CCT College Dublin

Assessment Cover Page

|  |  |
| --- | --- |
| Module Title: | Advanced Data Analytics  Big Data Storage and Processing |
| Assessment Title: | Integrated CA1 Sem 2 MSc in Data Analytics |
| Lecturer Name: | David McQuaid  Muhammad Iqbal |
| Student Full Name: | Mohamed Noor |
| Student Number: | SBS22070 |
| Assessment Due Date: | 03/10/2022 |
| Date of Submission: | 02/10/2022 |

Declaration

By submitting this assessment, I confirm that I have read the CCT policy on Academic Misconduct and understand the implications of submitting work that is not my own or does not appropriately reference material taken from a third party or other source. I declare it to be my own work and that all material from third parties has been appropriately referenced. I further confirm that this work has not previously been submitted for assessment by myself or someone else in CCT College Dublin or any other higher education institution.

Image Recognition for Learning Arabic Words

Mohamed Noor   
*Information and Communications Technology (ICT) Faculty   
CCT College Dublin*Dubln, Ireland  
sbs22070@student.cct.ie

*Abstract*—IEEE format requires you to include an abstract at the start of your paper, followed by a list of keywords. In the “Styles” section in Microsoft Word, you can find the appropriate styles for all the different sections and headings in the paper, which are already applied here. For example, the “abstract” style is applied to this text, the “keywords” style to the next section. Note that the titles “Abstract” and “Keyword” should remain as they are written here: italicized and followed by an em dash.

Keywords—image classification, Arabic, big data

# INTRODUCTION

Level 1 and 2 headings (as well as the paper title) should be written with title case capitalization, while level 3 and 4 headings are written in sentence case.

With more than 7,000 languages currently spoken [1], their study can be complex and biased by researchers’ preferences and the availability of funding. In general, languages can be grouped into families, of which there are more than 100 [2]. Arabic and English are spoken by about 360-400 million native speakers [3], [4] and are official United Nations languages [5]. However, English is a *lingua franca* through the prevalent inter-connected global economic structure [6]. Arabic is a part of the Afro-Asiatic family [7] while English is an Indo-European language [8], yet historical contact through a series of commerce and wars [9] has allowed for an intermingling (borrowing) of words. Arabic words, such as ‘sine’ and ‘alkali’ have penetrated English while words such as بلاستيك (‘bi-laastik’, plastic) have an English origin [10].

Unlike English, Arabic is an unusual language. For example, there are 200 words that refer to lion in Arabic [11]. The definite article ‘the’ in Arabic (‘al’, ال) never stands alone. Yet, when joined with another word will change its morphology (for example, lion (أسد) will become الأسد). There are also regional variations, which is why learners of Arabic usually tend to follow the Modern Standard Arabic [12].

Given the complexity of Arabic compared to English, it would be helpful for learners of Arabic, especially those who are not native, to have the ability to instantaneously convert images to Arabic words. The ubiquitous smartphones can play a significant role in helping these learners by snapping images of new objects that can be classified rapidly in Arabic.

Image recognition through deep learning is a particularly useful method for classifying images. Two types of solutions are relevant in the context of helping Arabic learners – (i) multi-class algorithms where each image is assumed to only take one label and the entire dataset consists of multiple labels, or (ii) a multi-label problem where each image can take multiple labels. Taking an image of a woman walking with a dog as an example, the image can be labelled either as a woman or a dog, but not both (multi-class), or as a woman and a dog (multi-label).

## Convolutional neural network (CNN)

CNN is a neural network that consists of one or more convolutional layer, non-linearity layer, pooling layer and finally fully-connected layer [13]. As the relative position of an object in an image is less relevant, the convolution layer allows for the extraction of features irrespective of object orientation. In fact, small datasets may be augmented by artificial shearing and rotation. Unlike recurrent neural networks, CNN is highly parallelisable especially when graphical processing units (GPUs) are used.

