Lung Cancer Cell Detection using YOLOX

MENG Yuanyuan, WANG Jue, WANG Tingsheng, PENG Qiulei

MSBD5012 Final Project
The Hong Kong University of Technology and Science

Abstract

Lung cancer is one of the most common disease in the world and until now, detecting lung cancer cells is still a challenging problem. In recent years, many studies using deep learning methods have made progress in distinguishing cancer cells from normal cells. In this paper, we use YOLOX model to realize the recognition of different lung cancer cells, with the technique of data augmentation to improve the model training.Finally, the accuary in YOLOX can reach 40.2%, compared to 14.8% in YOLOv3 model.

Keywords: Lung Cancer, YOLOX, Machine Learning, Medical Science, Darknet53, Cancer Cell Detection.

1 Introduction

In 2015, China's tumor registered data showed that in China, the number of new malignant tumor cases was nearly 3.929 million, and over 2 million patients succumb to the malignant tumor[1]. In all kinds of malignant tumors, lung cancer morbidity and mortality are both ranked first. According to GLOBOCAN2020[2], the incidence and death of lung cancer in China accounted for 37.0% and 39.8% of the global total, respectively. Not only smoking but also exposure to toxic chemicals such as radon, asbestos and arsenic can cause lung cancer, which is one of the most common cancers on the planet. Lung cancer can be divided into two categories: small cell lung cancer(SCLC) and non-small cell lung cancer(NSCLC). Furthermore, AC(adenocarcinoma) and SCC(squamous cell carcinoma) are the two most prevalent types of non-small cell lung cancer, NSCLC.

Yet, detection of lung cancer cells can be time-consuming and arduous due to the lack of well-defined histological characteristics. The detection is an essential process for diagnostic and treatment choices, which can be different for NSCLC and SCLC[3]. To enable faster and more accurate classification, machine learning is widely used in helping analyze lung cancer whole-slide images. In the paper[4], Yu et al. used several methods, including support vector machine(SVM) and Naive Bayes classifiers to classify normal cells and lung cancer cells, and achieve an AUC of nearly 85%. They also classified AC and SCC cells, with 75% accuracy. In the past few years, deep learning models have shown excellent results in the field of tumor slides classification. In the paper[5], Dua et al. applied three deep learning models to classify Computed Tomography (CT) images of lung cancer, as well as Corona Virus disease(COVID-19). They used a Darknet-53 network, which is the backbone of YOLOv3 model, ResNet50 and VGG19, and in the result shown in the paper, Darknet-53 achieved an accuracy with 100%, which is an impressive performance. In the meanwhile, ResNet50 and VGG19 models could reached accuracies of 80%, and 77% respectively.

Here, we demonstrate that how the efficiency of detecting lung cancer cells can be improved by applying YOLOX model. Zheng et al.[6] have carried out experiments on YOLO series and proposed a new model which overperformed with an AP of 47.3%, which exceeds the best performance so far of YOLOv3 by their improvement of YOLOv3 architecture.

However, most algorithms in YOLO series are computing based on anchor boxes preprocessed by some algorithms like clustering. During the training, the net would output a prediction of the bounding box using the predefined anchor box, and compare the predicted box with ground truth to calculate the difference between prediction and true label for further training process. In this way, initialization of anchor box plays an important role in the training of the

model.

In this paper, the experiment is designed to reduce the significance of anchor box such that the model could be able to learn without anchor box. Under the discovery from Zhang[6], we consider conducting an experiment to apply YOLOx to the Lung cancer detection research to improve the efficiency of detecting lung cancer cells by using the concept of dropping anchor boxes in the data preprocessing section.

2 Data Preprocessing

Datasets This project will use lung cancer imaging datasets from the HUAWEI MindSpore competition. Guangzhou Anbiping (LBP) Medical Sci.& Tech provided many samples of pathological imaging data for this competition, which are pathological images marked for the diagnosis of peripheral lung diseases to train a robust and accurate AI pathology model as well as ensure the privacy of biopsy cell classification.

Specifically, bronchoscope brushing technology is a method of sampling these pathological images to find cancer and cell changes that may cause cancer. Conventional liquid slides show darkly stained cell clusters, which are difficult to interpret regardless of microscopy or AI recognition. Liquid-based cytology is an innovative cytology slide preparation technology launched in recent years. Compared with traditional smears, liquid-based cytology stores specimens through storage solutions removes impurities in the specimens through a series of processes and forms a transparent glass slide under the microscope. The diagnostic accuracy and sensitivity of this cytology have been greatly improved.

Cancer cells are significantly different from normal cells in histopathology and cytopathology, but different types of cancer cells are very similar and difficult to classify. There are two main types of lung cancer datasets as shown in Figure1, non-small cell lung cancer (NSCLC) and small cell lung cancer (SCLC). There are two main types of NSCLC: squamous cell lung cancer (SCC) and adenocarcinoma (AC).

