Semantic Image-Based Profiling of Users' Interests with Neural Networks

Szymon Wieczorek, Dominik Filipiak, and Agata Filipowska

Department of Information Systems
Poznań University of Economics and Business
Poland
szm.wieczorek@gmail.com, {dominik.filipiak,agata.filipowska}@ue.poznan.pl

Abstract. We propose a method for creating a semantic profile of user's interests emerging from pictures by application of neural networks trained for object recognition. We use BabelNet, an online encyclopaedic dictionary, to generalise object names into categories of interests. Our method is evaluated with ground-truth data based on social tagging mechanism. Experiments are conducted entirely on original data containing 60,000 images crawled from Flickr, evenly distributed among 300 users. Results show that object recognition methods combined with object category generalisation can be effectively used to predict user's interests. The accuracy of the presented method seems to change with the neural network used for object recognition (5 NN tested in total), therefore it has a strong potential for further development. ResNet-50 turned out to be the most accurate network in our experiment.

Keywords: online social networks · user characteristics · activity recognition and understanding · semantic tagging · deep learning

1 Introduction

Online Social Networks (OSN) have billions of users across the world. The majority of them voluntarily provide all sort of demographic information, such as name and age, interests, or their real-life social connections and behaviours. Due to the richness of data, OSN users have become a target of numerous studies. Several scholars have sought to understand phenomena such as complete profiles [14, 19], language [5, 33], or specific user activities, such as *likes* on Facebook [18, 23]. Recently, the growing popularity of image based-services, including Instagram or Flickr, can be observed. User-generated content has changed – from the form of text to images, sometimes accompanied by textual information. With the expansion of image-based communication, the approach to experiments performed on OSN datasets has changed. It is believed that images can provide as much information about a user as text or in some cases even more. At the same time, the field of computer vision has experienced a rapid growth, especially when it comes to implementations of machine learning techniques, namely neural networks (hereafter NN-s). Convolutional neural networks (hereafter CNN-s) have become a popular solution to numerous image-based tasks, such as object classification or semantic segmentation.

The recent line of research on user profiling tends to take advantage of the growing popularity of visual content and computer vision. For instance, Segalin et al. [35] studied connections between images users interacted with in Flickr and their personality described with the Big Five model. They did that by extracting visual features of images and juxtaposing them to a set of self-assessed and attributed personality traits. Wei and Stillwell [39] tried to examine how intelligent Facebook users are thought to be by other people, based on their profile image. Liu et al. [24] also focused on analysing profile pictures, trying to answer the question if there is a connection between a type of a personality and the choice of a profile picture. Recent research has tended to focus mainly on psychological aspects like personality traits, intelligence or how OSN users are perceived by others. Although some attention was paid to profiling users interests, this topic is still not described as good as these mentioned before. A major obstacle in performing such experiments is obtaining an appropriate dataset. It should contain images from OSN and data about the interests of users who interacted with those images. The lack of ground-truth data forces us to experiment with various possible sources of information.

In this paper, we propose a technique of estimating OSN user's interest, based on visual content that a user interacted with, and validated with tags-based profiles. We use CNN-s (pre-trained on the ImageNet dataset) for the object recognition task to create a user interest profile. CNN-s are based on well-established architectures [20, 37, 38, 16]. We use data collected from Flickr, which contains 200 images marked as *favourite* for each of 300 randomly chosen users, along with tags those images were described with. In total, the dataset contains 60,000 images. As for categories of interests, we use 34 domain labels from BabelNet. In the final stage of the experiment, we calculate F_1 scores to compare the similarity between image-based and tags-based interest profile. The results show that our method can be used to predict user's interests from images.

2 Related Work

The increasing popularity of social networks made them a target of intensive research and experiments. Numerous studies on various topics were performed, thanks to the data provided by OSN users. At the very beginning, mostly textual data were examined (information included in the profiles, user posts) along with the social connections (like online friends or followed brands). Expansion of image-based services has been an incentive to conduct studies on visual, user-generated content. Fast progress in computer vision allowed researchers to use new, advanced techniques of image analysis. Especially convolutional neural networks, with their promising results in object recognition tasks, gathered a lot of attention. Profiling users with image analysis connected these two developing fields of research. This section describes the main work in object recognition with

CNN-s, OSN users profiling and social tagging (since we employ this technique to create the dataset for evaluation of our method).

