Automated Detection of COVID-19 Cases in Digital Medical Images Using Deep Learning Approaches

Seminar presentation in Accordance with Partial Fulfillment of Requirements of Menoufia University for the Degree of Master in Computers and Information

Presented by:
Mahmoud Zaky Fetoh

Supervisors:
Prof: Khalid M. Amin
Dr: Ahmed M. Hamad

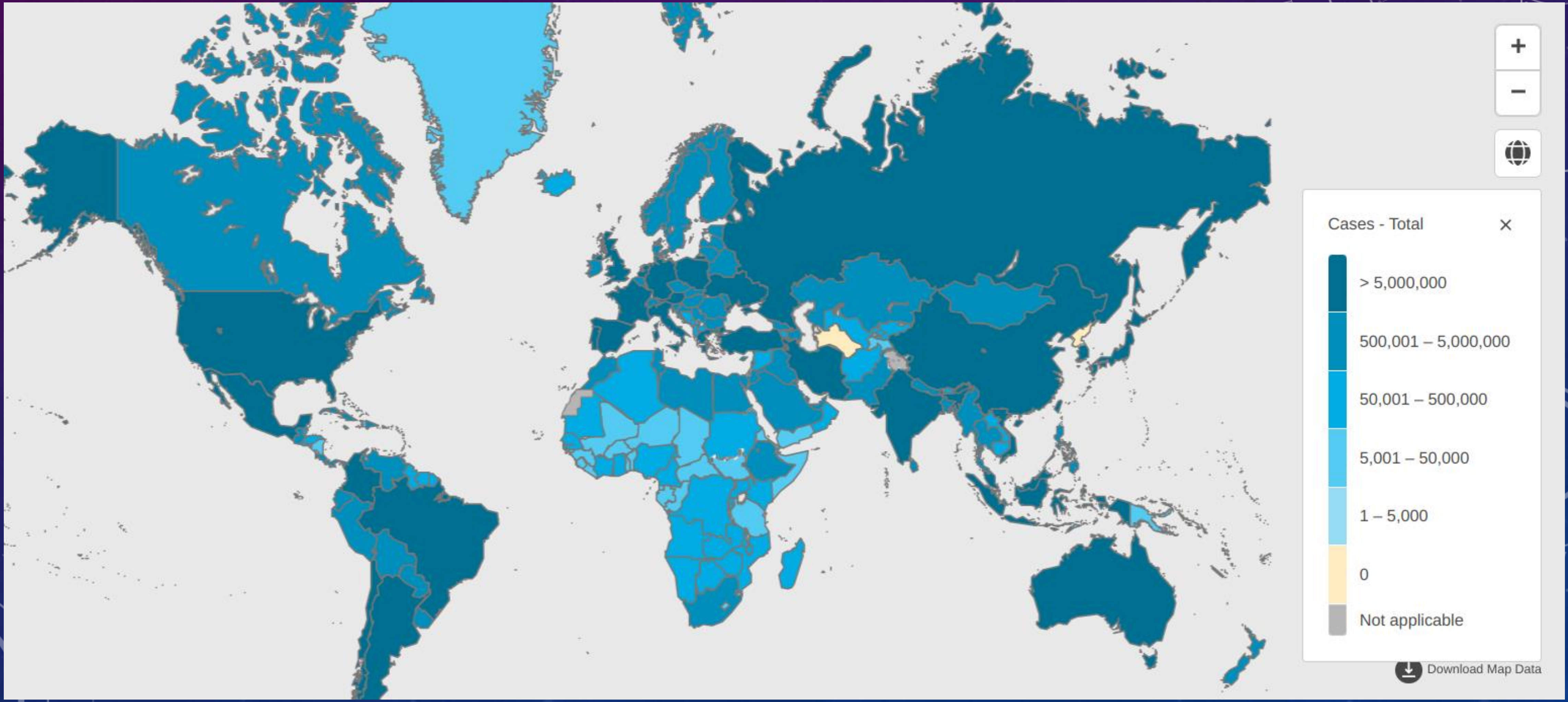
ACKNOWLEDGEMENT

I would like to thank my supervisors and my teachers Dr. Ahmed M. Hamad and Prof. Khalid M. Amin for their help and guide during my Study.

TABLE OF THE CONTENT

INTRODUCTION

- Covid-19 is a contagious disease caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2).



INTRODUCTION

- Typical diagnostic tools for COVID-19 are :

Detection Approach	Description	Accuracy	Time to Results
RT-PCR	Detects genetic material of the virus through amplification of RNA	Highly accurate	Within a few hours
Rapid Antigen Test	Detects specific proteins on the surface of the virus	Less accurate than RT-PCR	Within minutes
Antibody Test	Detects the presence of antibodies against the virus in a blood sample	Not used to diagnose active infections, but can indicate past infections	Within a few hours to a few days
Chest Imaging	Uses chest X-rays or CT scans to detect COVID-19 pneumonia	Not used as a primary diagnostic tool, typically used in combination with other methods	depends on imaging procedure

INTRODUCTION

How Hospitals react to CoViD-19?

CBC Test

low Lymphocytes

Chest X Ray

Covid19 pneumonia (Human involved)

PCR test

Covid19 (P/N)

Conclusion:

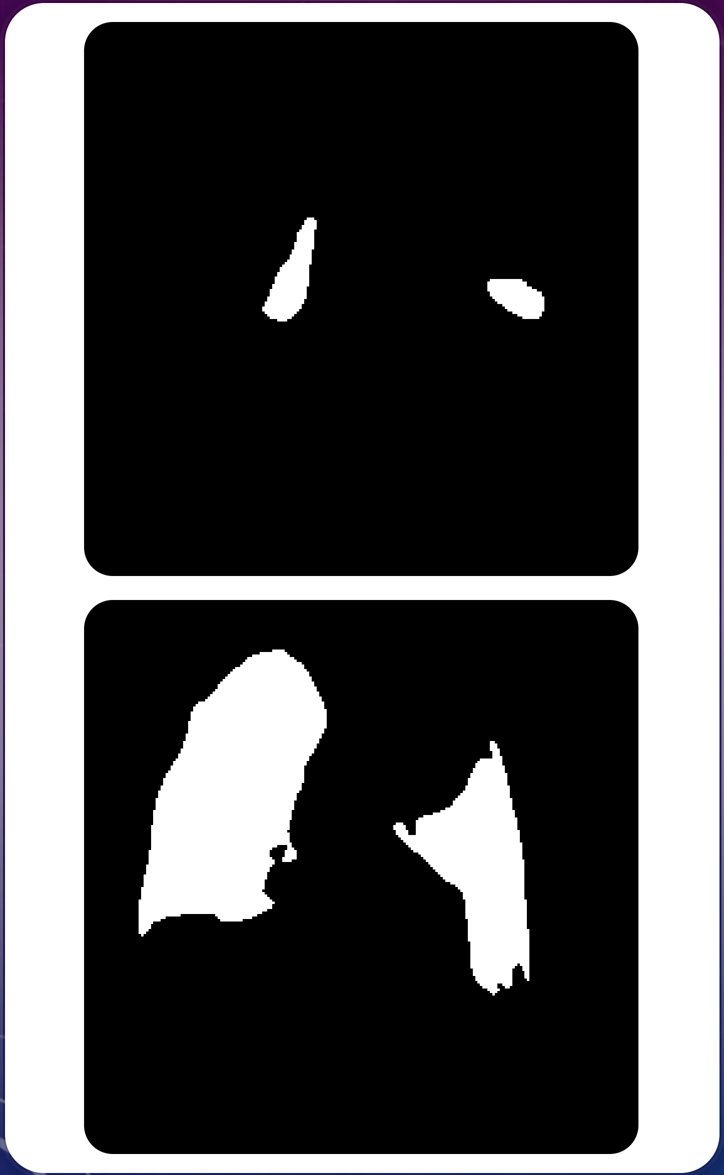
- Chest X-Ray diagnosis is time Consuming phase.
- Is there a chance of reducing that time ?

INTRODUCTION

CXR vs CT

Chest X-ray	CT scan
Quick and easy to acquire from the imaging device	More time-consuming and complex
Uses a small amount of radiation	Involves higher levels of radiation exposure
Less sensitive than CT scans in detecting abnormalities	More sensitive than chest X-rays in detecting abnormalities
Good for initial screening and detecting a wide range of abnormalities	Better for detecting smaller abnormalities and providing detailed information about their size, shape, and location
Less expensive	More expensive
May miss small lesions or nodules	More likely to detect small lesions or nodules

TABLE OF THE CONTENT



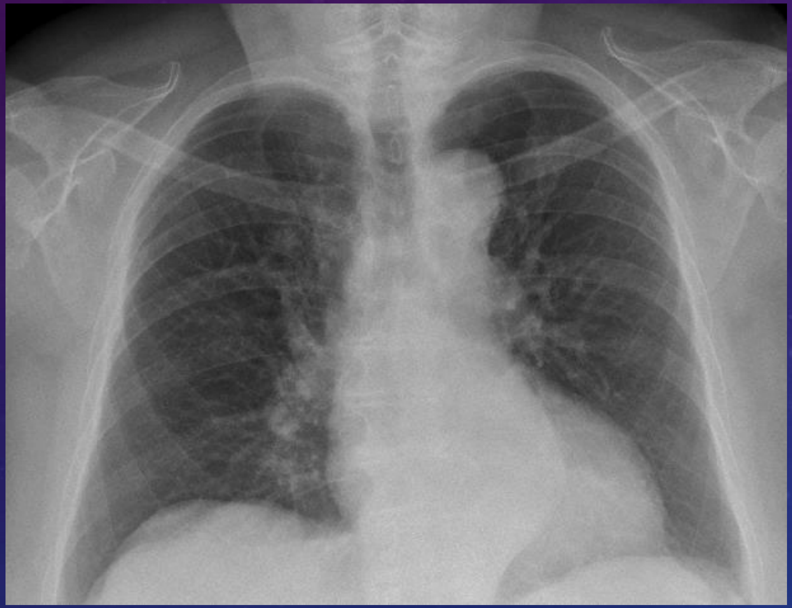
PROBLEM STATEMENT

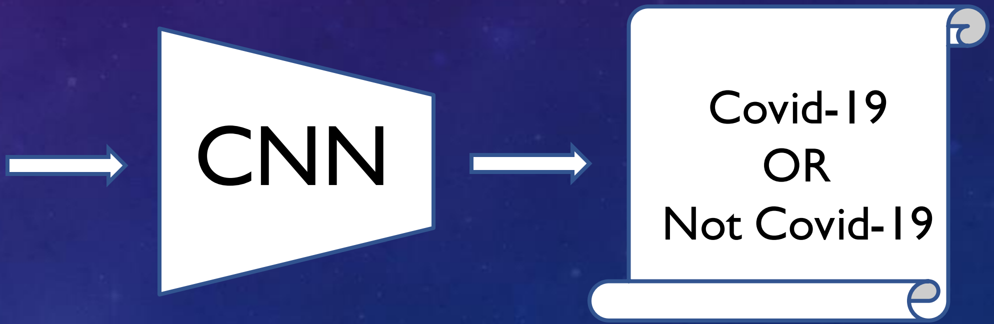
- Detection Challenges:
  - Viewpoint variation.
  - Illumination.
  - Deformation.
  - Occlusion.
  - Scale.

Szeliski, Richard. *Computer vision: algorithms and applications*. Springer Science & Business Media, 2010.

PROBLEM STATEMENT

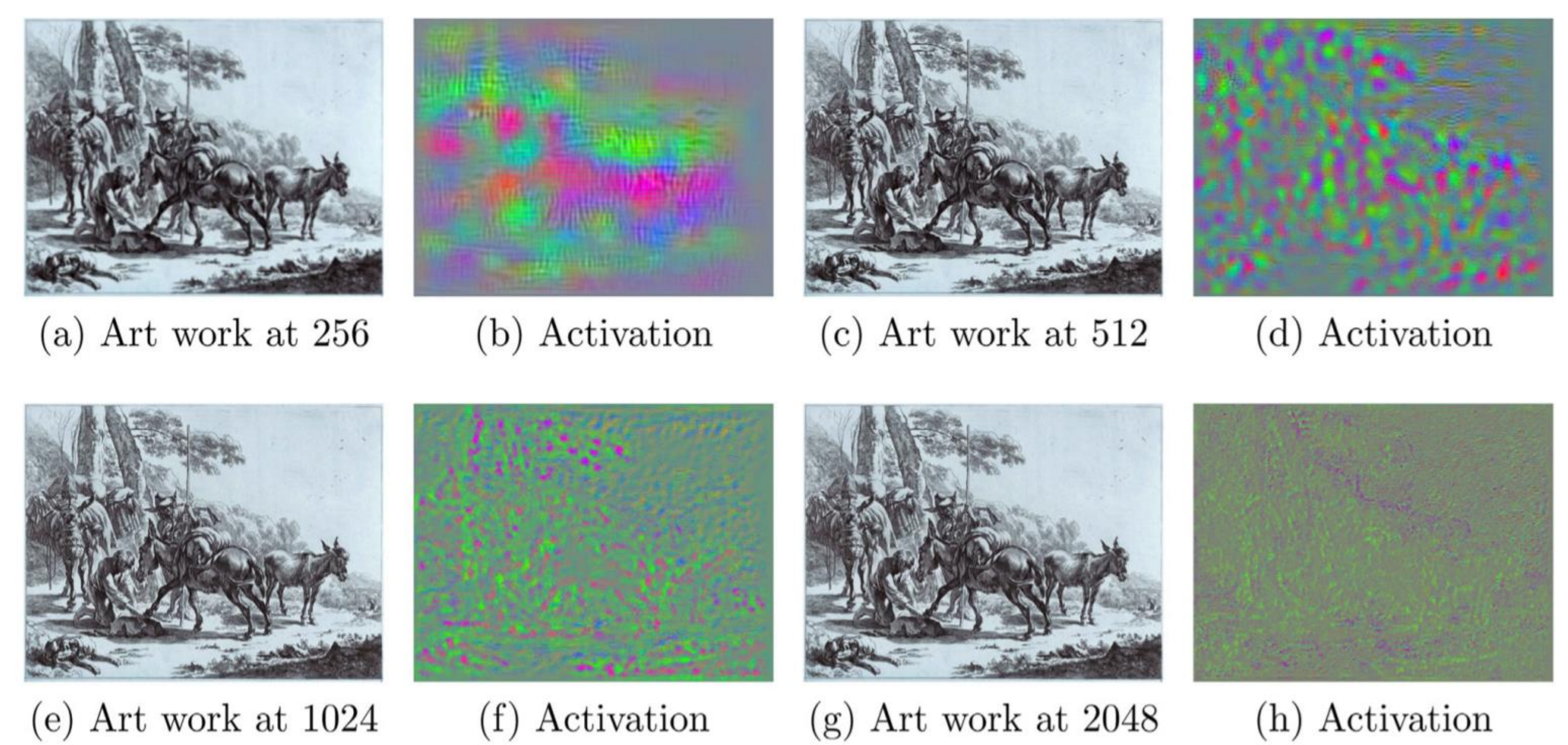
- Covid-19 detection using Convolutional Neural networks