There is a serious shortage of cytopathologic around the world, but the progress of computer science makes it possible to realize cytology artificial intelligence. The target of this project is to do multi-label object detection for lung cancer cells. There are a total of 1000 Lung Cancer pictures in the datasets. The table below records the label for every picture. For instance, the first sample is SCLC which is labeled as 1 in this column.

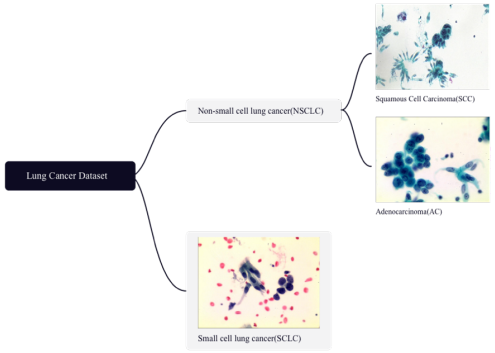

Figure 1: Categories of cancer cells.

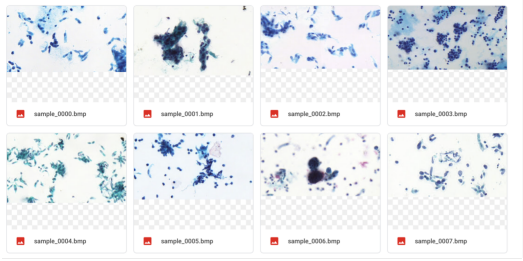

Figure 2: Some data samples.

image_id	xmin	ymin	xmax	ymax	scc	ac	sclc	nsclc
sample_0000	0.348	0.4096	0.4405	0.7364	0	0	1	0
sample_0000	0.1613	0.4398	0.2263	0.5858	0	0	1	0
sample_0000	0.0113	0.8072	0.0804	0.9548	0	0	1	0
sample_0001	0.1642	0.2093	0.453	0.8419	1	0	0	1
sample_0001	0.5055	0.2259	0.6596	0.872	1	0	0	1
sample_0005	0.14300000000000000	0.0889	0.4314	0.366	0	0	1	0
sample_0005	0.0054	0.0497	0.0954	0.2575	0	0	1	0
sample_0005	0.4972	0.3645	0.8432	0.8208	0	0	1	0
sample_0006	0.5334	0.5166	0.6043	0.6506	0	0	0	1
sample_0007	0.8904	0.2334	0.9858	0.375	0	0	0	1
sample_0007	0.2321	0.0015	0.4085	0.2786	0	0	0	1

Figure 3: Label information provided with datasets

Preprocessing To make the dataset suitable for model, there are some changes made based on original data. Firstly, large size of original image results in time costly model training process, and reducing the image sizes could fasten it to a great extent. Every image is resized to thunmbnail seize in proportion using Python Imaging Library. And then, it is requisite to generate a JSON files to be compatible to both model training and testing processes, which contains the size, bounding box, Image_ID, and Category_ID for all images. While the original dataset only contains a CSV file that

has related records, the updated JSON files are generated based on this file and the newly resized dataset.

3 Models

Data augmentation To improve model performance, data augmentation is introduced to this experiment. Image merging and image mask are processed respectively to enlarge the original datasets. Image merging is to merge two images arbitrarily selected into one image horizontally, and it could generate up to 499,500 new images. The technique used in image mask is to use an unmeaningful grey block to mask a certain part of the images in order to add noises to the dataset and improve model performance.

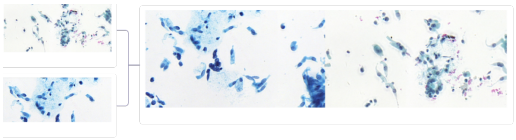

Figure 4: Image merging

Darknet53 In this paper, the backbone we use is called Darknet53[7], which is similar to ResNet50, yet without BottleNeck(BN). In each residual unit, there are two convolutional layers. The first one is 1×1 and the second one is 3×3. Each convolutional layer is followed by a batch normalization layer and a LeakyReLU layer. There are no fully connected layers or pooling layers except the layers in the last. In Darknet-53 network, downsampling is based on the 3×3 convolutional layers where the stride is 2, replacing the pooling layers. The utilization of residual connection makes it possible to design a deep network and alleviates the problem of gradient disappearance during training, making the model easier to converge.

YOLOv3 In YOLOv3, there are only convolutional layers, and the size of the output feature map is based on adjusting the convolution step size. Therefore, there is no special restriction on the input image size. Besides, YOLOv3 draws on the idea of pyramid feature maps. Small-size feature maps are used to detect large-size objects, while large-size feature maps are used to detect small-size objects.
YOLOv3 outputs 3 feature maps in total, the first feature map is down-sampled 32 times, the second feature map is down-sampled 16 times, and the third is down-sampled 8 times.

