The Distinction between R-CNN and Fast R-CNN in Image Analysis:
A Performance Comparison

Maad M. Mijwil¹, Karan Aggarwal², Ruchi Doshi³, Kamal Kant Hiran⁴, and Murat Gök⁵

¹ Computer Techniques Engineering Department, Baghdad College of Economic Sciences University
Baghdad, Iraq
Corresponding author’s email: mr.maad.alnaimiy [AT] baghdadcollege.edu.iq

² Electronics and Communication Engineering Department, Maharishi Markandeshwar (Deemed to be University)
Mullana, Ambala, India
Email: karan.170987 [AT] gmail.com

³ Department of Computer Science and Engineering, Universidad Azteca
Chalco, Mexico
Email: ruchi.doshi [AT] univ-azteca.edu.mx

⁴ Department of Computer Science and Engineering, Sir Padampat Singhania University
Udaipur, India
Email: Kamal.hiran [AT] spsu.ac.in

⁵ Department of Computer Engineering, Yalova University
Yalova, Turkey
Email: murat.gok [AT] yalova.edu.tr

ABSTRACT— Deep learning techniques have become vital in many fields in the modern era because they are excellent at analysing and predicting real big data to act in different situations. Although it is marvellous in many aspects, it is prone to misinterpretation of data, so teams of experienced specialists cannot be dispensed with in following up on the execution stages of data analysis. Convolutional Neural Network is one of the most significant deep learning techniques. It is widely employed in visual image analysis. In this article, R-CNN and Fast R-CNN are summarised and compared and are the best in image analysis. This article concluded that the most suitable performance is for Fast R-CNN in testing and training.

Keywords— Artificial intelligence, Deep learning, Convolutional neural network, Artificial neural networks

1. INTRODUCTION

Image processing techniques or applications are utilised in many domains in military manufacturing, machine vision, monitoring and tracking (vehicle tracking) and many more [1-6]. Moreover, the medical sector is one of the sectors in which image processing techniques are utilised, such as chest x-rays, magnetic brain imaging, etc [7-11]. These techniques are utilised to eliminate the risk of misdiagnosis, which may lead to the death of the patient [12-17]. Thanks to artificial intelligence and the growth of algorithms and devices, the error rate of diagnosis and human intervention has been reduced in the modern era [18-25]. Deep learning is one of the essential techniques of artificial intelligence and a branch of machine learning. Many people ask what is the difference between machine learning and deep learning [26-32]. To clarify the difference between them, we must know that deep learning is part of machine learning techniques that have the ability to predict, diagnose and analyse data, but what distinguishes deep learning from other machine learning techniques is if one of the machine learning techniques is given a wrong prediction, the expert must intervene and make revisions [33-37]. In deep learning, the algorithm can determine on its own whether a prediction is accurate or not. Artificial intelligence techniques depend in their work on transferring a dataset to a computer, after which an analysis of the dataset is performed [38-41].

The execution of deep learning techniques can be used to classify a large set of images, no matter how complex, for instance, in a study issued by Seo et al. [42] in the use of hierarchical convolutional neural network technology to categorise images of a fashion set. Fashion images are among the most difficult images because fashions are a pack of colours, additions, different characteristics and depth of classification. In addition, deep learning techniques have been employed to analyse images of COVID-19 patients to a large extent, as these techniques have contributed to diagnosing
the condition of patients and determining the infection rate with the virus [43–48]. In recent years, many algorithms have been designed that contribute to analysing a wide range of complex medical images with satisfactory time periods [49–52]. One of the most famous algorithms is the convolutional neural network, which is one of the widely used and involved deep learning algorithms [53–55]. In other words, this algorithm depends on the number of database samples. The higher its execution and better results in training and testing the data.

2. CONVOLUTIONAL NEURAL NETWORK

A convolutional neural network, which is named CNN, is one of the most famous and widely accustomed algorithms that have the power to process, classify, and segment images. It is one of the deep learning techniques that bring an input image and assign weight to different objects in the same image, then distinguish objects from each other and elicit relationships between objects. Moreover, this algorithm displayed satisfactory performance in processing 2D data, such as video clips and images. The image is a two-dimensional array containing values from 0 to 255, where numbers near zero represent darker areas, and numbers near 255 represent lighter areas. To be clear, computers or other devices read the image based on its pixel values since each image is a combination of pixel values. Any modification in pixel values will also change the image data. In general, image classification is performed by detecting the presence of high-level features such as categories of people, objects, et [56–59]. In other words, if we can detect a nose, an eye, or a mouth, then it can be said that it is the image of a human's face. Figure 1 The public structure of a convolutional neural network with an example of inserting images containing an object (bird) and how the type of object in the object is determined.

Figure 1: The structure of Convolutional Neural Network with simple example.

Figure 2: deep learning is utilised to classify images of COVID-19 patients.

