing: exploit emotional impact of an image. In *Proceedings of the 20th ACM international conference on Multimedia*, pages 1365–1366. ACM, 2012.

[21] C. Li, A. Sun, and A. Datta. Twevent: segment-based event detection from tweets. In *Proceedings of the 21st ACM international conference on Information and knowledge management*, pages 155–164. ACM, 2012.

[22] J. Li, S. Roy, J. Feng, and T. Sim. Happiness level prediction with sequential inputs via multiple regressions. In *Proceedings of the 18th ACM International Conference on Multimodal Interaction*, pages 487–493. ACM, 2016.

[23] J. Machajdik and A. Hanbury. Affective image classification using features inspired by psychology and art theory. In *Proceedings of the international conference on Multimedia*, pages 83–92. ACM, 2010.

[24] T. Mikolov, K. Chen, G. Corrado, and J. Dean. Efficient estimation of word representations in vector space. *arXiv preprint arXiv:1301.3781*, 2013.

[25] W. Mou, H. Gunes, and I. Patras. Automatic recognition of emotions and membership in group videos. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 27–35, 2016.

[26] B. O'Connor, R. Balasubramanyan, B. R. Routledge, and N. A. Smith. From tweets to polls: Linking text sentiment to public opinion time series. *ICWSM*, 11(122-129):1–2, 2010.

[27] 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)*, pages 1–42, April 2015.

[28] S. Siersdorfer, E. Minack, F. Deng, and J. Hare. Analyzing and predicting sentiment of images on the social web. In *Proceedings of the international conference on Multimedia*, pages 715–718. ACM, 2010.

[29] A. Sun and S. S. Bhowmick. Quantifying tag representativeness of visual content of social images. In *Proceedings of the international conference on Multimedia*, pages 471–480. ACM, 2010.

[30] B. Sun, Q. Wei, L. Li, Q. Xu, J. He, and L. Yu. Lstm for dynamic emotion and group emotion recognition in the wild. In *Proceedings of the 18th ACM International Conference on Multimodal Interaction*, pages 451–457. ACM, 2016.

[31] V. Vonikakis and S. Winkler. Emotion-based sequence of family photos. In *Proceedings of the 20th ACM international conference on Multimedia*, pages 1371–1372. ACM, 2012.

[32] V. Vonikakis, Y. Yazici, V. D. Nguyen, and S. Winkler. Group happiness assessment using geometric features and dataset balancing. In *Proceedings of the 18th ACM International Conference on Multimodal Interaction*, pages 479–486. ACM, 2016.

[33] Y. Wang, Y. Hu, S. Kambhampati, and B. Li. Inferring sentiment from web images with joint inference on visual and social cues: A regulated matrix factorization approach. In *Ninth International AAAI Conference on Web and Social Media*, 2015.

[34] J. Wu, Z. Lin, and H. Zha. Multi-view common space learning for emotion recognition in the wild. In *Proceedings of the 18th ACM International Conference on Multimodal Interaction*, pages 464–471. ACM, 2016.

[35] 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.

[36] Q. You, J. Luo, H. Jin, and J. Yang. Robust image sentiment analysis using progressively trained and domain transferred deep networks. In *The Twenty-Ninth AAAI Conference on Artificial Intelligence (AAAI)*, 2015.

[37] 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.

[38] B. Zhou, A. Lapedriza, J. Xiao, A. Torralba, and A. Oliva. Learning deep features for scene recognition using places database. In *Advances in Neural Information Processing Systems*, pages 487–495, 2014.