Convolutional neural network is a deep learning architecture readily applied to image classification tasks, due to its high accuracy on large datasets. The ground-breaking moment for CNN was the 2012 ILSVRC challenge, won by the AlexNet architecture [20]. Since then, CNN-s became a worldwide standard in computer vision. In following years, several works describing methods of improving its performance were published. Zeiler and Fergus [42] presented a new architecture, based on AlexNet, but with significant changes to the layers connections and sizes of filters. Simonyan and Zisserman [37] described VGG, a deeper architecture based on relatively small (3x3) convolution filters, which achieved the best performance in the ILSVRC 2014 object recognition challenge. The Inception network proposed by Szegedy et al. [38] was also focused on extending the depth of the network, in order to increase the classification accuracy. Further studies resulted in new approaches to CNN-s such as R-CNN-s, introduced by Girschick et al. [13]. R-CNN concern a CNN modified to work with region proposals extracted with selective search method. R-CNN achieved 30% relative improvement over the previous state-of-the-art results in the object detection task on PASCAL VOC 2012. This technique was later improved – Fast R-CNN [12] and Faster R-CNN [32] used new approaches to R-CNNs, achieving significant improvements mainly in the case of the speed of the network. The most recent work, called Mask R-CNN, presented by He et al. [15] extends Faster R-CNN by adding a mask segmentation feature, which detects all pixels belonging to an object instead of just marking it with a bounding box. Mask R-CNN outperformed all other existing work on COCO 2016 dataset. Clevert et al. [7] proposed the ELU activation function, which can be a substitute for ReLU or LReLU and speed up the learning process in deep neural networks in order to improve the classification accuracy. Authors achieved better results than ReLU-based neural networks on CIFAR-100 and improved learning speed on ImageNet. A different approach to CNNs was presented by Lin et al. [22] in the "Network in network" structure. It replaces the convolution layer with mini-neural-network and increases the accuracy on MNIST, CIFAR and SVHN datasets. CNN-s were proved to be very flexible when it comes to implementation and that makes them a fine choice for different tasks in the computer vision. Recently, almost every new CNN architecture comes with state-of-the-art results in different fields of research. Higher accuracies in computer vision tasks allow researchers to achieve better results in research that make use of them. Some of the papers mentioned in the following subsection are a good candidate for user profiling tasks.

OSN data has been successfully used in a large number of studies. Balduzzi et al. [4] identified 1.2 million user profiles using only their email addresses. Publicly available profiles were crawled and used for automated profiling. In an immense study performed by Kosiński et al. [19], OSN data were used to estimate various demographic information, like gender, relationship status, race, political preferences etc. Findings presented by [14] show that OSN are a valid source of information about a user personality. In this case, data acquired from

Facebook profiles were used to estimate Big Five personality traits. Sensitive information including the user's home location can be also extracted from social network data. Li et al. [21] developed two methods to predict Twitter users location. Both methods used Twitter users' profiles, their tweets and following relationships as input data. Chaabane et al. [6] presented a method for tracking OSN users' internet movements. Web browsing data can be used for profiling – especially user interests. There are many works focused on profiling users with computer vision methods, but only very few of them were specifically targeting user interests. User-generated visual content was successfully used to tasks like identifying gender [41] or even the atmosphere of places [31]. Lovato et al. [25] achieved high precision in identifying the user, basing on image aesthetics preferences. To this end, aesthetic differences between 200 Flickr users were modelled using features extracted from 40,000 different images. Segalin et al. [34] have shown that there is a correlation between image aesthetic features and attributed personality traits. A similar study was later conducted, this time with the use of a convolutional neural network to extract the features [35]. This approach outperformed previous state-of-the-art results. You et al. [40] proposed a method for profiling user interests with visual content posted on social networks. They approached this problem as an image classification task and used a CNN to predict labels for images of 748 Pinterest users. Category labels, assigned by users themselves, were used as a ground-truth interests dataset. This technique proved that CNN-s can be effectively used to profile users with visual content.

