Understanding Image Virality

Arturo Deza[1], Devi Parikh[2]
[1]UCSB. [2]Virginia Tech.

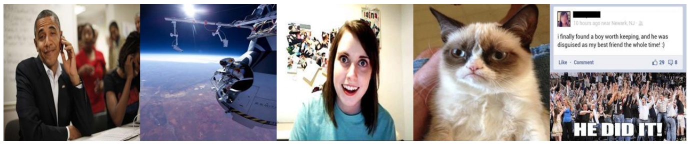

Figure 1: Example viral images from Reddit.

Virality of online content on social networking websites is an important but esoteric phenomenon often studied in fields like marketing, psychology and data mining. In this paper we study viral images from a computer vision perspective. We introduce three new image datasets from *Reddit* and define a virality score using *Reddit* metadata. We train classifiers with state-of-the-art image features to predict virality of individual images, relative virality in pairs of images, and the dominant topic of a viral image. We also compare machine performance to human performance on these tasks. We find that computers perform poorly with low level features, and high level information is critical for predicting virality. We identify the 5 key visual attributes that correlate with virality: animal, synthetically generated, (not) beautiful, explicit, sexual. We create an attribute-based representation of images that can predict relative virality with 68.10% accuracy (SVM+Deep Relative Attributes – better than humans at 60.12%. Finally, we study how human prediction of image virality varies with different "contexts" in which the images are viewed, such as the influence of neighbouring images, images recently viewed, as well as the image title or caption. This work is a first step in understanding the complex but important phenomenon of image virality. Our datasets and annotations will be made publicly available.

Reddit is the main engine of viral content around the world. We focus only on the image content. These images are sometimes rare photographs, or photos depicting comical or absurd situations, or Redditors sharing a personal emotional moment through the photo, or expressing their political or social views through the image, and so on. Each image can be upvoted or downvoted by a user. Viral content tends to be resubmitted multiple times as it spreads across the network of users. Viral images are thus the ones that have many upvotes, few downvotes, *and* have been resubmitted often by different users.

Let score S_h^n be the difference between the number of upvotes and downvotes an image h received at its n^{th} resubmission to a category. Let t be the time of the resubmission of the image and c be the category (*subreddit*) to which it was submitted. $\bar{S}_c^t$ is the average sore of all submissions to category c at time t. We define A_h^n to be the ratio of the score of the image h at resubmission n to the average score of all images posted to the category in that hour [3].

$$A_h^n = \frac{S_h^n}{\bar{S}_c^t} \tag{1}$$

We add an offset to S_h^n so that the smallest score $\min_h \min_n S_h^n$ is 0. We define the overall (across all categories) virality score for image h as

$$V_h = \max_n A_h^n log\left(\frac{m_h}{\bar{m}}\right) \tag{2}$$

where m_h is the number of times image h was resubmitted, and $\bar{m}$ is the average number of times any image has been resubmitted. If an image is resubmitted often, its virality score will be high.

We use $\sim 132K$ submissions from Reddit [3] to create a dataset of 10,078 images. We then subselect a dataset of 500 images containing the 250 most and least viral images each using Equation 2. This stark contrast in the virality score of the two sets of images gives us a clean dichotomy to explore as a first step in studying this complex phenomenon. In contrast with the clean dichotomy represented in the dataset above, we also create a dataset of pairs of images where the difference in the virality of the two images in a pair is less stark. We pair a random image from the 250 most viral images with a random image from $> 10k$ images with virality lower than the median virality. Similarly, we pair a random image from the 250 least viral images with a random image with higher than median virality. We collect 500 such pairs. We call this our 500_p dataset. Training was done on the the remaining 10k images by pairing above-median viral images with below-median viral images.

Classification Method	Performance
Chance	50%
Khosla et al. Popularity API [2]	51.12%
SVM + image features	58.49%
Human	60.12%
Human annotated Atts.-5	65.18%
SVM + Deep Attributes-5	**68.10%**

Table 1: Relative virality prediction accuracies on the 500_p dataset.

Prelimarinary evaluations showed that current image features have chance like performance for machine prediction of virality. We also found that human virality prediction is better if humans are asked to make a relative judgement on pairs of images, rather than absolute judgements. However their performance is not very high (60.12%). Note though, that humans are quite good at extracting relative attributes of image pairs [4]. This motivated us to annotate 52 relative attributes on our dataset, and discover a characterization of vitality in terms of these semantic attributes.

We use our annotated dataset and identify the top 5 attributes that when combined maximize virality prediction performance. These are: Animal, Synthetically Generated (SynthGen), (not) Beautiful, Explicit and Sexual. Note that these top-5 attributes are visual.

We then train relative attribute predictors for each of these attributes with DECAF6 deep features [1] and an SVM classifier to obtain relative attribute predictions on all image pairs. The relative attribute prediction accuracies we obtain by testing on our (500_p) dataset are: Animal: 87.91%, Synthgen: 67.69%, Beautiful: 81.73%, Explicit: 65.23%, Sexual: 71.13% (Chance: 50%). Combining these automatic attribute predictions to inturn (automatically) predict virality, we get an accuracy of 68.10% (Chance: 50%), better than human performance (60.12%) at predicting relative virality directly from images. Using our deep relative attributes, machines can predict relative virality more accurately than humans! This is because humans do not fully understand what makes an image viral (hence the need for a study like this and automatic approaches to predicting virality). If we use ground truth relative attribute annotations for these 5 attributes we achieve (65.18%) accuracy, lower than using automatic attribute predictions. This may be because the attribute classifiers trained by the machine may have latched on to biases of viral content. The resultant learned notion of attributes may be different from human perception of these attributes.

Reddit images are very different from standard images used in most previous automatic image understanding research. In fact, we found that an SVM with DECAF6 can classify Reddit images from SUN images with 90.38% accuracy. This wide variability of images on Reddit, and the poor performance of state-of-the-art image features indicates that automatic prediction of image virality will require advanced image understanding techniques. We believe computer vision researchers are the most likely to make progress on this new, challenging and important problem

[1] Jeff Donahue, Yangqing Jia, Oriol Vinyals, Judy Hoffman, Ning Zhang, Eric Tzeng, and Trevor Darrell. Decaf: A deep convolutional activation feature for generic visual recognition. *arXiv preprint arXiv:1310.1531*, 2013.

[2] Aditya Khosla, Atish Das Sarma, and Raffay Hamid. What makes an image popular? In *International World Wide Web Conference (WWW)*, Seoul, Korea, April 2014.

[3] H. Lakkaraju, J. McAuley, and J. Leskovec. What's in a name? understanding the interplay between titles, content, and communities in social media. *ICWSM*, 2013.

[4] Devi Parikh and Kristen Grauman. Relative attributes. In *ICCV*, 2011.

This is an extended abstract. The full paper is available at the Computer Vision Foundation webpage.

www.reddit.com, Reddit is considered the main engine of virality around the world, and is ranked 24^{th} among the top sites on the web by Alexa (www.alexa.com) as of March 2015.