2021 International Nuclear Atlantic Conference – INAC 2021
Virtual meeting, Brazil, November 29 – December 2, 2021

ISBN: 978-65-594-1256-3

IMAGE CAPTIONS GENERATION FOR NON-PROLIFERATION CONTROL WITH ARTIFICIAL INTELLIGENCE

André Luis Ferreira Marques[1]

[1]*alfmarques@usp.br, Universidade de São*
Paulo – EPUSP- Engenharia de Computação
Av. Prof. Luciano Gualberto, 380 - Butantã,
São Paulo - SP, 05508-010

1. Introduction

The image or photograph description using Artificial Intelligence (AI) makes part of the Natural Language Processing (NLP), and it has gathered more importance as the computational means get stronger and specialized, with the Recurrent Neural Networks (RNN) and Convolutional Neural Networks (CNN) [1]. Applications can be seen on journalism, educational textbooks, web services, among others. A recent frontier focuses the use of this digital feature to help blind people, with the application of the Braille code with Deep Learning (DL) [2]. Equally important, NLP techniques play an advance contribution in system automation, supporting human activities [3].

This work presents an AI application to the nuclear non-proliferation control, aiding to identify/detect ‘dual use’ or specific items, in ports, airports and borderline customs systems, in man computer interfaces. The paper focuses the text generation from an image, based on selected pictures. Although the control agents have a specific recognition/identification skill, after training, the idea here is to aid them to confirm what it may be found during inspections, with reference to the nuclear non-proliferation, as a hint about the hardware. Moreover, as an initial hypothesis, the control agents may not sustain the suitable training for this task. The software proposed has the technical requirement to run in a standard laptop personal computer CPU, with an output within 30 seconds. Therefore, following a cheap pathway, the neural network application uses Python code, a ‘Jupyter’ notebook and Anaconda environment, based on [4].

2. Methodology

The overall software structure follows a format used in Data Science (DS) applications. The first branch loads up the math software libraries to be processed, which deal with the RNN and CNN models and their layers, embedding and dropout functions, Keras and Tensorflow frameworks, among others. In this work, the Long Short-Term Memory (LSTM) was used for text generation, being a type of RNN, recommended when it is key to retain memory knowledge to associate items within several categories.

The VGG16 neural network architecture was also used, due to its simple configuration, applied to image classification, object detection, general classification, image super-resolution, and it is already installed into the Keras package. The VGG16 network has been used for image detection and classification, and it can work with CPU and GPU as well, requiring at least 533 MB [5]. Figure 1 presents the overall scheme of this neural network, while handling the mathematical steps to process the image captured by a digital camera or surveillance system. In brief words, from the left to the right, the digital image goes throughout a set of math matrices operations, according to the neural network configuration, transforming digital pixels into numeric arrays. The numbers indicate the size of the matrices considered in the present model.

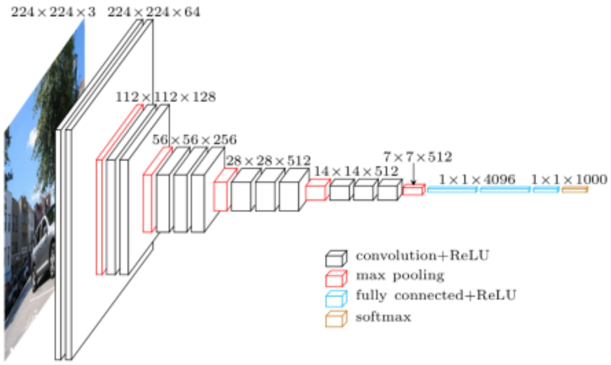


Figure 1: VGG16 neural network general configuration.

The input dataset to train and test the neural network was based on open-source photographs from the world wide web. Differently from other datasets, such as Flickr8k and ImageNet [6], which have thousands/millions of pictures about many subjects/categories, we chose to build up the input file, because no free access reference has been available. Therefore, a set of 72 pictures were picked up about: hexafluorine cylinders for enrichment process, ultracentrifuge technologies, mechanical and electronic bearings, uranium pellets for nuclear power reactors, nuclear fuel pins for pressurized water reactors (PWR), nuclear fuel bundles for CANDU reactors, vacuum pumps, and hexafluorine cylinder cannisters. The pictures have the same kind of format (.jpg) but with different configurations in terms of pixels, dimensions, and description attributes. The size of the input image dataset is 3.8 MB, and it may be enlarged for future software improvements.

Associated to each picture, there are five captions describing what can be identified technically: the hardware identification with standard industry names, or labels, number of items, the hardware colors, the presence of operators, the order of magnitude of dimensions, the type of materials involved and others. The quality of captions generated by the neural network comes from the accuracy of the terms associated to the images. The more diverse/varied the input captions, the more suitable the output can be, due to the math operations based on Bayesian techniques [7]. Table 1 presents an example of the associated captions to an image number.

Table 1 – Example of associated captions

Image number	Captions #	Captions
30051.jpg	0	Nuclear power reactor fuel element
30051.jpg	1	Nuclear fuel pin element for power reactor
30051.jpg	2	Metal nuclear fuel element made with pins
30051.jpg	3	Metal spacers and grids for nuclear fuel pins
30051.jpg	4	Metal hardware for nuclear power reactor core

The input text needs to be pre-processed to extract typing details worthless to the captions generation, like special characters (e.g., @, # etc.), word spacing, capital letters and others. After that, the words go into an embedding process to be transformed into math vectors linked to the digital image. Due to the initial requirements above, the maximum length for the description was bounded by 80 characters.