Metadata can carry a lot of valuable information, especially when it is manually added by a user in a form of tags. Tags have become a very popular solution for organising and cataloguing content produced by users in OSN. In some services, such as Instagram or Twitter, tags are more important than in the others, e.g. Facebook. Even in the latter, tagging systems are present in some form or they are desired by users, though. Aggarwal [2] distinguishes several motivations for tagging and various types of tags. Some of these motivations are: attracting the attention of other users, expressing opinions, presenting the user himself or simply describing a resource. Tags can be i.a. personal, organisational, subjective but also context- and content-based. Motivations behind tagging and meaning of tags have been well described in numerous works [3, 36]. Some researchers tried to examine the usefulness of tags in profiling users. Michlmayr & Cayzer [27] presented an original algorithm, *Add-A-Tag*, used to construct user profiles from tags. Profiles could be used e.g. to facilitate web browsing by making suggestions based on users' interests extracted from tags. Hung et al. [17] proposed a method for user profiling with tags associated with his own OSN account and a collection of tags specified by contacts from the network. These profiles were used for media recommendation. Although the accuracy of the method was rather low, the authors claimed that this measure is not an adequate representation of the real efficiency of this method. Firan et al. [10] studied music recommendation systems and presented their own approach based on Last.fm database. In one case the authors managed to achieve better results with tags-based method than with collaborative filtering. Oramas et al. [30] presented a hybrid approach to music recommendation, using tags and collaborative filtering. The authors proved that their method can result in better recommendation results that collaborative filtering algorithms used on their own.

This study was undertaken to determine whether it is possible to leverage social tagging for semantic OSN user profiling with computer vision techniques. While several attempts have been made to investigate similar issues, they focus mostly on the process automation or validation on the tag level [9, 11]. We approach this problem differently, since we use BabelNet[1] to extract broader categories from tags and predictions. Such an approach should yield more accurate and broader results, though we believe this perspective is enough to briefly describe users' interests. We use BabelNet, a well-known online multilingual semantic framework, which can be also perceived as an encyclopaedic dictionary [28, 29]. Our study can be also perceived as an NN and BabelNet evaluation against social tagging. The details of this approach are described in the next section.

3 Dataset and Research Method

As mentioned in Section 1, currently there is no existing dataset which connects images user interacted with in social networks (e.g. posted, *liked*, or shared images) and their interests. Therefore, there is a need to collect appropriate data. We decided to use Flickr as our data source. Alongside images, we also collected tags that each image was labelled with (Figure 1). Required data can be acquired from Flickr with custom Python scripts, using Flickr API.

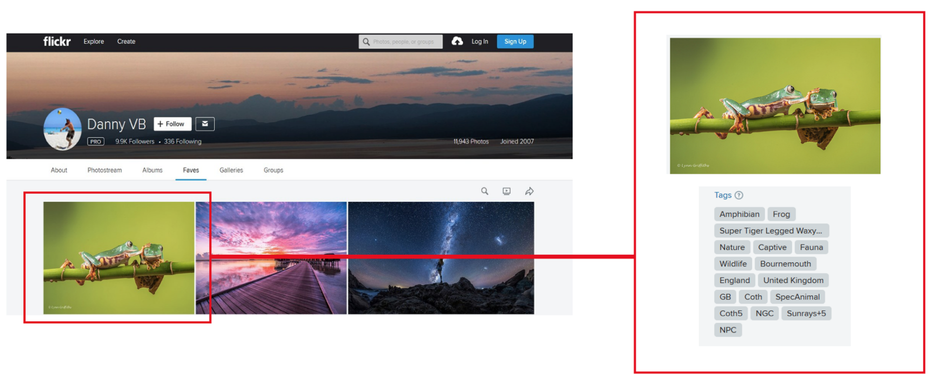

Fig. 1. Building a dataset. Source: `flickr.com`

Firstly, we create a list of randomly chosen users. We decided to pick users at random, instead of using other techniques such as snowballing, due to the