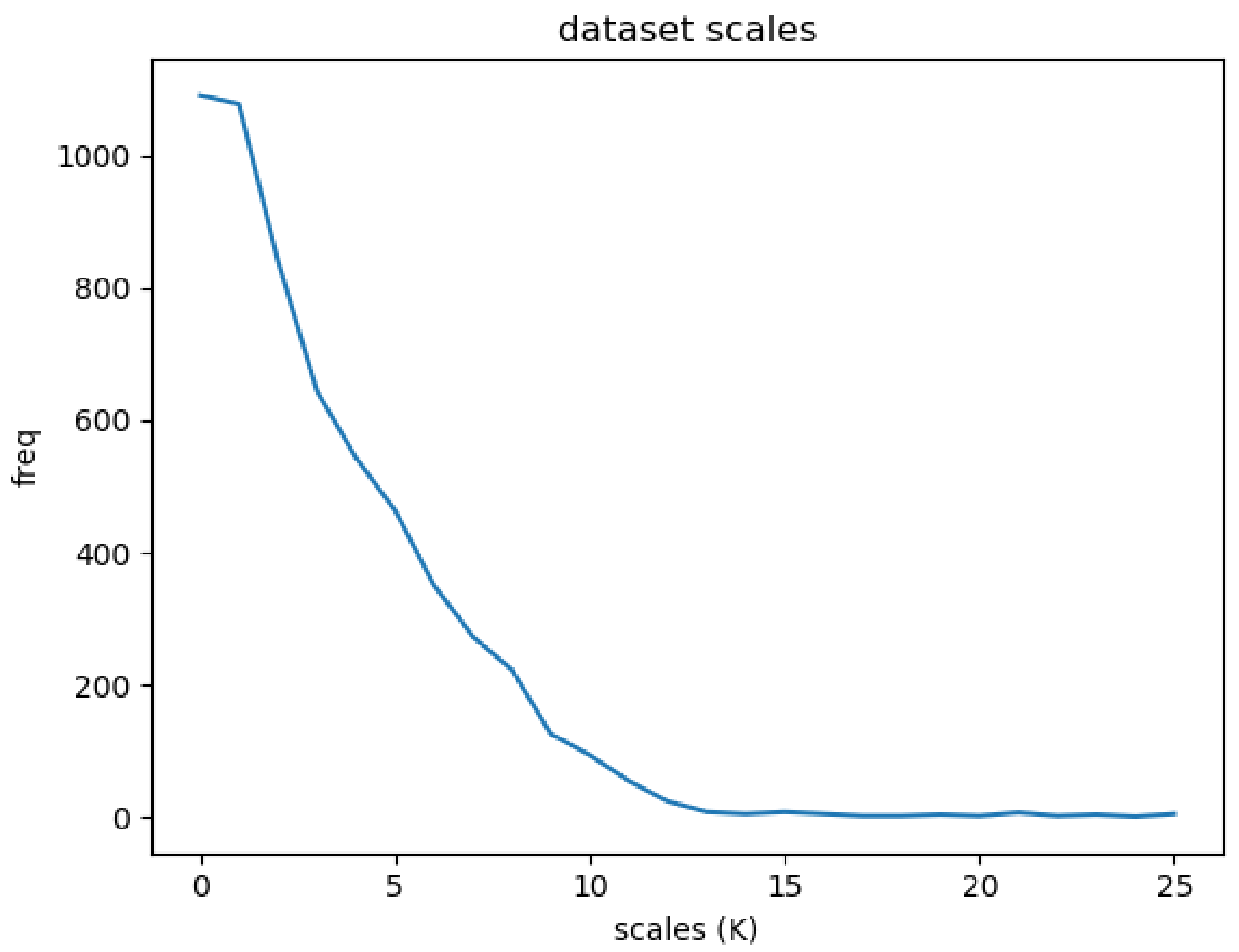
PROBLEM STATEMENT

- Convolutional Neural Network is Scale variant



(a) Art work at 256 (b) Activation (c) Art work at 512 (d) Activation

(e) Art work at 1024 (f) Activation (g) Art work at 2048 (h) Activation



PROBLEM STATEMENT

STATISTICAL DISTRIBUTION
OF THE PNEUMONIA SCALE

PROBLEM STATEMENT

Convolutional Neural Networks Sizes

Name CNN	Parameter Number
Alex-Net	62.3M
VGG-16	138M
ResNet-152	60M
DenseNet-121	7.2M
MobileNet	13M

RESEARCH OBJECTIVE

- Reduce human power gap between covid19 patients and number of radiologists with high accuracy low analysis time.

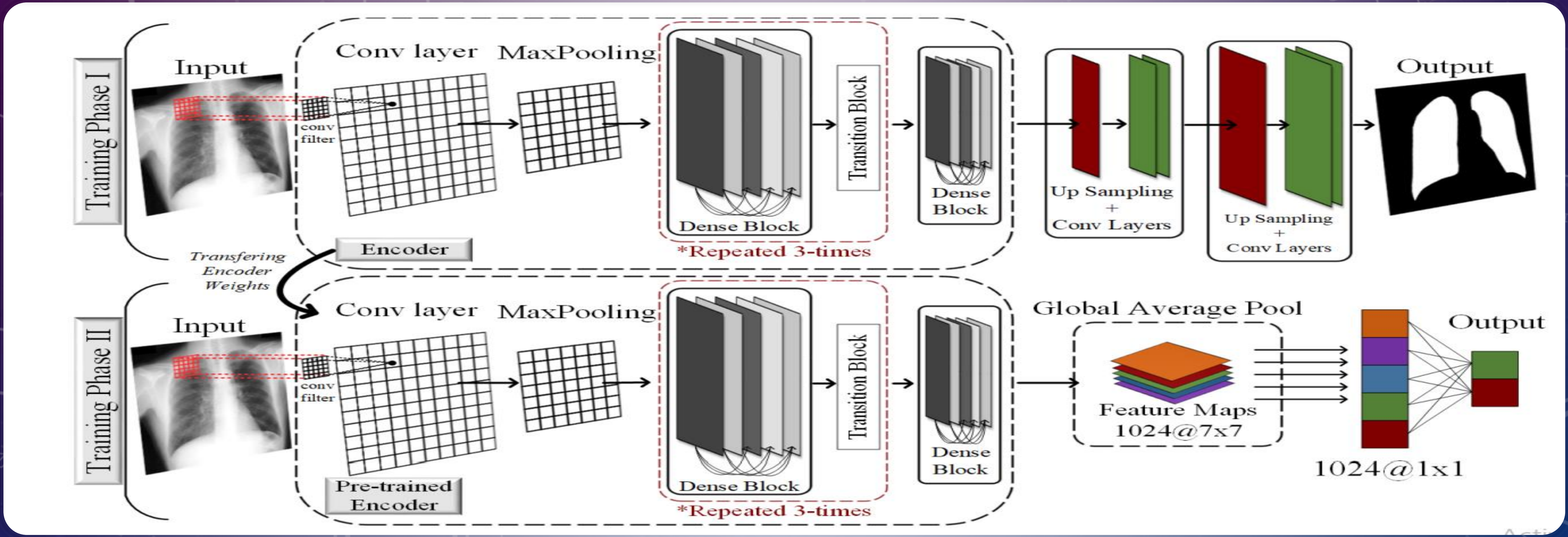
Accuracy

Analysis time

TABLE OF THE CONTENT

RELATED WORK

1) RECOVNET (ARCHITECTURE)



Degerli, Aysen, et al. "Reliable COVID-19 Detection Using Chest X-ray Images." *2021 IEEE International Conference on Image Processing (ICIP)*. IEEE, 2021.

RELATED WORK

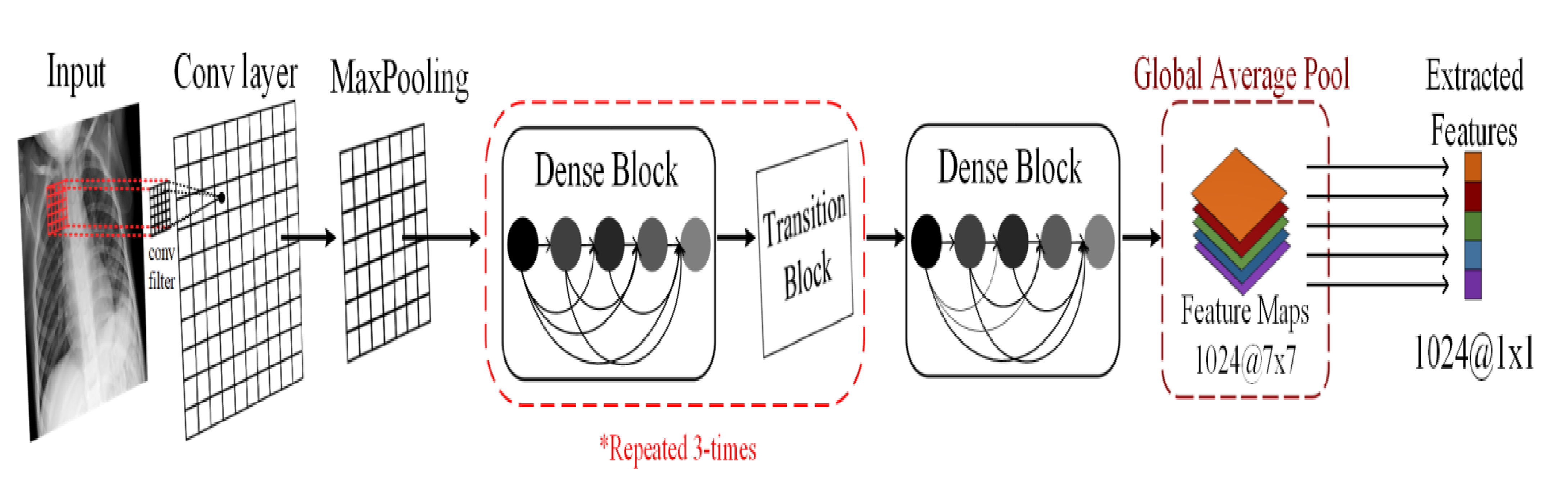
1) RECOVNET (RESULT)

Model	Sensitivity	Specificity	Precision	F1-Score	F2-Score	Accuracy
ResNet-50	96.571	99.953	98.734	97.641	96.996	99.828
Inception-v3	94.762	99.821	95.307	95.033	94.870	99.634
Inception-ResNet-v2	94.286	99.803	94.828	94.556	94.394	99.599
DenseNet-121	97.429	**99.974**	**99.320**	**98.365**	**97.801**	**99.880**
ReCovNet-v1	97.810	99.901	97.438	97.624	97.735	99.824
ReCovNet-v2	**98.571**	99.770	94.262	96.369	97.678	99.726

Degerli, Aysen, et al. "Reliable COVID-19 Detection Using Chest X-ray Images." *2021 IEEE International Conference on Image Processing (ICIP)*. IEEE, 2021.

RELATED WORK

2) Ahishali et al. (Architecture)



M.Ahishali et al., "Advance Warning Methodologies for COVID-19 Using Chest X-Ray Images," in IEEE Access, vol. 9, 2021

RELATED WORK

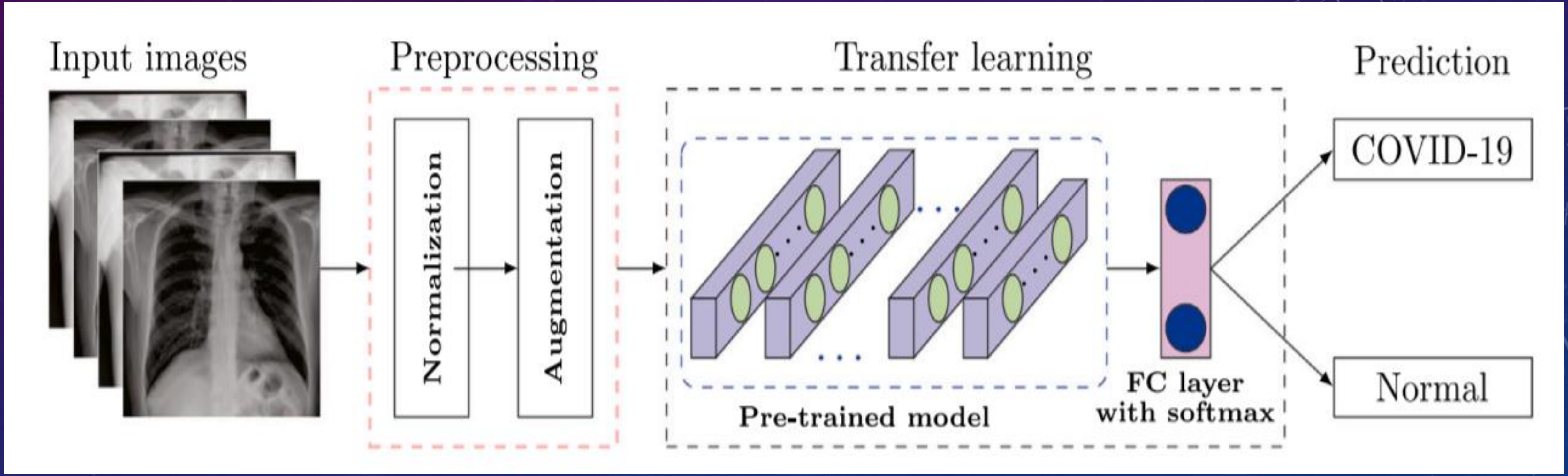
2) Ahishali et al. (Results)

Method	Accuracy	Sensitivity	Specificity
SRC-Dalm	0.9818 ± 0.006	0.9371 ± 0.036	0.9867 ± 0.006
SRC-Hom.	0.9481 ± 0.010	0.8171 ± 0.057	0.9626 ± 0.009
CRC-light	0.9783 ± 0.007	0.9486 ± 0.033	0.9816 ± 0.007
CRC	0.9823 ± 0.006	0.9657 ± 0.027	0.9842 ± 0.006
CSEN1	0.9635 ± 0.009	0.9886 ± 0.016	0.9607 ± 0.010
CSEN2	0.9248 ± 0.012	0.9943 ± 0.011	0.9171 ± 0.014
ReconNet	0.9424 ± 0.011	0.9943 ± 0.011	0.9367 ± 0.012
MLP	0.9584 ± 0.009	0.9371 ± 0.036	0.9607 ± 0.010
SVM	0.9681 ± 0.008	0.9657 ± 0.027	0.9683 ± 0.009
k-NN	0.9458 ± 0.011	0.9257 ± 0.039	0.9481 ± 0.011

M.Ahishali et al., "Advance Warning Methodologies for COVID-19 Using Chest X-Ray Images," in IEEE Access, vol. 9, 2021

RELATED WORK

3) Nayak, Soumya Ranjan, et al. (Architecture)



Nayak, Soumya Ranjan, et al. "Application of deep learning techniques for detection of COVID-19 cases using chest X-ray images: A comprehensive study." *Biomedical Signal Processing and Control* 64 (2021)

RELATED WORK

3) Nayak, Soumya Ranjan, et al. (Results)

Architectural descriptions of the pre-trained CNN models used in this study.

