

Extending ImageNet to Arabic using Arabic WordNet

Abdulkareem Alsudais

College of Computer Engineering and Sciences, Prince Sattam bin Abdulaziz University
Al-Kharj, Saudi Arabia
A.Alsudais@psau.edu.sa

Abstract

ImageNet has millions of images that are labeled with English WordNet synsets. This paper investigates the extension of ImageNet to Arabic using Arabic WordNet. The objective is to discover if Arabic synsets can be found for synsets used in ImageNet. The primary finding is the identification of Arabic synsets for 1,219 of the 21,841 synsets used in ImageNet, which represents 1.1 million images. By leveraging the parent-child structure of synsets in ImageNet, this dataset is extended to 10,462 synsets (and 7.1 million images) that have an Arabic label, which is either a match or a direct hypernym, and to 17,438 synsets (and 11 million images) when a hypernym of a hypernym is included. When all hypernyms for a node are considered, an Arabic synset is found for all but four synsets. This represents the major contribution of this work: a dataset of images that have Arabic labels for 99.9% of the images in ImageNet.

1 Introduction

ImageNet is a dataset comprised of 14 million images (Deng et al., 2009; Russakovsky et al., 2015). Each image in the dataset is labeled with a WordNet (Miller, 1995) synset representing the identifying object in the image. The fall 2011 release of the dataset has a total of 21,841 unique synsets that are used to label images. The dataset is organized by dividing these synsets into several major subtrees. Moreover, ImageNet is structured in a way that maintains the semantic hierarchical structure of synsets in WordNet, where each image is also linked to branches of hypernyms (Figure 1). ImageNet is one major reason for recent advances in computer vision research and deep learning (Cetinic et al., 2018; Stock and Cisse, 2018; Kornblith et al., 2019).



Synset in ImageNet	Arabic Synset from AWN
0. Siberian husky	<i>not available</i>
1. Sled dog: Direct hypernym for “siberian husky”.	<i>not available</i>
2. Working dog: Hypernym for “sled dog”.	كلب عامل، كلب الشغل
3. Dog: Hypernym for “working dog”.	كلب

Figure 1: Images in ImageNet for the synset “Siberian husky”. Although an Arabic synset from AWN is not available for the synset or its direct hypernym, one is available for the hypernym of the hypernym.

While computer vision research has seen significant progress in recent years, the focus has been on English. Limited work exists to extend research to other languages, including Arabic. This lack of research may present challenges to scientists, researchers, and practitioners who seek to address problems related to computer vision in Arabic. Furthermore, the unavailability of a large dataset of images labeled in Arabic may prevent the development of solutions that address challenging tasks, such as visual question answering and image classification in Arabic. Therefore, a large dataset of images labeled with Arabic has the potential to progress research in Arabic computer vision. Moreover, scholars studying Arabic natural language processing often develop methods specifically designed for Arabic. Thus, it is possible that similarly unique methods are needed for Arabic computer vision.

The primary objective of this paper is to investigate the effectiveness of extending ImageNet to Arabic using Arabic WordNet (AWN) by searching in AWN for all the synsets used in ImageNet. AWN was originally developed in 2006 (Black et al., 2006). Since then, several authors have attempted to extend it by improving its coverage or quality (Alkhalifa and Rodríguez, 2009; Abouenour et al., 2013; Bond and Foster, 2013; Regragui et al., 2016; Batita et al., 2019). The possibility of using AWN to extend ImageNet has been experimented with in one paper (Alsudais, 2019). In the paper, the author tested using AWN to find Arabic synsets for a small sample of 100 images from ImageNet and indicated that Arabic synsets were found for only six synsets. However, the author did not attempt to discover if Arabic synsets were available for hypernyms of these synsets. This paper attempts to overcome the problems of limited availability for direct matches by also searching branches of hypernyms in AWN. In summary, this paper makes three major contributions:

- It investigates the possibility of extending ImageNet to Arabic using the Arabic WordNet (AWN).
- It adds to the limited work in computer vision research in Arabic.
- It generates a new, large dataset of images with Arabic labels. This dataset includes at least one label for all but four of the 21,841 synsets used in ImageNet.

