

KOSI- Key Object Detection for Sentiment Insights

Daniel Dimanov¹ and Shahin Rostami²

Faculty of Science & Technology, Bournemouth University, Bournemouth, BH12 5BB,
United Kingdom

¹i7461730@bournemouth.ac.uk

²srostami@bournemouth.ac.uk

WWW home page: <https://research.bournemouth.ac.uk/project/ciri/>

Abstract. This paper explores an original approach of using computer vision, data mining and an expert system to facilitate marketers and other interested parties to take automated data-driven decisions with the use of actionable insights. The system uses a state-of-the-art algorithm to retrieve all the images of a desired Instagram user profile. Then, the data is passed through a combination of different convolutional neural networks for object detection and a rule-based translation system to determine the interests of this profile user. Further, using a separately trained convolutional neural network with an original dataset developed as part of this study, personality insights are derived. The results from the conducted experiments yield a satisfactory prediction of interests and not very promising results for the personality prediction.

Keywords: Image Analytics, Data Mining, Convolutional Neural Networks

1 Introduction

The rising interest in social media marketing [23] has become the reason many companies are exploring new audience engagement tools in aid of gaining exposure and achieving better customer reach [8]. Many marketing companies struggle to broadcast their advertisements and messages to their desired target group, determined by their common interests or other common denominators. Currently, this is done by looking through fan pages, searching keywords in shares, or even looking at other key influencers in the given industry and targeting their followers in pursuit driven by the prejudice that these people are interested in a certain type of merchandise or product [19]. Some other strategies focus more on location, gender and hashtags, which, however, are very broad and inefficient for narrowing down the specific target audience. As a consequence, more software systems with a deeper level of analytics and insights are starting to emerge, like IBMs Watson Analytics ¹. Marketing agencies, as well as any other businesses

¹ <https://www.ibm.com/watson-analytics>

interested in marketing, need a system that can aid narrowing down the target group, utilising in an optimal way most of the data already available on social media. Data mining, even though dating back to more than two decades ago, has become one of the most popular tools for gathering and further analysing data in many different sectors [7]. The amounts of digital pictures taken and shared are drastically increasing [24]. Nearly five trillion photos are stored online at a rising rate of more than a trillion a year [18]. A huge part of the available data is left out when only text is considered for information extraction [28]. Since these valuable pieces of information are usually left out and there are a rising number of social media platforms focused purely on pictures and videos, this paper combines the concepts of data extraction, image processing and analysis to provide the end user with valuable insights for a chosen social media profile. The aim of this paper is to explore an algorithm that can predict interests and personality traits insights for Instagram users based purely on their posted images. Such a system would be able to facilitate the conversion of peoples social media data into highly valuable information to enable marketers and other interested parties to make data-driven decisions in two aspects. First, to have quantitative measures to select potential influencers and have clear facts for sponsorship negotiation. Secondly, to quantify the effectiveness of their campaigns.

The remainder of the paper is organised in the following way. Section 2 introduces the concept of computer vision and machine learning approaches used in the algorithm. Section 3 discusses the overall structure of the algorithm and presents insights about how the algorithm works. Section 4 examines the performance of the algorithm in terms of accuracy, speed and fit for purpose. Section 5 is a discussion of what the algorithm capabilities are, what can it achieve, future possible applications and significance of the developed algorithm. Finally, section 6 explores what future work can be done to build on the algorithm and surpass the state-of-art.

2 Computer Vision

Computer vision is an increasingly popular field of machine learning [25]. Computer vision is the translation of raw data input (still picture or video) into valuable information. The value of this information depends on the context it is used in and on the requirements [2]. Computer vision is the term used to refer to the ability of the machines (computers) to perceive entities like a human being. This term is often used to denominate the combination between concepts like image processing, image analysis, object detection and the deep learning regularly used to achieve the given goals of a project [26]. Convolutional Neural Networks(CNNs) are crucial for computer vision since these algorithms imitate the processes occurring in the visual cortex in humans brain [15]. CNNs are a type of artificial multi-layer neural network that applies convolutions on the input layer of the network [29]. It shares the concepts of the neuron and the adjustable weights and overall are similar to Artificial Neural Networks(ANNs), but also different, because of the additional convolutional, pooling and flattening