Model	Layers	Parameters (in million)	Input layer size	Output layer size
AlexNet	8	60	(224,224,3)	(2,1)
VGG-16	16	138	(224,224,3)	(2,1)
GoogleNet	22	5	(224,224,3)	(2,1)
MobileNet-V2	53	3.4	(224,224,3)	(2,1)
SqueezeNet	18	1.25	(224,224,3)	(2,1)
ResNet-34	34	21.8	(224,224,3)	(2,1)
ResNet-50	50	25.6	(224,224,3)	(2,1)
Inception-V3	42	24	(299,299,3)	(2,1)

Nayak, Soumya Ranjan, et al. "Application of deep learning techniques for detection of COVID-19 cases using chest X-ray images: A comprehensive study." *Biomedical Signal Processing and Control* 64 (2021)

RELATED WORK

3) Nayak, Soumya Ranjan, et al. (Results)

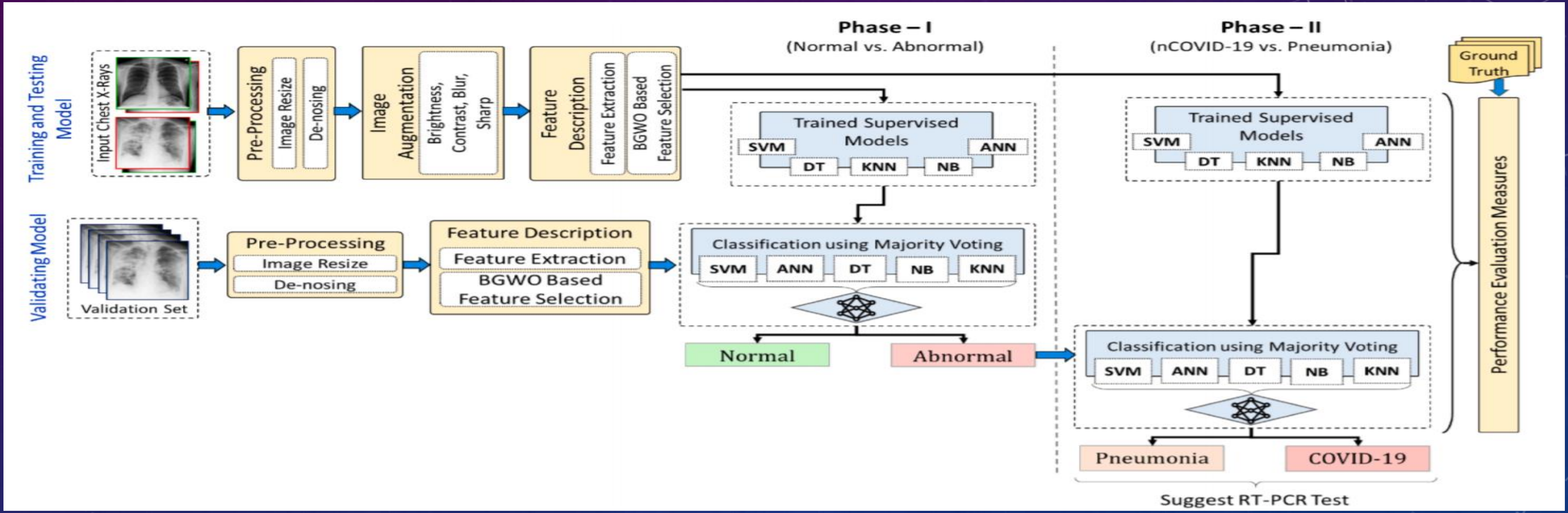
Classification results comparison of all eight CNN models.

Model	Pre (%)	Sen (%)	Spe (%)	F1-score	Acc (%)	AUC
ResNet-34	96.77	100.00	96.67	0.9836	98.33	0.9836
ResNet-50	95.24	100.00	95.00	0.9756	97.50	0.9731
GoogleNet	96.67	96.67	96.67	0.9667	96.67	0.9696
VGG-16	95.08	96.67	95.00	0.9587	95.83	0.9487
AlexNet	96.72	98.33	96.67	0.9752	97.50	0.9642
MobileNet-V2	98.24	93.33	98.33	0.9573	95.83	0.9506
Inception-V3	96.36	88.33	96.67	0.9217	92.50	0.9342
SqueezeNet	98.27	95.00	98.33	0.9661	96.67	0.9705

Nayak, Soumya Ranjan, et al. "Application of deep learning techniques for detection of COVID-19 cases using chest X-ray images: A comprehensive study." *Biomedical Signal Processing and Control* 64 (2021)

RELATED WORK

4) Chandra, Tej Bahadur, et al. (Architecture)



Chandra, Tej Bahadur, et al. "Coronavirus disease (COVID-19) detection in Chest X-Ray images using majority voting based classifier ensemble." Expert systems with applications 165 (2021)

RELATED WORK

4) Chandra, Tej Bahadur, et al. (Results)

Two class (normal vs. abnormal) performance comparison of the proposed method with the state of art methods.

Articles	Class	Algorithms/techniques	ACC (%)
Panwar et al. (2020)	2	nCOVnet	88.10
Hemdan, Shouman, and Karar (2020)	2	VGG19, DenseNet201, ResNetV2, InceptionV3, InceptionResNetV2, Xception, MobileNetV2	90.00
Maghdid, Asaad, Ghafoor, Sadiq, and Khan (2020)	2	AlexNet, Modified CNN	94.00
Narin et al. (2020)	2	ResNet50, ResNetV2, InceptionV3	98.00
Ozturk et al. (2020)	2	DarkNet	98.08
Proposed Method (Phase-I)	**2**	**Majority vote based classifier ensemble**	**98.06**

Three class (normal, nCOVID-19 and pneumonia) performance comparison of the proposed method with the state of art methods.

Articles	Class	Algorithms/techniques	Total number of images	ACC (%)
Ozturk et al. (2020)	3	DarkNet		87.02
L. Wang et al. (2020)	3	COVID-Net		93.30
Abbas et al. (2020)	3	DeTraC	196	95.12
Chowdhury et al. (2020)	3	AlexNet, ResNet18, DenseNet201, SqueezeNet	3487	97.94
Ucar and Korkmaz (2020)	3	Deep Bayes-SqueezeNe	5957	98.26
Nour et al. (2020)	3	CNN, SVM, DT, KNN	3670	98.97
Toğaçar et al. (2020)	3	MobileNetV2, SqueezeNet, SVM	458	99.27
Proposed Method (Overall)	**3**	**Majority vote based classifier ensemble**	**2346**	**93.41**

Chandra, Tej Bahadur, et al. "Coronavirus disease (COVID-19) detection in Chest X-Ray images using majority voting based classifier ensemble." Expert systems with applications 165 (2021)

RELATED WORK

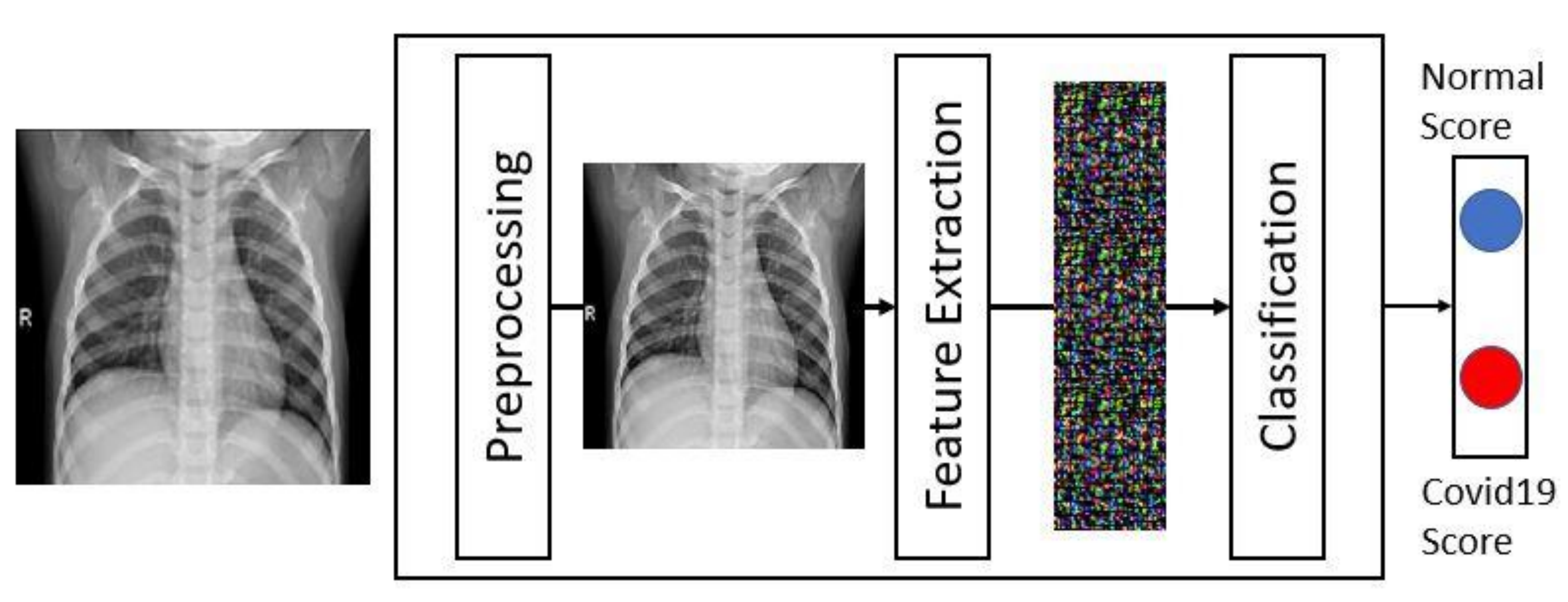
Disadvantages :

- Does not handle scale variant problem.
- Computationally expensive (High Analysis Time).
- Increase Accuracy for imbalanced dataset.

TABLE OF THE CONTENT

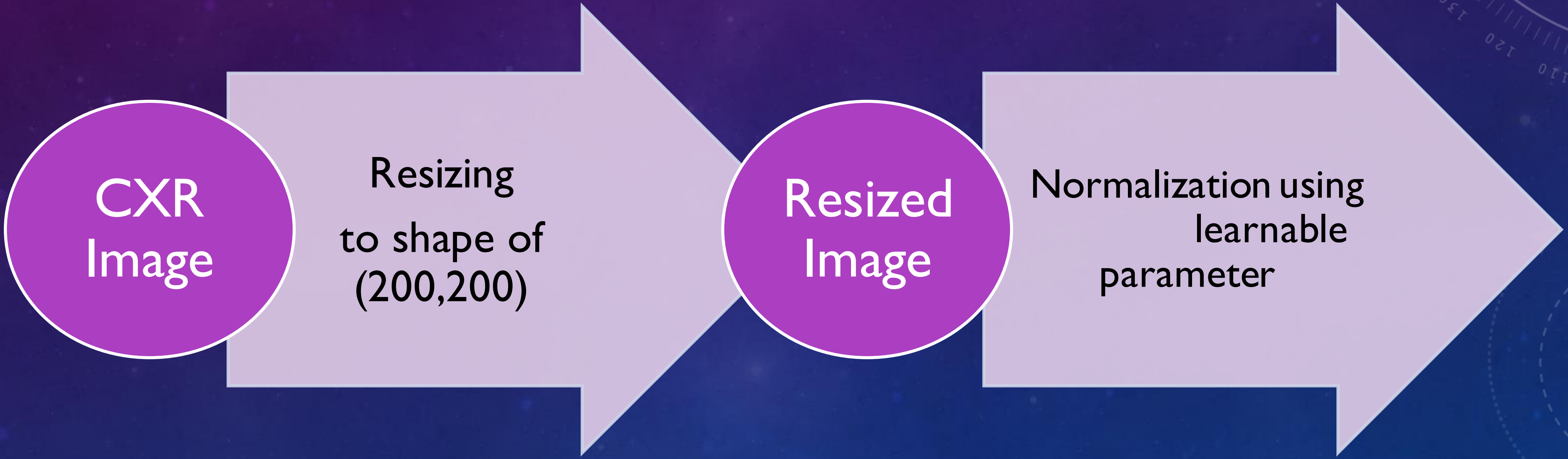
PROPOSAL I

Methodology



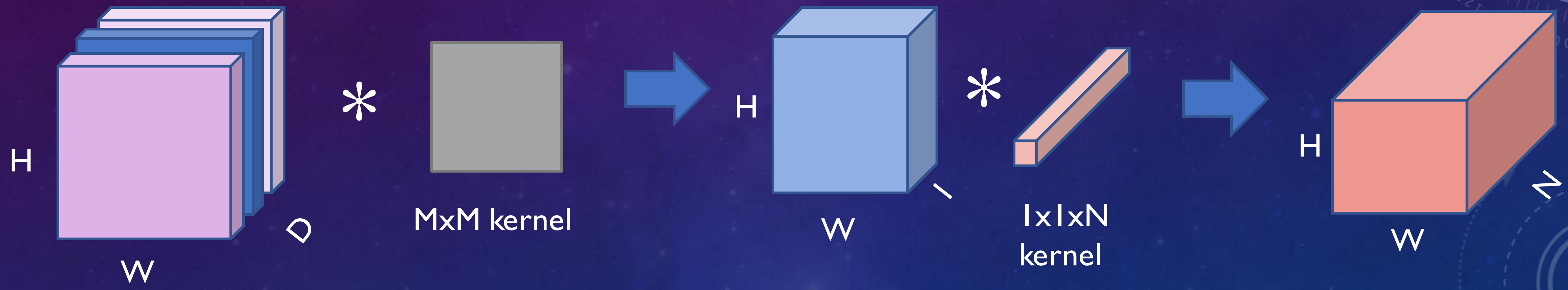
PROPOSAL I

Methodology (preprocesing)



PROPOSAL I

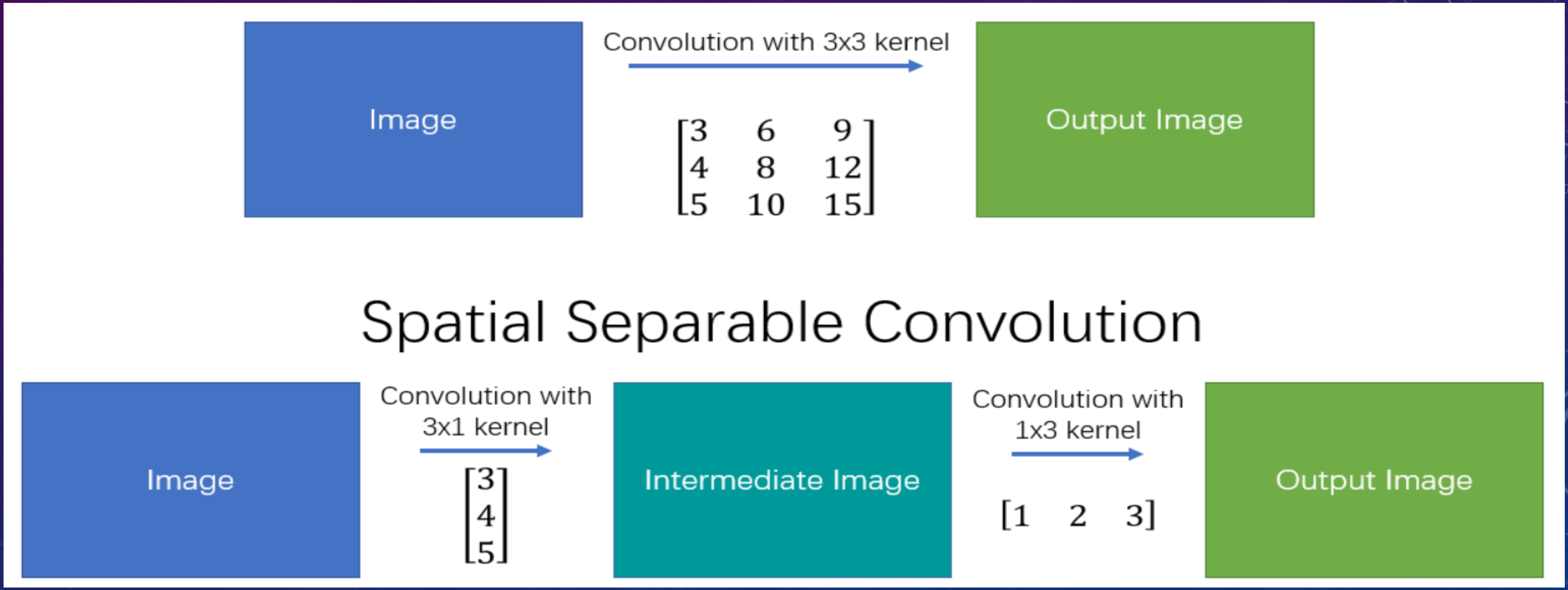
Methodology (Feature Extraction)



