Learning Subjective Attributes of Images from Auxiliary Sources

Francesco Gelli
National University of Singapore
francesco.gelli@u.nus.edu

Tiberio Uricchio
Università degli Studi di Firenze
tiberio.uricchio@unifi.it

Xiangnan He
University of Science and Technology of China
xiangnanhe@gmail.com

Alberto Del Bimbo
Università degli Studi di Firenze
alberto.delbimbo@unifi.it

Tat-Seng Chua
National University of Singapore
chuats@comp.nus.edu.sg

ABSTRACT

Recent years have seen unprecedented research on using artificial intelligence to understand the subjective attributes of images and videos. These attributes are not objective properties of the content but are highly dependent on the perception of the viewers. Subjective attributes are extremely valuable in many applications where images are tailored to the needs of a large group, which consists of many individuals with inherently different ideas and preferences. For instance, marketing experts choose images to establish specific associations in the consumers' minds, while psychologists look for pictures with adequate emotions for therapy. Unfortunately, most of the existing frameworks either focus on objective attributes or rely on large scale datasets of annotated images, making them costly and unable to clearly measure multiple interpretations of a single input. Meanwhile, we can see that users or organizations often interact with images in a multitude of real-life applications, such as the sharing of photographs by brands on social media or the re-posting of image microblogs by users. We argue that these aggregated interactions can serve as auxiliary information to infer image interpretations. To this end, we propose a probabilistic learning framework capable of transferring such subjective information to the image-level labels based on a known aggregated distribution. We use our framework to rank images by subjective attributes from the domain knowledge of social media marketing and personality psychology. Extensive studies and visualizations show that using auxiliary information is a viable line of research for the multimedia community to perform subjective attributes prediction.

KEYWORDS

Subjective Attributes, Probabilistic Modeling, Computational Marketing, Personality Psychology

ACM Reference Format:
Francesco Gelli, Tiberio Uricchio, Xiangnan He, Alberto Del Bimbo, and Tat-Seng Chua. 2019. Learning Subjective Attributes of Images from Auxiliary Sources. In *Proceedings of the 27th ACM International Conference on Multimedia (MM '19), October 21–25, 2019, Nice, France.* ACM, New York, NY, USA, 9 pages. https://doi.org/10.1145/3343031.3350574

MM '19, October 21–25, 2019, Nice, France

ACM ISBN 978-1-4503-6889-6/19/10…$15.00
https://doi.org/10.1145/3343031.3350574

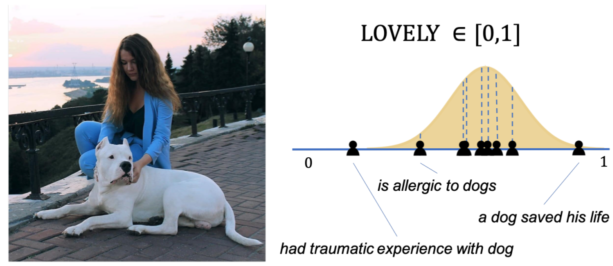


Figure 1: Example of subjective attribute "Lovely". If asking people how much they perceive that attribute on a scale from 0 to 1, we can expect a probabilistic distribution centered on a consensus

1 INTRODUCTION

Studying objective attributes of images is now commonplace in computer vision and multimedia community, with existing algorithms even reaching a sufficient level of maturity for use in commercial products. These algorithms model objective facets of objects, such as their geometry, parts, color and their different overall appearance. Going forward, the recent years have seen growing interest in subjective attributes, which are not objective properties of the content of an image, but are highly dependent on the perception of the viewer.

Subjective attributes are extremely valuable since in many applications images are tailored to the needs of a large group, which consists of many individuals with inherently different ideas and preferences. For instance, marketing experts choose images to establish specific associations in the consumers' minds [22, 23], news creators select images to impress audience while psychologists look for images with adequate emotions for therapy [24]. As a result, experts in such fields would largely benefit from an automatic approach to rank images according to subjective attributes.

However, current solutions are still not able to fully understand the subjective properties. In fact, the current state-of-the-art models for image understanding have strong limitations. Deep learning based approaches require a large quantity of annotated data to produce hard mappings between input images and labels [26]. They are not designed to provide multiple interpretations of an input.

Multiple instance learning (MIL) models [31], or domain adaptation [3] together with the recent developments with deep learning techniques [40] are a promising first step towards solving this shortcoming, but still need labels at the image level.

Following this paradigm, researchers have tried to obtain labels of the subjective attributes at the image level [4, 33], with multiple votes from multiple people and multiple contexts. This approach can work as long as images share some common cultural meanings that enable consensus between people, but controversial attributes are still hard to model [15]. For example, a photo of a person with a dog (Fig. 1) can inspire loveliness to the vast majority of people (a consensus between people) but can also evoke pain in people that had a traumatic experience with a dog.

Hence, we believe that subjective attributes come particularly into play when entities in real life (such as people or organizations) interact with images, like for example when brands share images on social media or users share tweets with images. These entities may have additional information which is reflected in the choice of images with which they interact, i.e. they select images according to their subjective attributes. For example, brands choose images for their marketing campaigns to communicate their brand personality to consumer [16], while users interact with images on social network websites according to their personality traits. These aggregated interactions can serve as auxiliary information to help infer the subjective attributes of images. Other examples of auxiliary information are photographers and hashtags, in which the interactions respectively correspond to shooting the picture and including the hashtag. The contribution of these entities and interactions can be exploited to infer the subjective attributes of images.

In this paper we propose a method to transfer such information from the domain of the auxiliary source to the image domain, allowing us to rank images by subjective attributes. By assuming that the interpretation of subjective attributes follow a certain distribution such as that shown in Figure 1, we exploit a probabilistic framework which permits us to model the uncertainty of the attributes of images while at the same time ensuring the transfer of information from the auxiliary source domain. We use our framework to rank images by subjective attributes from the domain knowledge of social media marketing and personality psychology. In the former application we learn to rank images for brands according to a set of subjective attributes defined in the literature of marketing research, while in the latter we transfer attributes of personality traits from users of a social network to image posts.

Our contribution is threefold: i) we use the information about subjective attributes of auxiliary sources to transfer them to image domain; ii) we propose an optimization algorithm that is able to infer subjective attributes from sets of auxiliary images with a probabilistic framework to model the uncertainty of such attributes; and iii) we apply our method to the domains of social media marketing and personality psychology, validating it with extensive studies and visualizations.