Image recognition is a well-studied domain of artificial intelligence with wide-ranging applications in real life. As such, complex pre-trained models have been published and are easily accessible within the TensorFlow ecosystem [14]. The use of pre-trained models hinges on the ability to adapt deep learning models across different use cases [15]. This ability, termed transfer learning, is based on how humans learn to identify new objects. A model trained on pictures of animals for a binary classification (‘is it an animal or not?’) can be used to classify the type of animal (‘what animal is it?’). Transfer learning will not be appropriate if the training dataset is completely different than the original dataset. When the datasets are similar, however, the use of pre-trained models can result in highly accurate models at a fraction of time needed for full training.

Another advantage of deep learning is that it requires no feature engineering. In the context of image classification, no annotation will be required to identify the legs or ears of an animal. This becomes useful when human cannot pre-identify features of high relevance to learning. Neural networks also lend itself to modular and complex architectures, including the ability for branching (multi-headed) training that can take advantage of distributed training (model parallelism). In this mode, one GPU may be used for one branch while another can be simultaneously used for the other branch, further reducing the total training time.

Several architectures of deep neural networks have been proposed for the use of image classification as summarised in TABLE 1. These include various MobileNet architectures [16] that originate from Google researchers, including MobileNetV2 [17] and MobileNetV3 Large [18]. These models are targeted for mobile and embedded devices where the computational power is more limited compared to personal computers. Although deeper and complex networks may achieve higher accuracy, they also require longer time and power for training. Through MobileNet architectures, the researchers aimed to create more efficient networks by using depthwise separable convolutions, which themselves are a combination of 3x3 depthwise convolution and 1x1 (pointwise) convolution [16]. MobileNetV2 builds on the original MobileNet work by including an inverted residual structure with shortcut connections between thin bottleneck layers [17]. Overall, the use of depthwise separable convolution significantly reduces computation – for instance, MobileNetV2 requires about 8 to 9 times less power than standard convolution [16]. The bottleneck layers accept a low-dimensional compressed representation, which is then expanded to a high dimension and filtered with depthwise convolution (Fig. 1).

In deep learning, it can be difficult to judge whether one architecture will be better than the other. The use of network architecture search (NAS) is essentially ‘a neural network to design a neural network’ as it automated the manual search for the optimal architecture [19]. MobileNetV3-Large is a product of NAS and network design, with the resulting increase of 3.2 % in accuracy on ImageNet classification with 20 % reduction of latency compared to MobileNetV2 [18].

Diagram

Description automatically generated

Fig. 1. Evolution of MobileNet networks through the use of separarable and bottleneck layers. Diagonal hatches indicate linear layers. The final (lightly coloured) layer is also the beginning of the next block. Upon stacking, the blocks in (c) and (d) are equivalent. Figure taken from [17].

Another family of relation architectures is the Inception network, including InceptionV3 [20] and InceptionResNetV2 [21]. This family is also aimed at simplifying network structure for its use in mobile and other low performant devices. The Inception networks arose from architectural considerations around convolutional layers as well as balancing of the network width and depth. In general, convolutions with spatial filters larger than 3x3 tend to be more computationally expensive. Nevertheless, they can capture long-range dependencies [20]. As a workaround, a multi-layer network with 3x3 convolutions was proposed with a resulting network depth of 42 layers for InceptionV2. InceptionResNetV2 extends the thinking further by incorporating residual connections, which was first introduced in [22] to reduce training time.

The NasNetMobile architecture [23] also uses NAS to design the best model where accuracy was increased by 1.2 % with a concomitant reduction in computation of 28 %. The NAS search space was limited to only the best cell structure instead of re-designing the entire architecture. Further limits were also placed on activation (ReLU only) and no batch normalisation/ReLU for depthwise separable convolution. The VGG19 (Visual Geometry Group 19 weight layers) [24] is one of the two models that take advantage of depth as the main driver for image recognition algorithms through the use of very small (3x3) spatial filters. In contrast, the Xception network [25] was propositioned as an interpretation of the Inception module, forming an intermediate between regular convolution and depthwise separable convolution. Both InceptionV3 and Xception have the same number of parameters but performance gains in the latter were achieved through a better use of parameters.