Depth-wise Separable Convolution

PROPOSAL I

Methodology (Feature Extraction)

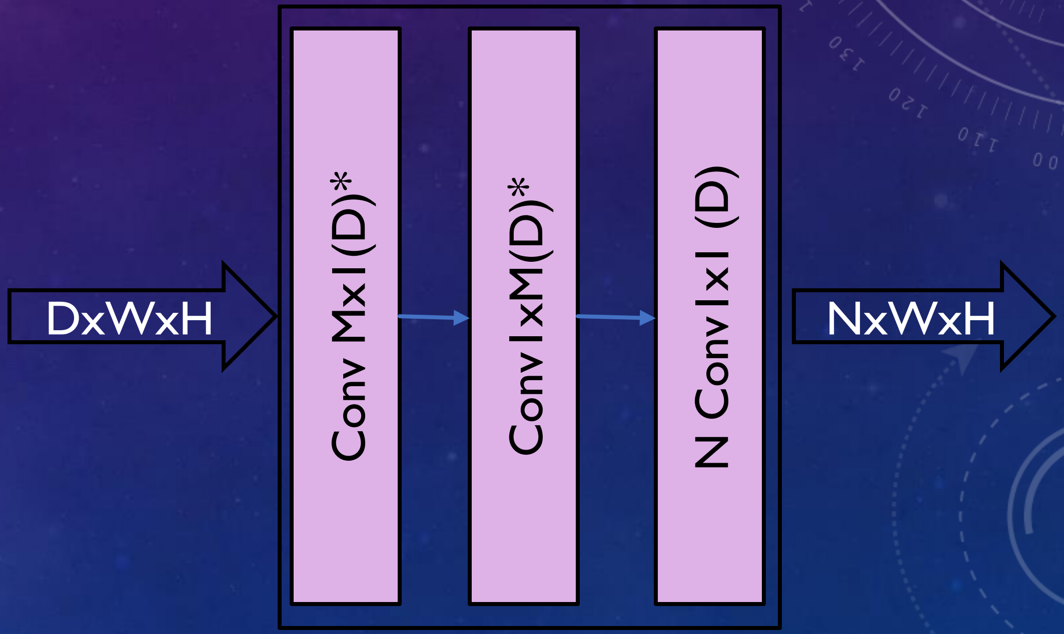


Spatial Separable Kernels

PROPOSAL I

Methodology (Feature Extraction)

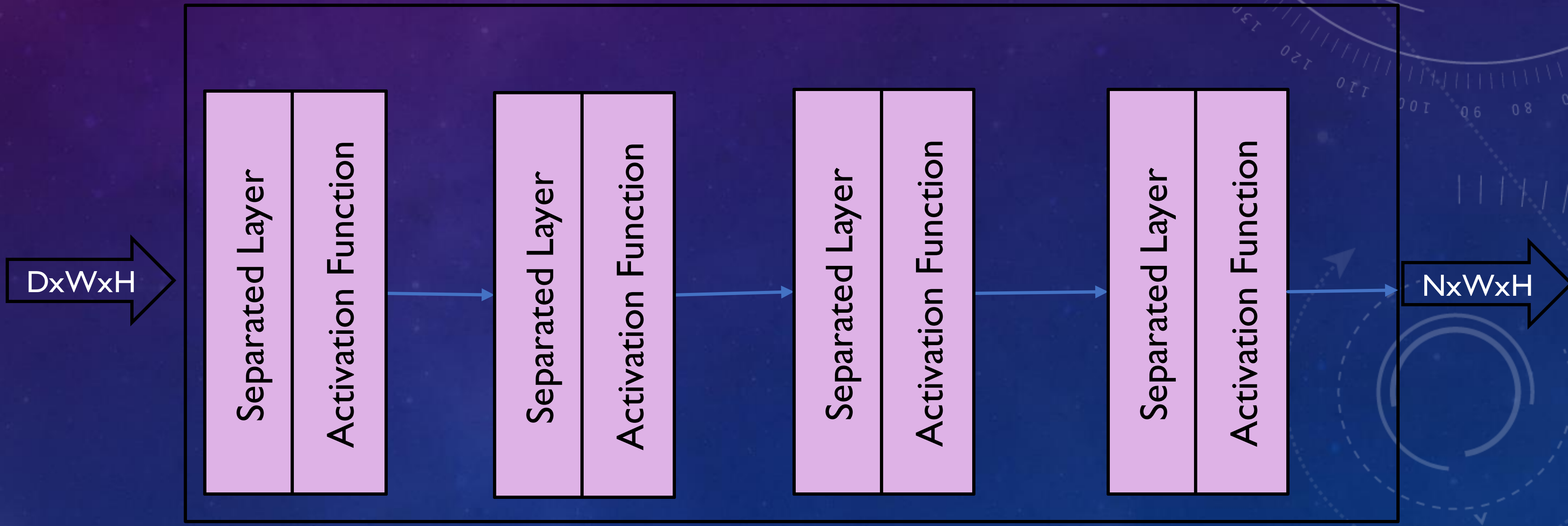
- Advantage
  - Apply kernel separability and depthwise separable convolutions
    - Reduces model overfitting by forcing the model to learn separable kernels.
    - Fast processing
  - Lower parameter Count.
    - Reduces model overfitting
- Drawback
  - Too Simple to detect Complex pattern



PROPOSAL I

Methodology (Feature Extraction)

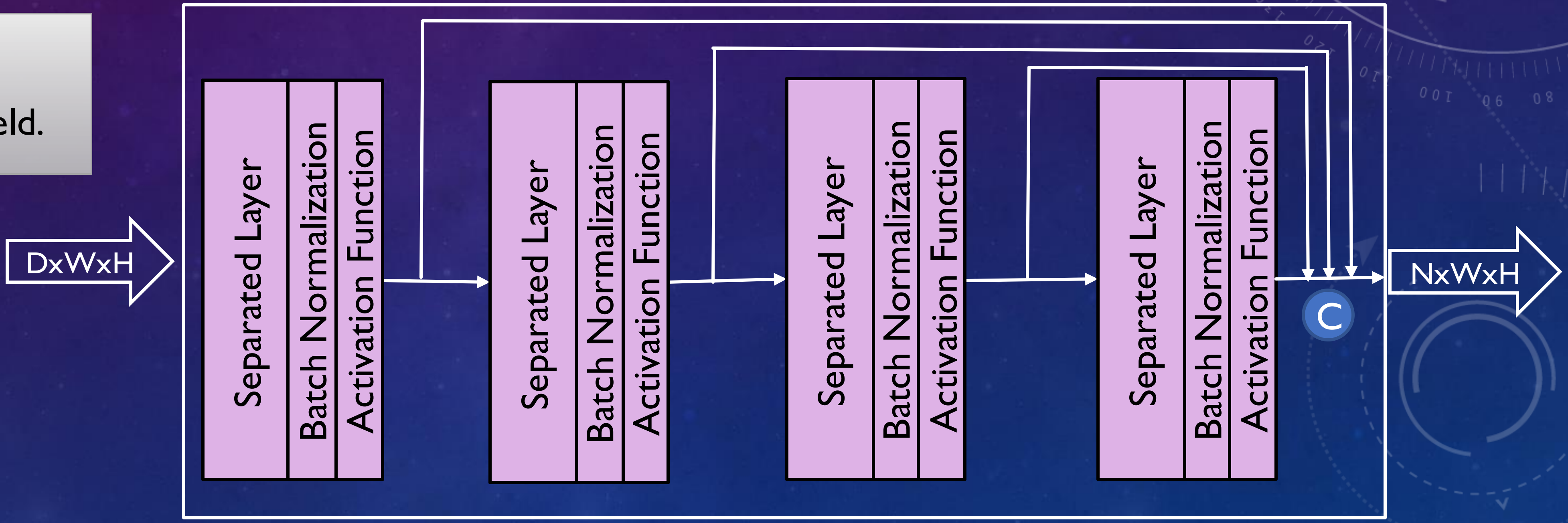
- Drawback
  - High Covariate Shift
  - Risk of model overfitting



PROPOSAL I

Methodology (Feature Extraction)

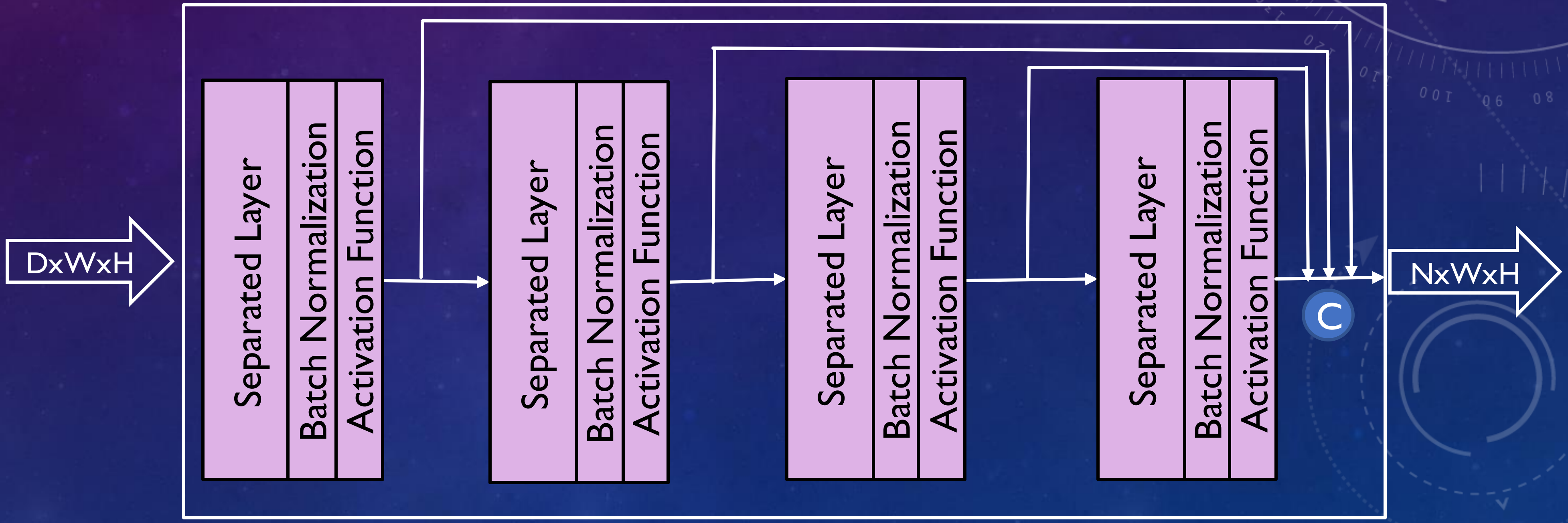
- Drawback
    - Small Receptive field.



PROPOSAL I

Methodology (Feature Extraction)

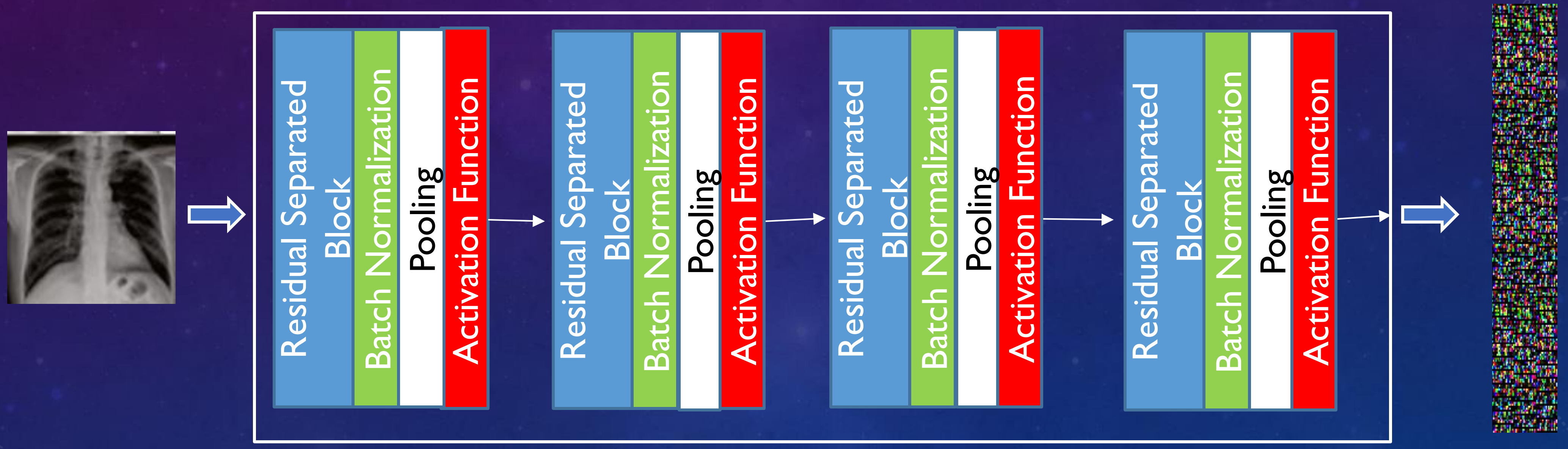
- Advantage
  - Large Receptive field.
  - Reduces overfitting.
  - Reduces the vanishing gradient problem.
  - Reduces covariate shift.



residual separated block
Kernel Size for each Separated Layer is 3, 5, 7, and 9 Respectively
C : is a channel wise concatenation

PROPOSAL I

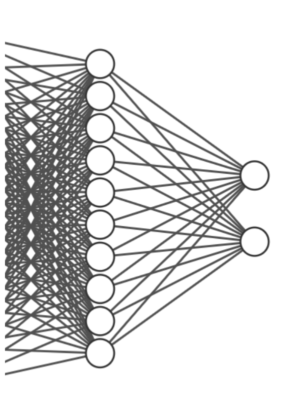
Methodology (Feature Extraction)



PROPOSAL I

Methodology (Feature Extraction)

Layer Number	Layer Size	Activation Function
RSBLayer1	4×16	LeakyReLU
RSBLayer2	4×23	LeakyReLU
RSBLayer3	4×64	LeakyReLU
RSBLayer4	4×64	LeakyReLU
RSBLayer5	4×64	LeakyReLU
RSBLayer6	4×16	LeakyReLU
Flatten The Feature maps to 1D 576 feature vector		
LinearLayer1	64	LeakyReLU
LinearLayer2	64	LeakyReLU
LinearLayer3	2	Softmax



