Neural Architecture Search for Pneumonia Diagnosis from Chest X-Rays

Abhibha Gupta, Parth Sheth, Pengtao Xie

Neural Architecture Search

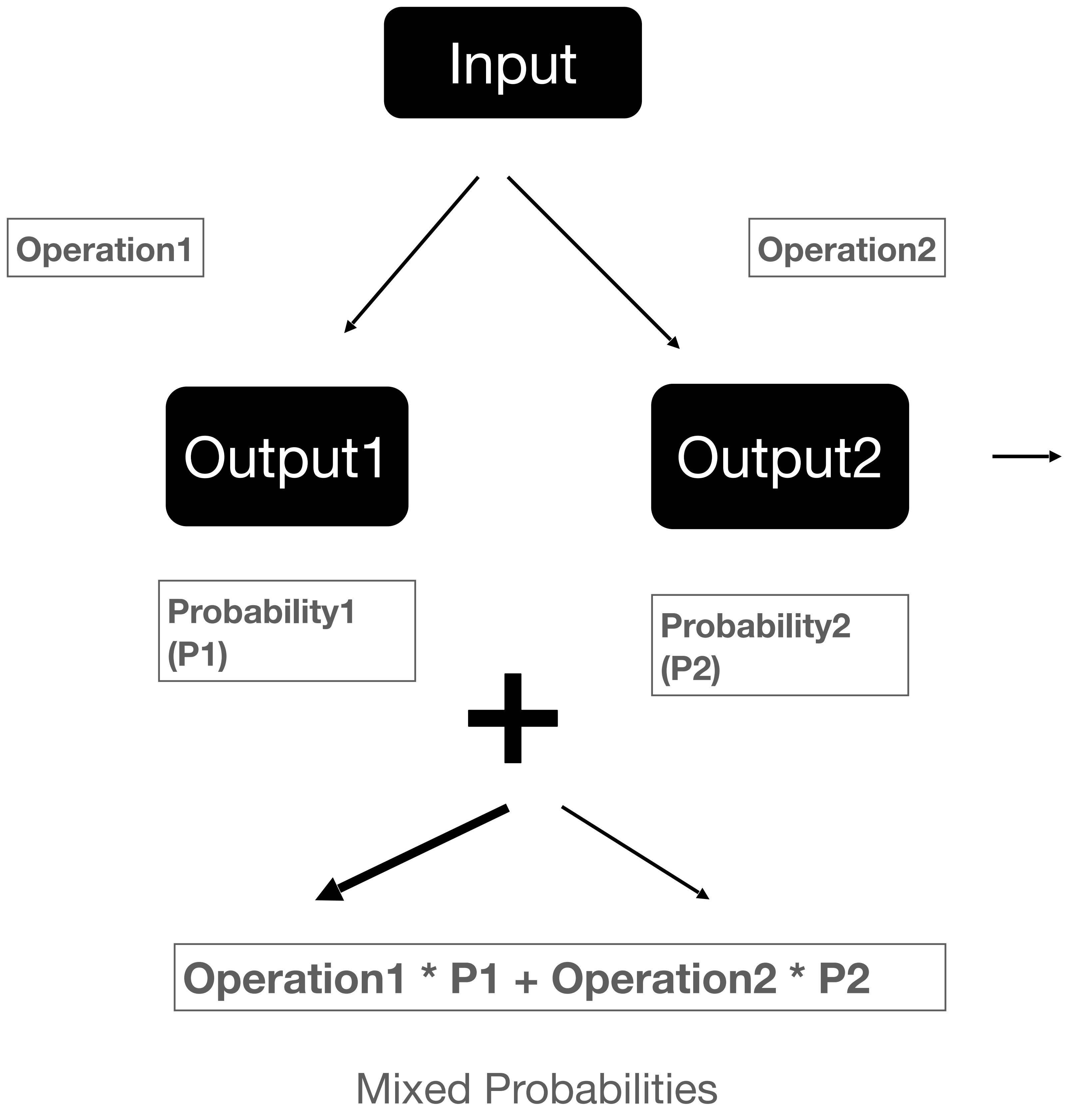
Aim -> To search for the optimal architecture and weights

Differential Architecture Search (DARTS)

How is it better?

Candidate Operations

1) 1x1 conv
2) 5x5 max pooling
3) 3x3 depth wise separable conv
4) ….