2 RELATED WORK

2.1 Subjective attributes in images

Significant efforts were spent in studying the relations between images and various subjective attributes, such as sentiment [5, 19, 46], aesthetics [7, 8, 21, 28, 30, 41], wellness [10, 35], memorability [20, 38] and interestingness [11]. Most of these works follow the paradigm of obtaining labels of the subjective attributes at the image level and adopt the same annotation-dependant methodologies that are commonly used in the case of objective attributes [4, 33].

On another line of research, more customized solutions were devoted to analyze the much less tangible properties such as visual persuasiveness [32] or attributes from other disciplines such as marketing and psychology. Focusing on the domain of politics, Huang and Kovashka developed the first model to infer the persuasive power of an image of a politician [17]. Ye and Kovashka designed visual features such as concepts and memorability to analyze non-literal relationships between persuasive images and text [45]. However, their method is limited to predict if image and text are aligned to convey the same message, which is only indirectly related to persuasiveness. Hussain et al. developed a method to understand the message behind image advertisement based on Q&A [18]. The authors proposed a recurrent neural network based on LSTM to generate answers to questions such as: "What should the viewer do according to this ad?". The work was later extended in [43] to learn relations between image regions and symbols. Finally, in the domain of personality, several works found significant correlations between the personality of users and visual concepts of the images they interacted with in a social network website, suggesting the existence of a link between image content and personality traits [12, 36]. Other works attempted to predict the personality traits of users from their activity on images on social networks [25, 39]. However, most of these solutions either rely on a set of custom features which belong to the specific domain, or on costly image annotations [43].

2.2 Relations to other models

Our approach shares some similarities with other models that are typically used to infer objective attributes. Convolutional neural networks (CNN), multiple instance learning and domain adaptation models are the closest technologies related to our task.

First and foremost, deep CNN have in recent years become the most popular class of algorithms for image related tasks since they outperformed previous hand crafted methods [26]. These models are not well suited by themselves to emit multiple interpretations of the same input, thus they are limited in absence of a consensus like that of the highly subjective attributes. Since we consider additional information where we are able to estimate a consensus in attributes, we use these models to predict the final attributes.

Multiple instance learning (MIL) methods work on bag of instances which are individually labeled [31]. A bag is labeled as negative if all the instances in it are negative. On the other hand, a bag is labeled as positive if it contains at least one positive instance. The task is to learn a concept that will label individual instances correctly. Similarly to this work, they can be used to transfer information from groups to the single entities. However, different from the proposed approach, they are not designed to the agreement of the instances of a bag (a consensus), but rather for the presence of at least one positive instance.

Domain adaptation approaches regard the learning of models from a source data which is related but different from a target data

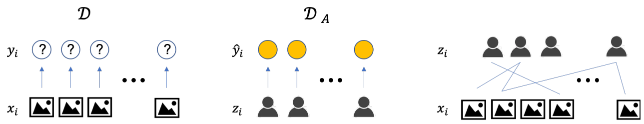


Figure 2: The three datasets are: images with latent attributes, auxiliary entities with known attributes and image-entity interactions

distribution [3]. The relation between source and target is usually the adoption of same labels on both data, with the existence of pairs of images with same labels from both domains. Our approach is similar in which the distribution of the auxiliary source can be seen as the source data and the image instances as the target. However, our setting differs in that our image instances and auxiliary source are not paired and are different entities.

3 PROPOSED METHOD

In this section, we describe the details of the method we use to learn subjective image attributes from auxiliary sources. After formulating the problem and introducing the notations, we describe the neural network model followed by the learning algorithm.

3.1 Notations & Problem Formulation

We adopt bold capital letters for matrices and bold lowercase letters for vectors (e.g. $\mathbf{A}$ and $\mathbf{x}$), while sets and set elements are represented with calligraphic capital and regular lower case letters respectively (e.g. $\mathcal{D}$ and d). Probability distributions are indicated with cursive capitals (e.g. P). Whenever elements of a set are indexed by two or more values, like in $x_{a,b}$, we use the notation " : " to denote the vector of all elements along that specific dimension. For example, $\mathbf{x}_{a,:}$ indicates a vector with components $x_{a,1}, x_{a,2}, \ldots$

Given an image dataset with $\mathcal{D} = \{(\mathbf{x_i}, y_i)\}, i = 1, \ldots, n$, where $\mathbf{x_i}$ is an image feature with a fixed dimension, and y_i are unknown scalar labels. The goal of our task is to use the information of an auxiliary source $\mathcal{D}_A$ to train a function $nn : \mathcal{D} \mapsto \mathbb{R}$ such that for each pair $\mathbf{x_i}, \mathbf{x_j} \in \mathcal{D}$:

$$sign(nn(\mathbf{x_i}) - nn(\mathbf{x_j})) = \begin{cases} 1 & \text{if } y_i > y_j \\ -1 & \text{otherwise} \end{cases} \tag{1}$$

In other words, nn which in our work is implemented by a neural network, must use the information of the auxiliary dataset to learn to rank images by a certain subjective attribute. Such dataset is indicated as $\mathcal{D}_A = \{(z_i, \hat{y}_i)\}, i = 1, \ldots, m$, where z_i are called auxiliary entities and $\hat{y}_i$ are the known label assigned to z_i corresponding to a subjective attribute. For example, $\mathcal{D}_A$ can be a set of users z_i of a social media with the $\hat{y}_i$ being their respective amount of extroversion. We denote the interactions between images and auxiliary entities as $\mathcal{D}_I = (\mathbf{x_i}, z_i), i = 1, \ldots, l$. In the same example, such set indicates which images $\mathbf{x_i}$ were re-posted by such users z_i. The three datasets are illustrated in Fig. 2. For the rest of the paper we refer to the set of images $\mathbf{x}$ such that $(\mathbf{x}, z) \in \mathcal{D}_I$ as the bag of linked images of z, indicating that entity z interacted with image $\mathbf{x}$. Vice versa, the set of auxiliary entities z such that $(\mathbf{x}, z) \in \mathcal{D}_I$ will be referred to as the bag of linked entities of image $\mathbf{x}$.