EXPERIMENTAL RESULTS I

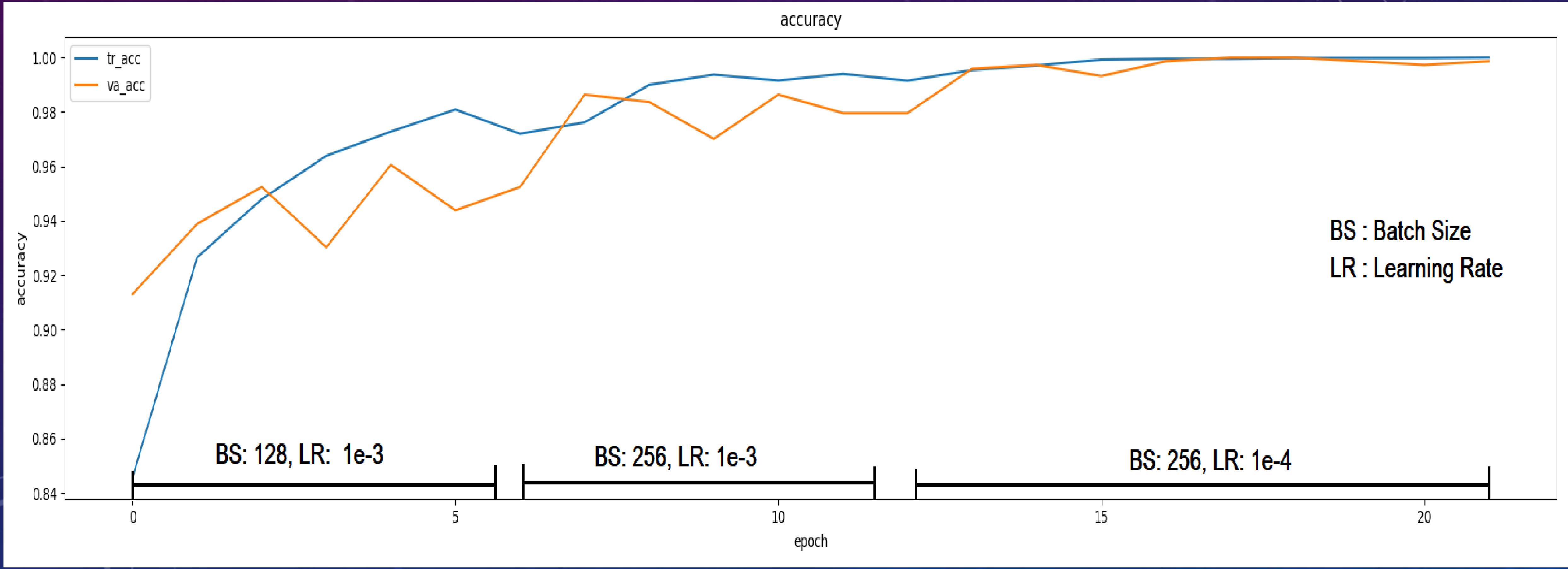
Data Set and Machine specification

- benchmark QaTa-COV19 dataset is used:
  - 8284 Chest X-Ray image is used for training
    - first 4142 normal Chest X-Ray samples
    - All 4142 Covid-19 samples
  - 922 Chest X-Ray image is used for testing
    - 461 normal Chest X-Ray samples
    - 461 Covid-19 samples

- Experiments are conducted on :
  - Lenovo Z50-70,
  - Intel CORE i7-4510U CPU 2.00 GHz,
  - 8GB RAM,
  - NVIDIA GeForce 840M GPU;
  - and with python and PyTorch library.

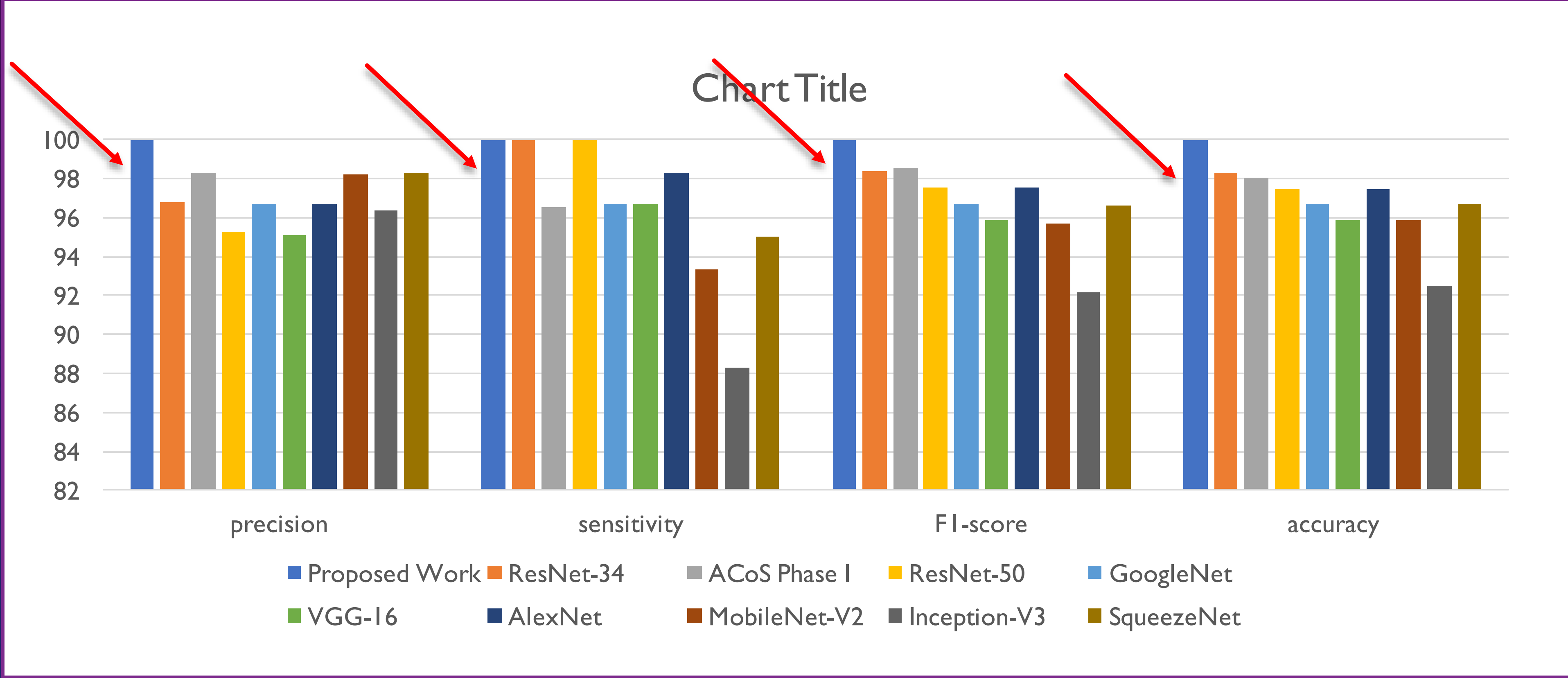
EXPERIMENTAL RESULTS I

Training



EXPERIMENTAL RESULTS I

Model Evaluation



EXPERIMENTAL RESULTS I

Model Evaluation

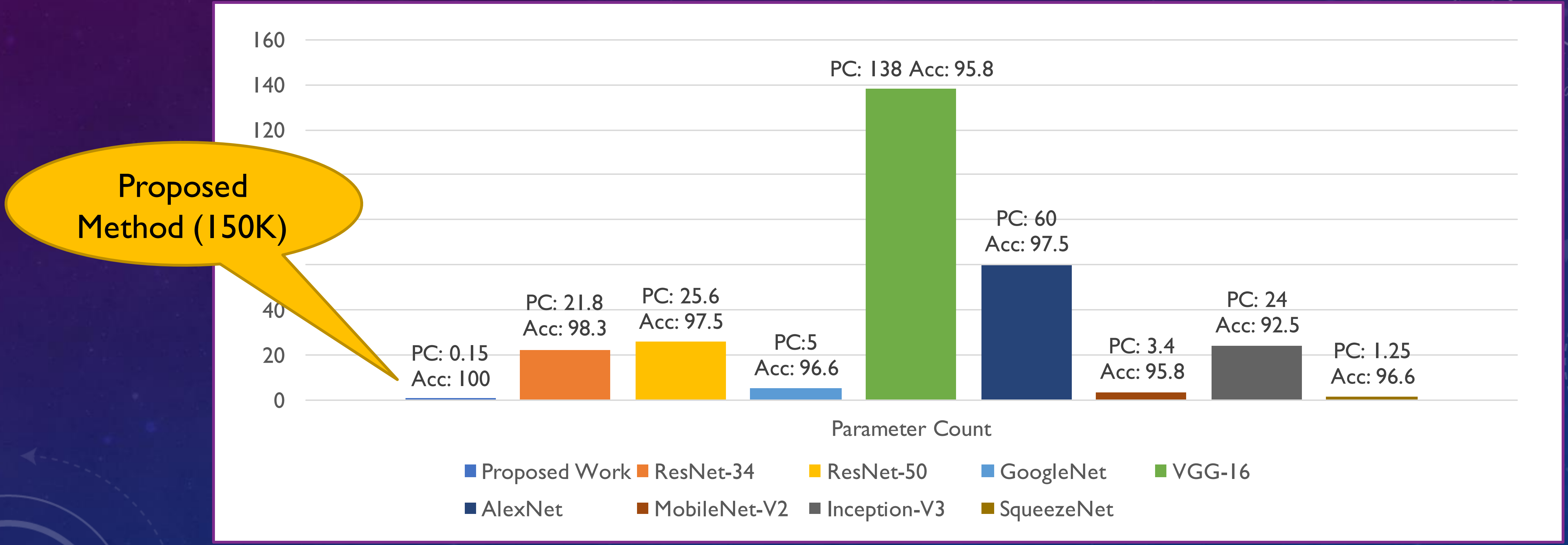
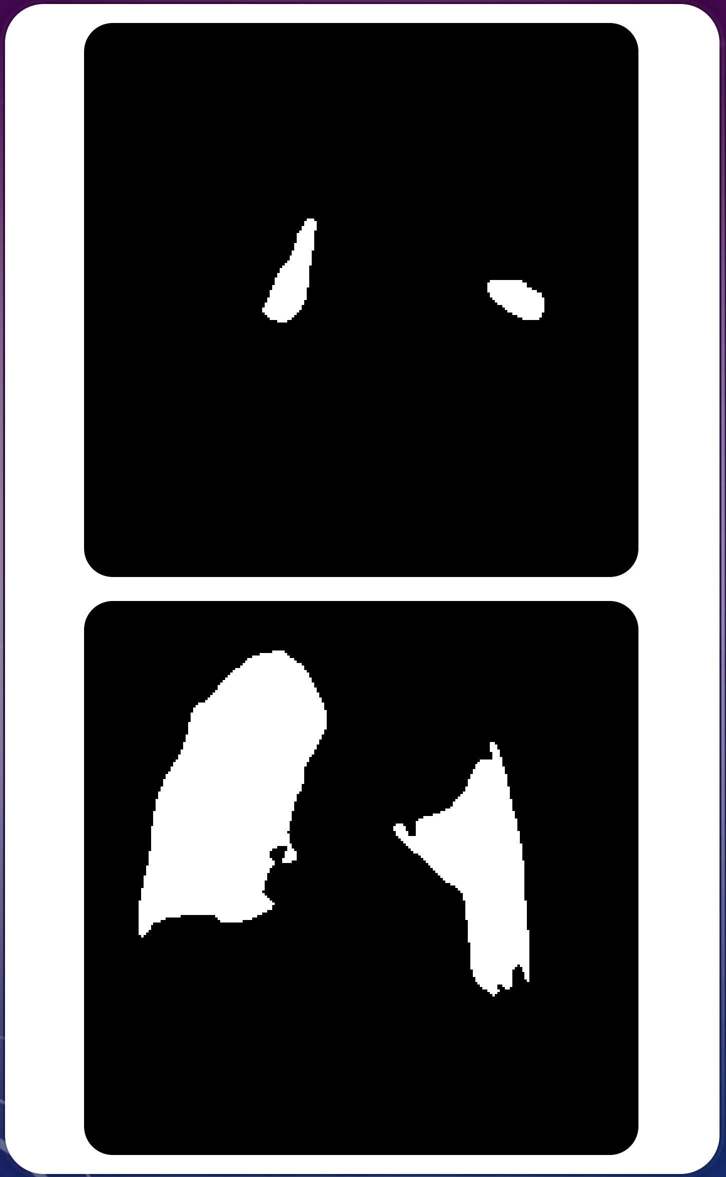


TABLE OF THE CONTENT

- Introduction
- Problem Statement
- Related Work
- Proposed Work I
- Proposed Work II
- Conclusion

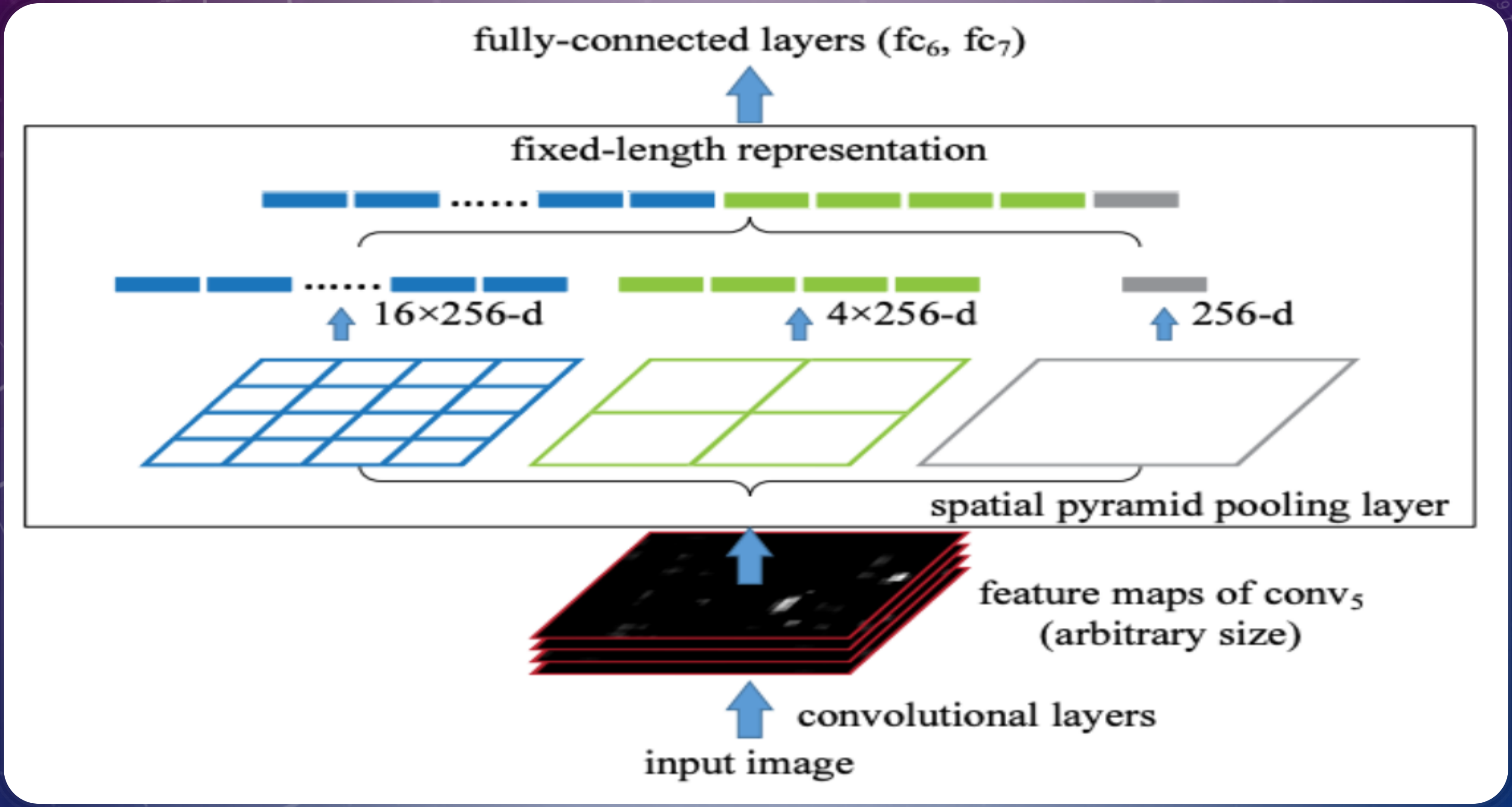


OBJECTIVE OF PROPOSAL II

- Pneumonia scale change from one sample to other.