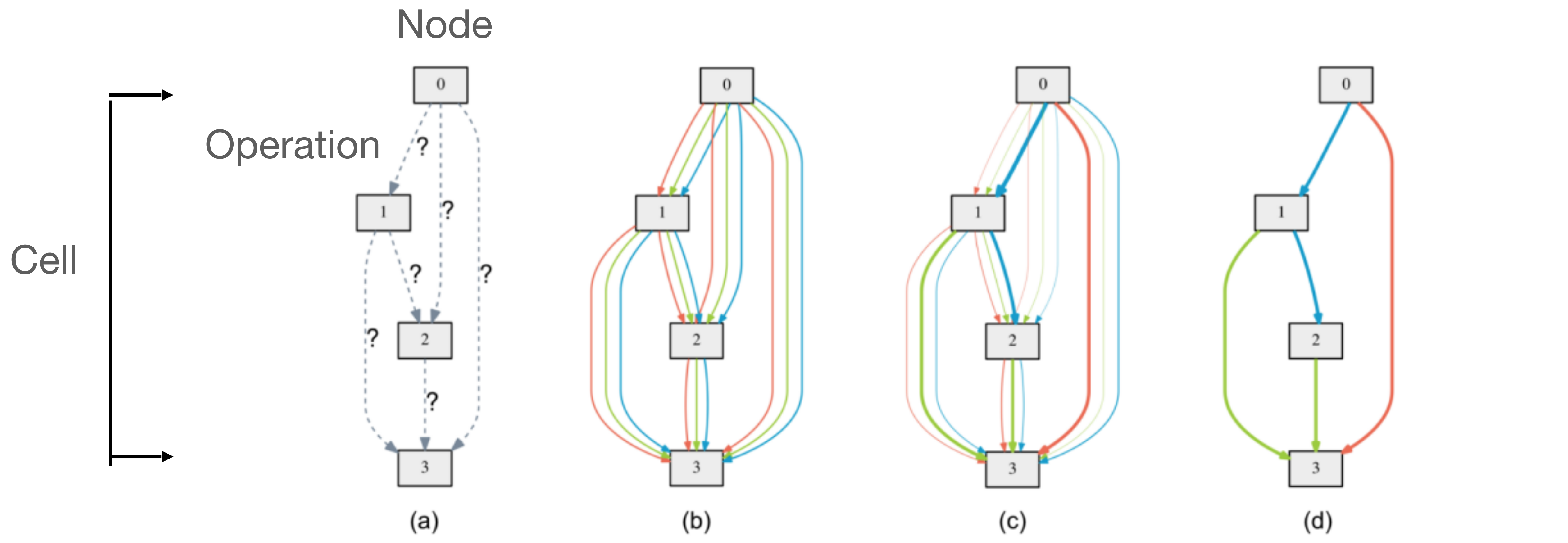


Figure 1: An overview of DARTS: (a) Operations on the edges are initially unknown. (b) Continuous relaxation of the search space by placing a mixture of candidate operations on each edge. (c) Joint optimization of the mixing probabilities and the network weights by solving a bilevel optimization problem. (d) Inducing the final architecture from the learned mixing probabilities.

Partially Connected - Differential Architecture Search (PC-DARTS)

How is it better than DARTS?