In the upsampling layer, to generate a large-size image from a small-size feature map, interpolation and other methods are applied. For illustration, use the nearest

	Type	Filters	Size	Output
	Convolutional	32	3 × 3	256 × 256
	Convolutional	64	3 × 3 / 2	128 × 128
1×	Convolutional	32	1 × 1	
	Convolutional	64	3 × 3	
	Residual			128 × 128
	Convolutional	128	3 × 3 / 2	64 × 64
2×	Convolutional	64	1 × 1	
	Convolutional	128	3 × 3	
	Residual			64 × 64
	Convolutional	256	3 × 3 / 2	32 × 32
8×	Convolutional	128	1 × 1	
	Convolutional	256	3 × 3	
	Residual			32 × 32
	Convolutional	512	3 × 3 / 2	16 × 16
8×	Convolutional	256	1 × 1	
	Convolutional	512	3 × 3	
	Residual			16 × 16
	Convolutional	1024	3 × 3 / 2	8 × 8
4×	Convolutional	512	1 × 1	
	Convolutional	1024	3 × 3	
	Residual			8 × 8
	Avgpool		Global	
	Connected		1000	
	Softmax			

Figure 5: Darknet53 Architecture

neighbor interpolation algorithm to transform an 8*8 image into an 16*16 image. In addition, the upsampling layer does not change the number of channels of the feature map.

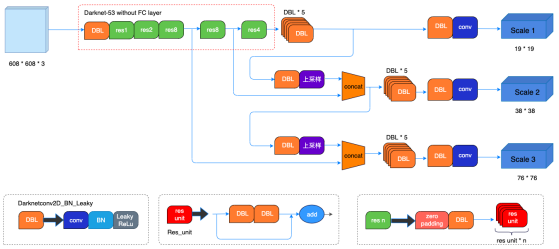

Figure 6: YOLOv3 Architecture

YOLOX architecture Our experiment adopts the architecture of YOLOX. It combines YOLOv3 with Darknet53 as the baseline model. YOLOX adopts the architecture of Darknet53 backbone and the YOLOv3-SPP layer. The paper introduced two main changes made comparing with yolo series from YOLOv3 v5. First one is Decoupled head, and the other one is anchor free,which are going to explain below.

Decoupled head The goals of this experiment is detecting lung cell and classifying them into correct categories. Therefore, classification and object detection are the main task for the model. Comparing with the model architecture of previous YOLO series,

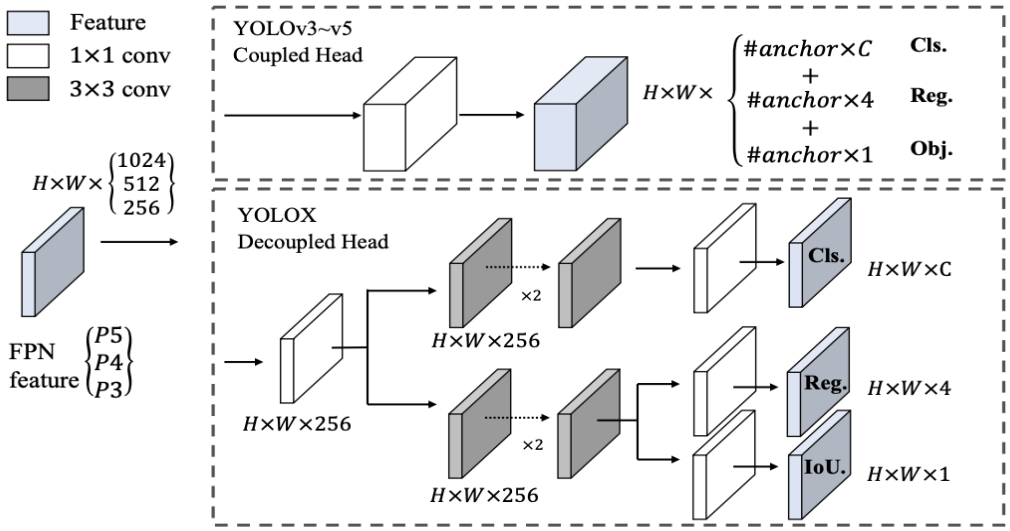

Figure 7: Architecture comparison for previous YOLO series and YOLOX

most of them are taking a 1×1 convolution layer for each level of FPN feature and the classification and localization head keeps coupled in one feature map. YOLOX introduced a new way to decouple heads. For each level of FPN, it first adopt a 1×1 convolution layer and reduce the feature channel to 256. After that, two parallel branches with two 3×3 convolution layers are added, one is for classification task, and the other one is for regression task. Besides, in the regression part, YOLOX separates this part into two section, which are regression part and IoU branch used for calculation IoU loss for bounding box respectively. Please see the clear architecture diagram in Figure 7.