PROPOSAL II

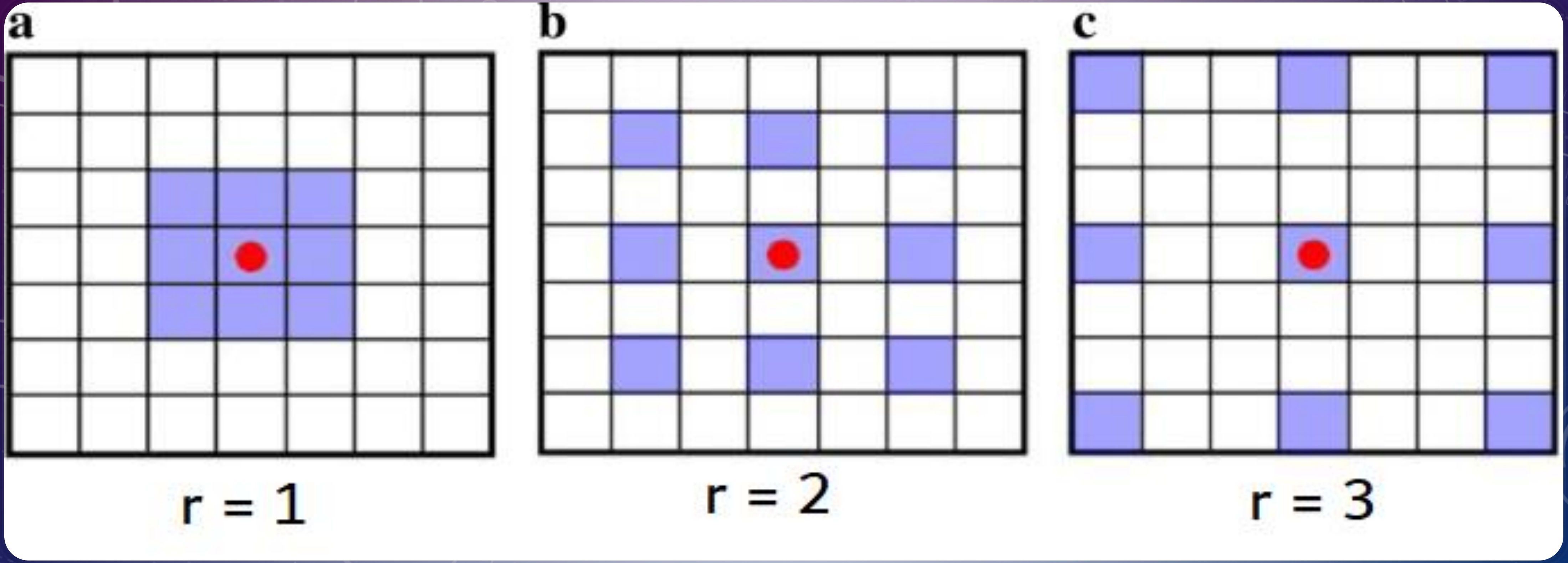
SPATIAL PYRAMID POOLING NETWORK



He, Kaiming, et al. "Spatial pyramid pooling in deep convolutional networks for visual recognition." *IEEE transactions on pattern analysis and machine intelligence* 37.9 (2015)

PROPOSAL II

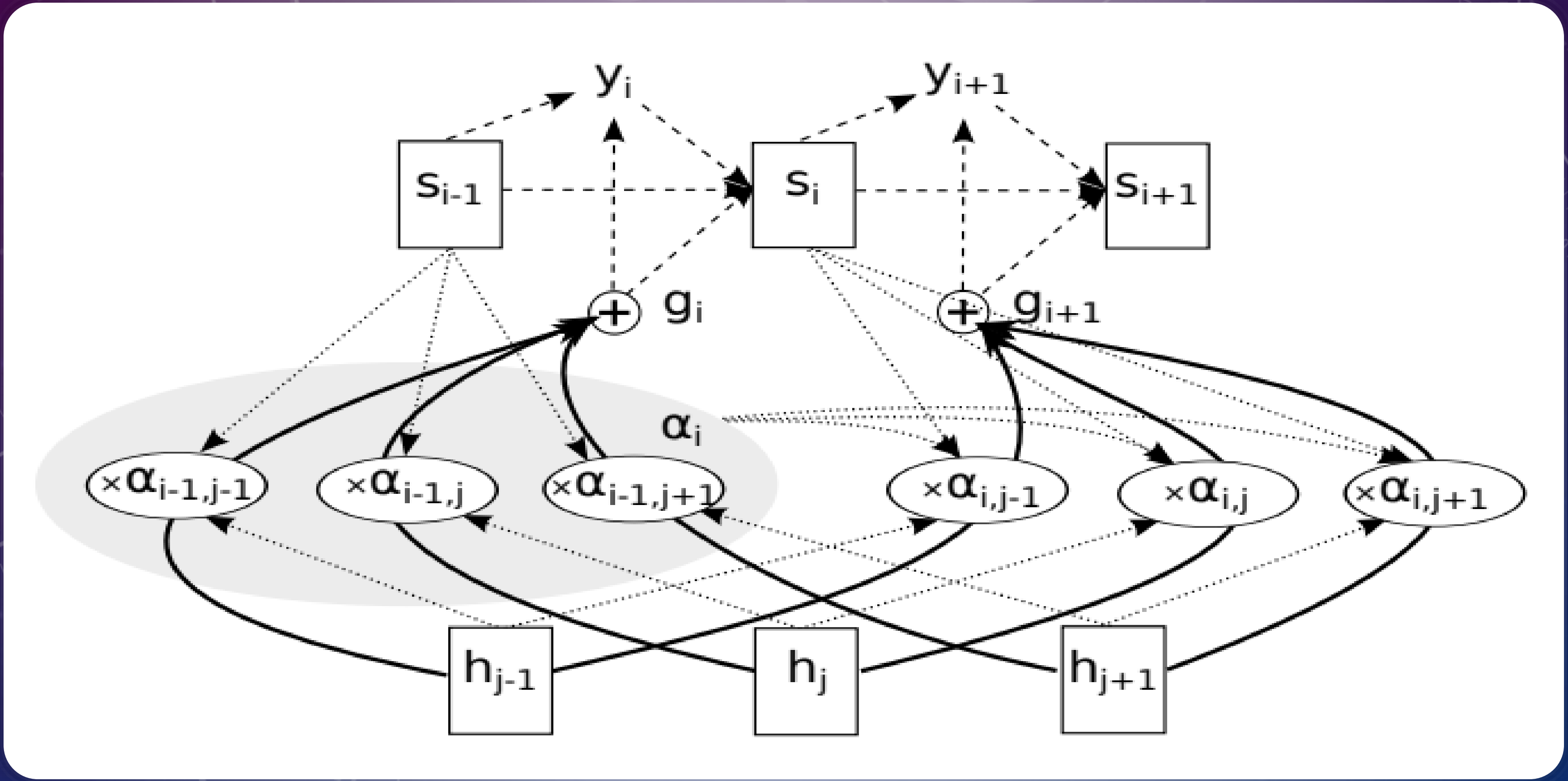
DILATED FILTERS (ATROUS CONVOLUTION)



Chen, Liang-Chieh, et al. "Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs." *IEEE transactions on pattern analysis and machine intelligence* 40.4 (2017)

PROPOSAL II

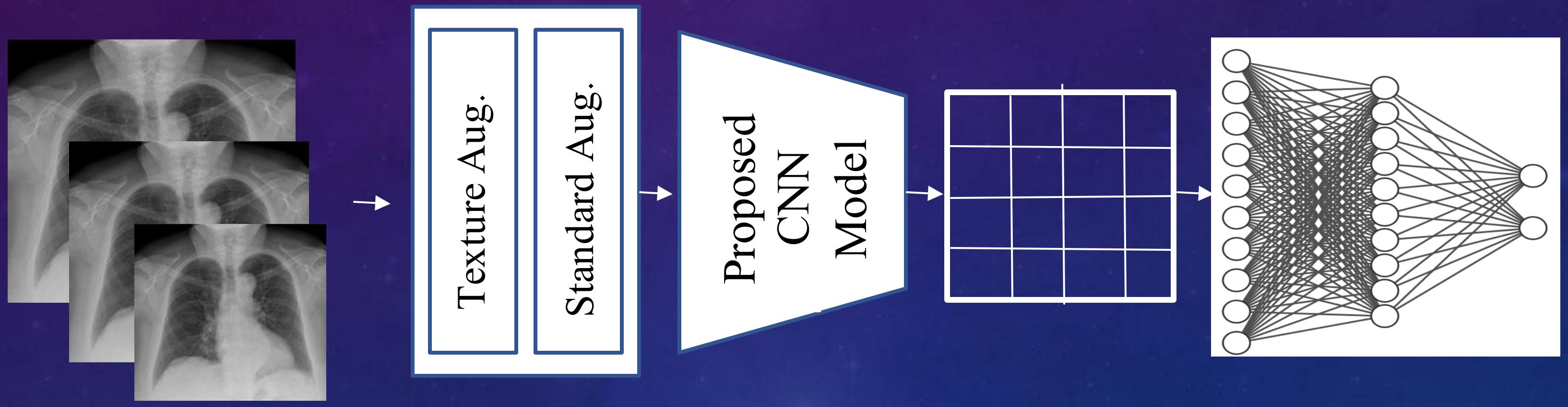
ATTENTION SEQ2SEQ



Chorowski, J. K., Bahdanau, D., Serdyuk, D., Cho, K., & Bengio, Y. (2015). Attention-based models for speech recognition. *Advances in neural information processing systems*

PROPOSAL II

Methodology



PROPOSAL II

Methodology (Texture Augmentation)

PROPOSAL II

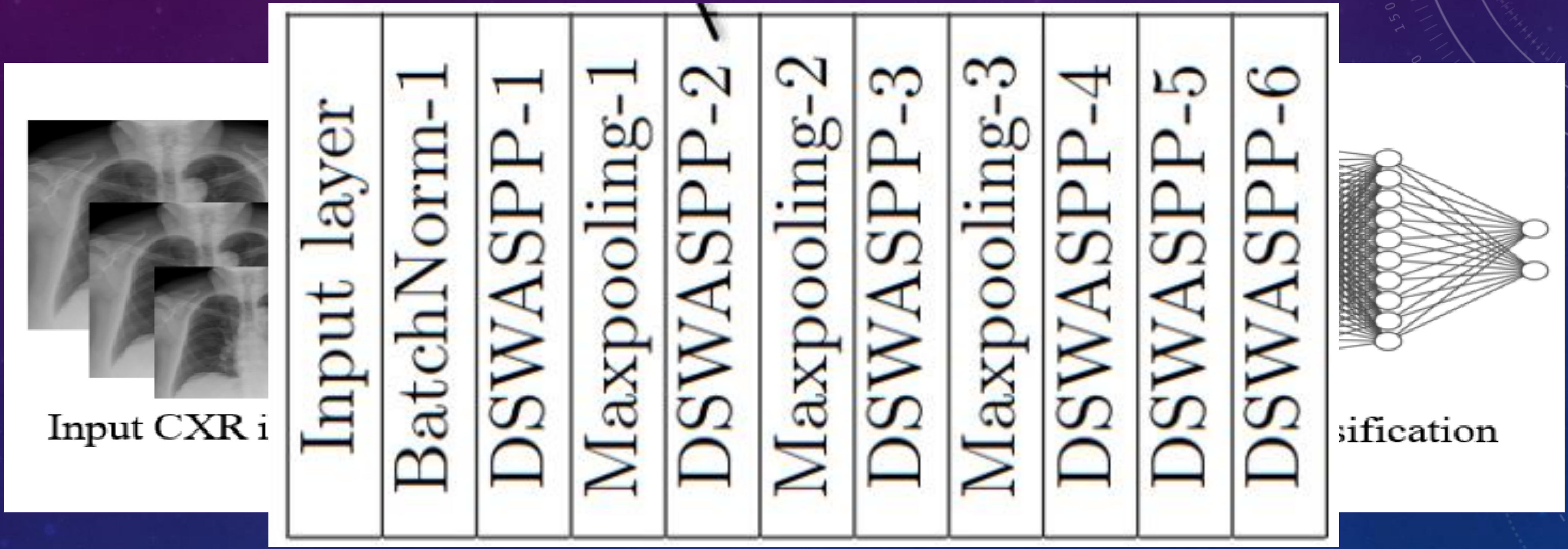
Methodology

- Network design principle
  - Any linear combination is followed by Batch Norm.[9]
  - ReLU is used for as activation function.[10]
  - For the same block Dense Residual Connection is included.[11]
  - Dropout layer is added as a first layer of FullyConn layers.[12]
  - Max Norm Constrain.
  - Thin architectures and small Kernel size (3x3).