3.2 Neural Network Model

For this task we used the simple but effective multi-output architecture, based on a series of multilayer perceptrons (MLP). We use such design to predict several image attributes with a single model. Inspired by the basic architectures of multi-task learning [6], we first learn a hidden image feature x_h which is a common representation for all output attributes. We compute x_h with the multilayer perceptron MLP_1 with input the visual feature x from a pre-trained CNN:

$$\mathbf{x_h} = LL_3(\xi(LL_2(\xi(LL_1(\mathbf{x}))))) \tag{2}$$

where LL_1,LL_2 and LL_3 are linear layers and ξ is a Leaky ReLU activation function [29]. The feature x_h is then processed by multiple multilayer perceptrons, one for each output:

$$o_1, o_2, \ldots, o_N = MLP_{m1}(x_h), MLP_{m2}(x_h), \ldots, MLP_{mN}(x_h) \tag{3}$$

where each of the attribute-specific multilayer perceptron consists in a single linear layer and a sigmoid activation.

3.3 Probabilistic learning

The goal of the optimization is to learn the parameters of the neural network nn, which are initialized from a normal distribution.

In order to develop a probabilistic framework, we need to make assumptions on the distribution of the latent variables that we are attempting to model. Therefore we assume that the latent image labels y_i follow a known distribution whose statistics can be inferred from the bag of linked entities z_i (similar to Figure 1). Specifically, we expect the consensus of latent labels y_i being close to the label $\hat{y}_i$. According to our notation, given a auxiliary entity $(z_i, \hat{y}_i) \in \mathcal{D}_A$ and the bag of linked images x_i such that $(\mathbf{x_i}, z_i) \in \mathcal{D}_I$, we assume that the corresponding image labels y_i follow a one-dimensional normal distribution centered in $\hat{y}_i$ in which the standard deviation is unknown.

Based on such an assumption, the intuition behind the optimization algorithm is to forecast the attribute $o_i = nn(\mathbf{x_i})$ for multiple images belonging to the bag of linked entity z_j, and then learn nn such that its outputs follow a normal distribution centered on the entity label $\hat{y}_j$. That is $o_i \sim \mathcal{N}(\hat{y}_j, \sigma_j)$, where the standard deviation σ_j can be learned during training together with the parameters.

During each epoch, our training procedure iterates over all auxiliary entities $z_i \in \mathcal{D}_A$, using mini-batches of size B. The optimization algorithm for one batch is shown in Algorithm 1 and illustrated in Figure 3. For each point z_b in the mini-batch, K elements $x_{b,k}$ are randomly sampled from the bag of linked images and propagated through the neural network nn, producing $B \times K$ output values which we denote as $o_{b,k}$.

We adopt a pairwise ranking loss same as [13] to incorporate the relative ordering for pairs of images. For this reason, in each mini-batch we sample a second set of B auxiliary entities z'_b. If $\hat{y}_b \geq \hat{y}'_b$ such loss enforces $o_{b,k}$ to be greater than $o'_{b,k}$ and vice versa for $\hat{y}_b \leq \hat{y}'_b$.

A second loss term ensures that the outputs generated by nn for each entity z_i follow the expected distribution, centered on the consensus $\hat{y}_b$. For each entity z_b in the mini-batch we use the K output values $o_{b,k}$ to fit a normal distribution Q_b, like in Figure 4, where two sets of points in the same batch are used to estimate the two normal distributions. We then use the Kullback-Leibler (KL)

Algorithm 1 Optimization Algorithm

procedure TRAIN BATCH(z_b,z'_b)
  $B \leftarrow$ batch size
  $K \leftarrow$ bucket size
  for b=1,...,B **do**
    sample K points $\mathbf{x_{b,k}}$ st $(\mathbf{x_{b,k}}, z_b) \in \mathcal{D}_I$
    sample K points $\mathbf{x'_{b,k}}$ st $(\mathbf{x'_{b,k}}, z'_b) \in \mathcal{D}_I$
    for k=1,...,K **do**
      $o_{b,k} \leftarrow nn(x_b, k)$
      $o'_{b,k} \leftarrow nn(x'_b, k)$
  $loss \leftarrow 0$
  for b=1,...,B **do**
    $P_b, P'_b \leftarrow$ reference normal distributions with mean $\hat{y}_b, \hat{y'}_b$ and stds σ_b, σ'_b
    $Q_b, Q'_b \leftarrow$ fit normal distributions from $o_{b,:}, o'_{b,:}$
    $loss \leftarrow$loss+$\mathcal{L}_{KL}(P_b \,||\, Q_b) + \mathcal{L}_{KL}(P'_b \,||\, Q'_b)$
    $loss \leftarrow$loss+$\mathcal{L}_{PW}(\mathbf{o_{b,:}}, \mathbf{o'_{b,:}}, sign(\hat{y}_b - \hat{y}'_b))$
    $nn, \sigma_b \leftarrow$ update parameters with back-propagation

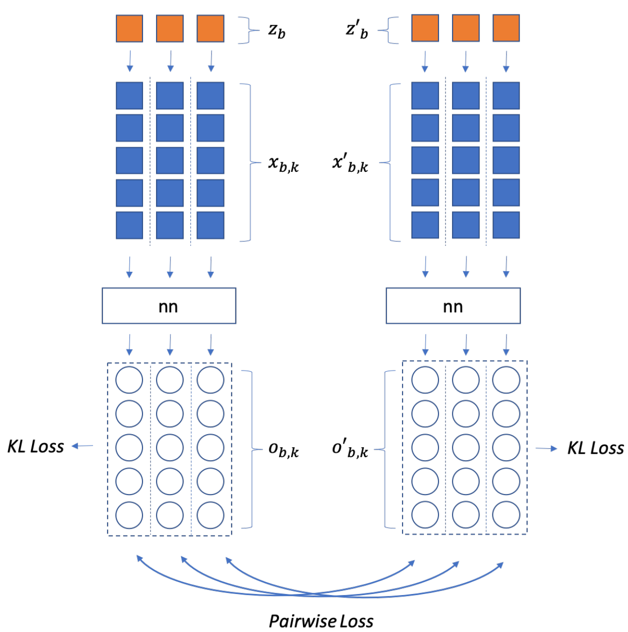


Figure 3: Optimization. In this example B=3 and K=5.

divergence to compute how well Q_b approximates the ground truth distributions P_b, which is centered on the entity label $\hat{y}_b$:

$$D_{\text{KL}}(P_b \,||\, Q_b) = \sum_i P_b(i) \log_2 \frac{P_b(i)}{Q_b(i)} \tag{4}$$