layers[11] [29]. CNNs consist of five or three layers of operations depending on the level of detail and complexity of each layer. In this paper, they will be split into five step-wise operations from the raw input to the output, which will not be included as separate layers. The first layer is the convolutional layer [5], followed by the hidden layer(ReLU activated usually [14]), Pooling [21] , Flattening [12] and fully connected layers(ANN)[27] where the input for every layer is the output of the previous one [29]. These powerful algorithms rely on classification to recognise complex patterns or in the case of computer vision - objects.

Machine learning algorithms are used in a wide range of areas, such as medical science [1], concealed weapon detection [20] , and retail forecasting [17].

An example of an implementation of this concept is the Single Shot MultiBox Detection(SSD) algorithm ²[16]. SSD uses a single deep neural network trained to recognise objects in images and their position in the image. The power of SSD comes from convolving the images many times before detecting features with filters. Thus, the algorithm can find abstract features, using which it recognises what the object in the corresponding box is [4]. In the training process, it is fed with images, which have labelled boxes defining the ground truth on what objects are present in the image. In the paper of Liu et al.[16] the training process is based on Pascal VOC and COCO experiments. Moreover, in the paper a comparison is drawn with the relative state-of-the-art systems and SSD, even though simpler, proves to be superior and faster. (Example of usage: Fig. 1)

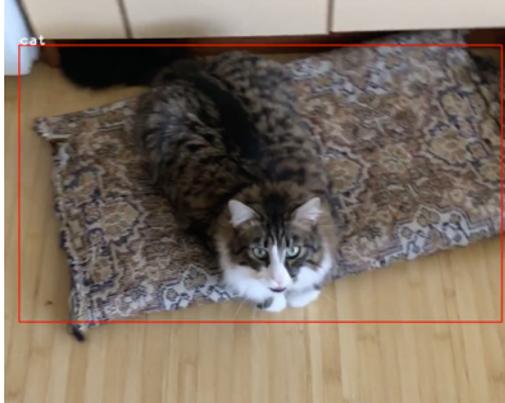


Fig. 1. SSD algorithm used to recognise a cat in video frame

3 KOSI Algorithm

² <https://github.com/amdegroot/ssd.pytorch>

Result: Target user interests and personality predictions

```

while authentication is rejected do
  if authentication is successful then
    Create user directory;
    Open sub-process executing API calls;
    Collect all user photos;
    Store only salient data (no descriptions or tags);
  else
    prompt for new credentials;
  end
end
Initialise counters and start timer;
Pass all images to the SSD algorithm;
load SSD model  $\leftarrow$  pre-trained_weights;
VGG16  $\leftarrow$  pretrained VGG16 on ImageNET;
dir  $\leftarrow$  user's directory
for each file in [dir] do
  if video then
    array  $\leftarrow$  split_in_frames(video);
    for each image in [array] do
      image  $\leftarrow$  SSD.add_bounding_boxes(image);
      dir  $\leftarrow$  dir  $\cup$  {image};
    end for
  end if
  image  $\leftarrow$  base_transformations(image);
  image  $\leftarrow$  rescale_and_reshape(image);
  initialise objects_found[]
  for each bounding_box do
    confidence  $\leftarrow$  bounding_box.get_confidence();
    if confidence > 0.6 then
      objects_found[object] + = 1;
    end if
  end for
  predictions[]  $\leftarrow$  VGG16(file) ;
  object  $\leftarrow$  decode_object()  $\leftarrow$  where predictions[id] > 60% ;
  objects_found[id] + = 1;
  predictions_ranked  $\leftarrow$  rank(predictions);
  if  $\sum_{i=1}^2$  predictions_ranked[object  $\leftarrow$  decode_object(i)] > 75% then
     $\sum_{i=1}^2$  objects_found[predictions_ranked[object  $\leftarrow$  decode_object(i)]] + = 1;
  end if
end for
interests  $\leftarrow$  Rule-based_system(objects_found);
personalities  $\leftarrow$  personalities_module(dir);