[9] Ioffe, Sergey, and Christian Szegedy. "Batch normalization: Accelerating deep network training by reducing internal covariate shift." *International conference on machine learning*. PMLR,
[10] Krizhevsky, Alex, Ilya Sutskever, and Geoffrey E. Hinton. "Imagenet classification with deep convolutional neural networks." *Advances in neural information processing systems* 25 (2012).
[11] Zhang, Yulun, et al. "Residual dense network for image super-resolution." *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2018.
[12] Srivastava, Nitish, et al. "Dropout: a simple way to prevent neural networks from overfitting." *The journal of machine learning research* 15.1 (2014)

PROPOSAL II

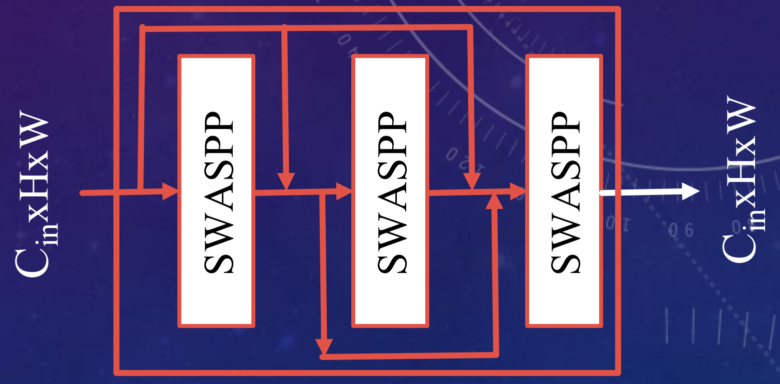
Methodology (Proposed Architecture)

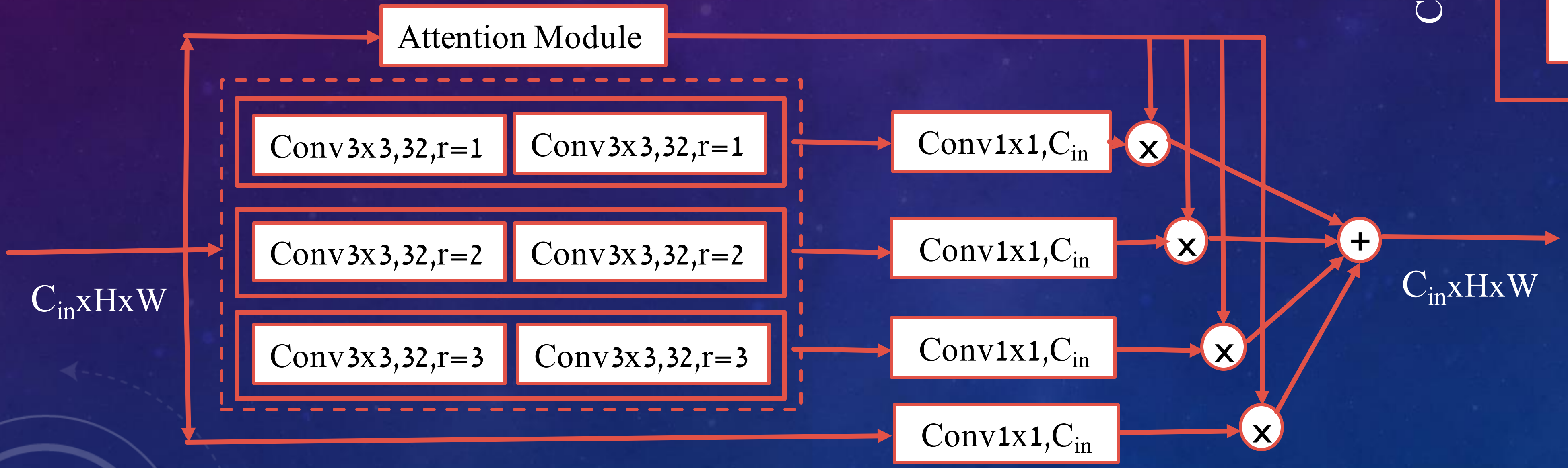


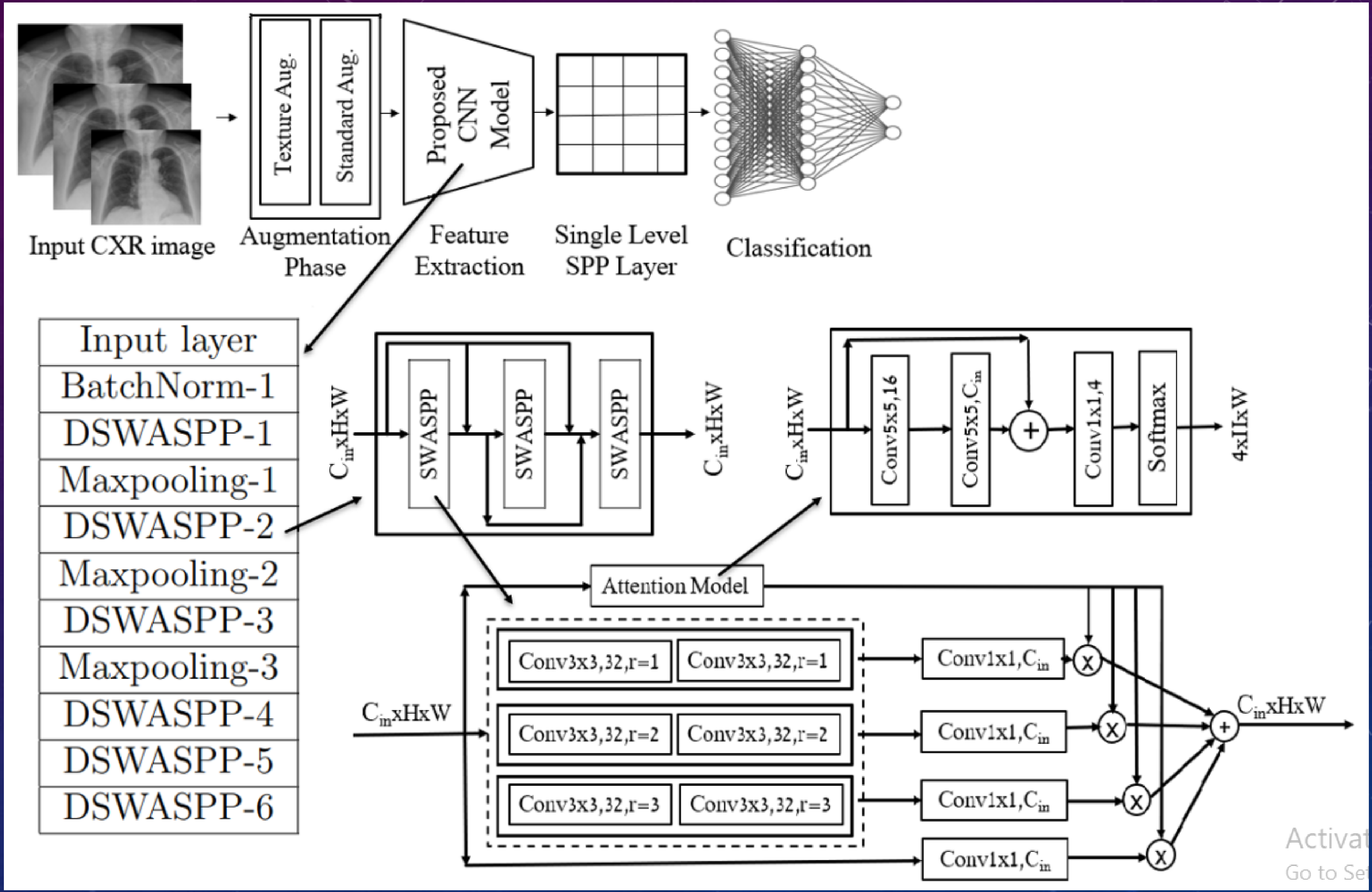
PROPOSAL II

Methodology (Proposed Architecture)

- Dense spatial weighted Atrous Spatial pyramaid pooling (DSWASPP)






Input CXR image
Texture Aug.
Standard Aug.
Augmentation Phase
Proposed CNN Model
Feature Extraction
Single Level SPP Layer
Classification
Input layer
BatchNorm-1
DSWASPP-1
Maxpooling-1
DSWASPP-2
Maxpooling-2
DSWASPP-3
Maxpooling-3
DSWASPP-4
DSWASPP-5
DSWASPP-6
C_{in}xHxW
SWASPP
SWASPP
SWASPP
C_{in}xHxW
C_{in}xHxW
Conv5x5,16
Conv5x5,C_{in}
+
Conv1x1,4
Softmax
4xHxW
Attention Model
C_{in}xHxW
Conv3x3,32,r=1
Conv3x3,32,r=1
Conv3x3,32,r=2
Conv3x3,32,r=2
Conv3x3,32,r=3
Conv3x3,32,r=3
Conv1x1,C_{in}
Conv1x1,C_{in}
Conv1x1,C_{in}
Conv1x1,C_{in}
X
X
X
X
+
C_{in}xHxW

PROPOSAL II

Methodology
(overall Architecture)

Layer Name	Proposed CNN Architecture		
	Input Shape	*Output Shape*	*Param. Count*
Input layer	-	$1 \times 320 \times 320$	0
BatchNorm-1	$1 \times 320 \times 320$	$1 \times 320 \times 320$	2
DSWASPP-1	$1 \times 320 \times 320$	$32 \times 320 \times 320$	121,035
Maxpooling-1	$32 \times 320 \times 320$	$32 \times 160 \times 160$	0
DSWASPP-2	$32 \times 160 \times 160$	$64 \times 160 \times 160$	298,236
Maxpooling-2	$64 \times 160 \times 160$	$64 \times 80 \times 80$	0
DSWASPP-3	$64 \times 80 \times 80$	$128 \times 80 \times 80$	604,956
Maxpooling-3	$128 \times 80 \times 80$	$128 \times 40 \times 40$	0
DSWASPP-4	$128 \times 80 \times 80$	$128 \times 80 \times 80$	784,092
DSWASPP-5	$128 \times 80 \times 80$	$128 \times 80 \times 80$	784,092
DSWASPP-6	$128 \times 80 \times 80$	$128 \times 80 \times 80$	784,092
SPP-1	$128 \times 80 \times 80$	12800	0
Dropout-1	12800	12800	0
FC-1	12800	128	1,638,528
FC-2	128	128	16,512
FC-3	128	64	8,256
FC-4	64	2	130
Softmax	2	2	0
Total Number of Parameter			5,040,571

Any linear combination is followed by BN and leakyReLU nonlinearity excluding re-projection layer of the SWASPP modules

EXPERIMENTAL RESULTS II

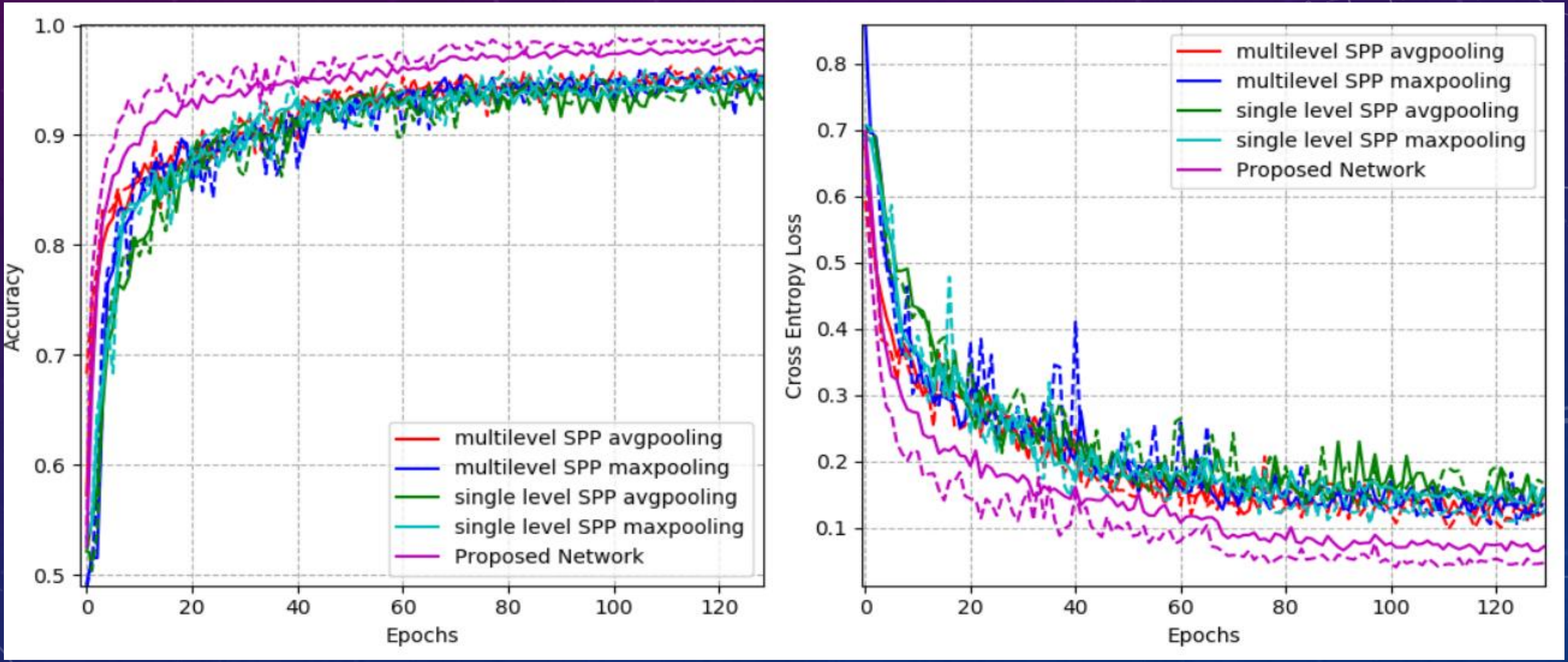
- Experiment is conducted on (Tesla P-100) 16GB GPU and 32 GB Ram using Pytorch Using QATA-COV-19 dataset (balanced Number of labels for classes)
- All models have the same settings
  - Num of epochs 200.
  - Adam optimizer with dynamic learning rate.
  - Batch size is 256.
  - Cross Entropy loss.
  - Train validation test is 0.6 0.1 0.3, respectively.

EXPERIMENTAL RESULTS II

- Base Networks for initial comparison.

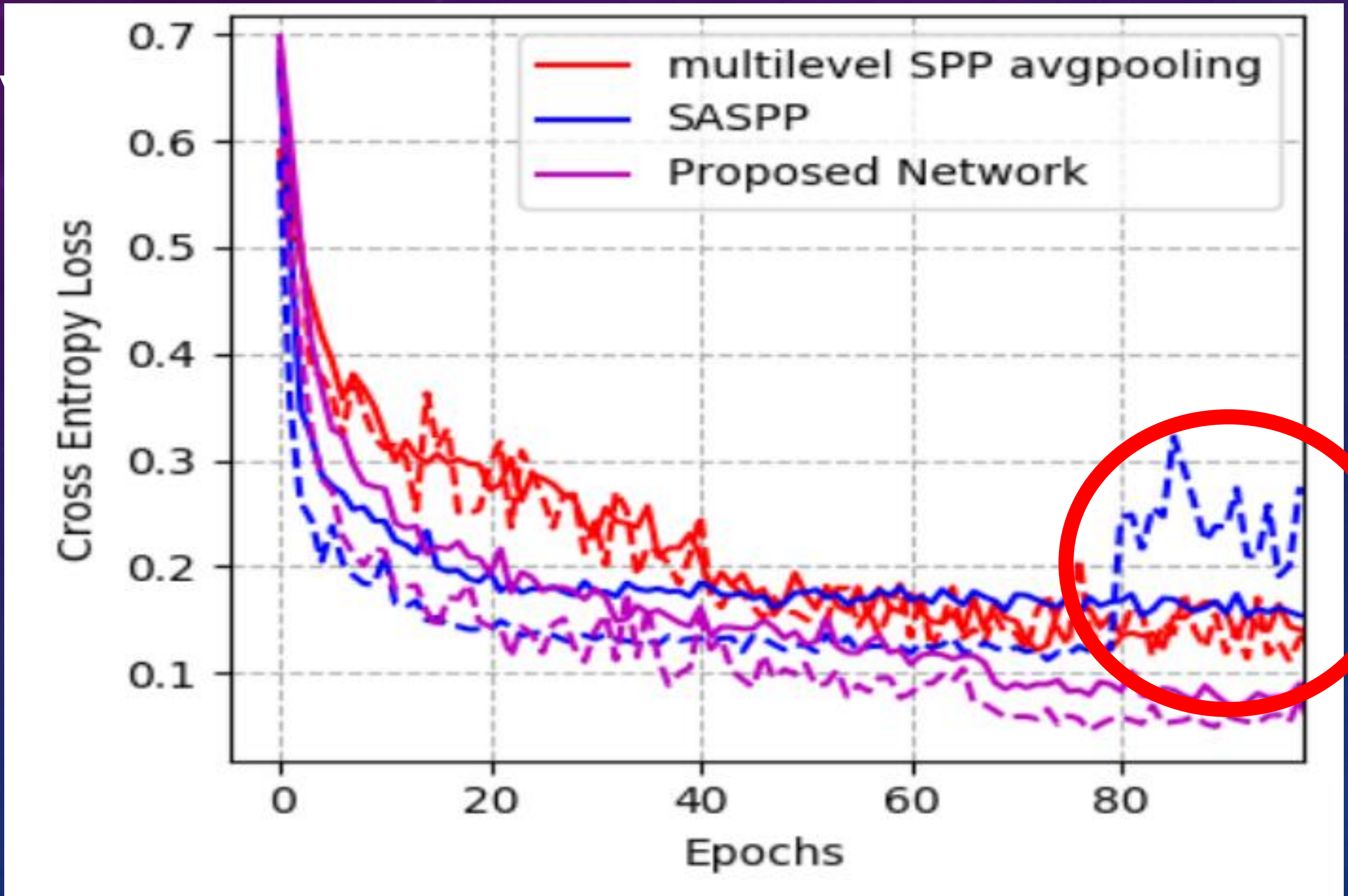
Model Type	Baseline CNN Architectures	
	Variant	*Param. Count*
SPP	ML Average pooling	14,916,420
	ML max pooling	14,916,420
	SL Average pooling	14,490,436
	SL max pooling	14,490,436
SASPP		13,031,841