[1] `http://babelnet.org`

possible similarity of interests between users that follow each other. It helps us to reassure that categories of interests will be different from each other, since the variety of interests is an essential factor for the study. We collected a list of 300 profiles. Then, for each of them, we downloaded 200 images that a given user marked as his *favourite*. These exact numbers of users and images were included in the PsychoFlickr dataset[8], which was used in other works, and we follow this convention. At the same time, we crawl tags for all images. As a result, every user has one file with all the tags that his favourite photos were described with. For profiling interests, there is no dataset with data about interests provided by users. You et al. [40] used category labels, assigned by users, as their ground-truth data. In our approach, a model of interests will be created with tags extracted from images' descriptions. Tags need to be processed before creating a profile. We remove all misspelt words, words describing the technical aspects of the photograph (like brand and model of the camera used to take the picture, type of the picture, colours etc.), and other, so-called *stopwords*, which carry no valuable information. Next, we use BabelNet to take an advantage of its semantic network that allows us to find a domain for every word in its English dictionary. BabelNet synset encapsulates 34 different domains:

- art, architecture and archaeology,
- biology,
- business, economics and finance,
- chemistry and mineralogy,
- computing,
- culture and society,
- education,
- engineering and technology,
- farming,
- food and drink,
- games and video games,
- geography and places,
- geology and geophysics,
- health and medicine,
- heraldry, honors and vexillology,
- history,
- language and linguistics,
- law and crime,
- literature and theatre,
- mathematics,
- media,
- meteorology,
- music,
- numismatics and currencies,
- philosophy and psychology,
- physics and astronomy,
- politics and government,
- religion, mysticism and mythology,
- royalty and nobility,
- sport and recreation,
- textile and clothing,
- transport and travel,
- warfare and defense.

Remaining tags are loaded to a custom script which, for each word found in the dictionary, returns one of the 34 domains distinguished by BabelNet. As some words can have more than one meaning, we firstly search for meaning classified by BabelNet as key for a given word. If there is more than one key meaning we extract a domain for each of them. Working under the assumption that the most used tags reflect user's interests in the topic to whom they refer, extracted domains serve as the ground-truth profile of interests for the user (Figure 2). Domains which occur most often are the most probable interests of a given user.

We decided to approach the problem of extracting interests from images as an object classification task. There is no need to extract aesthetic values of the image as the object included in this image describes its topic better than

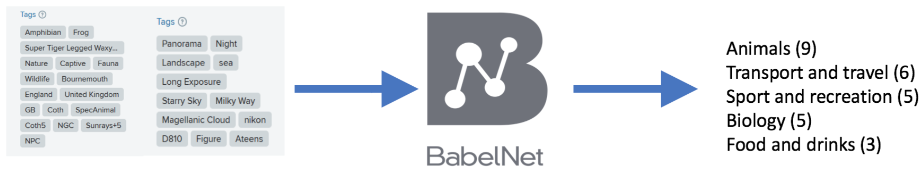

Fig. 2. Determining superior categories with BabelNet

colours, light distribution or other aesthetic measures. Knowing that existing neural network models achieve high accuracy in object recognition tasks, we decided to use Tensorflow [1] implementations of AlexNet [20], VGG-16 [37], VGG-19 [37], GoogleNet [38], and ResNet-50 [16]. All networks were modified to return a name of one object which is, with the biggest probability, included in the picture. Some classes have more than one name for the object assigned. In that case, we use only the first of them. This name constitutes the image's label. All CNN-s were trained on ImageNet dataset, which contains 1000 categories of objects. We used pre-trained Caffe models, listed in BVLC Model Zoo resources[2] and converted to Tensorflow. To work with our CNN-s, images needed to be prepared. Because the size of the input layer is fixed, we cropped images to 227x227 for AlexNet and 224x224 for the rest of the models. Resized images are then loaded to CNN-s and predicted labels are stored in a text file. When the object recognition process is done, predictions are cleaned accordingly to the rules that apply to tags. Cleaned labels are then loaded to the same script which extracts word domains from BabelNet. Five domains with the most occurrences serve as a predicted profile of user interests. The whole pipeline is summarised in Figure 3.