Since we have no knowledge of the real value of the standard deviation, we try two approaches to approximate it. The first aims to set each standard deviation to a fixed value $\overline{\sigma}$, such as: $P_q \sim \mathcal{N}(\hat{y}_b, \overline{\sigma})$. Despite its simplicity, this approach ignores the differences between auxiliary entities. For example, it could be the case that one entity interacts only with one specific kind of images, while the others may interact with images of various kinds. As a second approach we

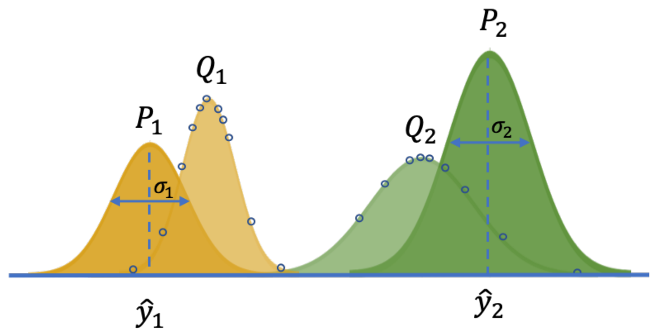


Figure 4: Probabilistic optimization for two auxiliary entities (in green and yellow). Q_1 and Q_2 are pushed by the KL divergence loss to match P_1 and P_2 respectively.

propose to learn the standard deviations during training, together with the model's parameters: $P_q \sim \mathcal{N}(\hat{y}_b, \sigma_b)$. In this case, there is no supervision on the standard deviation and the model is free to learn whatever value as long as the distribution is centered to the ground truth label. During back-propagation, both the parameters of nn and the standard deviations σ_b are updated to optimally fit the supervision labels $\hat{y}_b$.

The expected behavior of our training procedure is illustrated in Figure 4. In this example with two auxiliary entities, the KL loss is pushing Q_1 and Q_2 closer to P_1 and P_2 respectively, while at the same time adjusting the broadness of the latter. In traditional neural networks, images that are far away from the consensus (in the figure those small circles on Q_1 and Q_2 which are far from their mean) would act as noisy inputs and would try to fit them close to the label. However, by modeling distributions and not the points themselves, our learning method is not affected by such cases, and will only fit those images close to the consensus.

The final expression of the loss can be written as:

$$\begin{aligned} \mathcal{L} = \sum_{b=1,\ldots,B} & (\mathcal{L}_{PW}(\mathbf{o_{b,:}}, \mathbf{o'_{b,:}}, sign(\hat{y}_b - \hat{y}'_b)) \\ & + \alpha(\mathcal{L}_{KL}(P_b \,||\, Q_b) + \mathcal{L}_{KL}(P'_b \,||\, Q'_b))) + \beta||\theta||_2 \end{aligned} \tag{5}$$

where θ is the set of parameters and α and β control the importance of the regularization terms.

4 APPLICATIONS & EXPERIMENTS

In this section we describe the experiments that we performed to evaluate the proposed method. We separate this section into two parts, each describing a different application. We choose the two domains of social media marketing and personality psychology, given the importance of subjective attributes in these two domains, and report results of experiments and visualizations. We release the dataset and code to the public [1].

4.1 Subjective Attributes in Marketing

The first application is a real-life scenario from marketing, where designing data-driven machines to assist advertisers is becoming a popular research topic [13]. The task has been formulated as an

[1] https://github.com/GelliFrancesco/subjective_attributes

Table 1: Statistics and notations of datasets. $\mathcal{D}_I$ is split into training and testing set

	size $\mathcal{D}$	size $\mathcal{D}_A$	size $\mathcal{D}_I$		min links
Marketing	698,230	462	693,291	4,939	200
Personality	149,278	551	145,345	5,433	100

	z_i	$\hat{y}_i$	$(x_i, z_i) \in D_I$ if
Marketing	brands on IG	BAV	brand z_i posted x_i
Psychology	Twitter users	Big Five	user z_i retweeted x_i

image ranking problem. While previous works also ranked photographs according to relevance to a specific brand, they ignored the more complex marketing frameworks which utilize specific attributes defined by experts. Since such attributes are defined with respect to the consumer's perception and are largely available for brands through surveys conducted by marketing agencies, we believe that it is natural to employ our learning framework to learn such subjective attributes. Previous studies tried to predict such attributes from microblog and other textual sources [42], but to our knowledge the problem of linking such information to images is still unexplored. For these reasons we take advantage of our model to transfer a set of attributes in marketing research from brands to advertisement images.

One of the most widely adopted set of subjective attributes in marketing is Brand Asset Valuator (BAV), which is a popular framework by the established marketing agency Young & Rubicam [27]. Such framework expresses the opinion that consumers have towards a brand with the four orthogonal factors of *differentiation*, *relevance*, *esteem* and *knowledge*, which distinguish new and unfocused brands from growing, leaders and declining [2]. The four BAV factors can be aggregated in more complex metrics, such as *brand strength=differentiation×relevance*, *brand stature=esteem×knowledge* and *brand asset=brand strength×brand stature.*

4.1.1 Dataset. We collected an Instagram dataset with the posting history of 462 brands. For each of these brands we obtained brand attributes from the marketing study by Lovett et al. [27], which contains fundamental brand information, BAV data, Aaker's brand personality features [1] and other brand-image attributes [3].

We modeled this dataset to our framework as in Table 1. In this case, the auxiliary entities z_i are Instagram brands, their labels $\hat{y}_i$ being the marketing attributes, including BAV and survey metrics. The linked images to brand z_i are in this case those posted by the brand on its Instagram timeline.

As we can see in Table 1, the interactions are split into training and testing set, according to chronological order. Please note that for this dataset the total number of interactions equals exactly to the number of images, since each image can only be posted (or adopted) by a single brand.