ImageNet has been used to solve tasks in the intersection of language and vision research (Zhou et al., 2018; Chen et al., 2019; Davis et al., 2019; Vempala and Preot, 2019). For Arabic computer vision, limited related work currently exists. Several authors have worked on the generation of Arabic captions for images (Jindal, 2018; Al-muzaini et al., 2018). In another paper, a new dataset related to Arabic computer vision was built. The authors constructed a dataset of 3,000 clips that they classified with emotional labels such as “happy”, “sad”, or “angry” (Shaqra et al., 2019). The authors argued that emotional facial expressions may be different depending on the cultural context. In other papers, attempts to connect ImageNet to external resources were made by linking the synsets in ImageNet to items in Wikidata (Nielsen, 2018) and by extending a

sample of images in ImageNet to German using human subjects, which resulted in a dataset of 1,305,602 images (Roller and Schulte, 2013). In the only other closely related paper, the author investigated the possibility of generating Arabic labels for images in ImageNet using an online translator (Alsudais, 2019). In the paper, the author targeted a sample of 1,895 images from ImageNet and used an online translator to generate Arabic labels for the synsets. A human judge then evaluated the accuracy of the translations. The results indicated that the translations were accurate for 65% of the images, which represented 1,643 unique synsets and 1,910,935 images. This suggests that solely using a translator to translate labels of images in ImageNet may not produce highly accurate results.

2 Extending ImageNet to Arabic using Arabic WordNet

In ImageNet, each synset has a name and an ID. To begin exploring the possibility of finding Arabic synsets and labels for images in ImageNet using AWN, all the synsets’ IDs are retrieved from ImageNet. There are several releases for ImageNet. In this paper, the fall 2011 release is used. This release includes 21,841 unique WordNet synsets, and each is linked to one or many images. For example, the synset ID “n07873807” includes 1,296 images of “pizza”. There are also 1,186 images for “dish”, the direct hypernym for “pizza”. Moreover, all the images labeled with “pizza” can also be labeled with “dish”. Since “dish” is a hypernym for several other synsets used in ImageNet, such as “sushi” and “curry”, the number of images for “dish” extends to all images with a synset that is a hyponym (child) for “dish”. ImageNet’s data are downloaded directly from ImageNet’s website¹. ImageNet provides the URLs of the images. These URLs are viewed in order to access the images. Due to ongoing developments related to the removal of problematic images present in the “person” subtree, ImageNet no longer provides a method to download the full dataset directly (Yang et al., 2020). The dataset has a total of 14,197,122 images. Each synset has an average of 944 images directly assigned to the synset. The minimum number of images is 1 image, and the maximum is 2,382 images for a synset.

¹ <http://www.image-net.org/download.php>

2.1 Direct Arabic Synsets for Synsets in ImageNet

The first step is to investigate if Arabic synsets are available for each of the 21,841 synsets used in ImageNet. To complete this, all the synsets IDs are processed. For each synset, AWN is searched to find a direct match. There are several versions of AWN and several methods to access it currently exist. To gain additional knowledge on WordNet and AWN, and to determine a reliable method to access it, the online interfaces for both the Open Multilingual WordNet (OMW)² (Bond and Paik, 2012) and the Princeton WordNet³ (Princeton University, 2010) are tested. Additionally, the WordNet interface in the python library NLTK⁴ is tested. The interface includes the OMW, which has AWN (Black et al., 2006; Abouenour et al., 2013). This version of AWN has 9,916 Arabic synsets, which is less than the number of synsets used in ImageNet. This is the first indicator that it may not be possible to find direct matches for all synsets. Still, it is not clear if it is possible for a synset in ImageNet to be directly linked to several synsets in AWN. Based on this experimental phase, the NLTK interface is selected to access Arabic synsets in AWN. ImageNet uses WordNet 3.0, which has synsets IDs that are different than one used in WordNet 3.1. Therefore, the results in this paper are necessarily achieved by using WordNet 3.0.