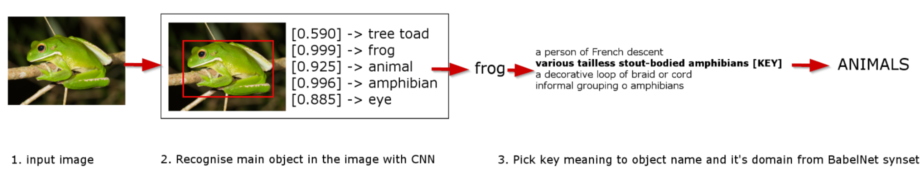

Fig. 3. Describing user's interests based on a single picture

For N users i, we have two sets of BabelNet word domains for each: A_i (tag-based) and $B_{i,t}$ (image-based). The tags-based set is the ground-truth profile of user interests. Image-based is built from results of the object recognition task (t denotes the used neural network). Given that, we can assume that the more similar tag- and image-based sets are, the better the capability of our method to

[2] http://caffe.berkeleyvision.org/model_zoo.html

predict interests is. For each user, we rank interests in these sets in descending order. After that, each set is narrowed down to the first five positions. In this way, we obtain one ground-truth predicted profiles (out of five most probable user interests). Predictions are evaluated by comparing the similarity between the samples of predicted and ground-truth interests for given user. We use F_1 score: $F_1(A_i, B_{i,t}) = \frac{2|A_i \cap B_{i,n}|}{|A_i|+|B_{i,n}|}$, where A_i and $B_{i,t}$ stands for the aforementioned sets of interests. The average F_1 score for a given network t is denoted as $\bar{F_1} = \frac{1}{N}\sum_{i=1}^{N} F_1(A_i, B_{i,t})$. The exact results are described in the following section.

4 Results

We report results for predictions based on five neural networks – AlexNet, VGG16, VGG19, GoogleNet, and ResNet50. In all cases, we calculated the F_1 score for each user. To show the distribution of accuracy, we counted the number of correct predictions $|A_i \cap B_{i,t}|$ for all users and denoted it as n. Overall accuracy of predictions made with a given network is expressed by the average number of matches $\overline{|A_i \cap B_{i,t}|}$ and corresponding average F_1 score $\overline{F_1}$. We compared prediction accuracies, mean scores, and score distributions for all neural networks. Table 1 contains these results (depicted in Figure 4).

Table 1. Experiment results, $N = 300$. Source: own research

$\vert A_i \cap B_{i,t}\vert$	AlexNet		VGG16		VGG19		GoogleNet		ResNet50	
	n	n/N	n	n/N	n	n/N	n	n/N	n	n/N
0	2	0.0067	**5**	**0.0167**	3	0.0100	4	0.0133	1	0.0033
1	38	0.1267	46	0.1533	48	0.1600	**51**	**0.1700**	41	0.1367
2	143	0.4767	**156**	**0.5200**	151	0.5033	151	0.5033	139	0.4633
3	107	0.3567	89	0.2967	94	0.3133	90	0.3000	**109**	**0.3633**
4	**10**	**0.0333**	4	0.0133	4	0.0133	4	0.0133	9	0.0300
5	0	0.0000	0	0.0000	0	0.0000	0	0.0000	**1**	**0.0033**
$\overline{\vert A_i \cap B_{i,t}\vert}$	2.28		2.14		2.16		2.13		**2.29**	
$\overline{F_1}$	0.4567		0.4273		0.4320		0.4260		**0.4580**	

Unfortunately, we managed to achieve 100% F_1 score for only one user in our dataset (with ResNet50). On the other hand, the number of completely false predictions (with accuracy equal to zero) is low and does not exceed 2% for any of the networks. We expected the accuracy achieved in experiments performed with more complex networks to be slightly higher. While this statement turned out to be true (ResNet50 had the best accuracy), our tests showed that the difference between networks is subtle and actually the relative frequencies follow a somewhat similar pattern. Surprisingly, AlexNet was almost as good as ResNet50. The most significant differences between predictions made with tested networks can be observed in the frequency plot (Figure 4). Clearly, using significantly more accurate object recognition solutions did not lead to much better results. It let us think, that the bottleneck of this approach is either the category generalisation or the lack of strongly labelled ground truth data. Our results can be validated with an online demo of our application[3] (it requires a BabelNet API key – it can be obtained for free for research purposes).