```

Algorithm 1: KOSI Algorithm

Result: Key objects are translated to interests

```

ic ← image_count ← dir
objects[] ← KOSI;
objects[] ← decode(objects[])
exceptions[] ← Special objects derived from user study including "Person",
"dog" (and other animals) , "volleyball" (and other sports),
"confectionery", "toyshop", "bookshop", "self_propeller_vehicle" ...;
interests[] ;
for each object in objects[] do
  synset ← object;
  if topic_domains ∈ synset then
    td ← topic_domains;
    for each t ∈ td do
      for each l ∈ t.lemmas[] do
        interests[l.name] += objects[object];
      end for
    end for
  else
    while synset.min_depth > 8 do
      if synset ∉ exceptions[] then
        hyp_n ← syn.hypernyms[0];
        interests[hyn.n.lemmas[0]] += object[object];
      end if
    end while
  end if
end for
categories[] ← 140 object categories (manually derived)
for each interest in interests[] do
  interests[] ← categorise(interests[], categories[]);
end for
interests_counter ← occurrence, interest ← interests[];
all_int_count ←  $\sum_i^{interests(interests\_counter)}$  interests_counter[i];
final_interests[]
for o, i ← interests_counter do
  if o ≥ all_int_count * 0.05 then
    if i = "Person" then
      if o > 2.5ic then
        i = "Group activities";
      else if 2.5ic ≥ o > 1.5ic then
        i = "Relationship"
      else if o > 0.6ic then
        i = "Appearance"
      else
        i = "People"
      end if
    end if
    final_interests[] ← i;
  end if
end for
interests ← Rule-based_system(objects_found);

```

Algorithm 2: Rule based system

The algorithm (Algorithm 1) can be explored as having four separate layers. The first one is data extraction, which is responsible for collecting all pictures of an Instagram user, specified by the user of the system. The second one is object detection, which uses computer vision and a combination of two CNNs to discover all objects in all collected photos. The third layer is sentiment analysis, where the found objects are categories using WordNET³ and an expert system is created and used to determine the potential interests of the profile's user based on the key objects found in the pictures and the number of their occurrences(Algorithm 2). The fourth one is the personality insights detection.

The first layer consist of a data mining algorithm, which collects all images that will be analysed next. In the context of the project these images are collected from the specified user's profile. For the second layer, CNNs are used in two different ways. The data first goes through an SSD neural network(Please refer to 2), which was developed to recognise over 30 objects using Pascal VOC and COCO classes. [9]. Pre-trained model on these classes is implemented together with an official pytorch implementation of the SSD model described in Liu's et al. paper[16] from Alexander De Groot [6]. Firstly, all images go through a series of transformations, values of which are the suggested ones from the study of Wei Liu [16] and are passed through the SSD model. SSD uses boxes to determine the objects in the images, this is useful, since it can detect all objects present in the image and their placement. It can also be used to detect objects in video frames, but that function was purposefully left out in the development because its a trade-off, which is performance. All objects and their count are stored in Counter variable and the same variable is further used for the next step in the object detection. Because of the relatively few classes the SSD is trained and developed for, a different approach was implemented as well to combine the quality of the SSD detections and to fill the gap in the left-out objects. A high-level neural networks API, written in python called Keras ⁴ [10] was used to develop a second sieve. In this step, VGG16 [22] model is used. It is used together with ImageNet⁵, which is an enormous dataset, containing millions of images in over a thousand classes. The drawback of this approach is that the images are not box-labelled, thus a prediction of the whole image is done and in the case of Instagram there is rarely just one object that the image features. Thus a threshold for certainty of the predictions is introduced, below which predicted objects are ignored. All other predictions or a combination of the them (based on predefined rules) are added to the Counter and further analysed using a rule-based expert system. The last part of the algorithm is the personalities insights extraction, which uses the raw input of images from the first layer and since parallelism is not yet implemented it runs last. A novel CNN is used to determine the possible profile user's personality according to 16personalities⁶.