2.2 Arabic Synsets for Hypernyms

Since ImageNet structures synsets based on the semantic structure of synsets in WordNet, a synset in ImageNet is essentially a node that is connected to a branch or several branches of hypernyms. The objective of this step is to discover if an Arabic synset in AWN is available for the list of hypernyms linked to a synset. To accomplish this, the parent-child (or hypernym-hyponym) pairs in ImageNet are downloaded from a webpage in ImageNet's website⁵. Algorithm 1 includes details of the steps followed in order to find direct matches as well as Arabic hypernyms for synsets. The algorithm relies on the use of a recursive function that looks for an Arabic synset for all the hypernyms connected to the synset at all levels. The stopping condition for this recursive function is when all possible hypernyms are processed.

Algorithm 1: Finding direct AWN synsets for ImageNet's synsets as well as AWN synsets for all the hypernyms linked to ImageNet's synsets.

Input: list of all synsets in ImageNet. "Synset" below refers to the synset from ImageNet.

Output: 1) Direct_AWN: list of Arabic synsets from AWN that are directly linked to a synset in ImageNet, and 2) Hyper_AWN: a list of Arabic synsets from AWN that are linked to a hypernym of a synset in ImageNet.

```
1: For synset in ImageNet:
2:   If find_in_AWN (synset) is True then
3:     Direct_AWN.add (synset, AWN_synset)
4:     Find_Hypers (synset, synset, 1)
5: func Find_Hypers (synset, hyper_synset, level):
6:   Hypers=get_hypers (hyper_synset)
7:   If length (Hypers) == 0:
8:     Stop #no more hypernyms to process
9:   For hyper_synset in Hypers:
10:    If find_in_AWN (hyper_synset) is True then
11:      Hyper_AWN.add (synset,
12:                    hyper_AWN_synset, level)
13:    Find_Hypers (synset, hyper_synset, level+1)
```

To complete this process, all hypernyms of a synset are retrieved. In most cases, a synset that is a leaf node has one or two direct hypernyms. Each hypernym is searched for in AWN in order to discover if an Arabic synset for the hypernym exist in AWN. If one is found, it is added to the set of Arabic synsets for the primary synset. The level of the hypernym is also saved. For example, since an Arabic synset is available for the synset "dish", which happens to be a hypernym for "pizza", the Arabic synset for "dish" is saved for the "pizza" synset. Additionally, the number "one" is saved because "dish" is a direct hypernym for "pizza". Similarly, the number "two" is saved for "nutriment", which is the hypernym for "dish". If a synset has two direct hypernyms, they are both saved as appearing in level one. Hypernyms are only saved when an Arabic synset is found.

This step results in a dataset of all AWN synsets in ImageNet, as well as the Arabic synsets available at each level of their hypernyms. Since it is possible for a synset to have two hypernyms, the objective of searching hypernyms is to indicate if *any* Arabic synsets exist for *any* of the hypernyms. It is unclear if using hypernyms to label images in Arabic will produce images with acceptable and meaningful labels. Future work should investigate the quality of generated Arabic synsets.

² <http://compling.hss.ntu.edu.sg/omw/>

³ <http://wordnet-rdf.princeton.edu/>

⁴ <http://www.nltk.org/howto/wordnet.html>

⁵ http://www.image-net.org/archive/wordnet.is_a.txt

3 Results

3.1 Direct Arabic Synsets

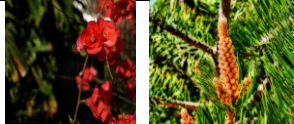
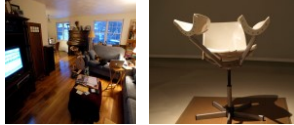
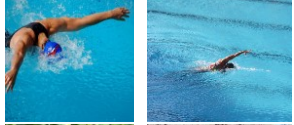

Images	ImageNet Synset	AWN Synset
	n00017222: plant.n.02	حياة نباتية، غرسة، نبات
	n03405725: furniture.n. 01	أثاث، قطعة أثاث
	n00442115: swimming. n.01	سباحة، عوم
	n13100156: poisonous_ plant.n.01	نبات سام

Table 1: Examples of images from ImageNet with their synsets and Arabic synsets, as found in AWN.

