

Raw Data-Based Filter for Image Collection and Classification

Abdul Mateen Rahmat*, Ahmed Aljanad*, Ala Abdulhakeem Abdulaziz Alreki*, Sami Naji*

*Department of Information Technology, American University of Afghanistan, Darulaman, Kabul, Afghanistan

mateen.rahmat@auaf.edu.af, aaljanad@auaf.edu.af, aabdulaziz@auaf.edu.af,
snaji@auaf.edu.af

Abstract— The current pathway in computer vision is heavily focusing on designing complex techniques for leveraging image classification accuracy, neglecting the importance of generating training image sets. In this work, we use the related tags provided by Flickr to construct a training image set for learning class models and perform classification. Our proposed filter generates a training set through combining related tags and the searched keyword, adopting CNN to extract features and perform classification using linear SVM. The experimental results show that we can outperform previously presented methods by 9% of average precision on the PASCAL VOC 2007 dataset.

Index Terms—Information Search and Retrieval, Metrics-complexity measures, performance measures General Terms Algorithms, Performance, Experimentation, Classification, Flickr, related terms, training set, classification

I. INTRODUCTION

Training image sets are the mainstay ingredient in which classifiers base their performances. Although previous research has focused on designing and developing semi-supervised techniques for learning representative class models and classifying images for any arbitrary class concept. Traditionally,

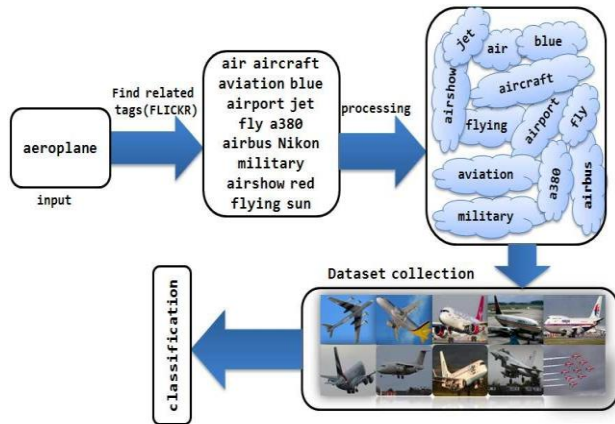


Fig. 1 Illustration of the textual information- based filter. The related tags pool from Flickr given the keyword as an input, process these terms (e.g., remove numeric terms such years, days, etc), construct bi-grams (e.g., 'jet+aeroplane'), query image search

normally ask humans to label images as positive if they contain the object of interest or negative otherwise. But recently this

practice has been shifted toward automatizing image collection by searching image search engines like Flickr for images, taking the retrieved results for extracting different low-level features such as color and applying any of the typical statistical learning techniques over these features to learn a class model. This model is then used to classify unseen images.

The primary focus of this work is to explore the usefulness of the textual information available, particularly related tags, on social websites like Flickr for obtaining semantically and visually related terms given the main keyword and taking them for collecting training images for a classifier. Reviewing some of the past literature, they usually utilize WordNet to obtain related terms such as 'Hypernymy' or 'Hyponymy' of a concept. Such lexical databases provide limited terms to a given concept and periodically updated. On the contrary, social websites like Flickr are enriched with terms and new concepts on a daily basis which can be considered as an advantage. In addition, this work will try to find out if the related tags are useful and indirectly comparable to WordNet in terms of generating training image sets through experiments.

So far, to the best of our knowledge, this has not been used for generating training image sets for learning representative class models for a classification task. The architecture of the proposed framework is shown in Fig. 1, illustrating the procedure of obtaining related terms to a given keyword donates object class name from Flickr. The combination is then used to obtain images that are less noisy and less separated from the main entity.

The rest of the paper is organized as follows: Section 2 summarizes some previously published research on the subject. Section 3 provides a brief description of Flickr related terms. Section 4 presents the proposed framework and Section 5 demonstrates the experiments setup and preliminary results. Conclusion and future directions are described in Section 6.

II. RELATED WORK

Various methods have explored the effectiveness of

when we learn a model for any specific concept, we

adding related terms to the keyword for improving image search results exploiting lexical databases [5], tags associated with images [7], big data [2]. Nevertheless, none of these methods have explored the power of related tags as semantically related terms for leveraging classification performances.