3. REGION-BASED CONVOLUTIONAL NEURAL NETWORK

We stated in the previous section that the image contains an object, whether a human or something, to determine the area of the object more accurately, using a region-based convolutional neural network algorithm [69–76]. The central concept of this algorithm is area proposals. In other words, this algorithm is employed to locate the object within the image. This algorithm is utilised to object localization, detection, and segmentation [74–81]. Although this algorithm is robust, it has drawbacks, including that it takes a huge amount of time to train the network as nearly 2000 area suggestions are sorted for each image and cannot be executed in real-time as it takes 47 seconds for each image. Figure 3 illustrates how this algorithm works.
4. FAST REGION-BASED CONVOLUTIONAL NEURAL NETWORK

An improved algorithm is developed for the R-CNN algorithm as a result of improving some errors in it, named the Fast R-CNN algorithm (See Figure 5 and 6). The enhanced algorithm depends on feeding the region suggestions to R-CNN. The improved algorithm function provides the input image to CNN to create a convolutional feature path through which the suggestions region is detected. Then it is squared using a grouping layer. Then, the image is assembled and shaped to a fixed size so that it can be inserted into a fully attached layer. Ultimately, softmax is used to estimate the proposed area class and bounding box displacement values. The main purpose of creating an improved algorithm for the R-CNN algorithm is that it does not actually need to feed 2000 area suggestions to the development neural network each time, but rather needs to perform the convolution only once for each image and a feature map is generated from it.

Figure 3: Object detection by R-CNN [82].

Figure 4: The structure of Fast R-CNN [Downloaded from MathWorks].

Figure 5: Object detection by Fast R-CNN [83].

Figure 6: The structure of Fast R-CNN [Downloaded from MathWorks].
5. THE EFFECTS

This article studies the methods of image analysis using deep learning techniques, as most of these techniques are based on artificial neural networks. Moreover, each level learns to transform the input data into a somewhat more abstract and syntax representation, which is the basis of the deep learning work. A convolutional neural network (CNN) is the most widely used class of deep neural networks applied to visual image analysis. CNNs use a variety of multi-layered visualisations designed to require minimal pre-processing. Also, CNNs use a variety of multitasking designed to reduce pre-processing. One of the problems with CNN is that to analyse an image, we may have to select several regions of that image, which will add to the computational workload. Therefore, R-CNN algorithms are designed to find and find these events quickly. Finally, instead of selecting a large number of image regions, the R-CNN method can pick a limited area of the image and use it using selective algorithms. The R-CNN method encountered several situations. It takes a lot of time to train the network, affects the computer's performance, and does not apply in real-time. The selective search algorithm is a fixed algorithm in many applications in visual image analysis. In other words, there is no learning in this matter, so the fastest R-CNN method is designed. Both R-CNN and Fast R-CNN utilise selective search to see location suggestions. Because selective search is a slow and time-consuming process, object identifier can be used instead of selective search. Table 1.2 and Figure 7 illustrates the performance of both R-CNN and Fast R-CNN. Table 2 indicates that the execution of both architectures is acceptable and satisfactory, and they can be used in image analysis and processing because they are positively efficient. Figure 8 illustrates the difference between the two architectures in the execution of image analysis.

Table 1: the performance between R-CNN and Fast R-CNN

<table>
<thead>
<tr>
<th>Performance Comparison</th>
<th>R-CNN</th>
<th>Fast R-CNN</th>
</tr>
</thead>
<tbody>
<tr>
<td>Test time per image</td>
<td>Less than a</td>
<td>Less than three</td>
</tr>
<tr>
<td>Speed</td>
<td>minute</td>
<td>seconds</td>
</tr>
<tr>
<td>1x</td>
<td>25x</td>
<td></td>
</tr>
</tbody>
</table>

![Figure 7: Training time and test time for both R-CNN and Fast R-CNN](image)

Table 2: Performance evaluation

<table>
<thead>
<tr>
<th></th>
<th>R-CNN</th>
<th>Fast R-CNN</th>
</tr>
</thead>
<tbody>
<tr>
<td>Performance</td>
<td>Very Good</td>
<td>Excellent</td>
</tr>
</tbody>
</table>

![Figure 8: The difference between the two architectures in the execution of image analysis](image)
6. CONCLUSIONS AND FUTURE WORK

The primary purpose of this article is to study the behaviour of R-CNN and Fast R-CNN and prove which is best for visual image analysis. This study demonstrated that Fast R-CNN is the fastest and best in image analysis and can be utilised in terms of speed and data training. In fact, there are many studies that use the original CNN in image analysis, especially medical images, because it is an important and valuable technique for all researchers and gives satisfactory results. In the future, more studies will be conducted on the original CNN and its application in the analysis of a range of medical images.

7. REFERENCES