Direct matches were found for 1,219 of the 21,841 synsets used in ImageNet. Some of these identified synsets are of higher-level categories, while others are of fine-grained ones. Table 1 includes examples of images where an Arabic synset was found in AWN. The table also shows both the English and Arabic synsets. Since each synset is linked to many images, the dataset of 1,219 synsets was extended to 1,150,651 images, which is 8.1% of ImageNet’s total number of images. This dataset represents a major contribution of the paper, as it can be used in several tasks related to Arabic computer vision. Since all the labels are of direct matches, the quality of the labels should be high. However, further examination and full evaluations are needed for confirmation.

3.2 Arabic Synsets for Hypernyms

To expand the dataset, hypernyms of synsets used in ImageNet were searched for in AWN. The result of this extension was the identification of Arabic synsets for all but four synsets used in ImageNet. These four synsets include only 1,366 images. This indicates that there are only 1,366 images in ImageNet without Arabic synsets in AWN for the synset or one of its hypernyms in its branch of hypernyms. A detailed summary of the results is presented in Table 2. In the table, “AWN’s synsets” refers to the number of Arabic synsets found for a synset in ImageNet at each level. The “AWN’s synset + previous” refers to the total number of Arabic synsets identified when the synsets found at level and the previous levels are combined. The “Images in ImageNet” refers to the total number of images found for each Arabic synset at each level.

When only the first and second level hypernyms were considered, the dataset included Arabic synsets for 79.8% of the synsets and 81.2% of the images in ImageNet. This represents a large dataset of 11,533,525 images, all labeled with an Arabic synset that is either the direct match for the synset used in ImageNet, the Arabic synset for the hypernym, or the Arabic synset for the hypernym of a hypernym. Although a synset in this subset (Row #5 in Table 2) was found for 17,438 of the synsets used in ImageNet, many of the identified Arabic synsets were used more than once since the total number of synsets in AWN is only 9,916. It is important to note that as the level of the hypernym increases, the hypernym become more abstract and general. For example, some of the 7th-level hypernyms include “entity”, “act”, and “event”. Therefore, the usability of Arabic synsets at higher levels requires additional investigation.

Level	AWN’s Synsets	AWN’s Synsets + Previous	Images in ImageNet	Images + Previous
Direct	1,219 (5.58%)	-----	1,150,651 (8.1%)	-----
<i>No person subtree</i>	993 (5.22%)		990,189 (7.6%)	
First Hypernym	10,400 (47.6%)	10,462 (47.9%)	7,113,853 (50.1%)	7,177,537 (50.5%)
<i>No person subtree</i>	8,846 (46.5%)	8,890 (46.7%)	6,419,390 (49.3%)	6,469,802 (49.7%)
2nd Hypernym	16,837 (77.0%)	17,438 (79.8%)	11,088,519 (78.1%)	11,533,525 (81.2%)
<i>No person subtree</i>	14,541 (76.5%)	15,064 (79.2%)	10,099,744 (77.6%)	10,508,392 (80.7%)
3rd Hypernym	19,490 (89.2%)	20,267 (93.7%)	12,697,838 (89.4%)	13,266,956 (93.4%)
4th Hypernym	20,671 (94.6%)	21,397 (97.9%)	13,365,182 (94.1%)	13,916,531 (98.0%)
5th Hypernym	20,816 (95.3%)	21,751 (99.5%)	13,483,006 (94.9%)	14,148,669 (99.6%)
6th Hypernym	19,910 (91.1%)	21,830 (99.94%)	12,865,510 (90.6%)	14,195,325 (99.8%)
7th Hypernym	18,191 (83.2%)	21,837 (99.98%)	11,729,718 (82.6%)	14,195,756 (99.9%)

Table 2: Summary of the number of synsets and number of images found.

While this dataset of Arabic synsets for images in ImageNet is likely reliable since it is based on the utilization of previously existing and evaluated datasets, certain characteristic of the Arabic language and naming decisions in AWN suggest that proper evaluation of the dataset's accuracy may be needed. For example, one of the synsets found in AWN was "nurse". Unlike English, Arabic nouns are gendered. Accordingly, in AWN the Arabic synset for "nurse" is "male nurse". Therefore, an automated image classification system that relies on this synset may suggest "male nurse" for images of female nurses.