Closely related work to ours has explored the usefulness of combining the textual tags come with images and different low-level features classifiers for enhancing the classification performance [7]. The authors build a textual feature vector coming with images relying on each term occurrences, build a visual vector from images, perform separate classification, and then combine both classifiers' confidence values using logistic regression to obtain the final classifier results. The presented method demonstrates the power of using textual tags with visual images in classification tasks. Chen et al. [1] introduce a generalized hierarchical matching framework, able to embed useful side information for improving classification performance.

III. HOW FLICKR RELATED TAGS CAN HELP LEVERAGING IMAGE CLASSIFICATION ACCURACY

When we search the image search engines for images, we usually use a single keyword and sometimes observe the first presented images are containing the object of interest while the later on images are somewhat scattered from the main object. This occurs because image search engines return images relying on the metadata associated with them which can be often ambiguous and polysemous (see Fig. 2).

Internet users in general when intend to obtain strongly desirable results, they selectively choose a related term to be attached to a search keyword which can act as a constraint to restrict the result to be within the context. To make this process adoptable by machines, we should develop a strategy to select related terms autonomously which should be visually related to the searched object.

To tackle this problem and to reduce the ambiguity and the polysemy of the search results we propose to use the related terms obtained from Flickr which constrain the search to return results that are visually similar and related to the object we are seeking. For example, related tags for a concept "airplane" are "jet aviation", "airport", etc.' Although there is no sufficient explanation on how these tags are generated but as described in the Flickr API they are generated.



Fig 3. Typing the keyword 'aeroplane' in Flickr, the results obtained present undesirable and contaminated images.: based on co-occurrence analysis.

IV. PROPOSED APPROACH

In this section, we will describe our new textual information related terms filter to automatically build a training image sets which classifier can learn a class model and then perform classification. The proposed approach consists of three main steps: **Related tags** (i.e, Flickr) to generate a set of related terms to the searched keyword; **Cleaning** to keep tags which are meaningful through removing short strings, and **dataset collection** to collect training image sets. Fig. 3 shows a schematic



Fig 2. A schematic representation of the proposed method representation of the proposed framework.

A. RELATED TAGS POOL

We adopt Flickr related tags to obtain related terms. We design an algorithm that takes the keyword donates the object class name C_n and asks Flickr to retrieve all the related tags RT . These related tags are generated based on the co-occurrence analysis or cluster usage analysis which tags are grouped on the basis of their frequent usage.

B. CLEANING MECHANISM

Flickr-supplied related tags RT remain noisy in a sense that there exist related tags RT which do not relate to the searched keyword or have no positive implications on the searched results. We design an algorithm that searches through the related tags RT and if it finds the tag x_i has no positive influence on the object class C_n , based on manually predefined words, then the algorithm excludes the tag x_i from the object class related terms CRT . For example, if the tag x_i is 'Nikon' in the CRT list does not contribute to recognizing the object class C_n 'aeroplane' in images neither is somehow related to the class C_n 'aeroplane', it removes the tag x_i from the CRT list. So, in our opinion, this should be done in an autonomous approach, capable of detaching such tags. However, in this work, we detach noisy tags in a semi-automatic way. We use publicly available English dictionary D_{en} that retains the tag x_i if x_i in D_{en} and therefore the tag x_i is considered meaningful. This involves removing short strings tags if the constrain ($x_i > 2$) is satisfied, non-English alphabets, removing plural x_i form of a tag x_i , split dashed/underscored tag if this rule is satisfied ($x + x_i$), prune out object class name.

Table 1 A set of classes and their top ten related tags

Classes	Top 10 related tags
aeroplane	air, aircraft, airplane, airport, fighter, flight, flying, jet, plane, sky
Bicycle	ride, race, city, cycle, cycling, people, red, road, street, urban
bird	animal, duck, gull, feathers, flying, nature, seagull, sky, water, wildlife
boat	beach, blue, lake, ocean, reflection, sea, ship, sky, sunset, water
bottle	alcohol, beer, drink, glass, green, light, macro, red, water, wine
bus	transport, school, England, London, night, people, road, street, traffic, urban
car	auto, city, classic, ford, night, old, red, road, street, Volkswagen
cat	animal, pet, cute, eyes, feline, kitten, kitty, tabby, whiskers, white
chair	black, blue, cat, girl, light, furniture, red, table, white, window
cow	white, animal, calf, cattle, milk, farm, field, grass, green, nature
dog	sleeping, black, canine, cute, sweet, funny, pet, animal, puppy, white horse
horse	animal, ranch, riding, farm, field, grass, pet, nature, pony, fence
motorbike	ride, police, motor, honda, race, classic, road, scooter, street, wheels
person	black, face, female, girl, man, people, human, smile, white, woman
sheep	animal, ewe, farm, field, grass, green, lamb, landscape, nature, wool
sofa	red, white, couch, sitting, furniture, girl, home, interior, living room, woman
train	bridge, locomotive, rail, railroad, railway, station, steam, subway, tracks, underground
tv	sony, black, blue, film, Germany, television, tower, watching, yellow, broadcast