³ <https://wordnet.princeton.edu>

⁴ <https://keras.io>

⁵ <http://www.image-net.org>

⁶ <https://www.16personalities.com>

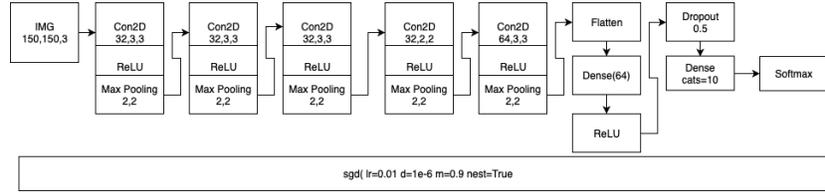


Fig. 2. CNN used for personalities detection

The personalities CNN (Figure 2) predicts the profile user’s personality based on the training set provided by a conducted study. The study is psychology based qualitative research in the form of personalised interviews with a group of 30 people, who were invited to participate. During the interviews, the participants were asked to take a personality test based on [3]. Since experience and expertise of psychology was required, an expert in the field of psychology was invited to supervise the interviews. The personality traits are motivated from the Big Five [13] study and 16 personalities test ⁶, used to determine the participants personalities.

4 Results

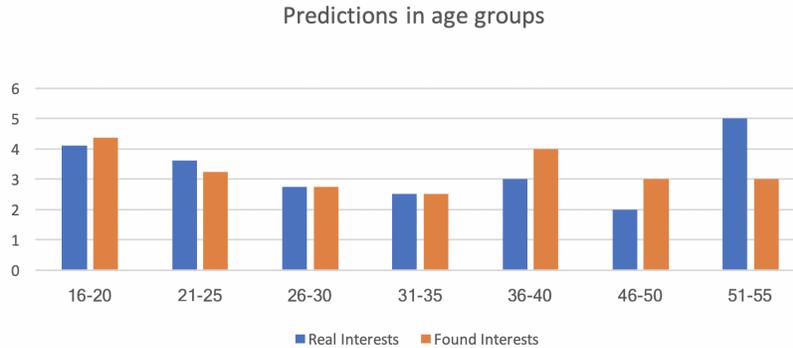


Fig. 3. Predictions in different age groups predicted vs. actual

This section presents and interprets some key findings of the paper as well as illustrate the performances of different modules. Based on the user study and further analysis of all the information the objects collected could be categorised into 584 categories, referring to 71 interests. The object detection algorithm was tested with a manually crafted test set, which

contains 244 different images with a total of 708 objects. 636 objects were correctly recognised, resulting in 89.83% performance. Out of the 244 images 214 were labelled correctly and the other 30 contained either false positive results or failed to recognise some of the objects, which is 87.70% accurate. Since the