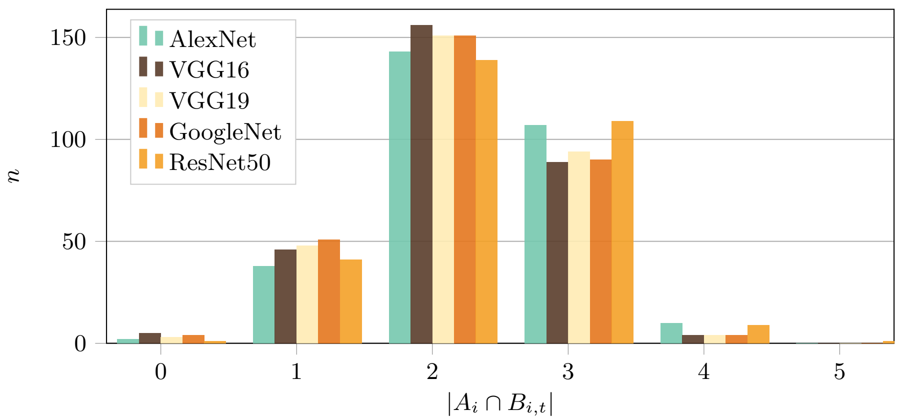

Fig. 4. Distribution of the number of matches n for different networks. Source: own research

5 Conclusions

Volumes of multimedia content generated by users of online social networks became so immense they can be no more processed by humans. Information contained in such content can be useful for various business purposes, such as user profiling. In this paper we presented an approach for extracting information about users' interests from images posted on online social networks. Our method combines image features extraction conducted with convolutional neural networks and semantic object category recognition and generalisation. Specifically, we used TensorFlow implementations of well-established CNNs to extract object features from Flickr images. We have chosen BabelNet for category generalisation task. The method was validated with data based on social-tagging mechanism.

Results of experiments performed on the dataset of 60,000 images crawled from Flickr show that multimedia content from online social networks can be effectively used to obtain information about user interests. Combination of neural network and BabelNet was capable of creating user's interests profile consisting

[3] `http://interestprofile.ml`

of general categories such as *animals*, *geography* and *travel* or *art and architecture*. ResNet50 turned out to be the most accurate network when validated with profiles based on tags annotated by authors of the photograph. Even though achieved accuracy is far from perfect, we believe that results are promising. Due to the error propagation, achieving better results relies on three factors: object recognition, semantic categorisation, and ground-truth data labels. The future work may include testing the method with state-of-the-art convolutional neural networks trained for object recognition. More importantly though, future tests can be also performed with different approaches to object category generalisation, e.g. with original ontologies or other sources of semantic connections (such as e.g. DBpedia Spotlight [26]). The presented solution can be also tested with different ground-truth datasets. It would be especially valuable, to confront the accuracy of the presented approach with other methods to validate its usefulness e.g. for content recommendation. Our method may be also modified and used for goals other than profiling interests e.g. for sentiment analysis. The main contribution of this paper is showing that semantic analysis of features extracted with CNNs can be simply applied to data analysis problems.

6 Acknowledgements

This work will be published as part of the book "Emerging Topics in Semantic Technologies. ISWC 2018 Satellite Events. E. Demidova, A.J. Zaveri, E. Simperl (Eds.), ISBN: 978-3-89838-736-1, 2018, AKA Verlag Berlin".