C. DATASET COLLECTION

Given the keyword donates the object class name C_n and a set related tags CRT, we develop an algorithm to crawl any image search engines for several defined images using a formulated query ‘tag+keyword’ ($x_i + C_n$). For each object class C_n , the algorithm takes N tags to be added.

V. EXPERIMENT AND RESULT

To evaluate the classification performance of our proposed filtering dataset, we construct 500 images per class for 18 classes taken from PASCAL VOC 2007,

which are collected from Flickr. Each class dataset is built using the first 10 related tags (see Table 1). As a testing set, we use all the images of PASCAL VOC 2007 for the chosen testing classes. In this work, Convolutional Networks (ConvNets) are used as a feature extractor [3] [1]; in particular, we use a feedforward multilayer perceptron, adopting publicly available pre-trained ImageNet deep learning model[6], taking the output of the 36 layers of the networks. We then feed the 4096-dimension sparse feature vector to learn the class model using linear SVM for each class. In the details, for each image class, we use 500 images, while for the negative class we consider a uniform sample from the other classes into play. As for the figures of merit, we adopt the PASCAL VOC’s interpolated average precision (AP). As for the direct competitor, the method in [7] is highly comparable particularly the way in which the training dataset is created. The authors utilize the metadata associated with an image together with visual features generated from Flickr to leverage the classification performance using the bag of word model and SVM. In addition, we show the results using our competitors that have been previously published in prestigious conferences in Table 2. As apparent, our textual information filter is comparable to some classification methods, which is considerably simple. Besides, related tags filter outperforms all the other methods, having very good performances on the "aeroplane, motorbike, bird, boat, bottle, bus, cat, chair, cow, dog, sheep ", showing comparable numbers on the other ones. In Table II the classification results are presented, showing that surprisingly our approach works better than other classification methods on PASCAL VOC 2007. In addition, it is interesting to observe the power of the related tags given by Flickr through figures. More in general, it is apparent that our textual information-based filter is gaining of about 19.2% over our strongest competitor while in comparison to other classification approaches our approach outperforms by about 9% on average. This result indicates the effectiveness of the related tags in image collection and image classification. of the images being retrieved by approach and compare to the result shown in Figure 3. More in general, it is apparent that Flickr retrieves more scattered and contaminated samples.

Table 2. A comparison of result obtained by our methods vs the rest

Model	plane	bike	bird	boat	bottle	bus	car	cat	chair	cow	dog	horse	motor	person	sheep	sofa	train	tv	mAP
[7]	76.7	74.7	53.8	72.1	40.4	71.7	83.6	66.5	52.5	57.5	51.1	81.4	71.5	86.5	55.3	60.6	80.6	87.8	50.5
INRIA [4]	77.5	63.6	56.1	71.9	33.1	60.6	78.0	58.8	53.5	42.6	45.8	77.5	64.0	85.9	44.7	50.6	79.2	53.2	60.9
LLC [8]	74.8	65.2	50.7	70.9	28.7	68.8	78.5	61.7	54.3	48.6	44.1	76.6	66.9	83.5	44.6	53.4	78.2	53.5	61.3
PROPOSED FILTER	90.5	86.2	85.6	86.1	54.7	76.6	82.5	77.4	55.7	70.5	72.0	76.0	69.8	67.6	62.3	34.0	72.7	34.4	69.7

VI. CONCLUSIONS

Training image set collection is recently catching the attention of many computer vision researchers, making it one of the interesting and challenging topics to handle. The intention of this paper is to emphasize, particularly to the computer vision community, that using related tags provided by Flickr, represent a significant source of information which can be even an alternative to WordNet that may enable us to obtain expressive and representative image training sets that is somewhat equivalent to hand-craft datasets like ImageNet.