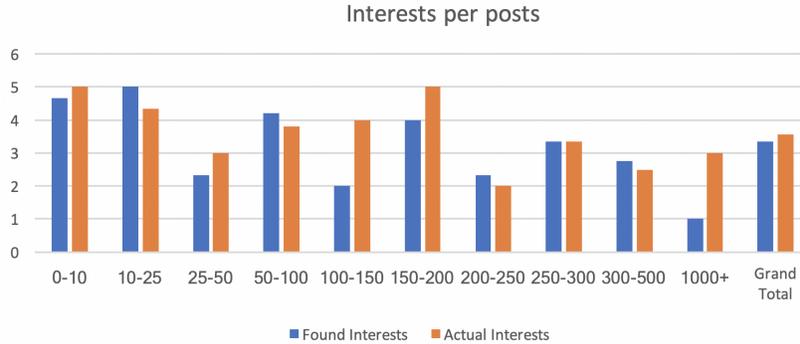


Fig. 4. Interests found per posts predicted vs. actual

test data is insufficient and with relatively low diversity to fully determine the algorithms accuracy with certainty, the accuracy of the object detection is yet to be finally determined. It can be argued that the findings in Fig. 3 do not necessarily explore the true performance for the corresponding age groups, because of the relatively small sample size of participants. The number of participants from each group has been attempted to be as evenly distributed and minor deviations do exist in the 30-50 groups, where results can be more easily swayed by outlier data, but overall the sample for the test set was carefully studied and is considered to be fairly representative for the problem domain. The rule-based system, although developed manually, may be regarded as a basis for improvement and automation using results from this exact algorithm. The categories devised from the system are based on human-logic, but the logic is transferable since a data-set could be developed and serve for training a semi-supervised algorithm. Nevertheless, an overall 93,48% accuracy is achieved in predicting interests, which is reduced to 81,13% after removal of false-positive results (Please refer to Fig. 3 and Fig. 4). A performance evaluation was done comparing the posts of an individual with the elapsed time of the system illustrated by Fig. 5.

5 Conclusions and Discussion

This paper explores an original approach of using computer vision, data mining and an expert system to take automated data-driven decisions with the use of actionable insights. The paper first introduces the problem and the proposed solution. It sets the aim, which is to convert peoples social media data into

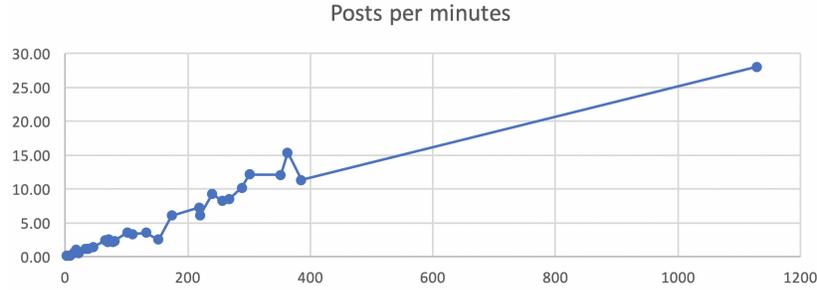


Fig. 5. Performance measured with posts successfully analysed compared to the time elapsed (in minutes)

actionable insights. The algorithm this paper discusses is constructed using a mining algorithm to gather the user images data, followed by two state-of-art CNNs to detect objects in the images and a rule-based system to further narrow down, categories and interpret the interests of this user. The rule-based system, even though manual can act as a basis for dataset creation, which can serve as a training and testing set for supervised or semi-supervised algorithm. As part of this study, two original datasets are created and original actionable insights are derived from the data.

6 Future Work

Based on the devised findings and analysis, as well as the whole methodology some limitations, recommendations and suggestions for future work are produced. Starting with one of the most imminent ones, in order for this project to comply with the new GDPR agreement and the depreciated Instagram API changes to the system seem to be required. Another gap in the system is that the rule-based system translates categories to interest using expert knowledge. Not only it misses some of the important objects and factors, but more data is needed for testing and understanding how objects and images relate to interests. However, the systems rule-based approach may be used for data-set creation for the training of a neural network. Therefore, a recommendation is that the manual decisions in the algorithm are substituted by a semi-supervised system in the future. Another recommendation is that bigger and much more diverse data-set is used when decisions are taken since all participants in the study are mainly aged between 18 and 29. The data-set created by the author is, although for future work more expertise is needed for deriving such conclusions, reliable, however, it is relatively small. For a convolutional neural network to be able to precisely determine such intended information a much larger data-set should be considered. Performance wise, all working algorithms and systems may be used asynchronously on different threads to conserve time and to utilise computing resources. Moreover, all results are achieved on MacBook Pro 2016 with Intel i7

and all training, as well as processing, is using only the CPU, because of inability to use the graphics card with CUDA and make it easy for integration. Thus, further optimisation may be done using GPU-acceleration, which will significantly speed up all computing processes and yield much better performance.