|  |  |  |
| --- | --- | --- |
| **Model** | **Number of parameters** | **Commentary** |
| InceptionV3 | 23.9M | Captures long-range relationships through deep small spatial filter layers, instead of large spatial filters. |
| InceptionResNetV2 | 55.9M | Incorporates residual connections to reduce training time. |
| MobileNetV2 | 3.5M | Adds bottleneck layers with separable layers. |
| MobileNetV3-Large | 1.4M | Used NAS to arrive at the optimal architecture. |
| NasNetMobile | 5.3M | Used optimised design space for NAS. |
| VGG19 | 143.7M | Deep CNN with 19 layers. |
| Xception | 22.9M | Improvement over InceptionV3. |

TABLE 1.

SUMMARY OF VARIOUS NEURAL NETWORK ARCHITECTURES FOR IMAGE CLASSIFICATION.

## Big image data

Success of deep learning research depends on the availability of large datasets. As shown in TABLE 1 above, these architectures consist of millions of trainable parameters, necessitating the use of large training datasets. Fortuitously, public image datasets are available, allowing for like-for-like comparison of various image classification algorithms. A popular dataset used for benchmarking is ImageNet,

ImageNet is XXX [ImageNet (image-net.org)](https://image-net.org/challenges/LSVRC/index.php) , a subset of which forms the ImageNet Large Scale Visual Recognition Challenge (ILSVRC) dataset. The full dataset contains XX images whereas the subset has YY images.

arXiv link: [[1409.0575] ImageNet Large Scale Visual Recognition Challenge (arxiv.org)](https://arxiv.org/abs/1409.0575)

<< some description of large image datasets, with a focus on ImageNet and MMID >>

<< image sources: Rationale for selecting the MMID – why this dataset and not the others???, ImageNet … >>

Was ImageNet human annotated and MMID was from Google Search, which itself is an algorithm that could be biased towards certain things. For example, searching for coffee could result in images of cups and mugs and not necessarily coffee powder.

MMID reference paper: [26]

## Study objectives

Learners of new languages are often advised to use flashcards that has an image of an object on one side and the word on the other [27]. To avoid confusion, multi-class algorithms may be of more benefit than multi-label for learners.

In this study, XXX

As a proof-of-concept, this paper discusses the use of pre-trained image classification algorithms for identifying images in Arabic. Pre-trained algorithms take advantage of the transfer learning concept in deep learning, where a model trained on a problem can be used in another problem if these are relatively similar. For instance, a model training on recognizing XXX can be used for YYY (give examples peer-reviewed).

## This Is a Level 2 Heading

### And this is a level 3 heading: Equations should be typed in either Times New Roman or Symbol font, or, if the equation is multileveled, inserted into your text as a graphic instead. On the far right of the line containing the equation, number it in parentheses, and use this number to refer to it in the text (1).

*a**b* 

### This is another level 3 heading: Lorem ipsum…

| Table Head | Table Column Head | | |
| --- | --- | --- | --- |
| Table column subhead | Subhead | Subhead |
| text | Texta |  |  |

#### And this is a level 4 heading: Make sure that the

1. This is a figure caption. It appears directly underneath the figure.

appropriate style is still applied to each section, reapplying styles if necessary.

#### This is another level 4 heading: It’s also possible to add bullet points when appropriate, using the “bullet list” style:

* Treat the word “data” as plural, not singular.
* For example, “the data indicate that …”

## This Is Another Level 2 Heading

A table heading (using the “table head” style) appears above a table. This will automatically number the table for you. Any footnotes appear below the table, using the “table footnote” style. Footnotes are indicated by superscript lowercase letters within the table. An example of a table can be seen in Table I, below.

# RESULTS

All the headings in the main body of your paper are numbered (automatically).

## Image selection

For this project, Python XX was used together with YYY.

## Programming environment

Python 3.9.12 was used for data exploration and deep learning. Given the size of the data, PySpark 3.3.0 was used. Other relevant packages are TensorFlow 2.10.0 (cite??), Pandas 1.5.0, Numpy 1.23.3 and Matplotlib 3.6.0.