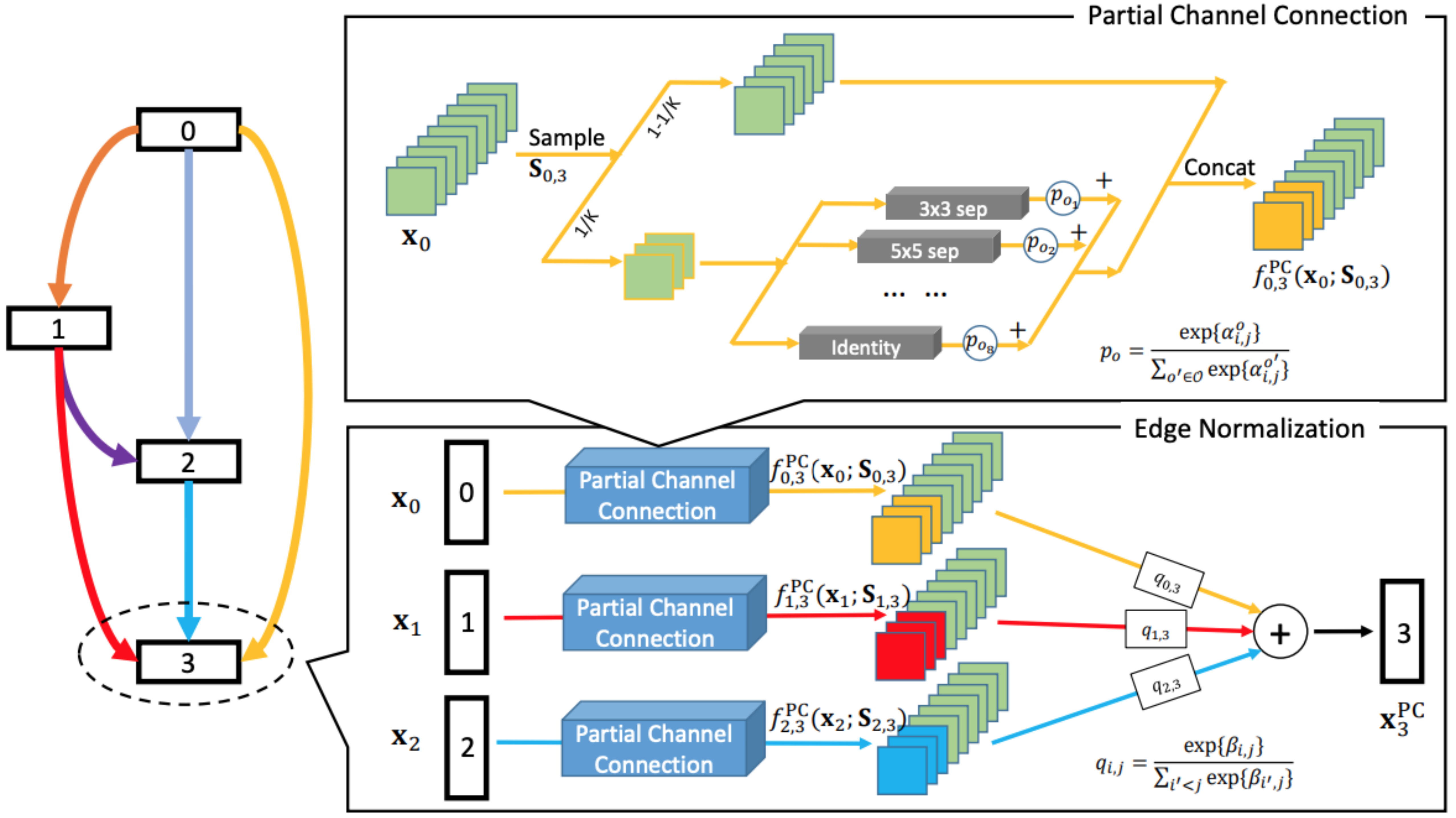
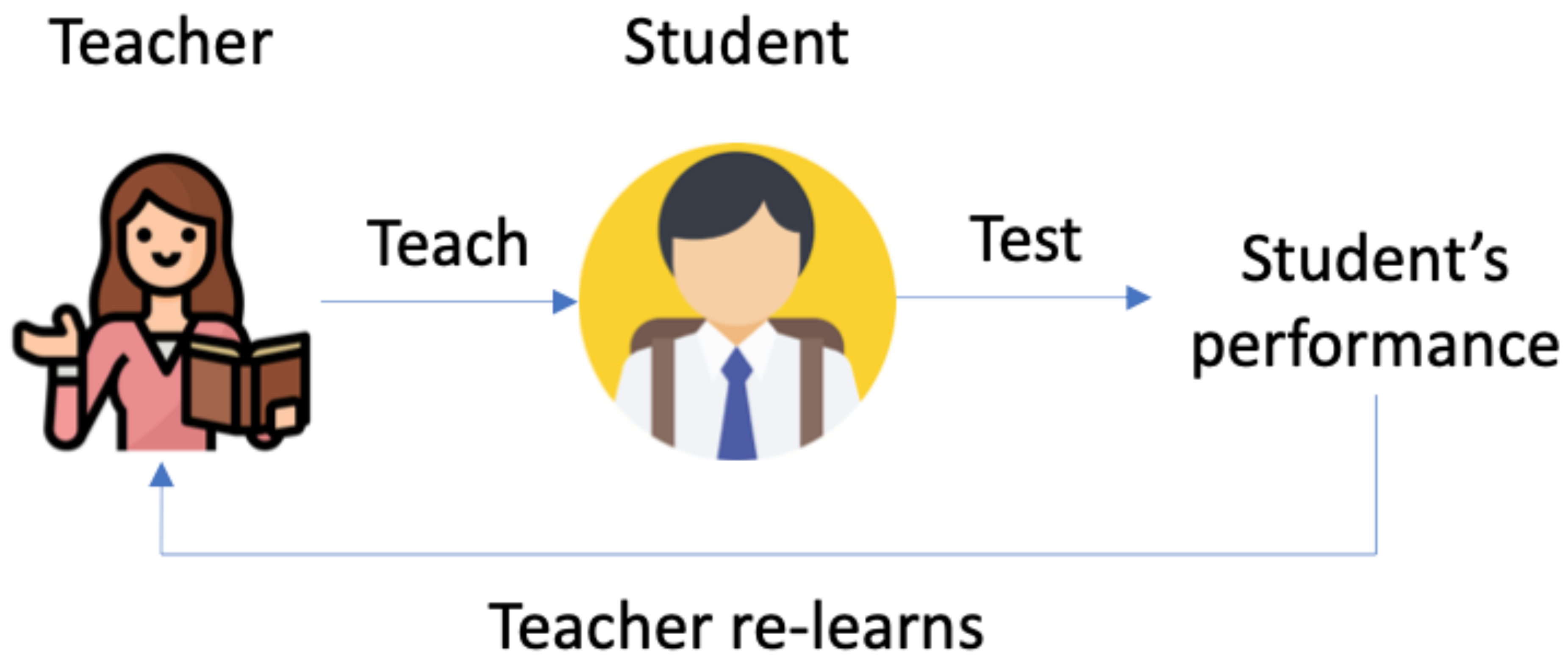