4.1.2 Experimental Settings. We trained our models with Adadelta [44] for 1000 epochs, with hyper-parameter $\alpha = 1$ and $\beta = 10^{-5}$ and mini-batch of 64 elements. Brand attributes were normalized

[2]https://www.bavgroup.com/about-bav

[3]brand-image is defined as the perception in the consumers' minds of a brand's total personality, built over time through advertising campaigns

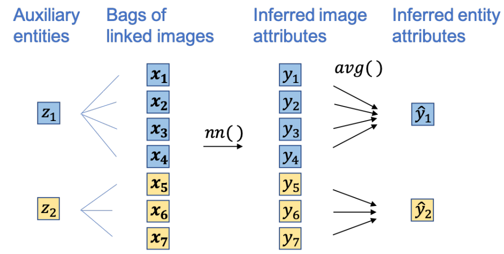


Figure 5: Evaluation methodology. Different colors correspond to bags of images linked to the same auxiliary entity

in [0, 1] using a *min-max* scaler. As in [13], we used the pre-trained image features extracted with VGG16 network [37]. For our learning with fixed standard deviation, we set the value of 0.1.

For this task it is impracticable to directly measure the accuracy of each image prediction, since the ground truth knowledge of image labels is missing. Since the only information available are the labels $\hat{y}_i$ on the auxiliary dataset, we indirectly measure the performance of our algorithm as in Figure 5. According to our assumptions that image labels y_i follow a normal distribution centered on the labels in the bag of linked entities, we average the predictions for each single brand and measure their Pearson's and Spearman's rank correlations. We tested the statistical significance of each correlation experiment we performed and consistently obtained a p-value of less than 0.01. Superior results than the baselines on this task indicate a more robust ability to learn not only the image attributes, but brand attributes as well. This is because the probabilistic framework is able to model images according to how far away they are from the consensus.

For each experiment, we compare against two baselines. *Naïve MIL* is a common naïve approach of multiple instance learning consisting of replicating the label of the bag to all of the included entities. We annotate each image with the label of the brand who posted it. For the second baseline, called *Average*, we design a simple learning baseline based on average pooling. It shares the same structure of our optimization algorithm, but instead of performing probabilistic learning with KL divergence, it averages all predictions $\frac{1}{B}\sum_b o_{b,k}$ and minimizes the mean square error with $\hat{y}_b$.

4.1.3 Ranking by BAV factors. In this study we assess the ability of our method in ranking images by the four BAV factors. We employed a multi-output neural network *nn* with four distinct outputs and performed our optimization on the training set. Finally, we ranked the testing image ads by the four factors independently as well as by the aggregated metric *brand asset.* Figure 6 shows the results as Pearson's correlations. We omit the Spearman's correlation, since our two results are consistently equivalent. From the chart we first observe that all of the tested methods and baselines produce significant positive correlations, highlighting that social media posts indeed reflect subjective brand attributes from marketing theory, remarking the importance of using data-driven solutions for designing effective digital campaigns on social media. Secondly, our methods, with fixed and learned standard deviations, consistently outperform the two baselines, indicating the importance of adopting the probabilistic learning framework. Finally, we notice that the esults are higher for the two attributes *differentiation*

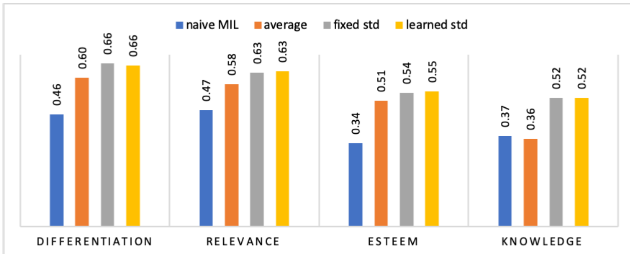


Figure 6: Results of ranking by BAV in terms of Pearson's correlations

Table 2: Results of ranking by brand asset

Brand Asset	naive MIL	average	fixed std	learned std
Pearson's	0.31	0.40	0.41	**0.44**
Spearman's	0.36	0.50	**0.56**	**0.56**

Table 3: Results of ranking by brand-image attributes

Upper Class	naive MIL	average	fixed std	learned std
Pearson	0.58	0.69	**0.74**	0.73
Spearman's	0.45	0.54	0.60	**0.62**

Fun	naive MIL	average	fixed std	learned std
Pearson	0.65	0.62	**0.75**	0.74
Spearman's	0.59	0.58	**0.72**	0.71

and *relevance*, which are the components of the aggregated metric *brand strength*, while the two metrics of *brand stature*, namely *esteem* and *knowledge* have inferior results. This result suggests that *brand strength*, which is an indicator of a brand's influence and appreciation in the consumers' minds, is more directly reflected in images than *brand stature*, which instead describes the scale of a brand in the whole consumers market.

We use the same model to test the ability to rank images by *brand asset*, which aggregates all the four attributes in a single performance indicator. Results are shown in Table 2. From the first row we notice that learning brand standard deviations from the data helps to achieve a higher Pearson's correlation, indicating the ability of the model to rank brands given their predicted value of *brand asset* with a linear relation. For the Spearman's metric, which assesses the ranking quality regardless of linearity, the performance is equivalent between both of our models. In all cases, our results outperform the baselines.

4.1.4 Ranking by Brand Image attributes. Beside BAV attributes, which are advanced concepts in marketing research, we also applied our model to two of the brand-image attributes included in our dataset. For this study we selected *Upper Class* and *Fun*, being among the easiest to comprehend by non-experts in marketing. For this purpose we adopted a similar methodology as the previous study, choosing instead a single-output *nn* to predict each single attribute. Table 3 shows the results for the two models, which are in this case trained independently from each other. We first notice that the attribute correlations obtained are higher than in the BAV case, probably because of their more tangible nature. Also our models perform much better than the baseline for this task. It is interesting to point out that learning with fixed standard deviations bring in sightly better results in some cases only. The reason could be that the latent standard deviations are, in these cases, not differentiated enough from each other to benefit from a more complex learning algorithm.

4.1.5 Visualization of Case Studies. We provide a qualitative visualization of a few case studies to better comprehend how our models rank images. Figure 7 shows three case studies for the attributes *Upper Class*, *Fun* and *Brand Asset*, obtained by ranking the images in the testing set.

We observe that the top *Upper Class* images contains sport cars and luxury bags. Since our dataset has high *Upper Class* values for most brands selling this kind of products, our model looks for the most evident image properties in their image posts, which in this case are the presence of such objects. Similarly, the least upper class brands in our dataset are fast food companies, resulting in images of hamburgers and fast food being placed in the bottom positions of the ranking. For similar reasons, *Fun* images feature babies, toys, cruises and computer games, while images about office life and cosmetic products are at the bottom of the ranking.

