Analyzing and Predicting GIF Interestingness

Michael Gygli*
Computer Vision Lab
ETH Zurich
Switzerland
gygli@vision.ee.ethz.ch

Mohammad Soleymani*
Swiss Center for Affective Sciences
University of Geneva
Switzerland
mohammad.soleymani@unige.ch

ABSTRACT

Animated GIFs have regained huge popularity. They are used in instant messaging, online journalism, social media, among others. In this paper, we present an in-depth study on the interestingness of GIFs. We create and annotate a dataset with a set of affective labels, which allows us to investigate the sources of interest. We show that GIFs of pets are considered more interesting that GIFs of people. Furthermore, we study the connection of interest to other features and factors such as popularity. Finally, we build a predictive model and show that it can estimate GIF interestingness with high accuracy. Our model outperforms the existing methods on GIF popularity, as well as a model based on still image interestingness, by a large margin. We envision that the insights and method developed can be used for automatic recognition and generation of interesting GIFs.

1. INTRODUCTION

Interest is related to curiosity and attention and as a result an important factor in driving multimedia users' online behavior [15]. It has various applications in multimedia such as video highlight detection [20] and media recommendation. Flickr, for example[1], highlights interesting photos shared on their website. Rather than analyzing the visual content, however, they rely on information such as the number of likes, comments, etc. [5]. Thus, they require users' feedback and social interactions in order to surface interesting content. However, social interactions and interestingness do not always co-exist [14], as we will show in our experiments.

The Graphics Interchange Format (GIF) is a graphical format that was introduced in 1987 by the Internet provider company CompuServe to depict moving pictures on the low-bandwidth Internet of that time. Recently, GIFs are making a comeback [22] and there are popular websites such as Tumblr[2] which is mainly used for posting and sharing GIFs.

*Equal contributions
[1]https://www.flickr.com/explore/interesting/
[2]https://www.tumblr.com/

MM '16, October 15 - 19, 2016, Amsterdam, Netherland

ISBN 978-1-4503-3603-1/16/10…$15.00

DOI: http://dx.doi.org/10.1145/2964284.2967195

Figure 1: Most and least interesting GIFs from our dataset. GIFs of action scenes, pets and accidents are considered most interesting, while most boring GIFs are of people.

GIFs have become a very popular media, due to their unique properties. In contrast to images, animated GIFs have better capabilities to show dynamic content, tell stories and convey emotions. At the same time they remain short and without sound, which makes them more discreet and easily consumable compared to videos, which require higher time and bandwidth commitment [1].

In this paper, we conduct an in-depth study on what makes GIFs interesting. Towards this goal, we constructed a database of GIFs annotated with interest and other affective attributes. 2739 GIFs were selected from the previous studies on GIFs [9, 1] and labeled on Amazon Mechanical Turk. A content-based GIF interestingness method is developed and evaluated. We obtain a prediction performance that is comparable to the ones on still images [18, 8, 7]. The contributions of this work are as follows: (i) We analyze the effect of emotional attributes on GIF interestingness and provide an in-depth analysis into what makes a GIF interesting; (ii) We study the relationship between GIF interestingness to GIF popularity as well as image interestingness. (iii) We build a computational model for automatic prediction of GIF interestingness.

2. BACKGROUND

Interest is a state which is related to curiosity and drives our attention. Silvia among other psychologist posited that interest is an emotion which has consistent subjective feelings and expressions in adults [15]. Halonen *et al.* [10] identified a set of characteristics that are related to visual interestingness through a qualitative study. The characteristics which were deemed relevant were aesthetics, affect, colors, composition, genre and personal connection. Chu *et al.* [6] studied the effect of familiarity of faces and images context

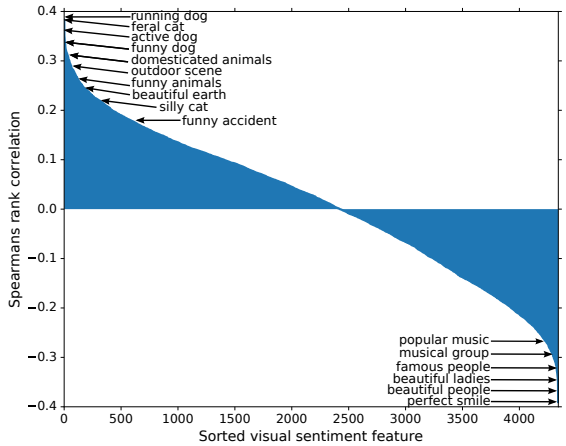

Figure 2: Correlations of visual sentiment attributes [13] with GIF interest. Some groups of contents stand out: pets, natural/aesthetic scenes and, on the negative side, music and people (see text).

on visual interestingness. They found that personally familiar faces and unfamiliar contexts are positively correlated with the interestingness in images.

