

Image Recognition for Pneumonia

Yunya Wang, yw4509

Cheng Zeng, cz1951

Kanshuai Wang, kw2606

Mingdi Zheng, mz2698

Ziyue Dong, zd780

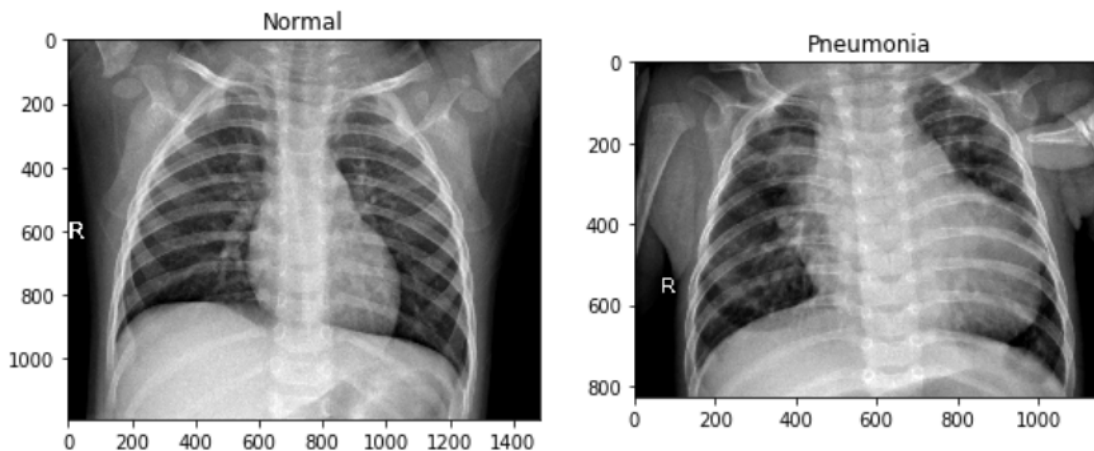
March 2020

1 The Question and Motivation

The risk of pneumonia has been immense for many, especially in developing nations where billions people face energy poverty and rely on polluting forms of energy. Estimated by the WHO, there are over 4 million premature deaths occurring annually from household air pollution-related diseases including pneumonia.

With the fast developing AI tools in neural networks (CNN) in image recognition, we could elaborate neural networks to help the doctors to speed up the pre-screening process for Pneumonia. Especially in developing countries, this technique may cause huge change in healthcare. For example, in Africa's 57 nations, there exists a gap of 2.3 million doctors and nurses.[2] This technique can effectively save time and money for those countries which are experiencing poverty.

2 Data set



The data is split into train, validation and test data containing X-rays of normal lungs and lungs with Pneumonia symptoms. In the training data-set, there are around 3900 samples labeled with Pneumonia and 1400 samples labeled as normal. We will elaborate Pytorch for the data processing and transfer them from pictures into 2D tensors for future analysis.

3 Proposed Algorithm

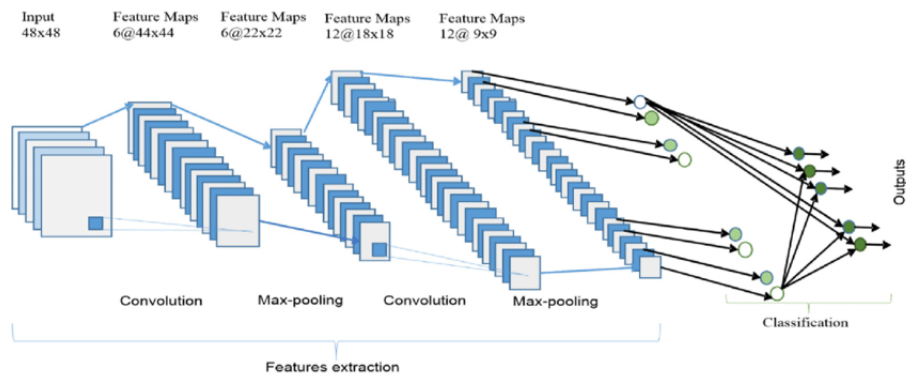
Nowadays, the development of convolutional neural network has allowed for significant gains in the ability to classify images[1]. Combining our needs and the properties of CNN, we decided to choose it as our main algorithm.

What I see



What a computer sees

08	02	22	97	38	15	00	40	00	75	04	05	07	78	52	12	50	77	91	08
49	49	99	40	17	81	18	57	60	87	17	40	98	43	69	48	04	56	62	00
81	49	31	73	55	79	14	29	93	71	40	67	53	88	30	03	49	13	36	65
52	70	95	23	04	60	11	42	69	24	68	56	01	32	56	71	37	02	36	91
22	31	14	71	51	67	43	89	41	92	36	54	22	40	28	46	33	13	80	
24	47	32	60	99	03	45	02	44	75	33	53	78	36	84	20	35	17	12	50
32	98	81	28	64	23	67	10	26	38	40	47	59	54	70	66	18	38	64	70
67	26	20	68	02	62	12	20	95	63	94	39	43	08	40	91	66	49	94	21
24	55	58	05	66	73	99	26	97	17	78	78	96	83	14	88	34	89	43	72
21	34	23	09	75	00	76	44	20	45	35	14	00	41	33	97	34	31	33	95
75	17	53	28	22	75	31	47	15	94	03	80	94	62	14	14	09	53	54	92
16	39	05	42	96	35	31	47	55	58	88	24	00	17	54	24	34	29	85	57
86	56	00	48	35	71	89	07	05	44	44	37	44	60	21	58	51	54	17	58
19	80	81	48	05	94	47	69	25	73	92	13	86	52	17	77	04	89	55	40
04	52	08	83	97	35	99	16	07	97	57	32	16	26	26	79	33	27	98	66
88	34	68	87	57	62	20	72	03	46	33	67	46	55	12	32	63	93	53	69
04	42	14	73	38	25	39	11	24	94	72	18	05	46	29	32	40	42	74	36
20	49	34	41	72	30	23	88	34	42	99	49	82	67	59	55	74	04	36	16
20	73	35	29	78	31	90	01	74	31	49	71	48	86	81	16	23	57	05	54
01	70	54	71	83	51	54	69	16	92	33	48	41	43	52	01	89	19	47	48



Then we can fit our model to the dataset and compute accuracy.

4 Proposed Methods of Optimization

For this problem, we are dealing with image processing: processing them from pictures to tensors. Also we are using CNN as our major algorithm and when the networks are really deep, the training time can take a really long time. In order to speed up and maximize our algorithm, we plan to include but not limited to the following methods of optimization:

1. Line_Profiler:

We use the line_profiler to estimate the consuming-time for different parts of the code and optimize based on it.

2. Cuda and Numba:

Given we are using GPU for most of our analysis, we will elaborate Numba to speed up our algorithm.

3. Cython:

For some of the functions used in the program, we will change it into Cython format.

4. Itertools:

We would take advantage of itertools to speed up the algorithm.

References

- [1] Krizhevsky, A., Sutskever, I., and Hinton, G.E. (2017). ImageNet classification with deep convolutional neural networks. Commun. ACM 60, 84–90.
- [2] Stephen, O., Sain ,M.,Maduh,U.J., and Jeong, D. An Efficient Deep Learning Approach to Pneumonia Classification in Healthcare. Journal of Healthcare Engineering, Volume 2019. ID 418094