Deep learning neural network learn to correlate a set of inputs to a set of outputs, based on training data, making then connections, based on descendent gradients, while the construction of the data links. The 'categorical cross entropy' was set as the evaluation index for the neural network application, and the lower this value, the better the computer model [8]. In a short way, with two possible options, the index reveals how close two discrete probabilities are from each other. In this work, two types of activation functions were used: 'Rectified Linear', or ReLU, and 'Softmax'. The option for ReLU focused the simplicity of the model, in the start and middle sectors, training easily, and with good outcomes. The 'Softmax' activation function was selected for the final stages of the neural network because its link to the 'categorical cross entropy' index.

The 'Bilingual Evaluation Understudy' (BLEU) score was selected to compare the generated caption against the input ones, where the higher the score, the better the text made, within 0 and 1, where this latter result is hard to achieve, meaning a perfect significance [9].

The test/training threshold was set as ¼ following common practices in DS. In the training phase of the neural network, the number of 'epochs', or neural network cycles, was evaluated. The larger the number of the epochs, the better the captions generated, but requiring more computing time. Initially, the number of epochs was set as 5, taking around 5 minutes of computing. Other scenarios were also done, with 10 and 15 epochs, demanding up to 15 minutes roughly.

3. Results and Discussion

Table 2 presents the smallest values of the categorical cross entropy outcomes of the three scenarios of epochs. The best score came from the larger number of epochs, as expected, because the computer model had more training.

Table 2 – Categorical cross entropy loss function for the 3 scenarios

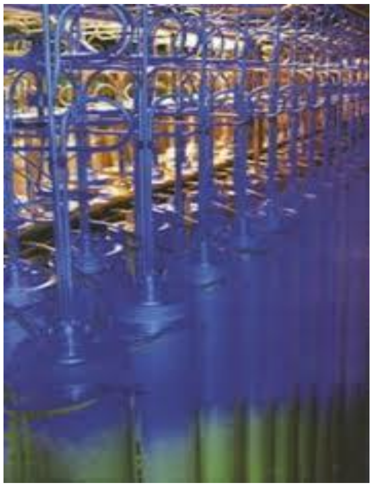
5 epochs	10 epochs	15 epochs
2.173	1.352	0.975

Table 3 summarizes the BLEU scores, and the order of magnitude looks compatible for this kind of application [10]. Scores close to 0.8 can be considered in the upper bound. As foreseen, the best result came from the larger number of epochs, due the same reason above.

Table 3 – BLEU scores for the 3 scenarios

5 epochs	10 epochs	15 epochs
0.347	0.563	0.742

One digital image was taken outside the input dataset, to validate the overall output. Figures 2 presents the photo and the generated captions, referring to ultracentrifuges for uranium enrichment [11]. The elapsed time for this task was 30 seconds approximately.



Uranium Enrichment Ultracentrifuge Industrial Aluminium

Figure 2: Example of captions output – uranium enrichment technology (image Urenco)

4. Conclusions

The neural network developed coped with the technical requirements and presented fair results. Figure 2 showed the generated captions coherent with the digital image. The model took around 30 seconds to generate the captions, which was considered acceptable for the items checking, under nuclear non-proliferation scope, running in a standard laptop CPU platform.

For future work, the results can be improved with the use of further NLP algorithms, such as "Attention" and "Bidirectional Encoder Representations from Transformers – BERT", and thus the computing tasks shall be revised [11,12]. The adoption of better computer hardware, such as GPU, will also boost the overall output. In addition, the dataset can be expanded to upgrade the learning process by the neural network, or to work in a more specific need.

References

[1] https://machinelearningmastery.com/develop-a-deep-learning-caption-generation-model-in-python/, accessed on Sept, 1st, 2021.

[2] Zaman, Sameia & Abrar, Mohammed Abid & Hassan, Mohammad & Islam, A N M Nafiul. (2019). A Recurrent Neural Network Approach to Image Captioning in Braille for Blind-Deaf People. https://www.researchgate.net/publication/344292153.

[3] J. P. S. Medeiros, A. C. da Cunha, A. M. Brito and P. S. Motta Pires, "Automating security tests for industrial automation devices using neural networks," 2007 IEEE Conference on Emerging Technologies and Factory Automation (EFTA 2007), 2007, pp. 772-775, doi: 10.1109/EFTA.2007.4416854.

[4] https://paperswithcode.com/method/vgg, accessed on Aug,26th, 2021.

[5] https://arxiv.org/pdf/1409.1556.pdf

[6] https://www.image-net.org/

[7] https://doi.org/10.1155/2020/3062706

[8] https://towardsdatascience.com/cross-entropy-for-classification-d98e7f974451, accessed on Aug, 26th, 2021.

[9] https://aclanthology.org/P02-1040.pdf

[10] https://towardsdatascience.com/foundations-of-nlp-explained-bleu-score-and-wer-metrics-1a5ba06d812b, accessed on Sept. 25th, 2021.

[11] https://www.world-nuclear.org/nuclear-essentials/how-is-uranium-made-into-nuclear-fuel., accessed on Aug, 26th, 2021.

[12] https://arxiv.org/pdf/1511.02793.pdf

[13] https://arxiv.org/pdf/1511.02793.pdf