Project Document
Distributed Image Recognition System

Mengdan Wu

19121727@bjtu.edu.cn

Note:

This document can help you understand the functions and implementation methods of current company projects.

Abstract

After entering the 1990s, Internet / www has developed rapidly all over the world, providing a lot of convenience for users. However, there are also shortcomings. On the one hand, it provides users with a wealth of information resources and various means of communication, but on the other hand, it makes it very difficult for users to find and process the information they need. How to use the huge computing resources on Intemet reasonably and effectively has become an important issue that researchers pay close attention to.

In this experimental project, we have realized the distributed image recognition project through the distributed system. We have three parts of servers. The first server is used to process the requests sent by the front end (we will become the intermediate server below), and the other two servers are Alibaba cloud server and Huawei cloud server, which are used to process the pictures sent by the front end. Image and video for image recognition, marking and returning to the front-end server. Using the scheme of distributed system, we realize the learning of Yolo image recognition machine. The distributed system perfectly solves the front-end massive data request processing problem.

Table of Content

Introduction

In the face of huge information to be processed, how to process information quickly and effectively to meet the needs of users is an important problem. Huge information is not enough to be processed on one server. We can distribute information on multiple servers at the same time, which also embodies the advantages of distributed system.

We use YOLO, a new approach to object detection. Prior work on object detection repurposes classifiers to perform detection. Instead, we frame object detection as a regression problem to spatially separated bounding boxes and associated class probabilities. A single neural network predicts bounding boxes and class probabilities directly from full images in one evaluation. Since the whole detection pipeline is a single network, it can be optimized end-to-end directly on detection performance. Our unfied architecture is extremely fast. Our base YOLO model processes images in real-time at 45 frames per second. A smaller version of the network, Fast YOLO, processes an astounding 155 frames per second while still achieving double the mAP of other real-time detectors. Compared to state-of-the-art detection systems, YOLO makes more localization errors but is less likely to predict false positives on background. Finally, YOLO learns very general representations of objects. It outperforms other detection methods, including DPM and R-CNN, when generalizing from natural images to other domains like artwork.

Theory

Next, we will introduce the professional terms involved in this project.

(1) Distributed System

A distributed system is a system of computer nodes that communicate through a network and coordinate work to accomplish common tasks. The emergence of distributed systems is to use a cheap, ordinary machine to complete computing and storage tasks that cannot be completed by a single computer. The goal is to use more machines to process more data.

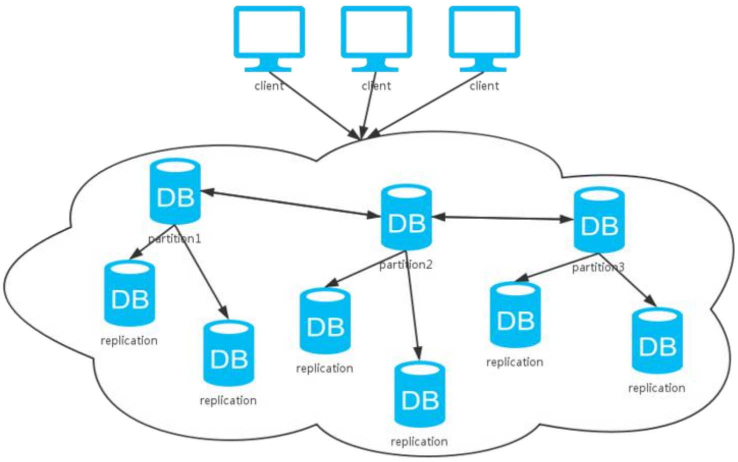

Figure 1: The Distributed System. Multiple clients, multiple servers processing user requests at the same time, Distributed operating processing system.

(2) Deep Learning

Deep learning is a direction in the field of machine learning. By learning the inherent laws and representation levels of sample data, machines can have analytical learning capabilities like humans, and can recognize text, images and sounds. The concept of deep learning comes from the research of artificial neural network. Multilayer perceptron with multiple hidden layers is a kind of deep learning structure. In-depth learning, by combining low-level features to form more abstract high-level representation of attribute categories or features, to discover the distributed feature representation of data. Figure 2 shows the structure of the neural network.

