Cloud Strategies for Image Recognition

Bernardo Botelho Antunes da Costa
Solvimm – Rio de Janeiro, RJ, Brazil
Email: bernardo.costa@solvimm.com

Pedro Silveira Pisa
Solvimm – Rio de Janeiro, RJ, Brazil
MídiaCom – Universidade Federal Fluminense (UFF) – Niterói, RJ, Brazil
Email: pedro.pisa@solvimm.com

***Abstract*—This paper presents two approaches for cloud applications with image recognition. Using a case study for labor safety in the field we compare the out-of-the-box services with the framework service from Amazon Web Services (AWS). We conclude that both recognize the helmet and the greater trade-off are the time to construct versus the assertiveness.**

I. Introduction

Image recognition is one of the most research problems in the Artificial intelligence (AI), especially because it deliveries results that enable a lot of business applications, such as in medical exam analysis and face authentication applications. The history of using AI to create an image-based application has been started between 1960 and 1970 when the mathematical theory about Neural Networks was started [1].

Computing capacity has grown following an exponential curve, while the cost and access to computing resources became cheaper and easy than ever. Initial investments in on-premise serves were replaced by the pay-as-you-go model created by Amazon Web Service (AWS) in 2006 and adopted by public cloud providers. All resources became accessible by almost everyone. Those things became deep learning accessible by daily-use applications [2].

II. Image Recognition

Image recognition is more than the classic problems usually appear on Internet (Digit Recognizer or Dogs vs. Cats). There are different approaches in which the techniques are used.

- Image Classification: We take a well-segmented image and classify it. Imagine a botanical garden wanted to build an application to enable the public to classify the pest on a plant. The audience takes a picture of a plant and the application says if the pest exists on the plant [3].
- Object Detection: We not only want to classify an object in an image but also to show where that object is. It as an extension of the above problem, in which the public now sends a photo of the vine and the application classifies each pest and mark where each one was found [4].
- Object Segmentation: The objective is to classify each pixel of the image, extracting the parts of the whole. In this technique, it is as if we take a colored pencil to outline and then paint the identified objects. With this type of technique, it is possible to delimit each part of the image. Still using the previous application, segmenting the objects means coloring where the pest was found [5].

III. Image Recognition in Public Cloud Providers

Considering Artificial Intelligence (AI) in the cloud, especially in the Amazon Web Services (AWS), Microsoft Azure, or Google Cloud Platform (GCP), we have two classes of services with a well-defined target audience. They are called by the AWS of AI Services and AI Platforms.

AI Services are cloud services that use AI under the hood requiring little or no understanding of AI theory from Developers. These services are very specific for solving a certain task and were designed to be used with API calls made directly to the specific service. Solutions for sentiment analysis in text, audio transcription, chatbot construction, and image/video recognition are some examples of AI services.

AI platforms are cloud services designed to be the working environment for Data Scientists. These services are best suited to be the place where Data Scientists build the entire AI life cycle in the cloud, which includes: data exploration, testing, model training, and deployment. Most AI Platform services are related to Amazon SageMaker, Azure Machine Learning Studio, or Google Cloud AI, protagonists of AI platforms.

That said, in the case of image recognition on AWS, AI service Amazon Rekognition are usually the first approach considered when dealing with problems with images and videos. AI services bring speed in the delivery of results, since we generally only need to consume the service's API, making it possible to reach satisfactory results very quickly, making this approach ideal for simple image recognition applications with Internet access.

When faced with more specific problems, AI services usually fail to perform well. Some scenarios require more specific techniques, such as to use a personalized neural network and applications with the requirement to operate embedded in places without an internet connection. In scenarios like these, using an AI Platform becomes the appropriate option.

In this paper, we tested both solutions using AWS AI technologies and compared the performance between them. In summary, some aspects to evaluate when deciding which approach to use are:

- Agility: if we need a proof of concept (PoC) or implement an initial solution as a starting point with less investment, using an AI service can be a great strategy;
- Customization: if the problem is more complex, we want a different algorithm, or we need to make finer adjustments and optimizations, a custom model starts to make more sense;

978-1-7281-9541-4/20/$31.00 ©2020 IEEE

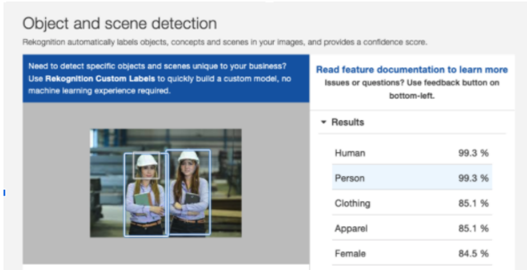

Fig. 1. Helmet identification with Amazon Rekognition.