Anchor-free In previous YOLO series, Anchor box is predefined for various datasets using some algorithm before training. The training process is predicting the bounding box based on anchor box provided. It increases the cost of data preprocessing in the first stage of building a model. Furthermore, anchor boxes generated by some clustering algorithm usually can only implied on the specific datasets. Rather than using anchor, YOLOX dropped anchor to let the model trained from scratch, which make the model is more flexible and simpler.

4 Experiment

Experiment In the very beginning, we try YOLOv3 with DarkNet53 and ResNet18 on the HUAWEI Cloud with MindSpore. In this approach, we load the images as DataSet class of MindSpore and preprocess it with a repeated rate of 10%, as well as the dropping rate of 10%. Then we define the YOLOv3 model based on MindSpore YOLOv3 framework and set the learning rate as 0.005, epoch size as 30, and batch size as 8. Specifically, when the epoch size is bigger than 30, the loss keeps increasing and when the learning rate is larger or smaller, the loss decreases too quickly or too slowly. In the evaluation part, not only we calculate the AP50 accuracy, but also the hitting rate of bounding box is taken into consideration. The best AP50 and hitting rate of YOLOv3 which train on MindSpore is 14.84375% and 24.21875%.

Then we attempted to train and test this lung cancer images data set on YOLOX, which is provided by MEGVII. In the beginning, we (preprocess, how to make the data formatted as COCO, and redefine the EXP class and ...). We also tried to make our own evaluate function to compute the AP50, which is not straightforward and useful as the evaluation of COCO. Instead of changing the basic learning rate, the configure learning rate equals to 0.01 / 64 as the settings in [6]. After refining parameters, the reasonable settings we found are shown below: the size of batches equals to 16, as well as the number of epoch equals to 300. Without data set expanding and random masking, the best result of our YOLOX model is AP50 26.5% and AP50-95 17.1%, Then, we applied random mask during the training period and generate expanding set of the raw training cancer images. To implement the ran-

dom masking during the training period except the last 50 epoches, with a probability of 1/8 in each batch, we change the pixel values in three channels to zero in a rectangle area of the image into black, and the width and height of the rectangle are both the 1/8 of raw images. The best AP50 and AP50-95 when applying random mask only are 40.2% and 26.1% respectively. As for the expanding, we concatenate both two images and their annotations horizontally to enlarge the training set. However, the accuracies by using expading decreased to 13.3% and 15.6% respectively.

5 Conclusion

In this report, YOLOX model is realized successfully in the Lung Cancer Dataset, and the performance of the average precision is better compared with the previous model YOLOv3. Nevertheless, due to the limitation of GPU storage capacity, and the images have to be compressed for the size being too large, as a result it would lose some important features in the training process. Thus, this model with present recognition accuracy is far from being practical in the medical field. Furthermore, tumor histomorphology examination in detail is the key to treatment for patients, each type of cancer has its particularity, and the determination of relevant prognostic markers is essential to personalize the cancer treatment. The application of deep learning in prognostic markers is still in its infancy. The supervised learning needs to locate the cancer cells carefully in the images, while the cost of obtain labelled data is relatively high. Pathological image recognition in Artificial Intelligence still has a very large upside potential.

References

[1] Zheng RS, Sun KX, Zhang SW, et al, Report of cancer epidemiology in China, 2015[J]. Chinese Journal of Oncology, 2019, 41(1):19-28

[2] Sung H, Ferlay J, Siegel R L, et al. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries[J]. CA: a cancer journal for clinicians, 2021, 71(3): 209-249.

[3] Hanna N H, Robinson A G, Temin S, et al. Therapy for stage IV non–small-cell lung cancer with driver alterations: ASCO and OH (CCO) joint guideline update[J]. Journal of Clinical Oncology, 2021, 39(9): 1040-1091.

[4] Yu, K.-H. et al. Predicting non–small cell lung cancer prognosis by fully automated microscopic pathology image features. Nat. Commun. 7, 12474 (2016).

[5] D. Hişam and E. Hişam, "Deep learning models for classifying cancer and COVID-19 lung diseases," 2021 Innovations in Intelligent Systems and Applications Conference (ASYU), 2021, pp. 1-4, doi: 10.1109/ASYU52992.2021.9598993.

[6] Ge Z, Liu S, Wang F, et al. Yolox: Exceeding yolo series in 2021[J]. arXiv preprint arXiv:2107.08430, 2021.

[7] Redmon J, Farhadi A. Yolov3: An incremental improvement[J]. arXiv preprint arXiv:1804.02767, 2018.