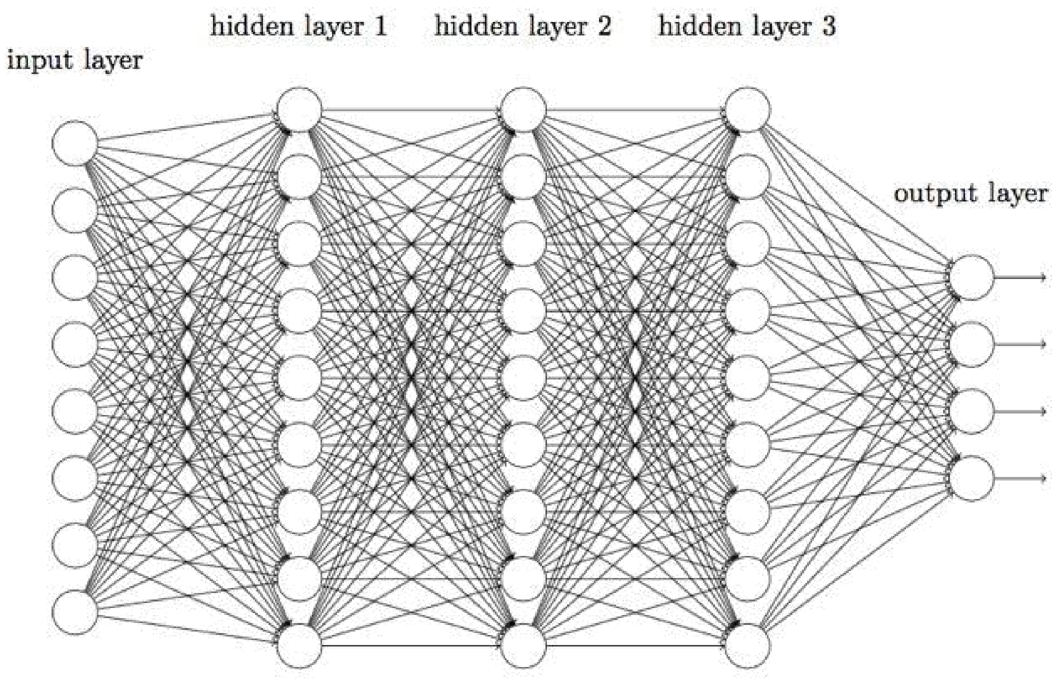

Figure 2: The Neural Network. Neural network has input layer, hidden layer and output layer, each layer has weight, threshold and activation function.

(3) Image Recognition

Image recognition is a technique that uses a computer to process, analyze, and understand images to identify targets and objects in various modes. Image recognition generally uses image acquisition, image preprocessing, and feature extraction for image recognition.

Image recognition, that is, to be able to recognize the objects in the picture just like people. We collect data samples, train data samples, extract feature values, let the machine learn and train, so that the machine can process different objects in the image. Image recognition present in Figure 3.

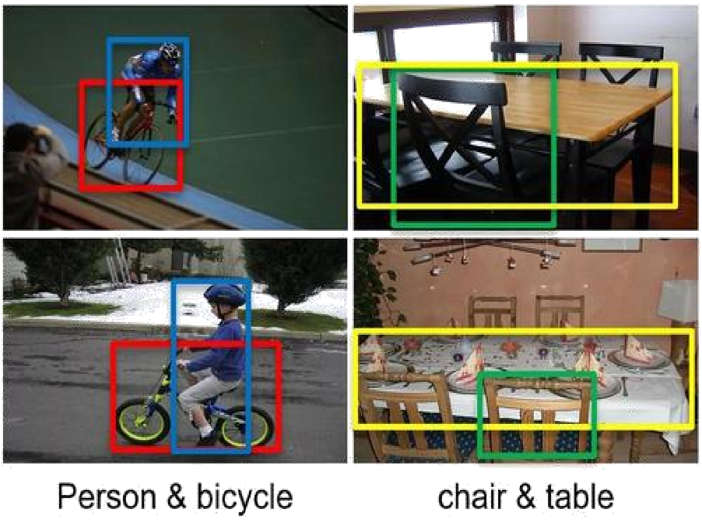

Figure 3: The Image Recognition. Different objects are detected by machines, which will play a very important role in the field of artificial intelligence.