This "nurse" synset is part of the "person" subtree, which is one of the major subtrees in ImageNet. This subtree includes several synsets that have been criticized for issues such as representation biases and offensive images (Shankar et al., 2017; Mehrabi et al., 2019). During this study, it was observed that images in the "Iraqi" synset were mostly war related. Additional issues were discovered for images in the "Syrian" synset. Recently, some of the scientists behind ImageNet addressed concerns regarding issues in the "person" subtree and indicated that an upgrade of ImageNet will be released with two changes: 1) only up to 158 of the 2,832 synsets in the "person" subtree will be kept, and 2) attention will be given to representation biases in images in the synsets that are not removed (Yang et al., 2020). In anticipation for these updates, Table 2 also includes results obtained when the 2,832 synsets in the "person" subtree were not included. These results suggest only a minor decrease in percentages of synsets found in each level.

4 Conclusion

In this paper, the possibility of extending ImageNet to Arabic is investigated. The following discoveries were made: 1) an Arabic synset in AWN exists for 1,219 of the synsets used in ImageNet, which represents 1,150,651 images, and 2) Arabic synsets in AWN exist for 99.9% of the images in ImageNet when the branches of hypernyms for synsets were considered. To improve the results found, several options are available. One option is to use the Extended Open Multilingual Wordnet (1.2), which enhances Arabic WordNet by utilizing Wiktionary and Unicode Common Locale Data Repository (Bond and Foster, 2013). This automatic extension of AWN increases the total number of unique synsets

available in AWN to 14,650 synsets. Using this version of AWN would likely find Arabic synsets for additional images used in ImageNet. Several other directions for future work exist. One important extension is to provide an extensive evaluation of the dataset. Another avenue for further research would involve investigating the availability of Arabic synsets for each of the synsets in ImageNet's 1,000 object classes, as this subset is often used in computer vision research.

References

- Lahsen Abouenour, Karim Bouzoubaa, and Paolo Rosso. 2013. On the evaluation and improvement of Arabic WordNet coverage and usability. *Language Resources and Evaluation*, 47:891–917.
- Musa Alkhalifa and Horacio Rodríguez. 2009. Automatically extending NE coverage of Arabic WordNet using Wikipedia. In *The 3rd International Conference on Arabic Language Processing CITALA2009.*, pages 20–36, Rabat, Morocco.
- Huda A Al-muzaini, Tasniem N Al-yahya, and Hafida Benhidour. 2018. Automatic Arabic Image Captioning using RNN-LSTM-Based Language Model and CNN. *International Journal of Advanced Computer Science and Applications*, 9(6):67–73.
- Abdulkareem Alsudais. 2019. Image Classification in Arabic: Exploring Direct English to Arabic Translations. *IEEE Access*, 7:122730–122739.
- Mohamed Ali Batita, Rami Ayadi, and Mounir Zrigui. 2019. Reasoning over Arabic WordNet Relations with Neural Tensor Network. *Computación y Sistemas*, 23(3):935–942.
- William Black, Sabri Elkateb, Horacio Rodriguez, Musa Alkhalifa, Piek Vossen, Adam Pease, and Christiane Fellbaum. 2006. Introducing the Arabic WordNet Project. In *Proceedings of the third international WordNet conference.*, pages 295–299.
- Francis Bond and Ryan Foster. 2013. Linking and Extending an Open Multilingual Wordnet. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics*, pages 1352–1362, Sofia, Bulgaria. Association for Computational Linguistics.
- Francis Bond and Kyonghee Paik. 2012. A Survey of WordNets and their Licenses. In *Proceedings of the 6th Global WordNet Conference (GWC 2012)*, pages 64–71.
- Eva Cetinic, Tomislav Lipic, and Sonja Grgic. 2018. Fine-tuning Convolutional Neural Networks for fine art classification. *Expert Systems With Applications*, 114:107–118.