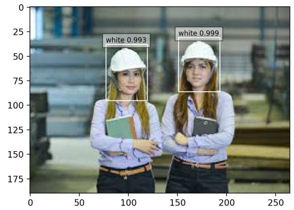

Fig. 2. Helmet identification in different position and colors.

- Internet access: if the location does not have constant Internet access, such as a new building in a remote location or an offshore operation on a platform or a ship, the use of AI platform is mandatory.

IV. Case Study

We take a classic problem in construction sites: the identification of safety helmets. This problem consists on ensure that workers are always using safety equipment when they are in the work. To build a solution, we have to define the approach used. Taking AWS as an example, a quick alternative to reach the first result and which requires almost none development is Amazon Rekognition. Using API calls for the service, it classifies the objects in the scene and returns the position of these elements in JSON format. If the return has a helmet with high confidence, as we can see on Figure 1, we know the worker has a safety equipment.

In this approach, the objects found are presented with the percentage of confidence of each one and their position. In this image, only one safety helmet was identified, with 73% confidence, instead of two. This approach maintains the standard AWS model, with no possibility of customizing the model.

Using AWS, to implement a custom model, we can use Amazon SageMaker, which is a managed service for creating, training, and deploying Machine Learning models. In this study case, a Neural Network was trained using the public dataset provided in the Wu et al. paper [6], whose authors benchmark some Deep Learning models to address the problem of detection of safety helmets. The dataset contains an image of several people, with and without a helmet, with 5 prediction classes: safety helmets in red, white, blue, yellow, and without a safety helmet. The results obtained for the same image can be seen on Figure 2. Although Amazon Rekognition did not identify the two safety helmets, the trained model responded with 99.3% and 99.9% confidence in the location of the helmets of the two women in the photo.

Both images on Figure 3 show environments where there are men without a safety helmet. In the images, none means that it was identified without a helmet. The men are on their sides, their backs, and some face them. The model must be able to interpret the safety helmets (or their absence) at different angles, with different positions, even with lower confidence, as was the case in the left image, where the man highlighted on the right had his helmet detected with the confidence of 69.3%, much lower than the others that exceeded 90% confidence. In situations like these, one of the main parameters used in these types of analyzes is the threshold. This parameter symbolizes the minimum percentage of confidence that we are willing to accept so that the prediction can be considered. For example, if the threshold is 70%, it indicates that predictions with a confidence rate below 70% will be ignored.

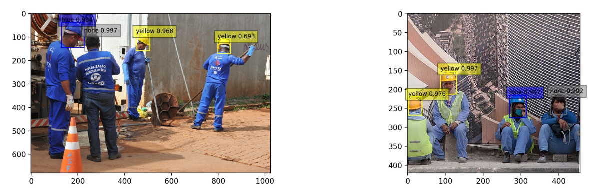

Fig. 3. Multiple people in different position and helmet colors.

V. Conclusion

There are several scenarios and approaches for the same challenge. When planning the solution, it is always important to clearly understand the objectives and the requirements to define the best strategy and select the appropriate tools. Also, testing the model with different angles, rotations and positions are very important for the model to be able to generalize sufficiently well the countless situations that can occur in a real environment. A good set of training data is key to making generalizations possible.

References

[1] J. Schmidhuber, "Deep learning in neural networks: An overview," *Neural networks*, vol. 61, pp. 85–117, 2015.

[2] J. Boulent, S. Foucher, J. Théau, and P.-L. St-Charles, "Convolutional neural networks for the automatic identification of plant diseases," *Frontiers in plant science*, vol. 10, 2019.

[3] D. C. Ciresan, U. Meier, J. Masci, L. M. Gambardella, and J. Schmidhuber, "Flexible, high performance convolutional neural networks for image classification," in *Twenty-second international joint conference on artificial intelligence*, 2011.

[4] C. Szegedy, A. Toshev, and D. Erhan, "Deep neural networks for object detection," in *Advances in neural information processing systems*, 2013, pp. 2553–2561.

[5] P. Voigtlaender and B. Leibe, "Online adaptation of convolutional neural networks for video object segmentation," *arXiv preprint arXiv:1706.09364*, 2017.

[6] J. Wu, N. Cai, W. Chen, H. Wang, and G. Wang, "Automatic detection of hardhats worn by construction personnel: A deep learning approach and benchmark dataset," *Automation in Construction*, vol. 106, p. 102894, 2019.