References

1. Bellegdi, S.A., Mohandes, S., Soufan, O.M., Arafat, S.: Computational intelligence for cardiac arrhythmia classification. *UKCI* **2011**, 93–97 (2011)
2. Bradski, G., Kaehler, A.: Learning opencv:[computer vision with the opencv library], [nachdr.] (2011)
3. Briggs, K.C.: Myers-Briggs type indicator. Consulting Psychologists Press Palo Alto, CA (1976)
4. Cai, Z., Vasconcelos, N.: Cascade r-cnn: Delving into high quality object detection. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 6154–6162 (2018)
5. Ciresan, D.C., Meier, U., Masci, J., Gambardella, L.M., Schmidhuber, J.: Flexible, high performance convolutional neural networks for image classification. In: Twenty-Second International Joint Conference on Artificial Intelligence (2011)
6. DeGroot, A., Brown, E.: Ssd pytorch. Github (2018)
7. Delen, D., Eryarsoy, E., Seker, S.: Introduction to data, text and web mining for business analytics minitrack. In: Proceedings of the 50th Hawaii International Conference on System Sciences (2017)
8. Erevelles, S., Fukawa, N., Swayne, L.: Big data consumer analytics and the transformation of marketing. *Journal of Business Research* **69**(2), 897–904 (2016)
9. Everingham, M., Van Gool, L., Williams, C.K., Winn, J., Zisserman, A.: The pascal visual object classes (voc) challenge. *International journal of computer vision* **88**(2), 303–338 (2010)
10. Gulli, A., Pal, S.: Deep Learning with Keras. Packt Publishing Ltd (2017)
11. Hussain, M., Bird, J.J., Faria, D.R.: A study on cnn transfer learning for image classification. In: UK Workshop on Computational Intelligence, pp. 191–202. Springer (2018)
12. Jin, J., Dundar, A., Culurciello, E.: Flattened convolutional neural networks for feedforward acceleration. arXiv preprint arXiv:1412.5474 (2014)
13. Judge, T.A., Higgins, C.A., Thoresen, C.J., Barrick, M.R.: The big five personality traits, general mental ability, and career success across the life span. *Personnel psychology* **52**(3), 621–652 (1999)
14. Kuo, C.C.J.: Understanding convolutional neural networks with a mathematical model. *Journal of Visual Communication and Image Representation* **41**, 406–413 (2016)
15. Lawrence, S., Giles, C.L., Tsoi, A.C., Back, A.D.: Face recognition: A convolutional neural-network approach. *IEEE transactions on neural networks* **8**(1), 98–113 (1997)
16. Liu, W., Wen, Y., Yu, Z., Yang, M.: Large-margin softmax loss for convolutional neural networks. In: ICML, vol. 2, p. 7 (2016)
17. Miller, S.M., Popova, V., John, R., Gongora, M.: Improving resource planning with soft computing techniques. Proceedings of UKCI 2008, De Montfort University, Leicester, UK pp. 37–42 (2008)

18. Paglen, T.: Invisible images: Your pictures are looking at you. *Architectural Design* **89**(1), 22–27 (2019)
19. Pradiptarini, C.: Social media marketing: Measuring its effectiveness and identifying the target market. *UW-L Journal of Undergraduate Research XIV* pp. 1–11 (2011)
20. Rostami, S., OReilly, D., Shenfield, A., Bowring, N.: A novel preference articulation operator for the evolutionary multi-objective optimisation of classifiers in concealed weapons detection. *Information Sciences* **295**, 494–520 (2015)
21. Scherer, D., Müller, A., Behnke, S.: Evaluation of pooling operations in convolutional architectures for object recognition. In: *International conference on artificial neural networks*, pp. 92–101. Springer (2010)
22. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556* (2014)
23. Stelzner, M.: *Social media marketing industry report* (2017)
24. Stenroos, O., et al.: *Object detection from images using convolutional neural networks* (2017)
25. Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., Wojna, Z.: Rethinking the inception architecture for computer vision. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 2818–2826 (2016)
26. Szeliski, R.: *Computer vision, texts in computer science* (2011)
27. Vo, D.M., Lee, S.W.: Semantic image segmentation using fully convolutional neural networks with multi-scale images and multi-scale dilated convolutions. *Multimedia Tools and Applications* pp. 1–19 (2018)
28. Welz, B., Rosenberg, A.: Exponential growth. In: *SAP Next-Gen*, pp. 31–35. Springer (2018)
29. Wu, J.: *Introduction to convolutional neural networks. National Key Lab for Novel Software Technology. Nanjing University. China* pp. 5–23 (2017)