References

1. Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., Corrado, G.S., Davis, A., Dean, J., Devin, M., et al.: Tensorflow: Large-scale machine learning on heterogeneous distributed systems. arXiv preprint arXiv:1603.04467 (2016)
2. Aggarwal, C.C.: An introduction to social network data analytics. Social network data analytics pp. 1–15 (2011)
3. Ames, M., Naaman, M.: Why we tag: motivations for annotation in mobile and online media. In: Proceedings of the SIGCHI conference on Human factors in computing systems. pp. 971–980. ACM (2007)
4. Balduzzi, M., Platzer, C., Holz, T., Kirda, E., Balzarotti, D., Kruegel, C.: Abusing social networks for automated user profiling. In: Recent Advances in Intrusion Detection. pp. 422–441. Springer (2010)
5. Burger, J.D., Henderson, J., Kim, G., Zarrella, G.: Discriminating gender on twitter. In: Proceedings of the Conference on Empirical Methods in Natural Language Processing. pp. 1301–1309. Association for Computational Linguistics (2011)
6. Chaabane, A., Kaafar, M.A., Boreli, R.: Big friend is watching you: Analyzing online social networks tracking capabilities. In: Proceedings of the 2012 ACM workshop on Workshop on online social networks. pp. 7–12. ACM (2012)
7. Clevert, D.A., Unterthiner, T., Hochreiter, S.: Fast and accurate deep network learning by exponential linear units (elus). arXiv preprint arXiv:1511.07289 (2015)

8. Cristani, M., Vinciarelli, A., Segalin, C., Perina, A.: Unveiling the multimedia unconscious: Implicit cognitive processes and multimedia content analysis. In: Proceedings of the 21st ACM international conference on Multimedia. pp. 213–222. ACM (2013)
9. Deng, Z.H., Yu, H., Yang, Y.: Image tagging via cross-modal semantic mapping. In: Proceedings of the 23rd ACM international conference on Multimedia. pp. 1143–1146. ACM (2015)
10. Firan, C.S., Nejdl, W., Paiu, R.: The benefit of using tag-based profiles. In: Web Conference, 2007. LA-WEB 2007. Latin American. pp. 32–41. IEEE (2007)
11. Fu, J., Rui, Y.: Advances in deep learning approaches for image tagging. APSIPA Transactions on Signal and Information Processing **6** (2017)
12. Girshick, R.: Fast r-cnn. In: Proceedings of the IEEE international conference on computer vision. pp. 1440–1448 (2015)
13. Girshick, R., Donahue, J., Darrell, T., Malik, J.: Rich feature hierarchies for accurate object detection and semantic segmentation. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 580–587 (2014)
14. Gosling, S.D., Gaddis, S., Vazire, S., et al.: Personality impressions based on facebook profiles. ICWSM **7**, 1–4 (2007)
15. He, K., Gkioxari, G., Dollár, P., Girshick, R.: Mask r-cnn. arXiv preprint arXiv:1703.06870 (2017)
16. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 770–778 (2016)
17. Hung, C.C., Huang, Y.C., Hsu, J.Y.j., Wu, D.K.C.: Tag-based user profiling for social media recommendation. In: Workshop on Intelligent Techniques for Web Personalization & Recommender Systems at AAAI. pp. 49–55 (2008)
18. Jin, X., Wang, C., Luo, J., Yu, X., Han, J.: Likeminer: a system for mining the power of'like'in social media networks. In: Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining. pp. 753–756. ACM (2011)
19. Kosinski, M., Stillwell, D., Graepel, T.: Private traits and attributes are predictable from digital records of human behavior. Proceedings of the National Academy of Sciences **110**(15), 5802–5805 (2013)
20. Krizhevsky, A., Sutskever, I., Hinton, G.E.: Imagenet classification with deep convolutional neural networks. In: Advances in neural information processing systems. pp. 1097–1105 (2012)
21. Li, R., Wang, S., Deng, H., Wang, R., Chang, K.C.C.: Towards social user profiling: unified and discriminative influence model for inferring home locations. In: Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining. pp. 1023–1031. ACM (2012)
22. Lin, M., Chen, Q., Yan, S.: Network in network. arXiv preprint arXiv:1312.4400 (2013)
23. Lipsman, A., Mudd, G., Rich, M., Bruich, S.: The power of "like". Journal of Advertising research **52**(1), 40–52 (2012)
24. Liu, L., Preotiuc-Pietro, D., Samani, Z.R., Moghaddam, M.E., Ungar, L.H.: Analyzing personality through social media profile picture choice. In: ICWSM. pp. 211–220 (2016)
25. Lovato, P., Bicego, M., Segalin, C., Perina, A., Sebe, N., Cristani, M.: Faved! biometrics: Tell me which image you like and i'll tell you who you are. IEEE Transactions on Information Forensics and Security **9**(3), 364–374 (2014)