- Zhenfang Chen, Lin Ma, Wenhan Luo, and Kwan-Yee K. Wong. 2019. Weakly-Supervised Spatio-Temporally Grounding Natural Sentence in Video. In Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, pages 1884–1894, Florence, Italy, July 28 - August 2, 2019. Association for Computational Linguistics.
- Christopher Davis, Luana Bulat, Anita Vero, and Ekaterina Shutova. 2019. Deconstructing multimodality: visual properties and visual context in human semantic processing. In Proceedings of the Eighth Joint Conference on Lexical and Computational Semantics (*SEM), pages 118–124, Minneapolis, June 6–7, 2019. Association for Computational Linguistics.
- Jia Deng, Wei Dong, Richard Socher, Li-jia Li, Kai Li, and Li Fei-fei. 2009. ImageNet: A large-scale hierarchical image database. In 2009 IEEE Conference on Computer Vision and Pattern Recognition, pages 248–255, Miami, FL.
- Vasu Jindal. 2018. Generating image captions in Arabic using root-word based recurrent neural networks and deep neural networks. In Proceedings of NAACL-HLT 2018: Student Research Workshop, pages 144–151.
- Simon Kornblith, Jonathon Shlens, and Quoc V Le. 2019. Do Better ImageNet Models Transfer Better? In Proceedings of the IEEE conference on computer vision and pattern recognition, pages 2661–2671.
- Ninareh Mehrabi, Fred Morstatter, Nripsuta Saxena, Kristina Lerman, and Aram Galstyan. 2019. A Survey on Bias and Fairness in Machine Learning. arXiv:1908.09635v2.
- George A Miller. 1995. WordNet: A Lexical Database for English. *Communications of the ACM*, 38(11):39–41.
- Finn Årup Nielsen. 2018. Linking ImageNet WordNet Synsets with Wikidata. In WWW '18 Companion: The 2018 Web Conference Companion, pages 1809–1814, Lyon, France.
- Princeton University. 2010. About WordNet. Princeton University, Princeton, NJ 08544, USA.
- Yasser Rezagui, Lahsen Abouenour, Fettoum Krieche, Karim Bouzoubaa, and Paolo Rosso. 2016. Arabic WordNet: New content and new applications Arabic WordNet. In Proceedings of the Eighth Global WordNet Conference, number January, pages 330–338.
- Stephen Roller and Sabine Schulte. 2013. A Multimodal LDA Model Integrating Textual, Cognitive and Visual Modalities. In Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing, number October, pages 1146–1157, Seattle, Washington, USA. Association for Computational Linguistics.
- Olga Russakovsky, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Sathesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, Aditya Khosla, Michael Bernstein, Alexander C Berg, and Li Fei-fei. 2015. ImageNet large scale visual recognition challenge. *International Journal of Computer Vision*, 115(3):211–252.
- Shreya Shankar, Yoni Halpern, Eric Breck, James Atwood, Jimbo Wilson, and D Sculley. 2017. No Classification without Representation: Assessing Geodiversity Issues in Open Data Sets for the Developing World. In NIPS 2017 workshop: Machine Learning for the Developing World.
- Ftoon Abu Shaqra, Rehab Duwairi, and Mahmoud Al-ayyoub. 2019. The Audio-Visual Arabic Dataset for Natural Emotions. In 2019 7th International Conference on Future Internet of Things and Cloud (FiCloud) The. IEEE.
- Pierre Stock and Moustapha Cisse. 2018. ConvNets and ImageNet Beyond Accuracy: Understanding Mistakes and Uncovering Biases. In ECCV (6), pages 504–519.
- Alakananda Vempala and Daniel Preot. 2019. Categorizing and Inferring the Relationship between the Text and Image of Twitter Posts. In Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, pages 2830–2840, Florence, Italy, July 28 - August 2, 2019. Association for Computational Linguistics.
- Kaiyu Yang, Klint Qinami, Li Fei-Fei, Jia Deng, and Olga Russakovsky. 2020. Towards Fairer Datasets: Filtering and Balancing the Distribution of the People Subtree in the ImageNet Hierarchy. In Conference on Fairness, Accountability, and Transparency (FAT* '20), pages 547–558, Barcelona, Spain. ACM.
- Mingyang Zhou, Runxiang Cheng, Yong Jae Lee, and Zhou Yu. 2018. A Visual Attention Grounding Neural Model for Multimodal Machine Translation. In Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, pages 3643–3653, Brussels, Belgium, October 31 - November 4, 2018. Association for Computational Linguistics.