(4) The YOLO Detection System

We reframe object detection as a single regression problem, straight from image pixels to bounding box coordinates and class probabilities. Using our system, you only look once (YOLO) at an image to predict what objects are present and where they are. YOLO is refreshingly simple: see Figure 4. A single convolutional network simultaneously predicts multiple bounding boxes and class probabilities for those boxes. YOLO trains on full images and directly optimizes detection performance. This unfied model has several benfits over traditional methods of object detection.

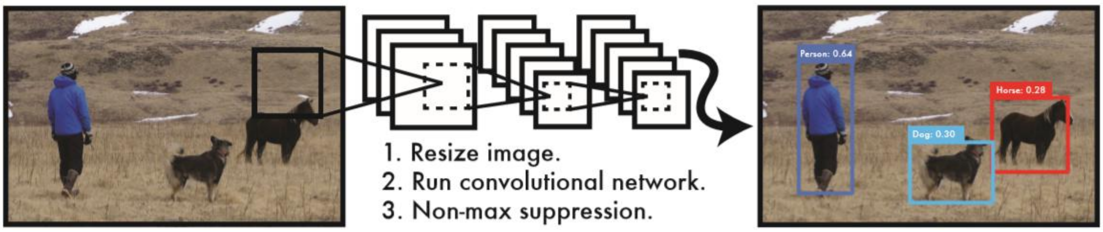

Figure 4: The YOLO Detection System. Processing images with YOLO is simple and straightforward. Our system (1) resizes the input image to 448×448, (2) runs a single convolutional network on the image, and (3) thresholds the resulting detections by the model's confidence.

Project Overview

This section explains the framework of a distributed system. In our distributed image recognition project, we use both Alibaba Cloud and Huawei Cloud.

We implemented deep learning image recognition based on distributed system architecture. By training twenty kinds of images of person, car, bus, track, motorbike and trafficlight and so on, and processing service requests through multiple servers, we obtained approximate ideal image recognition.

Figure 5 shows our distributed system architecture.

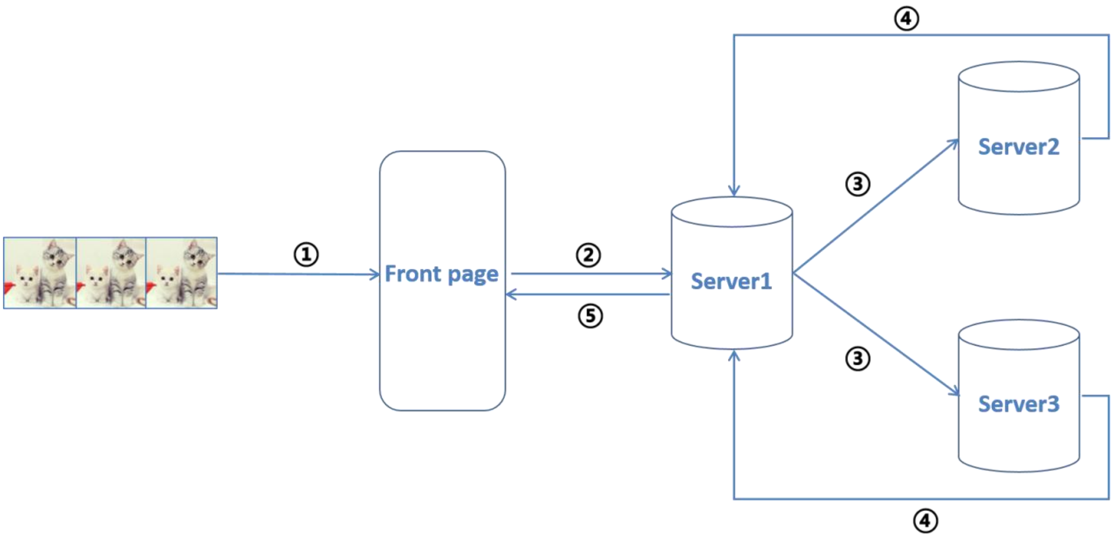

Figure 5: The Distributed System. The intermediate server is used to process the front-end user request, and the server 2 and the server 3 are respectively the Alibaba Cloud server and the Huawei cloud server, and participate in the image recognition work.

