

# Image Classification for Arabic: Assessing the Accuracy of Direct English to Arabic Translations

Ruyue Han

July 27, 2018

## Abstract

*Image classification is an ongoing research challenge. Most of the available research focuses on image classification for the English language, only few meathures focus on Arabic language. The purpose of this paper is to investigate a mehtod that generate Arabic labels for images of objects and test the accuracy of this method. The way this paper use is directly translating English labels that are available on ImageNet to Arabic labels. In this study, 2,887 labeled images were randomly selected from ImageNet. All of the labels were translated from English to Arabic using Google Translate. And the result shows that 65.5% of the Arabic labels were accurate.*

## 1. Introduction

In recent years, advances in computer vision and image classification have been significant. The goal of image classification is to generate an accurate label or a group of labels for an image that captures the content(s) of the image [1]. Because of the availability of large databases of labeled images recently, more and more accurate image classification algorithms are arising. Each image in the ImageNet database that used by Scale Visual Recognition Challenge is labeled with a wordnet synset for each object present in an image.

Most of the current research on image classification focuses on developing algorithms for labelling objects present in an image in English. So image classification for the Arabic language remains an unexplored problem. Because Arabic is one of the most spoken languages in the world, exploring image classification for Arabic is relevant and requires investigation.

Would the most accurate image classification algorithms be able to produce highly accurate results if the dataset used for training includes labels in Arabic? This is the real question. But author doesn't give anwser.haha. This paper focues on a smaller question that is measuring the accuracy of Ara-

bic labels generated by directly translating English lables already in ImageNet.

## 2. Related Work

**Image Classification** Image classification is the task of identifying and labelling an object or list of objects present in an image. Recent advances in the field are partly due to the availability of new large-scale image datasets such as ImageNet [3], and also due to the use of convolutional neural networks [4]. Identifying the objects present in an image could be used to many industries, like, generating image captions that consist of full sentences that describe the contents of an image [2] and visual question answering [5] or using in healthcare industry. But one limitation is that all these image classification applications focus on the English language.

**Image Classification for Arabic** Although image classification research has focused primarily on English and other Latin languages there are several remotely related works existing; for example, in one paper, the authors attempted to create a method that recognizes Arabic text in images [6].

## 3. Methodology

This studys procedure is in Figure 1.

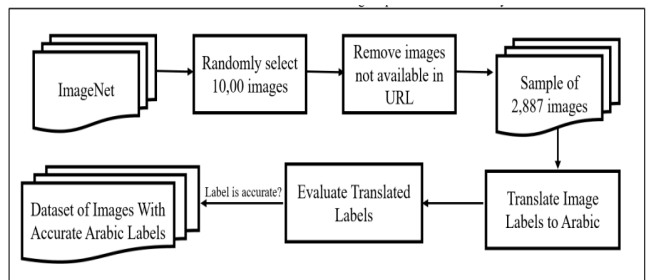


Figure 1. Overview of the methodological procedures for this study.

Table 1. The dataset and the number of images for each class

Class	Number of Images
Unigrams	1825
Bigrams	933
Ngrams	69
Total	2887

Table 2. Demographic Prediction performance comparison by three evaluation metrics.

Label	Number of Images			
	Unigrams	Bigrams	F1-Ngrams	Total
Accurate	1288 (71% )	576 (58%)	32 (45%)	1,895 (65.6%)
Inaccurate	355 (19% )	343 (34.5%)	30 (43%)	728 (25.2%)
Neutral	45 (2.5% )	26 (2.6%)	3 (4%)	74 (2.56%)
English	137 (7.5% )	48 (4.8%)	5 (7.2%)	190 (6.5%)
Total	1825	993	69	2887

## 4. Experiment

Author divided the dataset to three smaller sections based on the number of words in the object’s name to investigate if the textual structure of the labels has various effects on the performance of using a translation service to translate labels of images from English to Arabic. The first class of “unigrams” included labels that consist of only one word; the second class of “bigrams” included objects that consist of two words; the third class of “ngrams” included objects that consist of three or more words. The data we can see in table 4.

## 5. Results and Discussion

Results from the evaluation process showed that translated Arabic labels for 1,895 out of the 2,887 images were accurate. In other words, 65.6% of the image labels that were accurate. The result we can see from Table 2.

## References

- [1] Z. Akata, F. Perronnin, Z. Harchaoui, and C. Schmid. Good practice in large-scale learning for image classification. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 36(3):507–520, 2014. 1
- [2] S. Antol, A. Agrawal, J. Lu, M. Mitchell, D. Batra, C. Zitnick, and D. Parikh. Vqa: Visual question answering. in international conference on computer vision. *ICCV*, pages 425–2433, 2015. 1
- [3] J. Deng, W. Dong, R. Socher, L. Li, K. Li, and L. Fei-fei. Imagenet: A large-scale hierarchical image database. *IEEE Computer Vision and Pattern Recognition (CVPR)*, pages 2–9, 2009. 1
- [4] J. Huang, V. Rathod, C. Sun, M. Zhu, and a. K. M. A. Korrattikara, and A. Fathi. Speed/accuracy trade-offs for modern convolutional object detectors. *CVPR.IEEE*, pages 7310–7319, 2017. 1
- [5] A. Karpathy and L. Fei-fei. Deep visual-semantic alignments for generating image descriptions. *IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE*, 39(4):664–676, 2017. 1
- [6] F. Slimane, S. Kanoun, J. Hennebert, A. Alimi, and R. Ingold. A study on font-family and font-size recognition applied to arabic word images at ultra-low resolution. *Pattern Recognition Letters*, 34(2):209–218, 2013. 1