An interactive Python application was created using Visual Studio Code as an IDE through the Dash 2.6.2 and Plotly 5.10.0 packages.

## Exploratory data analysis

Images from public sources can be expected to vary in their dimensions. As CNN models can only deal with one input size, it is important to understand the distribution of image heights and widths. Some images can also be represented by symbols. For example, ‘+’ images 🡪 give some examples.

<<example of images: grid of 3x3>>

<< example of images of symbols >>

<< plots of image dimensions >>

|  |  |
| --- | --- |
| **Symbol** | **Example images** |
| $ | A picture containing text, clipart  Description automatically generatedA golden trophy with a white background  Description automatically generated with low confidenceA picture containing diagram  Description automatically generated |
| + | Icon  Description automatically generatedIcon  Description automatically generated |
| = | Graphical user interface  Description automatically generatedTable  Description automatically generatedText  Description automatically generated with medium confidenceGraphical user interface, application, table  Description automatically generated |
| ᵒ | Qr code  Description automatically generatedA picture containing tool, brush  Description automatically generated |

TABLE I

EXAMPLE IMAGES THAT ARE LABELLED AS SYMBOLS INSTEAD OF WORDS.

1. This Is the Heading for a Table
2. This is a table footnote.

Lorem ipsum….

##### Acknowledgments

The author is grateful for the provision of MMID dataset by Amazon Web Services for no charge.

##### References

[1] Eberhard, David M, Gary F. Simons, and C. D. Fennig, Eds., *Ethnologue: Languages of the World. Online version*, Twenty-Fifth edition. Dallas, Texas: SIL International, 2022. [Online]. Available: http://www.ethnologue.com.

[2] S. Malik-Moraleda *et al.*, “An investigation across 45 languages and 12 language families reveals a universal language network,” *Nat. Neurosci.*, vol. 25, no. 8, Art. no. 8, Aug. 2022, doi: 10.1038/s41593-022-01114-5.

[3] A. S. Kaye, “Arabic,” in *The World’s Major Languages*, 3rd ed., Routledge, 2018.

[4] “English Speaking Countries List | Lingoda Online English Language School,” *Lingoda*. https://www.lingoda.com/en/content/english-speaking-countries/ (accessed Oct. 01, 2022).

[5] U. Nations, “Official Languages,” *United Nations*. https://www.un.org/en/our-work/official-languages (accessed Oct. 01, 2022).

[6] B. Seidlhofer, “Language Variation and Change: The Case of English as a Lingua Franca,” in *English Pronunciation Models: A Changing Scene*, 2nd edition., K. Dziubalska-Kolaczyk and J. Przedlacka, Eds. Verlag Peter Lang, 2008.

[7] A. Howell *et al.*, “(No) variation in the grammar of alternatives,” *Linguist. Var.*, vol. 22, no. 1, pp. 1–77, Jan. 2022, doi: 10.1075/lv.19010.how.

[8] R. Guarasci, S. Silvestri, G. De Pietro, H. Fujita, and M. Esposito, “BERT syntactic transfer: A computational experiment on Italian, French and English languages,” *Comput. Speech Lang.*, vol. 71, p. 101261, Jan. 2022, doi: 10.1016/j.csl.2021.101261.

[9] A. Hashim, G. Leitner, and M. A. Aqad, “Arabic in contact with English in Asia: Linguistic, social and political influences of Arabic in the region,” *Engl. Today*, vol. 33, no. 1, pp. 25–32, Mar. 2017, doi: 10.1017/S0266078416000377.

[10] Y. Bader, “Semantic change in Arabic loanwords from English and French,” *Abhath Al-Yarmouk*, vol. 8, no. 2, pp. 33–48, 1998.

[11] P. Zemánek, “Two Hundred Ways to Call a Lion in Arabic: Names or Epithets?,” *Z. Dtsch. Morgenländischen Ges.*, vol. 171, no. 2, pp. 343–374, 2021, doi: 10.13173/zeitdeutmorggese.171.2.0343.