While such results are easily validated by our common sense, it is harder to understand why certain images are ranked higher or lower by the attribute *brand asset*. The first image of the third row shows a product with a colorful blue background. By an inspection of our dataset, we can see that established brands such as Google, Disney, Sony and Coca-Cola, which are the top *brand asset* brands, often adopt colorful backgrounds in their posts. The three following images show a distinctive visual style, with light colors and neat surfaces, which may be common for high *brand asset* brands which have a high budget for professional photographers and designers. It is interesting to observe the last images ranked by *brand asset*. Beside a few amateurish photographs, it is counter-intuitive to understand why the model learned that high *brand asset* brands do not post images about alcohol beverages. A possible explanation is that while alcohol brands are usually popular and well known, many people may have strong sentiment against such products, affecting the outcome of the BAV questionnaire.

4.2 Subjective Attributes in Psychology

Personality psychology is a very subjective discipline by nature, and most studies and observations are carried out via surveys. Two of the most common personality frameworks are MBTI, which assigns individuals to one of sixteen types of personality [34] and Big Five [14], which uses a combination of five traits to measure a person's imagination and creativity (*openness*), thoughtfulness and self-discipline (*conscientiousness*), ability to engagement with the external world (*extraversion*), willingness to be cooperative, altruistic and emphatic (*agreeableness*) and tendency to experience negative emotions and feelings (*neuroticism*).

Previous works investigated the relations between Big Five traits and online actions such as re-posting an image microblog, under the hypothesis that users with specific personality traits are stimulated by different visual content [12]. However, the authors ignored the large number of reasons behind people's re-posting actions

(a) Top 5 Upper Class

(b) Last 5 Upper Class

(c) Top 5 Fun

(d) Last 5 Fun

(e) Top 5 Brand Asset

(f) Last 5 Brand Asset

Figure 7: Visualization of ranking based on marketing attributes

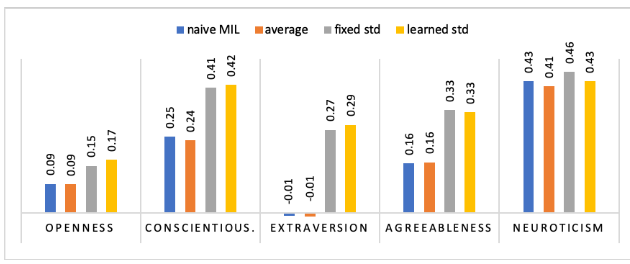


Figure 8: Results of ranking by Big Five in terms of Pearson's correlations

in social media and considered all the image microblog that the user interacted with with equal importance. We hence apply our probabilistic framework to learn to rank images by personality attributes.

4.2.1 Dataset. We adopted the Twitter dataset contributed by [12]. Since the dataset was originally used for an image recommendation task, we removed the negative samples and filtered those users that retweeted at least 100 images (see Table 1). Since in Twitter multiple users can retweet the same image, the number of interactions is larger than the number of images. For this dataset we model the users as auxiliary entities z_i, with the respective labels $\hat{y}_i$ being the Big Five traits and the two attributes of *age* and *leadership*. The bag of linked images of user x_i consists of those image tweets that the user retweeted on the social network website 1.

4.2.2 Ranking by Big Five. The goal of this analysis is to rank images given a specific personality trait. We train a multi-output neural network *nn* with five outputs, one for each trait, using the same experimental settings as in the marketing study, using the image features provided by [12].

Results are illustrated in Figure 8 for the Big Five traits separately. A first observation is that the trait *neuroticism* has the highest performance, followed by *conscientiousness*, confirming the findings in [12]. Secondly, our method outperforms the baselines for all traits, except for *extroversion*. For *neuroticism*, which indicates tendency to experience negative emotions and feelings, the naïve MIL baseline is also highly effective, suggesting that Twitter users are split between those who commonly re-post neurotic-friendly images and those who don't. If this is true, the unknown image labels $y_{b,k}$ for a single user b would be similar to each other most of the time, thus minimising the benefit of adopting a probabilistic modelling of such variable against setting it to the user's label $\hat{y_b}$. We see the opposite effect for the trait *extraversion*, suggesting that users, along with their time-span of activity on Twitter, do re-post both extrovert-friendly and non-friendly images.

Table 4: Results of ranking by user attributes

Age	naive MIL	average	fixed std	our
Pearson	0.30	0.29	**0.56**	**0.56**
Spearman's	0.34	0.33	**0.53**	**0.53**

Leadership	naive MIL	average	fixed std	our
Pearson	0.23	0.23	**0.29**	**0.29**
Spearman's	0.29	0.28	**0.33**	**0.33**

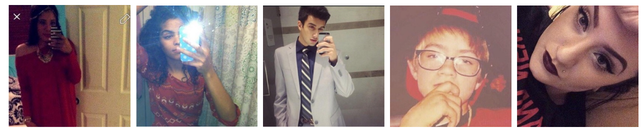
Finally, we tested our model on the two user attributes of *age* and *leadership* (Table 4). We immediately notice that *age* has a higher correlation than *leadership*. Such result is unsurprising, since demographic information (age and gender) are the most widely adopted criteria in ads targeting strategies.

4.2.3 Visualization of Case Studies. Visualization of the top results is presented in Figure 9, where each row correspond to one of the Big Five traits. A first glance shows that some of the images appear in the top results for multiple traits. This may be partially due to the Big Five traits being not orthogonal [9]. For example, *neuroticism* and *extraversion* are often found to be negatively correlated. Some of the results confirm the previous findings in [12]. For example,

