Deep learning method for Breast Cancer Classification

Libo Xu & Rajibul Islam

October 23, 2019

Overview

Background

- 523000 death and 2.4 million new patients / year
- 2nd major cause of cancer death after lung and bronchus cancer
- Most frequently diagnosis cancer among women in USA

Background cont...

- Early stage detection of breast cancer is curable
- Computer-aided diagnosis (CAD) system becoming popular
- ResNet deep learning approach for cancer detection is more reliable and effective

Dataset

- `pyimagesearch.com`
- # of patients 279
- total 277523 patches of 50×50 pixels images
- file format 8863_*idx*5_*x*1001_*y*801_*class*1.*png*

Dataset cont...

container	negative example (0)	positive example (1)	total data files	total files in %
train	143091	56725	199816	72 %
test	39748	15757	55505	20 %
validation	15899	6303	22202	8 %
total	198738	78785	**277523**	
dataset in %	71.71 %	28.39 %		

Method

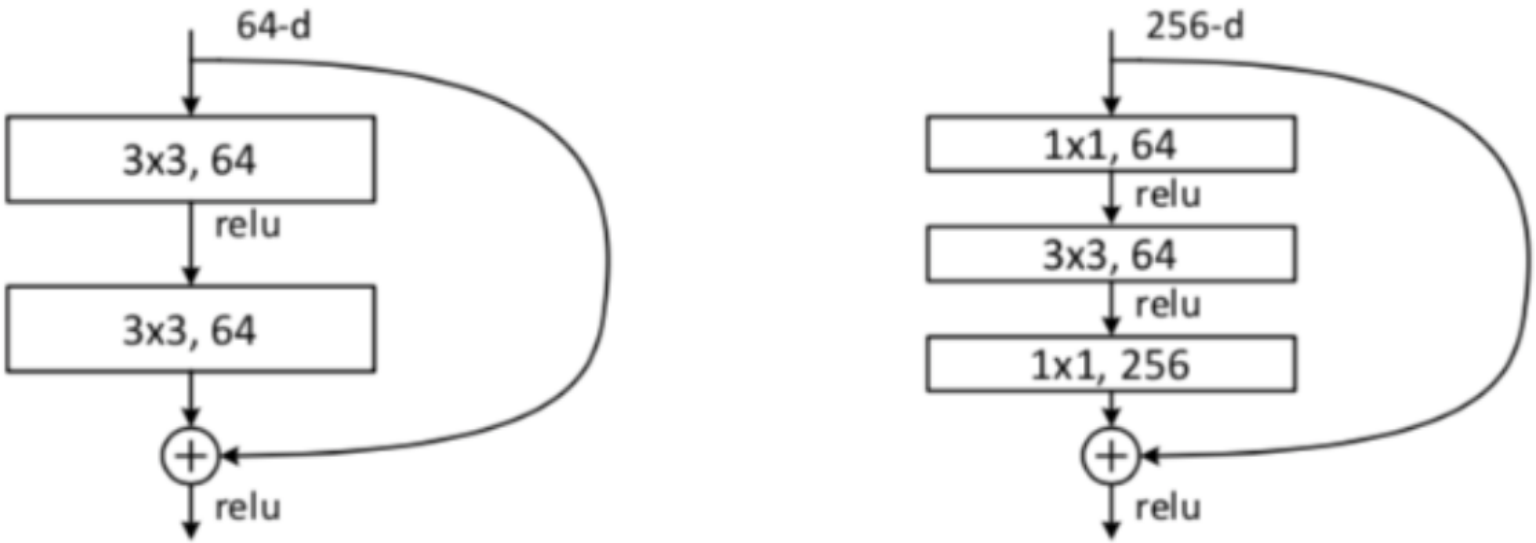

Figure: ResNet basic building block (left) & ResNet bottleneck building block (right)

Method cont...