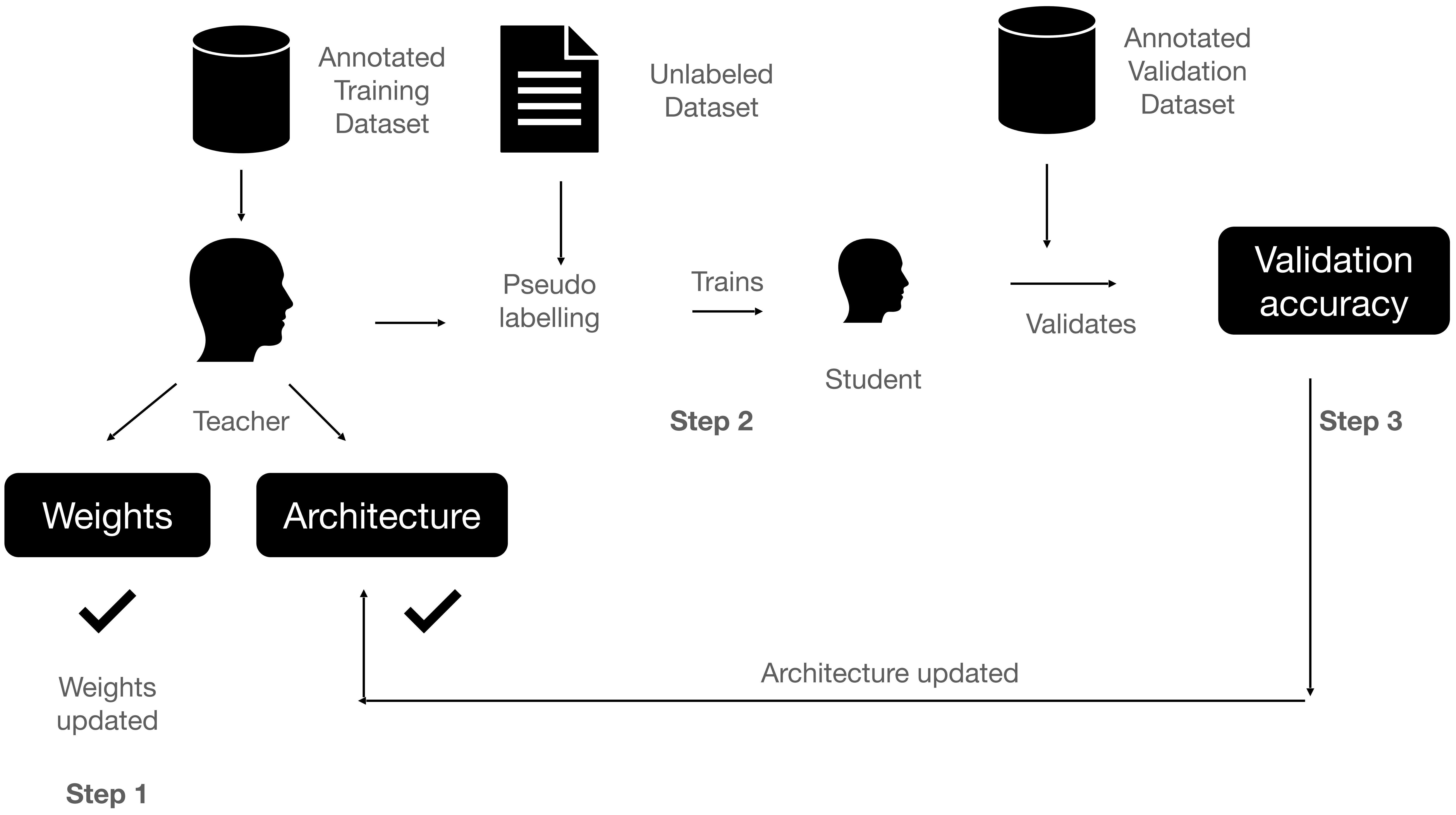
Figure 1: Illustration of the proposed approach (best viewed in color), partially-connected DARTS (PC-DARTS). As an example, we investigate how information is propagated to node #3, *i.e.*, $j = 3$. There are two sets of hyper-parameters during search, namely, $\{\alpha_{i,j}^o\}$ and $\{\beta_{i,j}\}$, where $0 \leqslant i < j$ and $o \in \mathcal{O}$. To determine $\{\alpha_{i,j}^o\}$, we only sample a subset, $1/K$, of channels and connect them to the next stage, so that the memory consumption is reduced by K times. To minimize the uncertainty incurred by sampling, we add $\{\beta_{i,j}\}$ as *extra* edge-level parameters.

Learning By Teaching (LBT)



Learning By Teaching (LBT)

Aim -> To search for the optimal architecture and weights of the Teacher using the Student.



Math behind it…

Step 1

$$T^*(A) = \min_T L(T, A, D_t^{(\mathrm{tr})}).$$

Step 2

$$S^*(T^*(A)) = \min_S L(S, D_s^{(\mathrm{tr})}) + \lambda L(S, D_{pl}(D_u, T^*(A))).$$

Human labeled data

Pseudo labeled dataset

Trade off parameter

Step 3

$$\min_A L(T^*(A), A, D_t^{(\mathrm{val})}) + \gamma L(S^*(T^*(A)), D_s^{(\mathrm{val})}),$$

Teacher model

Student model

Table 1: Notations in Learning by Teaching

Notation	Meaning
A	Architecture of the teacher
T	Network weights of the teacher
S	Network weights of the student
$D_t^{(\mathrm{tr})}$	Training data of the teacher
$D_t^{(\mathrm{val})}$	Validation data of the teacher
$D_s^{(\mathrm{tr})}$	Training data of the student
$D_s^{(\mathrm{val})}$	Validation data of the student
D_u	Unlabeled dataset

Related Work

Custom Architectures: Most related work concentrates on designing custom CNN architectures that are 12 - 18 layer deep and are 10-20 times larger in size as compared to our method. Siddiqui et al, etc.