In the light of the recent surge in the popularity of GIFs, Bakhshi *et al.* [1] studied the engaging factors in GIFs using a combination of interviews with GIF users and simple computational features. They found that animation, lack of sound, immediacy of consumption, low bandwidth and minimal time demands, the storytelling capabilities and utility for expressing emotions were the most important factors in making GIFs engaging. Jou *et al.* [12] built a tool for automatic prediction of perceived emotion in GIF viewers. Color histogram, facial expressions depicted in GIFs, visual features related to aesthetics and visual sentiment [2] were used as their content features. They found that facial expressions depicted in GIFs were the most effective features in determining perceived emotions in response to GIFs.

Gygli et al. and Grabner et al. [8, 7] showed that visual content features related to unusualness, aesthetics and general preference are effective in predicting visual interestingness. More recently, Soleymani [18] built a model for personalized interest prediction for images. He found that affective content, quality, coping potential and complexity have a significant effect on visual interest in images.

3. DATASET

We sampled 1729 GIFs from the Video2GIF dataset [9] and 1010 GIFs from the Tumblr dataset of [1]. We sampled randomly from Video2GIF, while for Tumblr, we undersampled GIFs with few likes, to have a more balanced distribution of likes and reblogs. We manually removed GIFs containing text and pornographic material. We deliberately removed the GIFs with subtitles to focus on visual interestingness of the content rather than language in the dialogues. In total, we selected 2739 GIFs to be labeled for our dataset, which is available online[3]. GIFs were in average 4.25 seconds long with the average frame rate of 11 frames/second.

[3]http://cvml.unige.ch/databases/gifInterest/

Scale	Kri. $\alpha \uparrow$
Interesting - Uninteresting	0.25
Boring - Exciting	0.32
Arousing - Soothing	0.31
Pleasant - Unpleasant	0.16
Appealing - Unappealing	0.13
Want full video - Don't want full video	0.17
Explicit - Non-explicit	0.30

Table 1: Krippendorff's alpha for different scales. All the scales were ordinal with the exception of explicitness which was nominal.

Scale	Arousal	Valence	Appeal	Curiosity
ρ	0.439	0.299	0.526	0.779

Table 2: Spearman's correlation coefficients (ρ) between the interestingness and emotional ratings.

GIFs were labeled through crowdsourcing on Amazon Mechanical Turk. Motivated by the affective dimension of interest and the previous findings on interestingness in images [18, 10], we labeled GIFs on interestingness, aesthetics, arousal and valence (pleasantness) on five point scales. We also labeled them on the presence of explicit or erotic content due to the strong effect of sexual arousal in emotional responses [3]. The scales were appealing-unappealing for aesthetics, arousing-soothing for arousal, pleasant-unpleasant for valence, interesting-uninteresting and boring-exciting for interest and "want to see full video"-"do not want to watch the full video" for curiosity. Each GIF received five labels on all these scales. 41 crowd-workers, 21 male, aged between 22 - 59 years old ($\mu = 35.0, \sigma = 9.6$) annotated the GIFs. They receive \$0.07 to annotate each GIF. Due to the culture specific references in the content of GIFs, we decided to restrict the tasks to the workers based in the USA. We calculated Krippendorff's alpha to measure the agreement of the labels adjusted to the chance agreement ($max(\alpha) = 1$). All the inter-rater agreements were in range of fair to moderate agreement similar to previous studies [18] (see Table 1).

Similar to [16], the interesting - uninteresting scale was combined with reverse boring - exciting to form the interestingness score which is used in the remaining of this paper.

4. ANALYSIS

While GIFs have gained immense popularity, they have not been analyzed from a computational perspective. One notable exception is the work of Baskshi *et al.* [1]. They however only analyze the relationship of low-level image features and GIF popularity. Instead, we use our new dataset to do an in-depth analysis of what makes a GIF interesting. Concerning the affective dimension of interest, the Spearman's rank correlation was calculated between interest and the emotional ratings (see Table 2). As expected, there is a strong correlation between interest and curiosity. All emotional ratings are significantly correlated with interest. Aesthetic appeal of GIFs has a higher correlation with interest compared to arousal or valence. In Figure 2, we relate visual sentiments, detected by [13], and interest, while Figure 3 shows a hierarchical clustering on the same features.