EXPERIMENTAL RESULTS II

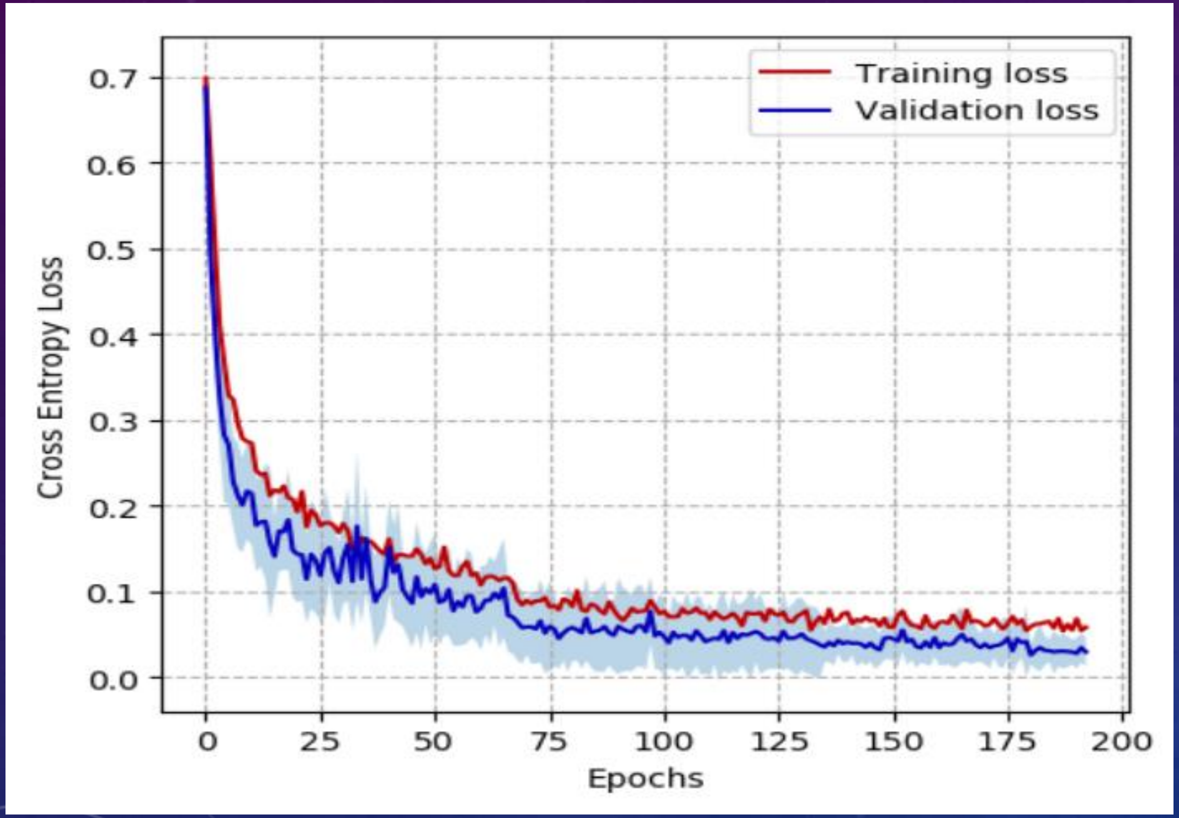


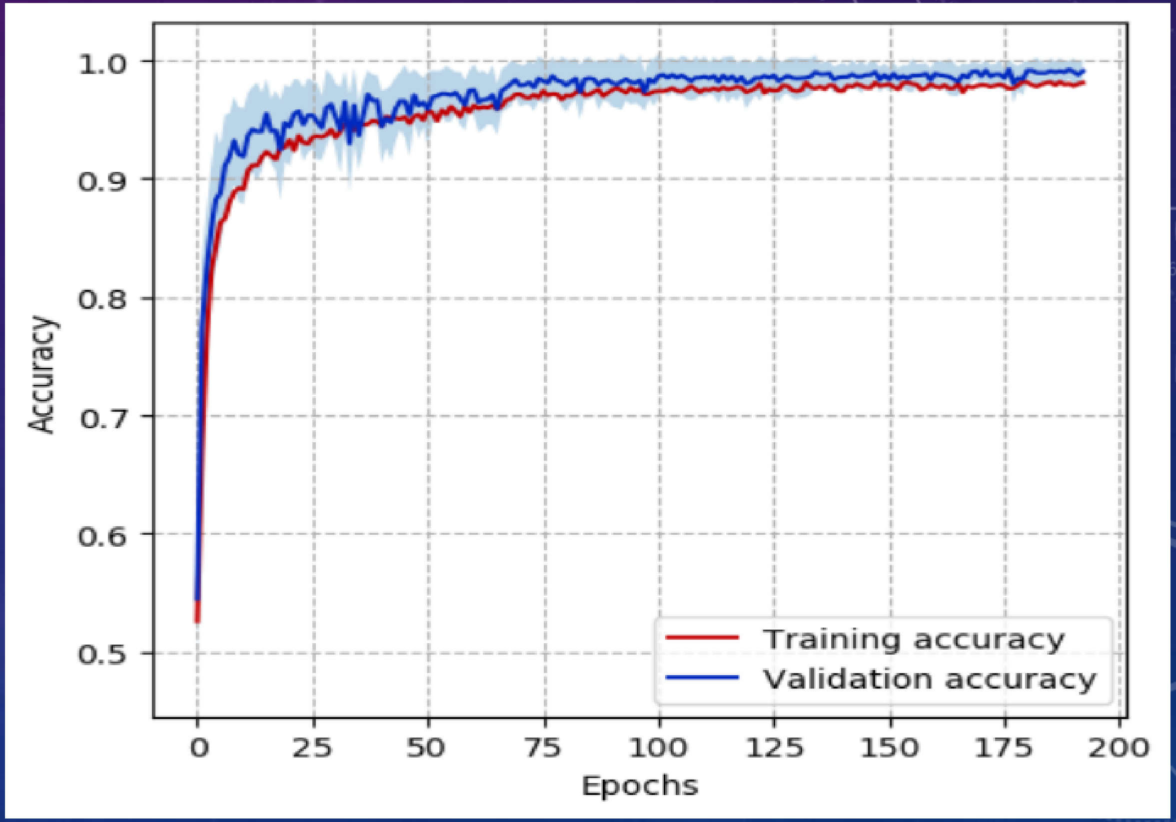
EXPERIMENTAL RESULTS II

- Base Net

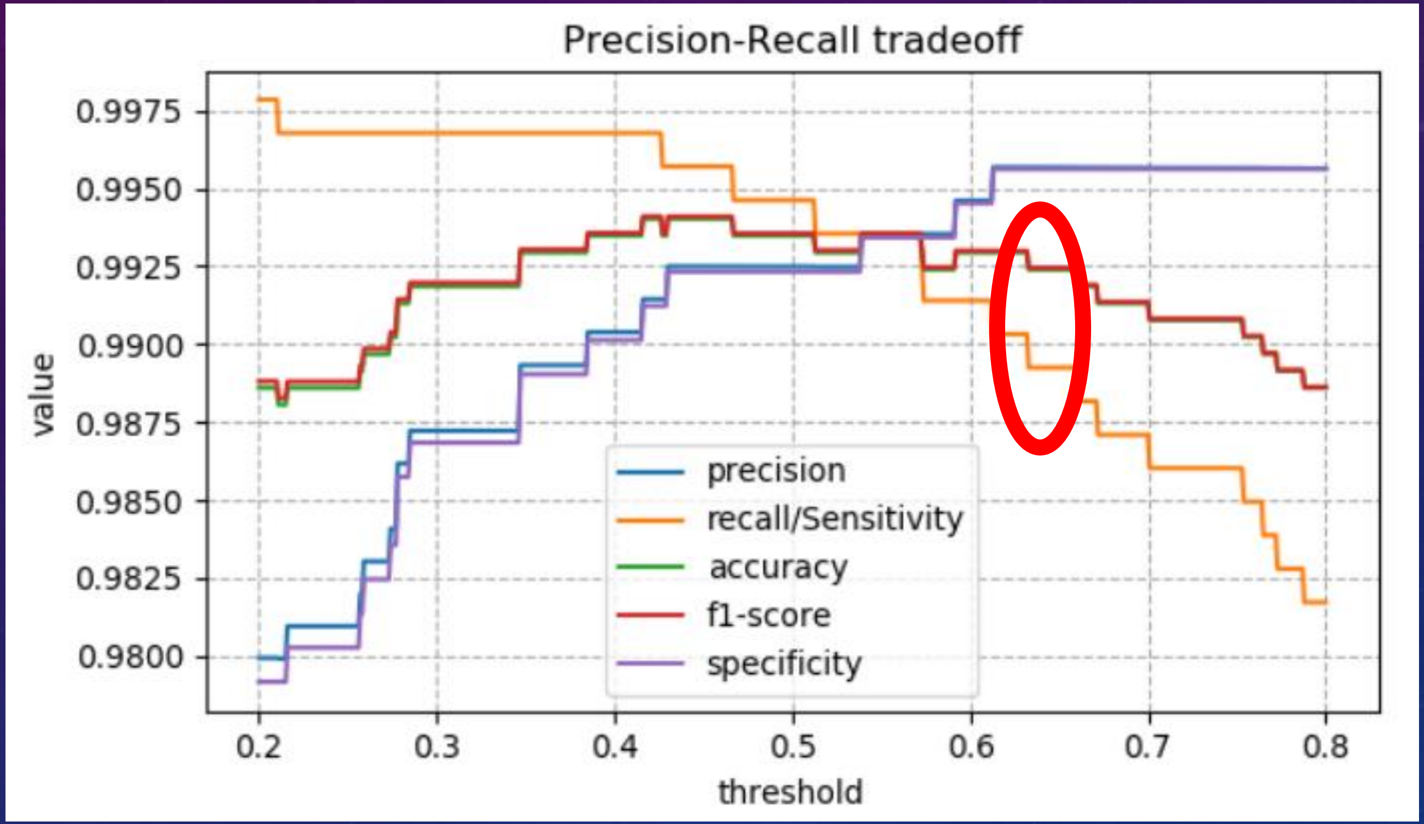


EXPERIMENTAL RESULTS II





EXPERIMENTAL RESULTS II



EXPERIMENTAL RESULTS II

Table 4: Comparison between Proposed network and Related works

Model Name	Accuracy	Sensitivity	Precision	Specificity	F1-score	Param. Count
Proposed	0.99294	**0.9903**	**0.9956**	0.9956	**0.9929**	**5,040,571**
SRC-Dalm[48]	0.985	0.886	-	0.993	-	-
SRC-Hom[48]	0.977	0.921	-	0.982	-	-
CRC-light[48]	0.973	0.955	-	0.974	-	-
DenseNet121*[48]	0.992	0.9714	-	0.9949	-	6,955,906
Inception-v3[48]	0.993	0.954	-	0.998	-	21,772,450
Modified MobileNetV2 [49]	0.98	0.98	0.97	-	0.97	-
ReCovNet-v2[47]	0.99726	0.98571	0.94262	0.9977	0.96369	-
ReCovNet-v1[47]	0.99824	0.9781	0.97438	0.99901	0.97624	-
DenseNet-121[47]	**0.9988**	0.97429	0.9932	**0.99974**	0.98365	6,955,906

TABLE OF THE CONTENT

CONCLUSION

- Novel light-weight and scale invariant CNN architecture for COVID19 classification.
- Lightweight architecture achieved a superior performance 100%
- Spatial separable kernels is used to reduce parameter count.
- Atrous Spatial pooling is used to construct the scale space.
- Trainable Attention module is used to guide the network across the scale space.
- proposed works outperformed prior work.
- proposed method II achieved a 99.03 for sensitivity.

CONCLUSION

Comparison Between Two the two methods

	Proposed Method I	Proposed Method II
Parameter Count	150K	5M
Scales Invariance	No (accuracy drops to 93%)	Yes (performance Not affected)
Training Time	Short (20 epoch)	Long (200 epoch)
Stable	No (affected by Acquisition Device)	Yes (texture and scale augmentation)
Fast Evaluation	Yes	No

REFERENCES

[1] Szeliski, Richard. *Computer vision: algorithms and applications*. Springer Science & Business Media, 2010.
[2] Van Noord, Nanne, and Eric Postma. "Learning scale-variant and scale-invariant features for deep image classification." *Pattern Recognition* 61 (2017): 583-592.
[3] Lowe, David G. "Distinctive image features from scale-invariant keypoints." *International journal of computer vision* 60.2 (2004): 91-110.
[4] Simonyan, Karen, and Andrew Zisserman. "Very deep convolutional networks for large-scale image recognition." (2014).
[5] He, Kaiming, et al. "Spatial pyramid pooling in deep convolutional networks for visual recognition." *IEEE transactions on pattern analysis and machine intelligence* 37.9 (2015): 1904-1916.
[6] Lin, Tsung-Yi, et al. "Feature pyramid networks for object detection." *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2017.
[7] Chen, Liang-Chieh, et al. "Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs." *IEEE transactions on pattern analysis and machine intelligence* 40.4 (2017): 834-848.

REFERENCES

[8] Qiao, Siyuan, Liang-Chieh Chen, and Alan Yuille. "Detectors: Detecting objects with recursive feature pyramid and switchable atrous convolution." *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2021.

[9] Ioffe, Sergey, and Christian Szegedy. "Batch normalization: Accelerating deep network training by reducing internal covariate shift." *International conference on machine learning*. PMLR, 2015.

[10] Krizhevsky, Alex, Ilya Sutskever, and Geoffrey E. Hinton. "Imagenet classification with deep convolutional neural networks." *Advances in neural information processing systems* 25 (2012).

[11] Zhang, Yulun, et al. "Residual dense network for image super-resolution." *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2018.

[12] Srivastava, Nitish, et al. "Dropout: a simple way to prevent neural networks from overfitting." *The journal of machine learning research* 15.1 (2014): 1929-1958.

[13] Degerli, Aysen, et al. "Reliable COVID-19 Detection Using Chest X-ray Images." *2021 IEEE International Conference on Image Processing (ICIP)*. IEEE, 2021.

REFERENCES

[14] M.Ahishali et al., "Advance Warning Methodologies for COVID-19 Using Chest X-Ray Images," in IEEE Access, vol. 9, pp. 41052-41065, 2021

[15] Yamac, Mehmet, et al. "Convolutional sparse support estimator-based COVID-19 recognition from X-ray images." *IEEE Transactions on Neural Networks and Learning Systems* 32.5 (2021): 1810-1820.

[16] Chorowski, J. K., Bahdanau, D., Serdyuk, D., Cho, K., & Bengio, Y. (2015). Attention-based models for speech recognition. *Advances in neural information processing systems*, *28*.

[17] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... & Polosukhin, I. (2017). Attention is all you need. *Advances in neural information processing systems*, *30*.

Thank You.