26. Mendes, P.N., Jakob, M., García-Silva, A., Bizer, C.: Dbpedia spotlight: shedding light on the web of documents. In: Proceedings of the 7th international conference on semantic systems. pp. 1–8. ACM (2011)
27. Michlmayr, E., Cayzer, S.: Learning user profiles from tagging data and leveraging them for personal (ized) information access (2007)
28. Navigli, R., Ponzetto, S.P.: Babelnet: The automatic construction, evaluation and application of a wide-coverage multilingual semantic network. Artificial Intelligence **193**, 217–250 (2012)
29. Navigli, R., Ponzetto, S.P.: Multilingual wsd with just a few lines of code: the babelnet api. In: Proceedings of the ACL 2012 System Demonstrations. pp. 67–72. Association for Computational Linguistics (2012)
30. Oramas, S., Ostuni, V.C., Noia, T.D., Serra, X., Sciascio, E.D.: Sound and music recommendation with knowledge graphs. ACM Transactions on Intelligent Systems and Technology (TIST) **8**(2), 21 (2017)
31. Redi, M., Quercia, D., Graham, L.T., Gosling, S.D.: Like partying? your face says it all. predicting the ambiance of places with profile pictures. arXiv preprint arXiv:1505.07522 (2015)
32. Ren, S., He, K., Girshick, R., Sun, J.: Faster r-cnn: Towards real-time object detection with region proposal networks. In: Advances in neural information processing systems. pp. 91–99 (2015)
33. Schwartz, H.A., Eichstaedt, J.C., Kern, M.L., Dziurzynski, L., Ramones, S.M., Agrawal, M., Shah, A., Kosinski, M., Stillwell, D., Seligman, M.E., et al.: Personality, gender, and age in the language of social media: The open-vocabulary approach. PloS one **8**(9), e73791 (2013)
34. Segalin, C., Perina, A., Cristani, M., Vinciarelli, A.: The pictures we like are our image: continuous mapping of favorite pictures into self-assessed and attributed personality traits. IEEE Transactions on Affective Computing **8**(2), 268–285 (2017)
35. Segalin, C., Cheng, D.S., Cristani, M.: Social profiling through image understanding: Personality inference using convolutional neural networks. Computer Vision and Image Understanding **156**, 34–50 (2017)
36. Sen, S., Lam, S.K., Rashid, A.M., Cosley, D., Frankowski, D., Osterhouse, J., Harper, F.M., Riedl, J.: Tagging, communities, vocabulary, evolution. In: Proceedings of the 2006 20th anniversary conference on Computer supported cooperative work. pp. 181–190. ACM (2006)
37. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556 (2014)
38. Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., Rabinovich, A.: Going deeper with convolutions. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 1–9 (2015)
39. Wei, X., Stillwell, D.: How smart does your profile image look?: Estimating intelligence from social network profile images. In: Proceedings of the Tenth ACM International Conference on Web Search and Data Mining. pp. 33–40. ACM (2017)
40. You, Q., Bhatia, S., Luo, J.: A picture tells a thousand words—about you! user interest profiling from user generated visual content. Signal Processing **124**, 45–53 (2016)
41. You, Q., Bhatia, S., Sun, T., Luo, J.: The eyes of the beholder: Gender prediction using images posted in online social networks. In: Data Mining Workshop (ICDMW), 2014 IEEE International Conference on. pp. 1026–1030. IEEE (2014)
42. Zeiler, M.D., Fergus, R.: Visualizing and understanding convolutional networks. In: European conference on computer vision. pp. 818–833. Springer (2014)