Machine Vision for Flags

Neural network and dataset

Chiming Ni

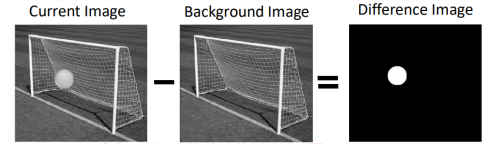
July 4, 2023

BSc, ZJU-UIUC Institute
chiming2@illinois.edu

Traditional Computer (Machine) Vision

Traditional computer vision is a broad term used to describe a variety of CV algorithms mostly proposed before 2011. These algorithms usually involves a series of *manually defined* procedures of manipulating images to get the result.

For example, the background removal algorithm from Prof. Yang's lecture is a typical example of traditional CV algorithm.



Ref: *Computer Vision Handouts, Harry Asada, MIT*

Everything Changes after 2011

In 2011, *Alex Krizhevsky* published a paper *ImageNet Classification with Deep Convolutional Neural Networks* which used a different approach called **neural network**. This approach was considered as a dead path decades ago but it turned out that their algorithm outperforms any other CV algorithms on record at that time.

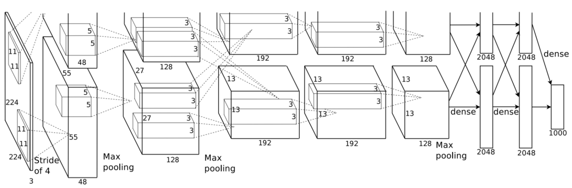


Figure 1: Architecture of AlexNet

What is New?

Traditional CV algorithms rely heavily on a good choice of procedures, which requires a lot of work even for the smartest brains in the world. This simply doesn't work for large-scale classification tasks. If we have 100 classes, then we need to consider the characteristics of all 100 classes when designing the procedures. The cost it takes is almost unacceptable.

This leads to an interesting insight. Rather than doing all the design work manually, why don't we find a way to automate the process of finding a good choice of procedures?

Machine Learning

The branch of computer science that focuses on this task is called *machine learning*. As its name suggests, machine learning is about letting machines learn how to do a certain task rather than telling them all the specific steps of that task.

The most famous approach of machine learning is *neural network*. But actually neural network worked pretty terrible in computer vision tasks decades ago when it was firstly proposed. So what has changed in the last 30 years?

Moore’s Law

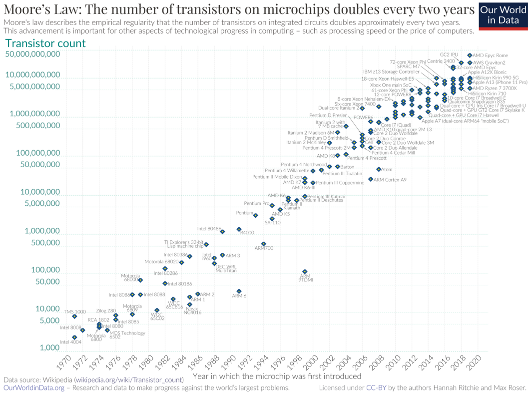


Figure 2: The number of transistors in an integrated circuit (IC) doubles about every two years.

ZJU-UIUC INSTITUTE
Zhejiang University-University of Illinois at Urbana-Champaign Institute
浙江大学伊利诺伊大学厄巴纳香槟校区联合学院

Moore's Law (Cont'd)

More transistors means faster computation, meaning the neural network can get larger and deeper. But it's not just the model itself is getting larger. The "homework" we use to train models is also getting larger, maybe even growing faster than the model itself.

Figure 3: ImageNet dataset contains more than 14 million images.

Power of Dataset

Dataset is an important factor in determining a neural network's performance, just like how homework is important to your grade. If a homework set is full of error or lacking diversity in topics, you're less likely to get a high score in the exam.

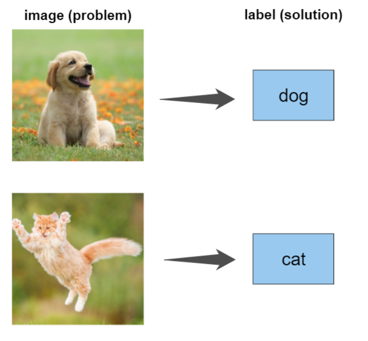


Figure 4: An example dataset

Project Task

In this project, you're required to identify three flags. Indonesia flag, China flag and the pirate flag (friendly flags can be chosen arbitrarily, but must contain two types). Of course this task can be done with traditional CV easily, but for this project we will use the neural network approach.

Don't worry, you don't have to train the model. We've already trained one for you, feel free to use it. But if you're interested in making your own dataset, listen carefully now.

Few Things to Notice

Capturing enough images to be annotated is the first step in making a dataset. You can capture images in any way you want, this includes

1. Take many photos with your cellphone.
2. Record a video then convert into a bunch of photos.

There're a few guidelines in preparing these images:

1. Keep diversity, your dataset should contain as many situations as possible. E.g., Indonesia+China, Indonesia+pirate, ...
2. The dataset should be large enough, with at least 500 images, our dataset uses over 1000 images.
3. Try different light conditions and point of views.
4. Balance the number of presence of each flag in dataset.

ZJU-UIUC INSTITUTE
Zhejiang University-University of Illinois at Urbana-Champaign Institute
浙江大学伊利诺伊大学厄巴纳香槟校区联合学院

Annotate Your Dataset

After acquiring enough images, it's time for annotation (labeling). There are many great platforms for annotation there. For this project, we will use roboflow to do the annotation.

Check out the tutorial on GitHub for more information:
`https://github.com/nice-mee/zjui_indonesia_summer_school_2023/wiki/Annotate-Your-Dataset`

If you're having any trouble in connecting to GitHub, check out the Gitee mirror:
`https://gitee.com/niceme_2020/zjui_indonesia_summer_school_2023/wikis/Annotate-Your-Dataset`

Send Your Dataset

Once you're done, send the code snippet to me via email:

chiming2@illinois.edu

I'll train the model for you on your dataset, then I'll send you back a 'yolov8n.onnx' file after the training is done. You can use it to replace the original 'yolov8n.onnx' in your project folder and build the TensorRT engine again.

Thank you!

ZJU-UIUC INSTITUTE
Zhejiang University-University of Illinois at Urbana-Champaign Institute
浙江大学伊利诺伊大学厄巴纳香槟校区联合学院