Pre-trained Models: Others Kermany et al, Stephen et al, Chouhan et al [ ] [ ] have fine tuned pre-trained models like InceptionV3, VGG 19, DenseNet121, AlexNet to detect pneumonia instances from normal instances. Das et al. [37] proposed a truncated Inception net for COVID-19 outbreak screening from chest X-rays. Some have reported high performance using ensemble based methods like Sirazitdinov et al. [28] who have used an ensemble approach which integrates RetinaNet and Mask R-CNN for pneumonia localization. The model size is still magnitudes larger than our method.

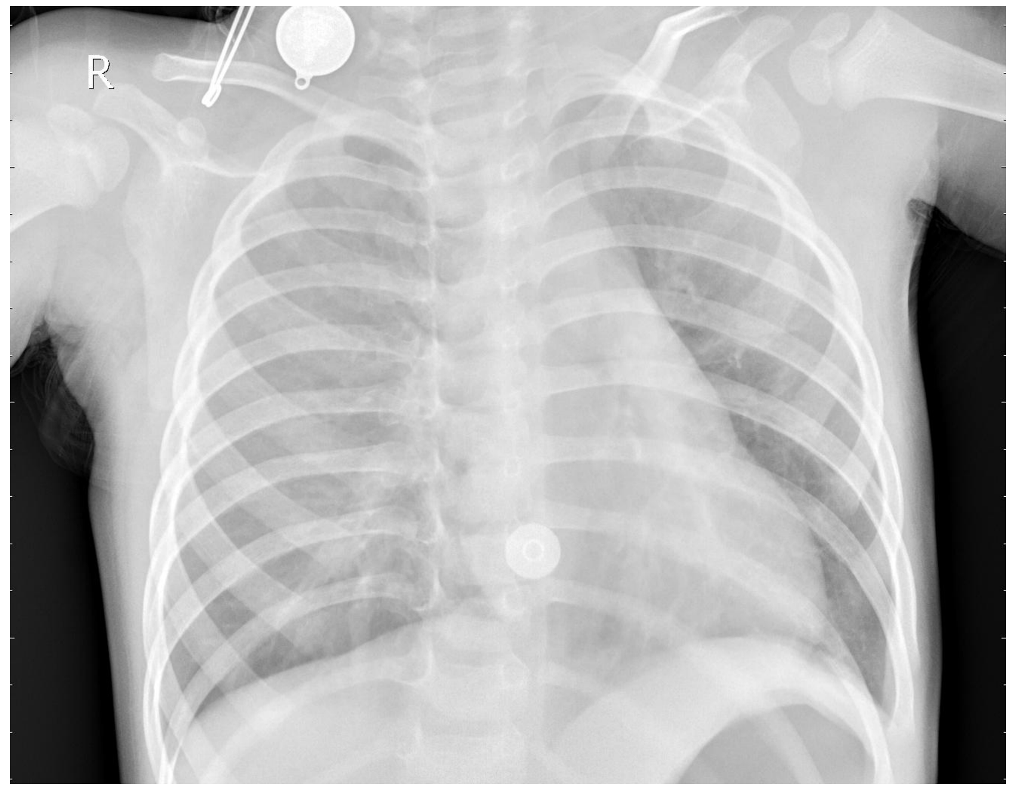
Preprocessing Methods: A variety of feature extraction techniques have been employed. For eg, Cha et al have used pretrained models as feature extractors and an attention based mechanism to select the features. Gu et al have used lung segmentation methods that extract that lung region using fully connected neural networks.

Other Methods: Mukherjee et al. [38] developed a unified deep neural network which leverages CT scans and chest X-rays simultaneously to detect COVID-19. Santosh and Antani [35] proposed to leverage lung region symmetry features for automated screening of pulmonary abnormalities from chest X-rays. Santosh et al. [36] perform edge map analysis of chest X-rays to automatically screen pulmonary abnormality.