Regarding the flow of data information in a distributed system, we will explain it through Figure 6.

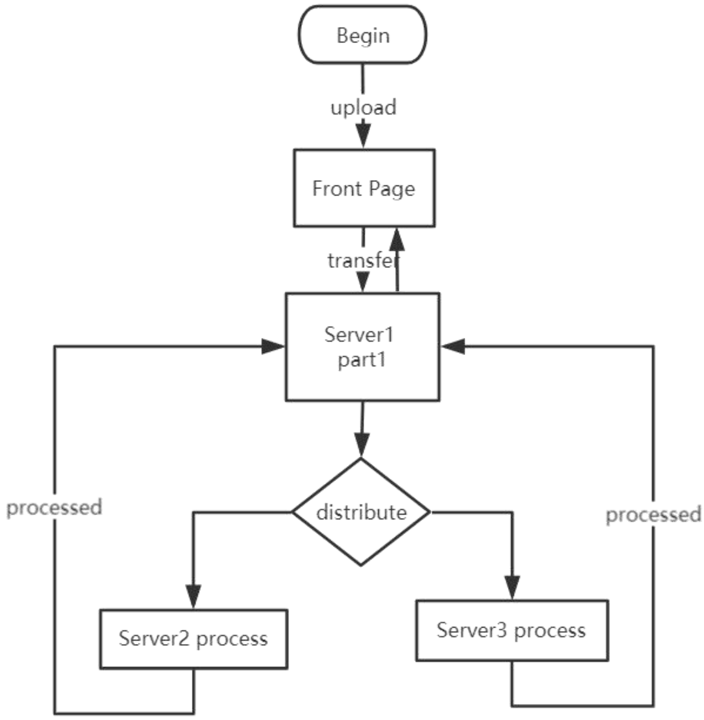

Figure 6: Data Flow Diagram. The front end transmits the image to be identified to the intermediate server, and the intermediate server determines which server the video or video should be forwarded through a load balancing policy.

System Structure

Based on the consideration of the real environment, we use part of Alibaba Cloud to process the image or video transmitted by the front end, and another part is used together with Huawei Cloud to process and return the data of the front end. The detailed process is illustrated by Figure 7.

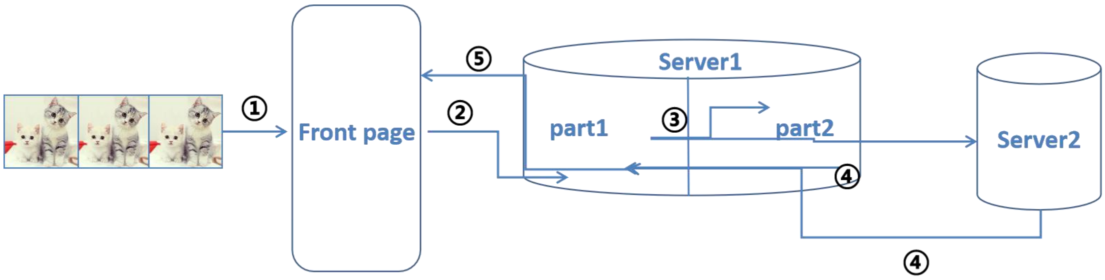

Figure 7: The Physical Distributed System. In the figure, we can see that the first server is divided into two parts, one for processing front-end data and the other for processing and returning data with another server.

The relevant data transfer process, we can see the detailed process in Figure 8.