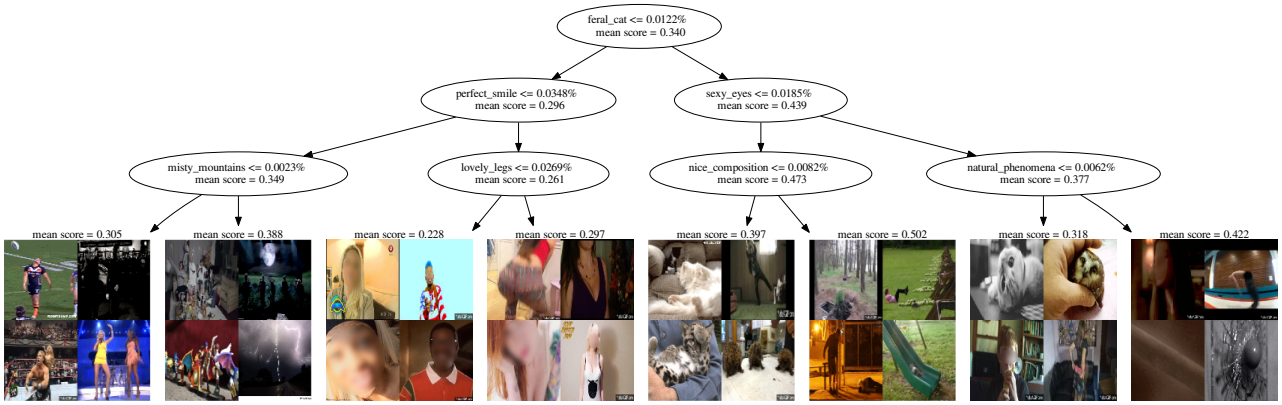

Figure 3: Hierarchical clustering of GIFs according to their interestingness scores by means of a CART tree [4] and visual sentiment features [13].

Thereby we observe several groups of attributes that are strongly correlated with interest. Namely GIFs of pets, lead by cats and dogs are typically considered of interest. Furthermore, beautiful and natural scenes, as well as funny content in general are deemed interesting. Negatively correlated attributes are mainly from two groups: music and people. Surprisingly, the presence of people has inverse correlation with GIF's interestingness, even though there are very popular memes of facial expressions. However, these are exceptions and the larger part of GIFs containing people are considered dull. This suggests that for finding interesting GIFs of people, a more fine-grained analysis of facial expressions is necessary. Thus, we extracted facial expressions for the subpart of GIFs containing a considerable amount of frames containing faces and correlated joyful, surprised, sad and angry expressions with interest. While joy remains negatively correlated (-0.129), the other emotions have a weak positive correlation with interest (surprise: 0.019, sorrow (sadness) 0.036, anger 0.096). Next, we put our annotated attributes in relation to the number of views and reblogs (Table 3). For this, we use the GIFs from Tumblr only, as it provides accurate social data [1]. In accordance with one's expectation, interest is related to the number of likes ($\rho = 0.279$). In fact, it is the most correlated among all the attributes we investigated. The number of times a GIF was re-blogged, however is not significantly correlated (p-value> 0.01) with any attribute, except the explicitness of content. Explicitness has a significant negative correlation, meaning that fewer Tumblr bloggers are re-sharing explicit GIFs compared to other content. While interest and the number of likes are related, their correlation is relatively weak. Popularity is not only influenced by the content but also by social factors such as uploaders' popularity. This effect was also observed and investigated by Schifanella *et al.* [14], who surface appealing photos among unpopular Flickr images.

5. INTEREST PREDICTION

In order to predict the interestingness of GIFs, we learn a model using a support vector regression (SVR) with an RBF kernel. As feature representation, we use a range of visual features, as described in the next Section.

Baselines. We compare our model with two related work, namely, image interestingness [8] and Video2GIF [9].

For image interestingness, we follow [8] and train a SVR with RBF kernel on the dataset of [11] with their interestingness labels. Rather than using low-level features, however, we use the outputs from the fully connected layer 6 (fc6) of a deep convolutional neural network, i.e., VGGNet-19 [17]. For reference, we evaluated this model on the test set of [8] and obtained a performance of $\rho = 0.70$, thus outperforming the original work of [8] ($\rho = 0.60$).

	rank correlation ρ with	
Attribute	**# of likes**	**# of reblogs**
Interest	0.279	-0.039
Arousal	0.056	-0.015
Curiosity	0.251	-0.069
Appeal	0.224	0.018
Valence	0.152	0.010
Explicit	-0.035	-0.149

Table 3: The annotated attributes in relation to popularity (views and reblogs)

Video2GIF is a method that learns to rank segments within a video according to their likelihood to be selected as GIFs. To this end, it uses pairs of video segments (positive vs. negative). A positive is a video segment that was selected to produce a GIF, while a negative segment is one that was never selected by users to produce a GIF. We compare two configurations: (i) A ranking learned from pairs (pos, neg) in the same video ("within") and (ii) a ranking learned from (pos, neg) in different videos ("across").