(a) Top 5 Openness

(b) Last 5 Openness

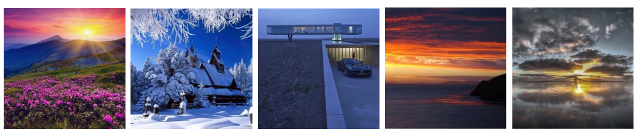
(c) Top 5 Conscientiousness



(d) Last 5 Conscientiousness

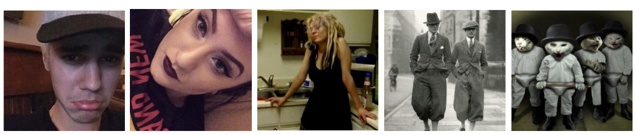
(e) Top 5 Extraversion

(f) Last 5 Extraversion

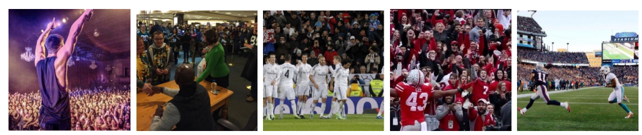


(g) Top 5 Agreeableness

(h) Last 5 Agreeableness



(i) Top 5 Neuroticism



(j) Last 5 Neuroticism

Figure 9: Visualization of ranking based on Big Five attributes

images of abandoned and desolated places, such as empty theater or factory are ranked high for the *openness* trait, which indicate inclination to adventure and challenges. Same as the previous study, we also found that being extrovert is highly correlated with actions on tweets about young models, while *neuroticism* is negatively correlated with crowds and sport. From our results we notice that selfies are ranked at the last positions for the trait *conscientiousness*, which is about thoughtfulness and self-discipline. It is possible that the model learns that this kind of self-portrait pictures are more often used by young and rebel teenagers. Finally, it is interesting to observe some unexpected associations learned by the model. While it may appear reasonable that images with religious content have a low *extraversion* score, we notice the same behaviour also for many pictures of animals, in particularly cats. Similarly, we can't find an intuitive reason for images of cakes, food, gym and text quotes being among the last in the ranking by *agreeableness*, which is a metric of being cooperative and altruistic. We encourage further studies by personality psychologists to shed light upon possible reasons behind such results.

5 CONCLUSIONS & FUTURE WORK

This work investigates the problem of learning subjective attributes of images, which are not objective property of the content, but are highly dependant on the perception of the viewer. Since existing methods rely on a large amount of annotated image data, we propose a method to learn subjective data from auxiliary data sources, which are entities that interact with images in the real world, such as users or brands in a social media websites. We propose an optimization algorithm to transfer the subjective attributes from the auxiliary source to the image domain and tested on the two applications of social media marketing and personality psychology.

Results obtained by our method show that image properties are not just a phenomena of the single, but can be interlaced to how they are perceived by larger communities aggregating multiple individuals. More specifically, there are different levels of objectiveness at different community scales [2]. For example, while objective attributes are perceived equally by all individuals, other more subjective attributes are widely shared across sub-communities, and others may also vary for each individual. According to the results of our experiments and these observations, in our future work we intend to take into account different levels of objectiveness by modeling communities at different scales.

ACKNOWLEDGMENTS

NExT research is supported by the National Research Foundation, Prime Minister's Office, Singapore under its IRC@SG Funding Initiative.

REFERENCES

[1] Jennifer L. Aaker. 1997. Dimensions of Brand Personality. *Journal of Marketing Research* 34, 3 (1997), 347–356.

[2] M. Batey. 2012. *Brand Meaning.* Taylor & Francis.

[3] Shai Ben-David, John Blitzer, Koby Crammer, Alex Kulesza, Fernando Pereira, and Jennifer Wortman Vaughan. 2010. A theory of learning from different domains. *Machine learning* 79, 1-2 (2010), 151–175.

[4] Damian Borth, Tao Chen, Rongrong Ji, and Shih-Fu Chang. 2013. Sentibank: large-scale ontology and classifiers for detecting sentiment and emotions in visual content. In *Proceedings of the 21st ACM international conference on Multimedia*. ACM, 459–460.

[5] Damian Borth, Rongrong Ji, Tao Chen, Thomas Breuel, and Shih-Fu Chang. 2013. Large-scale Visual Sentiment Ontology and Detectors Using Adjective Noun Pairs. In *Proceedings of the 21st ACM International Conference on Multimedia (MM '13)*. ACM, 223–232.

[6] Rich Caruana. 1997. Multitask Learning. *Machine Learning* 28, 1 (01 Jul 1997), 41–75.

[7] Ritendra Datta, Dhiraj Joshi, Jia Li, and James Z Wang. 2006. Studying aesthetics in photographic images using a computational approach. In *European conference on computer vision*. Springer, 288–301.

[8] Sagnik Dhar, Vicente Ordonez, and Tamara L Berg. 2011. High level describable attributes for predicting aesthetics and interestingness. In *CVPR 2011*. IEEE, 1657–1664.

[9] Heather E. P. Cattell. 1996. The original Big Five: A historical perspective. *European Review of Applied Psychology/Revue EuropÃl'enne de Psychologie AppliquÃl'e* 46 (01 1996), 5–14.

[10] Aleksandr Farseev and Tat-Seng Chua. 2017. Tweet Can Be Fit: Integrating Data from Wearable Sensors and Multiple Social Networks for Wellness Profile Learning. *ACM Trans. Inf. Syst.* 35, 4 (Aug. 2017), 42:1–42:34.

[11] Yanwei Fu, Timothy M. Hospedales, Tao Xiang, Shaogang Gong, and Yuan Yao. 2014. Interestingness Prediction by Robust Learning to Rank. In *Computer Vision – ECCV 2014*, David Fleet, Tomas Pajdla, Bernt Schiele, and Tinne Tuytelaars (Eds.). Springer International Publishing, Cham, 488–503.

[12] Francesco Gelli, Xiangnan He, Tao Chen, and Tat-Seng Chua. 2017. How Personality Affects Our Likes: Towards a Better Understanding of Actionable Images. In *Proceedings of the 25th ACM International Conference on Multimedia (MM '17)*. ACM, 1828–1837.

[13] Francesco Gelli, Tiberio Uricchio, Xiangnan He, Del Bimbo Alberto, and Tat-Seng Chua. 2018. Beyond the Product: Discovering Image Posts for Brands in Social Media. In *Proceedings of the 26th ACM International Conference on Multimedia*. ACM, 465–473.

[14] Lewis R Goldberg. 1990. An alternative "Description of personality": The Big-Five factor structure. *Journal of personality and social psychology* 59, 6 (1990), 1216.

[15] Emily M Hand, Carlos Castillo, and Rama Chellappa. 2018. Doing the best we can with what we have: Multi-label balancing with selective learning for attribute prediction. In *Thirty-Second AAAI Conference on Artificial Intelligence*.

[16] Ouwersloot Hans and Tudorica Anamaria. 2001. *Brand Personality Creation through Advertising*. Research Memorandum 015. Maastricht University, Maastricht Research School of Economics of Technology and Organization (METEOR).

[17] X. Huang and A. Kovashka. 2016. Inferring Visual Persuasion via Body Language, Setting, and Deep Features. In *2016 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*. 778–784.