layer name	output size	18-layer	34-layer	50-layer	101-layer	152-layer
conv1	112×112	7×7, 64, stride 2				
conv2_x	56×56	3×3 max pool, stride 2				
		$\begin{bmatrix} 3\times3, 64 \\ 3\times3, 64 \end{bmatrix}\times2$	$\begin{bmatrix} 3\times3, 64 \\ 3\times3, 64 \end{bmatrix}\times3$	$\begin{bmatrix} 1\times1, 64 \\ 3\times3, 64 \\ 1\times1, 256 \end{bmatrix}\times3$	$\begin{bmatrix} 1\times1, 64 \\ 3\times3, 64 \\ 1\times1, 256 \end{bmatrix}\times3$	$\begin{bmatrix} 1\times1, 64 \\ 3\times3, 64 \\ 1\times1, 256 \end{bmatrix}\times3$
conv3_x	28×28	$\begin{bmatrix} 3\times3, 128 \\ 3\times3, 128 \end{bmatrix}\times2$	$\begin{bmatrix} 3\times3, 128 \\ 3\times3, 128 \end{bmatrix}\times4$	$\begin{bmatrix} 1\times1, 128 \\ 3\times3, 128 \\ 1\times1, 512 \end{bmatrix}\times4$	$\begin{bmatrix} 1\times1, 128 \\ 3\times3, 128 \\ 1\times1, 512 \end{bmatrix}\times4$	$\begin{bmatrix} 1\times1, 128 \\ 3\times3, 128 \\ 1\times1, 512 \end{bmatrix}\times8$
conv4_x	14×14	$\begin{bmatrix} 3\times3, 256 \\ 3\times3, 256 \end{bmatrix}\times2$	$\begin{bmatrix} 3\times3, 256 \\ 3\times3, 256 \end{bmatrix}\times6$	$\begin{bmatrix} 1\times1, 256 \\ 3\times3, 256 \\ 1\times1, 1024 \end{bmatrix}\times6$	$\begin{bmatrix} 1\times1, 256 \\ 3\times3, 256 \\ 1\times1, 1024 \end{bmatrix}\times23$	$\begin{bmatrix} 1\times1, 256 \\ 3\times3, 256 \\ 1\times1, 1024 \end{bmatrix}\times36$
conv5_x	7×7	$\begin{bmatrix} 3\times3, 512 \\ 3\times3, 512 \end{bmatrix}\times2$	$\begin{bmatrix} 3\times3, 512 \\ 3\times3, 512 \end{bmatrix}\times3$	$\begin{bmatrix} 1\times1, 512 \\ 3\times3, 512 \\ 1\times1, 2048 \end{bmatrix}\times3$	$\begin{bmatrix} 1\times1, 512 \\ 3\times3, 512 \\ 1\times1, 2048 \end{bmatrix}\times3$	$\begin{bmatrix} 1\times1, 512 \\ 3\times3, 512 \\ 1\times1, 2048 \end{bmatrix}\times3$
	1×1	average pool, 1000-d fc, softmax				

Figure: ResNet Architecture

Method cont...

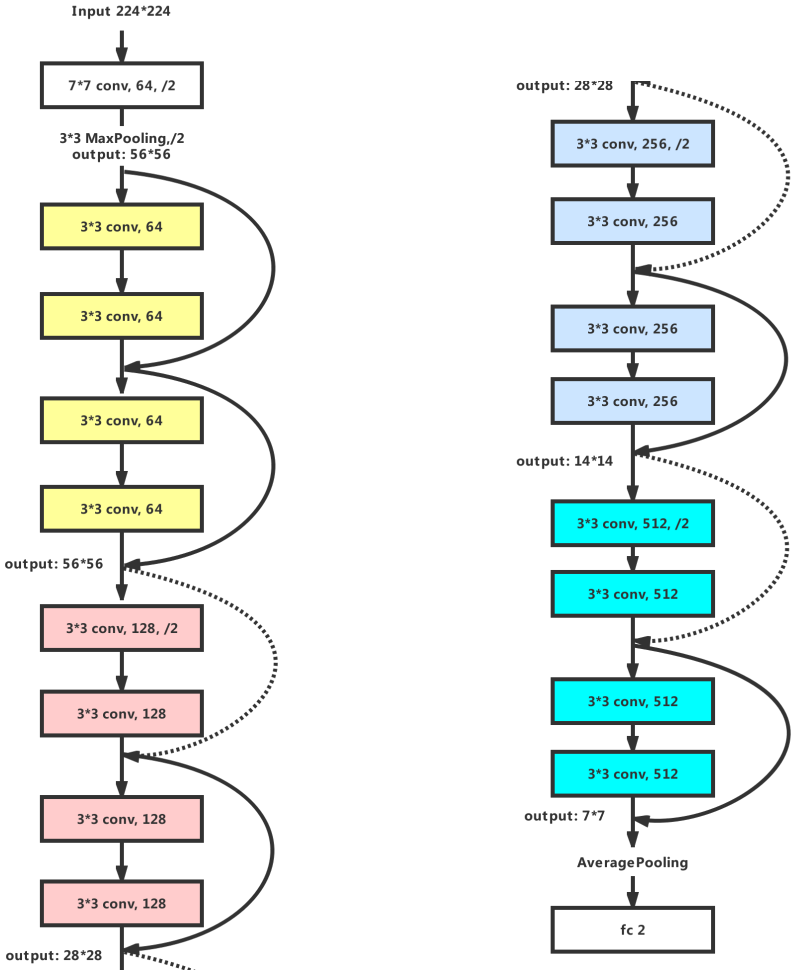

Figure: ResNet18

Method cont...