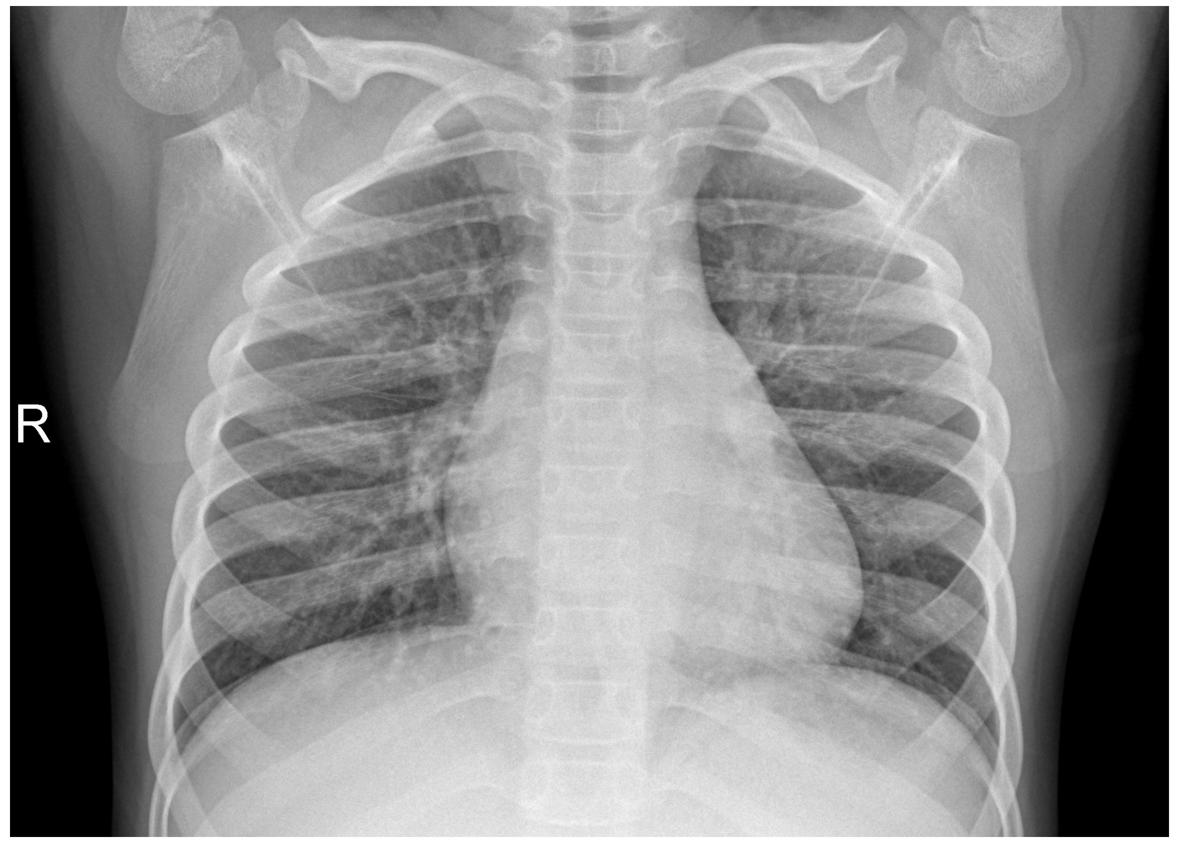
Dataset

5834 chest X-ray images were procured from pediatric patients aged one to five years from Guangzhou Women and Children's Medical Center.