[18] Zaeem Hussain, Mingda Zhang, Xiaozhong Zhang, Keren Ye, Christopher Thomas, Zuha Agha, Nathan Ong, and Adriana Kovashka. 2017. Automatic Understanding of Image and Video Advertisements. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.

[19] Jia Jia, Sen Wu, Xiaohui Wang, Peiyun Hu, Lianhong Cai, and Jie Tang. 2012. 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*. ACM, 857–860.

[20] P. Jing, Y. Su, L. Nie, H. Gu, J. Liu, and M. Wang. 2019. A Framework of Joint Low-Rank and Sparse Regression for Image Memorability Prediction. *IEEE Transactions on Circuits and Systems for Video Technology* 29, 5 (May 2019), 1296–1309.

[21] Yan Ke, Xiaoou Tang, and Feng Jing. 2006. The design of high-level features for photo quality assessment. In *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06)*, Vol. 1. IEEE, 419–426.

[22] Gunhee Kim and Eric P. Xing. 2013. Discovering Pictorial Brand Associations from Large-Scale Online Image Data. In *2013 IEEE International Conference on Computer Vision Workshops (ICDMW '15)*. 404–411.

[23] Gunhee Kim and Eric P. Xing. 2014. Visualizing Brand Associations from Web Community Photos. In *Proceedings of the 7th ACM International Conference on Web Search and Data Mining (WSDM '14)*. 623–632.

[24] Peter J. Lang. 1977. Imagery in therapy: an information processing analysis of fear. *Behavior Therapy* 8, 5 (1977), 862 – 886.

[25] Alixe Lay and Bruce Ferwerda. 2018. Predicting UsersâĂŹ Personality Based on Their âĂŸLikedâĂŹ Images on Instagram. In *2nd Workshop on Theory-Informed User Modeling for Tailoring and Personalizing Interfaces (HUMANIZE), 2018 : (CEUR Workshop Proceedings)*. CEUR-WS.

[26] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. 2015. Deep learning. *nature* 521, 7553 (2015), 436.

[27] Mitchell Lovett, Renana Peres, and Ron Shachar. 2014. A Data Set of Brands and Their Characteristics. *Marketing Science* 33, 4 (July 2014), 609–617.

[28] Wei Luo, Xiaogang Wang, and Xiaoou Tang. 2011. Content-based photo quality assessment. In *2011 International Conference on Computer Vision*. IEEE, 2206–2213.

[29] Andrew L Maas, Awni Y Hannun, and Andrew Y Ng. 2013. Rectifier nonlinearities improve neural network acoustic models. In *Proc. icml*, Vol. 30. 3.

[30] Luca Marchesotti, Florent Perronnin, Diane Larlus, and Gabriela Csurka. 2011. Assessing the aesthetic quality of photographs using generic image descriptors. In *2011 International Conference on Computer Vision*. IEEE, 1784–1791.

[31] Oded Maron and Tomás Lozano-Pérez. 1998. A framework for multiple-instance learning. In *Advances in neural information processing systems*. 570–576.

[32] Nikos Metallinos. 1998. Visual Persuasion: The Role of Images in Advertising. *Canadian Journal of Communication* 23, 2 (1998).

[33] Naila Murray, Luca Marchesotti, and Florent Perronnin. 2012. AVA: A large-scale database for aesthetic visual analysis. In *2012 IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, 2408–2415.

[34] Peter B. Myers Myers, Isabel Briggs. 1995. *Gifts Differing: Understanding Personality Type.* CPP, Inc.

[35] Nitish Nag, Vaibhav Pandey, Preston J. Putzel, Hari Bhimaraju, Srikanth Krishnan, and Ramesh Jain. 2018. Cross-Modal Health State Estimation. In *Proceedings of the 26th ACM International Conference on Multimedia (MM '18)*. ACM, New York, NY, USA, 1993–2002.

[36] Zahra Riahi Samani, Sharath Chandra Guntuku, Mohsen Ebrahimi Moghaddam, Daniel PreoÅčiuc-Pietro, and Lyle H. Ungar. 2018. Cross-platform and cross-interaction study of user personality based on images on Twitter and Flickr. *PLOS ONE* 13, 7 (07 2018), 1–19.

[37] Karen Simonyan and Andrew Zisserman. 2014. Very Deep Convolutional Networks for Large-Scale Image Recognition. In *International Conference on Learning Representations (ICLR '14)*.

[38] H. Squalli-Houssaini, N. Q. K. Duong, M. Gwenaelle, and C. Demarty. 2018. Deep Learning for Predicting Image Memorability. In *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. 2371–2375.

[39] Monica Therese Whitty, James Doodson, Sadie Creese, and Duncan Hodges. 2018. A picture tells a thousand words: What Facebook and Twitter images convey about our personality. *Personality and Individual Differences* 133 (2018), 109 – 114. Examining Personality and Individual Differences in Cyberspace.

[40] Jiajun Wu, Yinan Yu, Chang Huang, and Kai Yu. 2015. Deep multiple instance learning for image classification and auto-annotation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 3460–3469.

[41] Ou Wu, Weiming Hu, and Jun Gao. 2011. Learning to predict the perceived visual quality of photos. In *2011 International Conference on Computer Vision*. IEEE, 225–232.

[42] Anbang Xu, Haibin Liu, Liang Gou, Rama Akkiraju, Jalal Mahmud, Vibha Sinha, Yuheng Hu, and Mu Qiao. 2016. Predicting Perceived Brand Personality with Social Media.

[43] Keren Ye and Adriana Kovashka. 2018. ADVISE: Symbolism and External Knowledge for Decoding Advertisements. In *The European Conference on Computer Vision (ECCV)*.

[44] Matthew D. Zeiler. 2012. ADADELTA: An Adaptive Learning Rate Method. *CoRR* abs/1212.5701 (2012).

[45] Mingda Zhang, Rebecca Hwa, and Adriana Kovashka. 2018. Equal But Not The Same: Understanding the Implicit Relationship Between Persuasive Images and Text. In *British Machine Vision Conference 2018, BMVC 2018, Northumbria University, Newcastle, UK, September 3-6, 2018*. 8.

[46] Sicheng Zhao, Yue Gao, Xiaolei Jiang, Hongxun Yao, Tat-Seng Chua, and Xiaoshuai Sun. 2014. Exploring Principles-of-Art Features For Image Emotion Recognition. In *Proceedings of the 22Nd ACM International Conference on Multimedia*. ACM, 47–56.