Implementation details. We report results on the full dataset, using cross-validation: We split the dataset into 5 parts, and train on four parts and test on the fifth, in turn. As performance metrics, we report Spearman's rank correlation ρ as well as root mean square error (RMSE). When training the SVRs, we optimize C after cross-validation for each training split. For the RBF-kernel we set $\gamma = 1/d$, where d is the dimensionality of the feature representation.

5.1 Visual features

We use a range of visual features at varying levels of abstraction. Based on [1], a set of simple low-level features is extracted from every frame of GIFs. These simple features are pooled at the frame level by mean, standard deviation

	Feature	Description	Dimensionality
Simple features	Visual features [1]	Entropy, exposure, balance, brightness, compression quality [21], contrast, sharpness, uniformity, face count, face region, image asymmetry, motion energy, Contrast balance (Euclidean distance between the original image and the contrast-equalized image)	33
	Meta-information [1]	GIF frame rate, number of pixels, duration and aspect-ratio	4
	Loopiness	Distance between last and first frame	1
Google API	Text and face area	Ratio between text and face area and the full GIF[4]	2
	Spoof likelihood	Likelihood that a modification was made to the GIF to make it appear funny or offensive[4]	1
	Violence likelihood	Likelihood that the GIF depicts violent content[4]	1
	Facial expressions	Degree of the emotions sorrow, anger, surprise and joy[4]	4
	Visual sentiments [13]	Probabilities of adjective noun pairs related to sentiments	4342
	C3D [19]	Features from a CNN with spatio-temporal convolutions trained for action recognition in videos. We use the first fully-connected layer	4096

Table 4: Visual features, their descriptions and dimensionality.

and skewness to form a feature vector for each GIF. Similar to [1], GIF characteristics such as frame rate, number of pixels, duration and aspect-ratio are also extracted. We further extract attributes related to the content of frames using Google Cloud Vision API[4], namely the area covering faces and text, spoof and violence likelihood, as well as facial expressions. Features by Google Cloud API are pooled by averaging their values at frame level. Furthermore, we use two types of CNN-based features. (i) Spatio-temporal CNN features that were trained for action recognition and (ii) Visual sentiment attributes, which consist of probabilities for adjective noun pairs. A comprehensive list of features and their dimensionality is given in Table 4.

5.2 Results

The results for prediction GIF interest are summarized in Figure 5. Results demonstrate that our approach is able to predict GIF interestingness with high accuracy, compared to the baselines. It is interesting to observe that the sentiment features, while lacking temporal information, perform better than C3D. This indicates that the types of objects and their associated sentiments are more important than motion information. Our best performing model has a rank correlation of 0.53 showing that GIF interest can be predicted well from visual information only. When comparing the models trained for predicting GIF interestingness with previous methods, they perform much better than the baselines. This is not surprising, given that the baselines were not trained for this task directly. Somewhat surprisingly, Video2GIF trained to rank segments from the same video is actually uncorrelated with GIF interest (Video2GIF within). However, Video2GIF trained with samples coming from different videos is correlated. Its correlation (0.21) is even higher than that of image interestingness (0.15), despite not being trained with interest labels.

	Method	rank corr $\rho \uparrow$	RMSE $\downarrow$
As is	Image Interest [8]	0.1544	-
	Video2GIF [9] within	-0.0082	-
	Video2GIF [9] across	0.2055	-
Retrained SVR	Simple features [1]	0.3870	0.1818
	C3D [19]	0.4809	0.1746
	Sentiment features [13]	0.5219	0.1695
	Google API features	0.3806	0.1817
	C3D+Sentiments	0.5222	0.1694
	Full	**0.5308**	**0.1685**

Table 5: Predicting mean interest in GIFs

6. CONCLUSIONS

In this work, we analyzed and predicted human interest in GIFs. Towards this goal, we have introduced a new GIF dataset with interestingness annotation. We found the adjective noun pairs or sentiment features to be the most informative features, even though they are extracted from still images. GIFs depicting pets were in average found to be more interesting than the ones showing people. We also found that interestingness is associated with likability, i.e., number of likes a GIF receives. However, interestingness was not correlated with reblogging behavior or social popularity. Finally, we used our dataset to build a predictive model and showed its effectiveness in predicting interestingness. In the future, we would like to extend this model to predict personalized interest.