	Train	Test	Val
Normal	1341	184	58
Pneumonia	3874	340	58

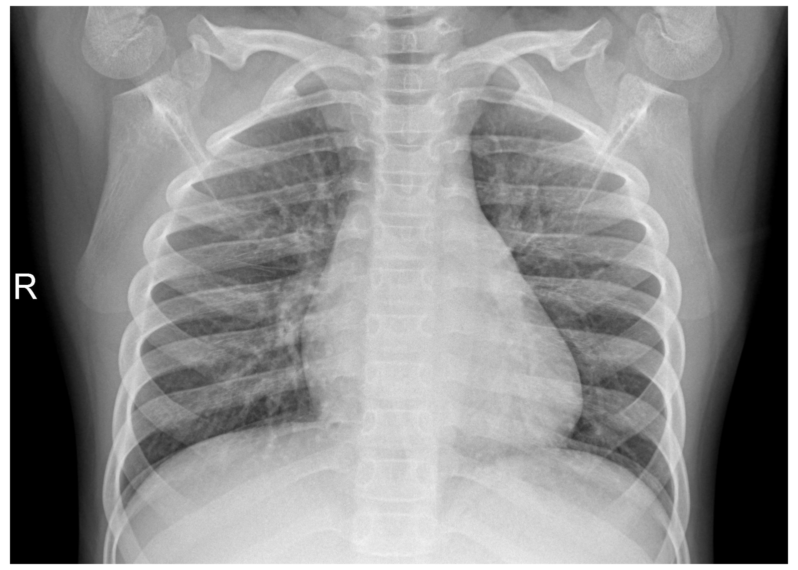


Pneumonia



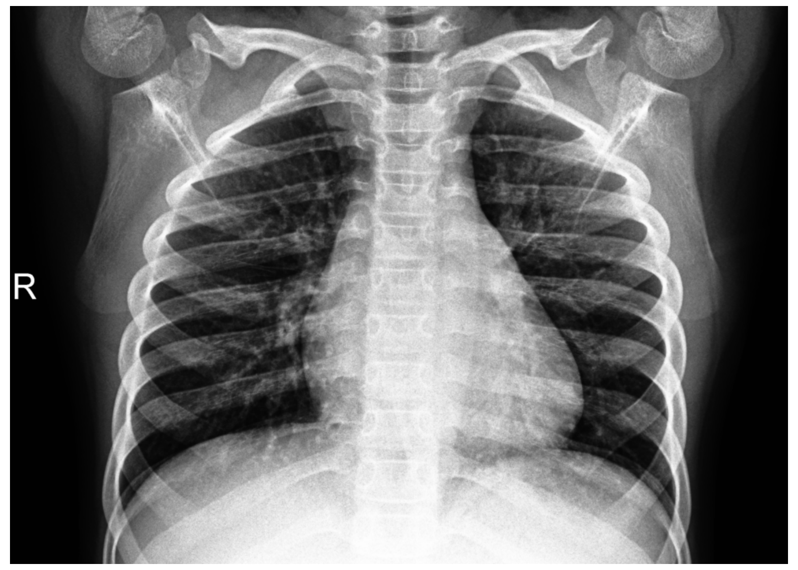
Normal

Pipeline



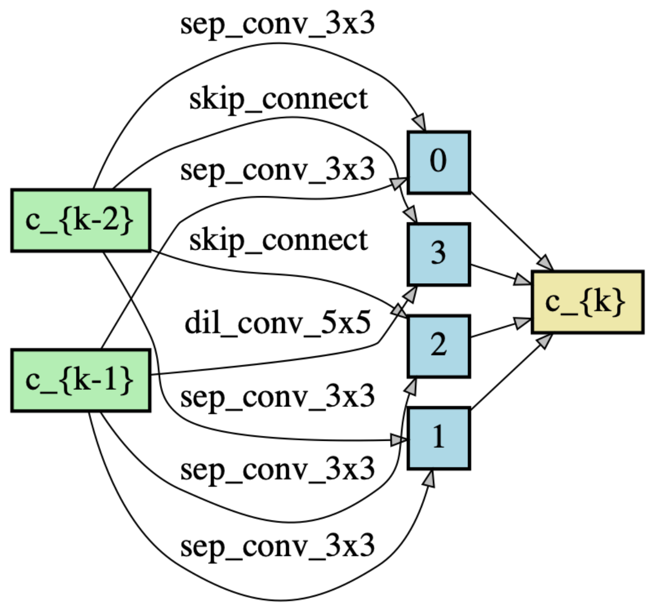
Input Image

Dynamic Histogram Equalisation



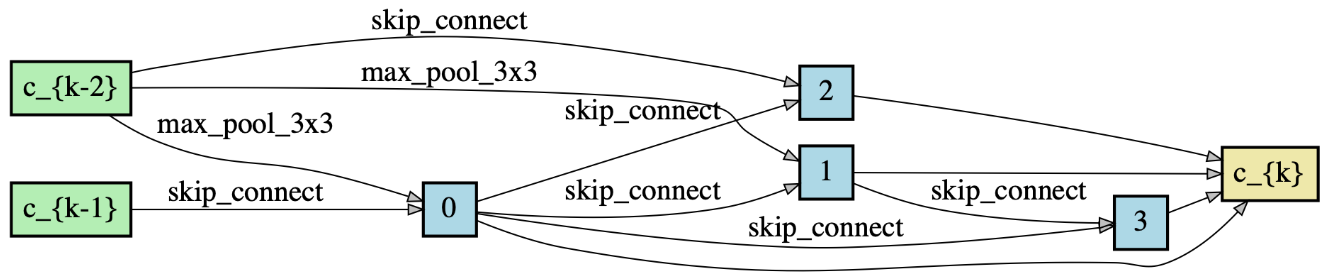
Preprocessed Image

LBT based NAS



Normal Cell

+



Reduction Cell

CXR dataset

Searched CNN

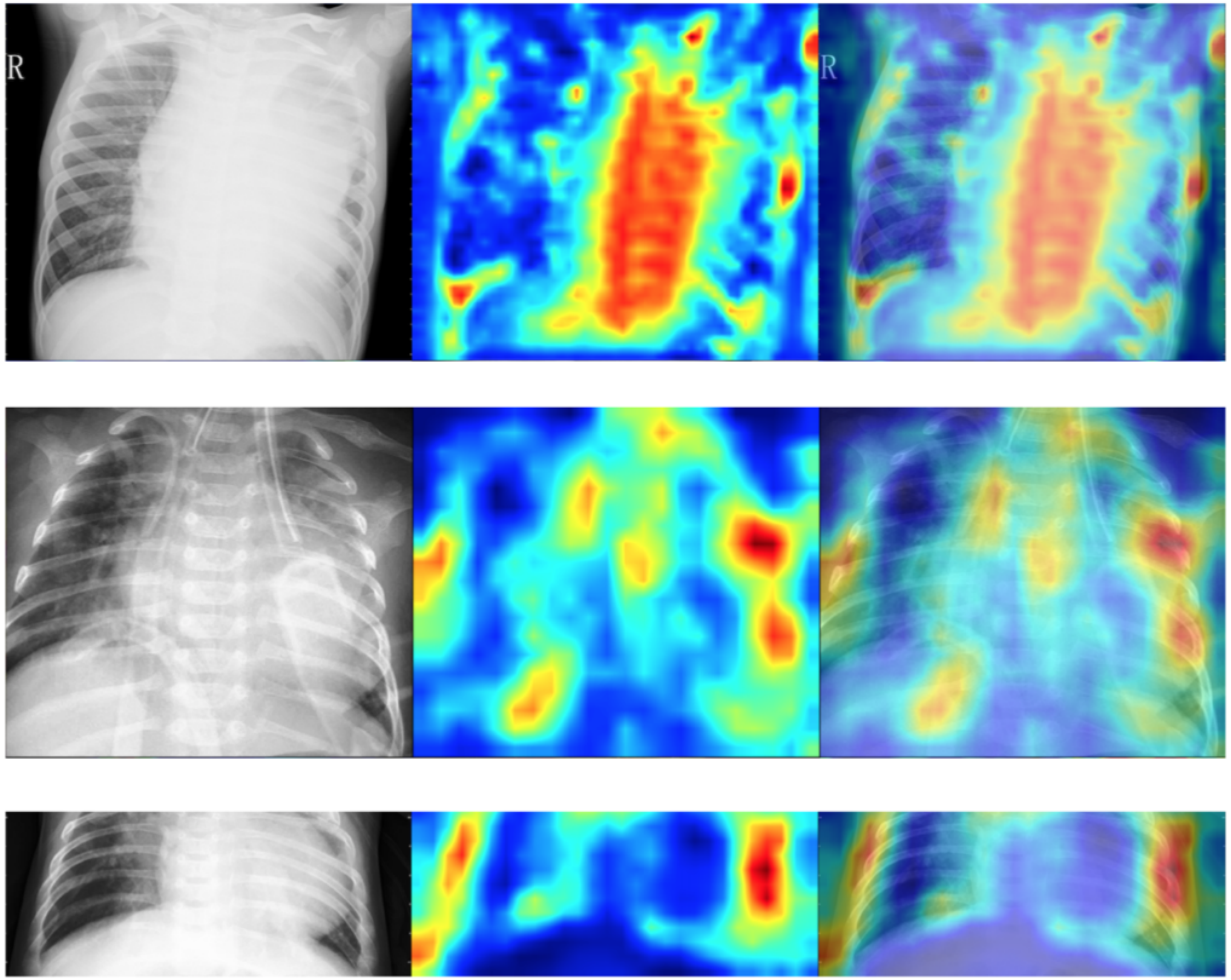
Trained on pneumonia detection using 5 fold cross validation

Test Accuracy

Results

Table 1. Comparison between our method and baselines. Model size is in MB.

Model	Sensitivity (%)	Specificity (%)	F1 (%)	AUC (%)	Accuracy (%)	Model size
VGG19 [51]	91.2±0.73	91.8±0.96	91.5±0.82	93.6±0.56	92.1±1.22	548
InceptionV3 [52]	90.5±0.84	91.5±0.92	90.8±1.17	92.2±0.39	91.7±1.05	292
DenseNet121 [52]	92.5±1.25	91.1±0.96	91.7±0.82	93.0±0.67	92.2±0.83	364
AlexNet [52]	91.8±1.33	91.1±1.05	91.5±0.94	92.8±0.37	92.1±0.70	244