REFERENCES

- [1] Q. Chen, Z. Song, Y. Hua, Z. Huang, and S. Yan. Hierarchical matching with side information for image classification. In *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*, pages 3426–3433, June 2012.
- [2] D. S. Cheng, F. Setti, N. Zeni, R. Ferrario, and M. Cristani. Semantically-driven automatic creation of training sets for object recognition. *Computer Vision and Image Understanding*, 131(0):56 – 71, 2015. Special section: Large Scale Data-Driven Evaluation in Computer Vision.
- [3] A. Krizhevsky, I. Sutskever, and G. E. Hinton. Imagenet classification with deep convolutional neural networks. In F. Pereira, C. Burges, L. Bottou, and K. Weinberger, editors, *Advances in Neural Information Processing Systems* 25, pages 1097–1105. Curran Associates, Inc., 2012.
- [4] M. Marszalek, C. Schmid, H. Harzall, and J. van deWeijer. Learning object representations for visual classification (AP in %) on VOC 2007. *The proposed filter outperforms some baselines classification methods*: object class recognition, Oct 2007. Visual Recognition Challenge workshop, in conjunction with ICCV.
- [5] F. Setti, D. Porello, R. Ferrario, S. A. Abdulhak, and M. Cristani. "tell me more": How semantic technologies can help refining internet image search. In *Proceedings of the International Workshop on Video and Image Ground Truth in Computer Vision Applications, VIGTA '13*, pages 3:1–3:6, New York, NY, USA, 2013. ACM.
- [6] A. Vedaldi and K. Lenc. Matconvnet – convolutional neural networks for matlab. *CoRR*, abs/1412.4564, 2014.
- [7] G. Wang, D. Hoiem, and D. Forsyth. Building text features for object image classification. In *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*, pages 1367–1374, June 2009.
- [8] J. Wang, J. Yang, K. Yu, F. Lv, T. Huang, and Y. Gong. Locality-constrained linear coding for image classification. In *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on*, pages 3360–3367, June 2010.



Abdul Mateen Rahmat is a senior computer science student at the American University of Afghanistan and he was working on various fields such as financial accounting, database developer, computer software and mobile application developer and technical analyst.

Since 2018, he was a research assistant at the American University of Afghanistan. His research interests include machine learning, artificial intelligence, computer vision, information, cryptocurrency, network security, software platform, and software. Mr. Rahmat's awards and honors include the Afghanistan National Science Competition golden medal, three president's honor awards and one president's high honor award from American University of Afghanistan.



Ahmed Aljanad received the B.Eng. (Hons.) degree from the Multimedia University, Malaysia, in 2011, the M. Eng. degree from UTeM, Malaysia, in 2014, and the Ph. D. degree from the National University of Malaysia (UKM), Bangi,

Malaysia, in 2018, all in electrical engineering. He works as a post-doctoral researcher fellow in the Department of Electrical Power Engineering, University Tenaga Nasional from November 2018 till September 2019. He is currently the Assistant professor of science, technology, engineering, and mathematics (STEM) at American University of Afghanistan (AUAF). His current research interests include artificial intelligence, power conversion systems for renewable energy and battery energy storage applications, power quality, power systems analysis.



Ala Alariki He completed his B.A. in Computer Science (Security technology at Multimedia University Malaysia (MMU) in 2009. His MITS in Computer Science (Information Security) was completed in 2012 at the

Universiti Teknologi Malaysia (UTM). He received his Ph.D. in Computer Science (Information Security) at the Universiti Teknologi Malaysia (UTM) in 2016. He is currently the Assistant Professor of Information Technology, Department of STM, American University of Afghanistan (AUAF). His current research interests includes, Information security, Artificial Intelligence, Networking and Data science

Sami Naji is an assistant professor at the department of



Information Technology and Computer Science at the American University of Afghanistan. He obtained his PhD from the University of Verona in computer science with dissertation title: Automatic training dataset creation

for object recognition, Verona, Italy. Sami obtained his master degree from Dongseo University, Busan, South Korea in computer science and obtained his B.Sc from Multimedia University in Malaysia. His main research interest focuses on machine learning, biometric, bioinformatics, natural language processing, deep learning, etc.