7. ACKNOWLEDGMENT

The work of Gygli was supported by the ERC project VarCity (#273940). Soleymani received support from the Swiss National Science Foundation (Ambizione and NCCR Affective Sciences).

[4]https://cloud.google.com/vision/

8. REFERENCES

[1] S. Bakhshi, D. Shamma, L. Kennedy, Y. Song, P. de Juan, and J. Kaye. Fast, Cheap, and Good: Why Animated GIFs Engage Us. In *Proceedings of the 34rd Annual ACM Conference on Human Factors in Computing Systems (CHI)*, 2016.

[2] 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 *Proceedings of the 21st ACM international conference on Multimedia*, pages 459–460. ACM, 2013.

[3] M. M. Bradley, M. Codispoti, D. Sabatinelli, and P. J. Lang. Emotion and motivation ii: sex differences in picture processing. *Emotion*, 1(3):300, 2001.

[4] L. Breiman, J. Friedman, C. J. Stone, and R. A. Olshen. *Classification and regression trees*. CRC press, 1984.

[5] D. S. Butterfield, C. Fake, C. J. Henderson-Begg, and S. Mourachov. Interestingness ranking of media objects, patent application 20060242139. (US 20060242139), 2006.

[6] S. L. Chu, E. Fedorovskaya, F. Quek, and J. Snyder. The effect of familiarity on perceived interestingness of images. In *Proc. SPIE 8651, Human Vision and Electronic Imaging XVIII*, 2013.

[7] H. Grabner, F. Nater, M. Druey, and L. Van Gool. Visual interestingness in image sequences. In *Proceedings of the 21st Annual ACM Conference on Multimedia*, 2013.

[8] M. Gygli, H. Grabner, H. Riemenschneider, F. Nater, and L. Van Gool. The Interestingness of Images. In *The IEEE International Conference on Computer Vision (ICCV)*, 2013.

[9] M. Gygli, Y. Song, and L. Cao. Video2GIF: Automatic Generation of Animated GIFs from Video. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016.

[10] R. Halonen, S. Westman, and P. Oittinen. Naturalness and interestingness of test images for visual quality evaluation. In *IS&T/SPIE Electronic Imaging*, 2011.

[11] P. Isola, J. Xiao, A. Torralba, and A. Oliva. What makes an image memorable? In *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on*, pages 145–152. IEEE, 2011.

[12] B. Jou, S. Bhattacharya, and S.-F. Chang. Predicting Viewer Perceived Emotions in Animated GIFs. In *Proceedings of the 22Nd ACM International Conference on Multimedia*, MM '14, pages 213–216, New York, NY, USA, 2014. ACM.

[13] B. Jou, T. Chen, N. Pappas, M. Redi, M. Topkara, and S.-F. Chang. Visual Affect Around the World: A Large-scale Multilingual Visual Sentiment Ontology. In *ACM International Conference on Multimedia*, number 1, 2015.

[14] R. Schifanella, M. Redi, and L. M. Aiello. An Image is Worth More than a Thousand Favorites : Surfacing the Hidden Beauty of Flickr Pictures. *International Conference on Web and Social Media (ICWSM)*, 2015.

[15] P. J. Silvia. *Exploring the psychology of interest*. Oxford University Press, 2006.

[16] P. J. Silvia, R. A. Henson, and J. L. Templin. Are the sources of interest the same for everyone? using multilevel mixture models to explore individual differences in appraisal structures. *Cognition and Emotion*, 23(7):1389–1406, 2009.

[17] K. Simonyan and A. Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014.

[18] M. Soleymani. The quest for visual interest. In *Proceedings of the 23rd Annual ACM Conference on Multimedia*, pages 919–922. ACM, 2015.

[19] D. Tran, L. Bourdev, R. Fergus, L. Torresani, and M. Paluri. Learning Spatiotemporal Features with 3D Convolutional Networks. In *The IEEE International Conference on Computer Vision (ICCV)*, 2015.

[20] B. T. Truong and S. Venkatesh. Video abstraction: A systematic review and classification. *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)*, 3(1):3, 2007.

[21] Z. Wang, H. R. Sheikh, and A. C. Bovik. No-reference perceptual quality assessment of JPEG compressed images. In *International Conference on Image Processing*, volume 1, pages I–477. IEEE, 2002.

[22] A. Williams. Fresh From the Internet's Attic. New York Times, Feb 13, 2013.