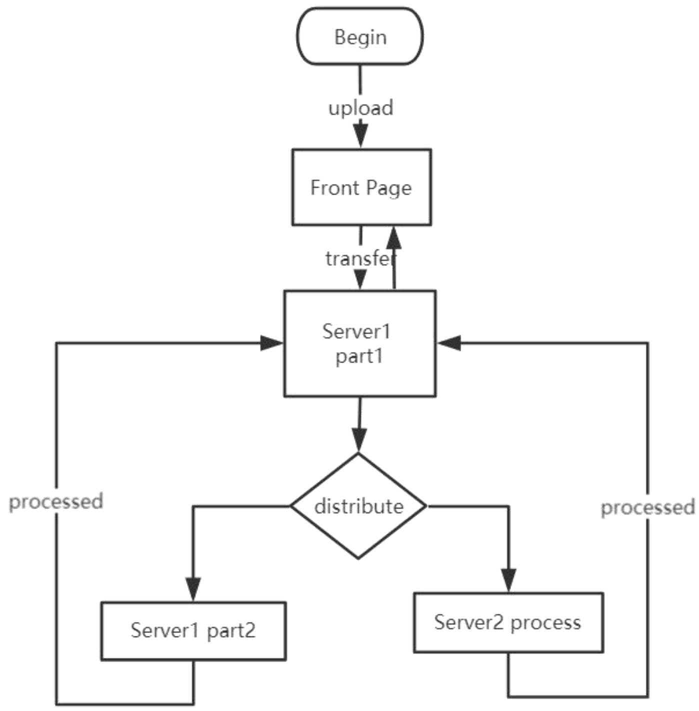

Figure 8: Physical Data Flow Diagram. The front end transmits the image to be identified to the first server, and the first server determines which server the video or video should be forwarded through a load balancing policy.

In this section, we have a general understanding of the overall framework of the system, we understand the structure of the system and the transmission of data in the system structure.

Model Construction

Below we will explain in detail the YOLO technology in the project.

The principle of YOLO technology is that divide the picture into cells (S=7 in the original text), and the subsequent output is in units of cells（Figure 9）:

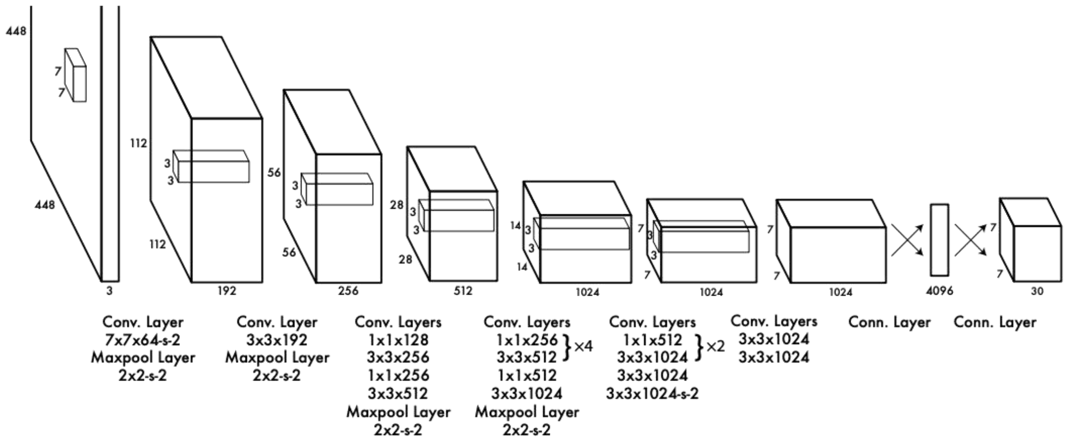

Figure 9: The YOLO Architecture. Our detection network has 24 convolutional layers followed by 2 fully connected layers. Alternating 1 × 1 convolutional layers reduce the features space from preceding layers. We pretrain the convolutional layers on the ImageNet classifification task at half the resolution (224 × 224 input image) and then double the resolution for detection.

1. If the center of an object falls on a cell, then the cell is responsible for predicting the object.

2. Each cell needs to predict B bbox values (bbox values include coordinates and width and height, B=2 in the original text), while

predicting a confidence score for each bbox value. That is, each cell needs to predict B $\times$ (4 + 1) values.

3. Each cell needs to predict C (the number of object types, the original C = 20, this is related to the database used) conditional probability values.