[12] A. Mahmoud, “Modern Standard Arabic vs. Non-Standard Arabic: Where Do Arab Students of EFL Transfer From?,” *Lang. Cult. Curric.*, vol. 13, no. 2, pp. 126–136, Jul. 2000, doi: 10.1080/07908310008666594.

[13] S. Albawi, T. A. Mohammed, and S. Al-Zawi, “Understanding of a convolutional neural network,” in *2017 International Conference on Engineering and Technology (ICET)*, Aug. 2017, pp. 1–6. doi: 10.1109/ICEngTechnol.2017.8308186.

[14] S. A. Sanchez, H. J. Romero, and A. D. Morales, “A review: Comparison of performance metrics of pretrained models for object detection using the TensorFlow framework,” *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 844, p. 012024, Jun. 2020, doi: 10.1088/1757-899X/844/1/012024.

[15] K. You, Y. Liu, J. Wang, and M. Long, “LogME: Practical Assessment of Pre-trained Models for Transfer Learning,” in *Proceedings of the 38th International Conference on Machine Learning*, Jul. 2021, pp. 12133–12143. Accessed: Oct. 01, 2022. [Online]. Available: https://proceedings.mlr.press/v139/you21b.html

[16] A. G. Howard *et al.*, “MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications.” arXiv, Apr. 16, 2017. doi: 10.48550/arXiv.1704.04861.

[17] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, “MobileNetV2: Inverted Residuals and Linear Bottlenecks.” arXiv, Mar. 21, 2019. doi: 10.48550/arXiv.1801.04381.

[18] A. Howard *et al.*, “Searching for MobileNetV3.” arXiv, Nov. 20, 2019. Accessed: Oct. 01, 2022. [Online]. Available: http://arxiv.org/abs/1905.02244

[19] T. Elsken, J. H. Metzen, and F. Hutter, “Neural Architecture Search: A Survey,” *J. Mach. Learn. Res.*, vol. 20, no. 55, pp. 1–21, 2019.

[20] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, “Rethinking the Inception Architecture for Computer Vision.” arXiv, Dec. 11, 2015. doi: 10.48550/arXiv.1512.00567.

[21] C. Szegedy, S. Ioffe, V. Vanhoucke, and A. Alemi, “Inception-v4, Inception-ResNet and the Impact of Residual Connections on Learning.” arXiv, Aug. 23, 2016. doi: 10.48550/arXiv.1602.07261.

[22] K. He, X. Zhang, S. Ren, and J. Sun, “Deep Residual Learning for Image Recognition.” arXiv, Dec. 10, 2015. doi: 10.48550/arXiv.1512.03385.

[23] B. Zoph, V. Vasudevan, J. Shlens, and Q. V. Le, “Learning Transferable Architectures for Scalable Image Recognition.” arXiv, Apr. 11, 2018. doi: 10.48550/arXiv.1707.07012.

[24] K. Simonyan and A. Zisserman, “Very Deep Convolutional Networks for Large-Scale Image Recognition.” arXiv, Apr. 10, 2015. doi: 10.48550/arXiv.1409.1556.

[25] F. Chollet, “Xception: Deep Learning with Depthwise Separable Convolutions.” arXiv, Apr. 04, 2017. doi: 10.48550/arXiv.1610.02357.

[26] J. Hewitt, D. Ippolito, B. Callahan, R. Kriz, D. T. Wijaya, and C. Callison-Burch, “Learning Translations via Images with a Massively Multilingual Image Dataset,” in *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, Melbourne, Australia, Jul. 2018, pp. 2566–2576. doi: 10.18653/v1/P18-1239.

[27] S. Rosenbloom, “Want to Learn French? Italian? Russian? There’s No Time Like the Present,” *The New York Times*, Apr. 28, 2020. Accessed: Oct. 01, 2022. [Online]. Available: https://www.nytimes.com/2020/04/28/travel/language-instruction-apps-television-youtube.html