- ImageDataGenerator
  Don't need to store data to memory every time
  Mark the label of each image automatically
  Image Augmentation to reduce the impact of minimization of dataset
  (We take 15000 images for training and 2250 for validation because it takes 2 hours/epoch if we use the whole dataset)

- Parameters setting

batch size	lr	epochs	loss	optimizer	metrics
32	1e-5	150	categorical_crossentropy	Adam	accuracy

- Imbalance data issue
  class_weight is calculated
  class_weight is [1., 2.] is used

Method cont...

- Evaluation

Confusion matrix

True positives (TP)	False negative (FN)
False positive (FP)	True negatives (TN)

Formulas:

$$\text{accuracy} = \frac{TP+TN}{TP+TN+FP+FN}$$

$$\text{sensitivity/recall (True positive rate)} = \frac{TP}{P} = \frac{TP}{TP+FN}$$

$$\text{specificity (True negative rate)} = \frac{TN}{N} = \frac{TN}{TN+FP}$$

$$\text{precision (Positive predictive value)} = \frac{TP}{TP+FP}$$

$$F_1 \text{ score} = \frac{2TP}{2TP+FP+FN}$$

Results

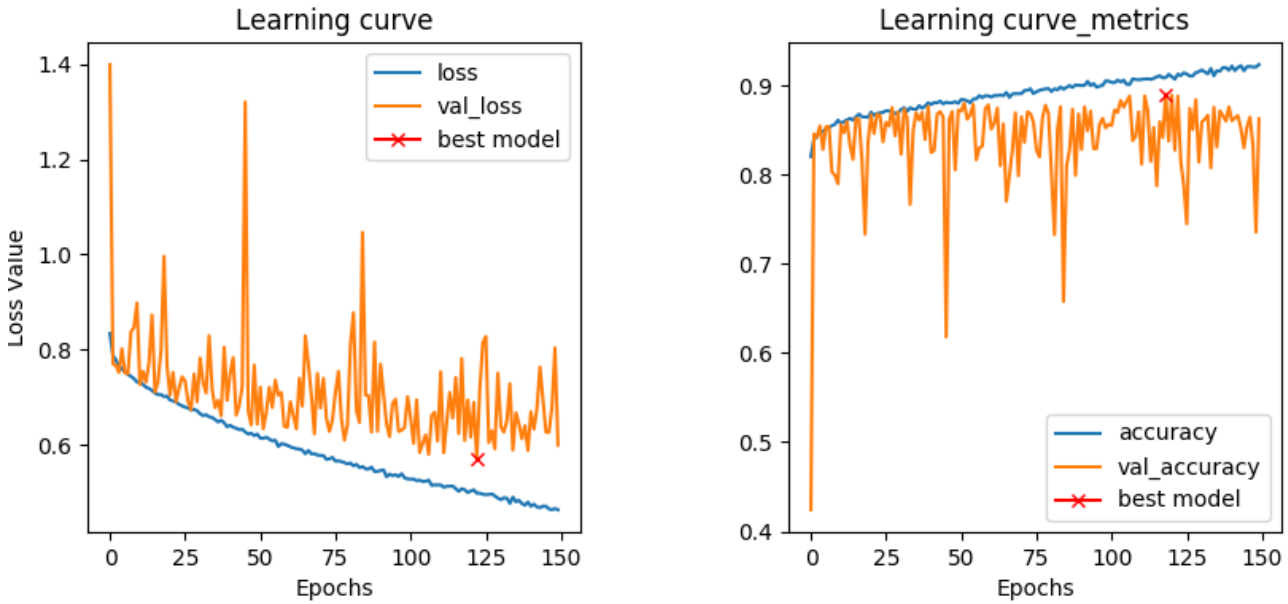

Figure: Learning curves of loss & accuracy

Results cont...

- Validation loss : 0.6163
  Validation accuracy : 0.8785

- Confusion matrix:

TP: 36480	FN: 3239
FP: 3679	TN: 12107

- Results comparison

Parameters	our results	pyimagesearch.com
Accuracy	0.8754	0.8483
Recall	0.9185	0.8503
Precision	0.9185	N/A
Specificity	0.7669	0.8470
F_1	0.9134	N/A

Future work

- Apply the decay of learning rate in the Adam optimizer
- Implementation of deeper ResNet to train the dataset, like ResNet-34 or even ResNet-152
- Train on the whole dataset of more than 2,70,000 images
- To avoid overfitting we could play more with the parameters setting

QUESTIONS !

Bejnordi BE; Veta M; et al. *Diagnostic Assessment of Deep Learning Algorithms for Detection of Lymph Node Metastase in Women eith Breast Cancer*, Dec 2017. 318(22): 2199-2210.

Syed JSG; Ahmed E; et al. *Breast Cancer Detection and Diagnosis Using Mammographic Data: Systematic Review*, Jul 2019. 21(7): e14464.