So, the final output dimension of the network is, although each cell is responsible for predicting an object (this is also a problem with this article, there may be problems when there are small objects), but each cell can predict multiple bboxes. value.

Integration Testing

The most effective way to prove that the system can correctly identify objects in photos or videos is to test them. Here we use two test methods. One is based on Web testing, and the other is based on WeChat applet testing.

(1) Web testing

The Web test passes through the Web front end, and the interface displayed to the user is a Web page. In this page, we can click the upload button to upload the image, which can be seen in Figure 9.

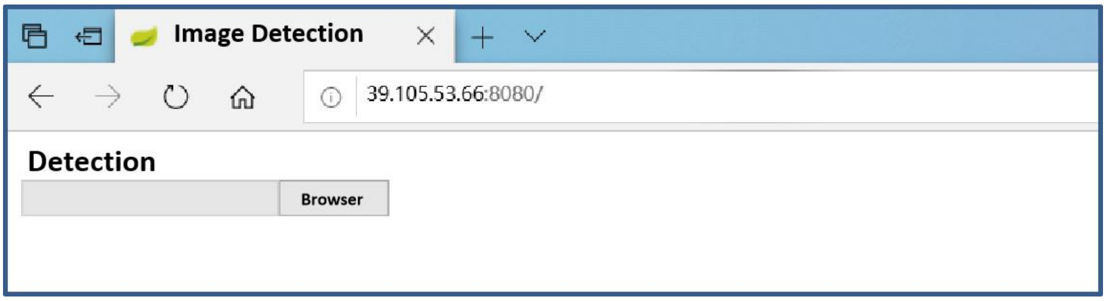

Figure 9: Web test page. We can access the front page through http://119.3.236.128:8080.

(2) WeChat applet testing

On the other hand, we use the WeChat applet for testing, which has a better user experience. Users can scan the code for experience and testing. The WeChat applet interface is shown in Figure 10.

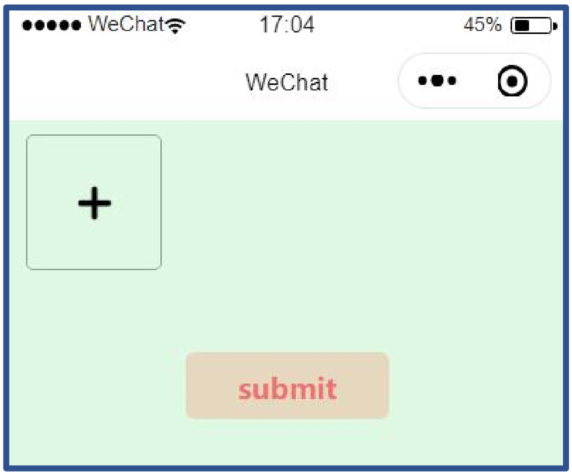

Figure 10: WeChat applet testing. By scanning the QR code on the left, you can enter the WeChat applet on the right.

In the WeChat program, the important point is how to upload the image. Here we have code interpretation of the image upload.

```
formSubmit: function () {
var that = this
var adds = that.data.img_arr;
for (var i = 0; i < this.data.img_arr.length; i++) {
wx.uploadFile({
    url: 'http://119.3.236.128:8080/upload',
    filePath: that.data.img_arr[i],
    name: 'file',//The corresponding name of the other server
    formData: adds,
    success: function (res) {
        console.log(res)
        if (res) {
            wx.showToast({
            title: 'Submitted for release!',
            //Upload success reminder
            duration: 6000
        });
      }
    }
})
}
```

Conclusion

In general, in this project, we used a distributed system architecture and used YOLO technology in deep learning image recognition to mark images. We download data from COCO Dataset for training, get the feature values of the data, training data, test data, and realize deep learning image recognition under the distributed system framework.

In the experimental data tested, we have obtained good experimental results. However, in the process of analyzing a large number of vehicle conditions, we found that our in-depth learning algorithm has certain drawbacks. In the case of very complicated vehicle conditions, Our algorithm has some error in the detection of traffic lights. In this regard, we will continue to improve our algorithms to achieve better experimental results.