VGG16 [51]	90.3±1.05	92.5±0.69	91.4±0.77	93.7±0.29	92.2±0.92	528
Xception [51]	90.0±0.85	91.6±0.91	91.1±0.87	92.7±0.46	91.5±0.84	88
GoogLeNet [52]	90.1±1.25	91.8±0.83	91.4±0.88	94.1±0.46	92.8±0.91	49
Kermany et al. [10]	92.4±0.87	90.6±0.92	91.2±1.03	93.1±0.37	92.2±0.77	292
Stephen et al. [30]	91.6±0.92	92.5±0.85	91.9±1.21	93.8±0.58	93.3±0.89	27
Siddiqi [31]	93.4±0.83	91.9±1.04	92.5±0.89	93.7±0.26	93.0±0.95	152
Liang et al. [27]	89.2±0.85	90.5±0.92	89.6±1.27	91.5±0.48	90.9±1.14	$\approx$ 98
Meta Pseudo Label [39]	90.0±0.96	91.5±0.73	91.2±0.79	92.7±0.35	91.5±0.88	33
Liu et al. [40]	90.8±0.75	92.1±0.98	91.6±0.64	92.9±0.29	91.8±0.95	14
Kundu et al. [53]	90.8±1.33	91.3±1.07	91.1±0.83	92.7±0.41	91.4±0.79	91
Cha et al. [54]	91.6±0.74	90.2±0.83	91.0±1.22	92.5±0.50	91.3±0.76	79
DARTS [7]	88.3±0.95	88.9±1.17	88.5±0.86	91.5±0.64	89.4±1.15	6.8
LBT-DARTS (ours)	91.8±1.22	92.7±1.04	92.4±1.17	93.9±0.59	92.6±1.07	6.6
PC-DARTS [8]	91.6±1.07	90.1±1.16	90.4±0.88	91.7±0.47	90.5±1.29	6.6
LBT-PC-DARTS (ours)	**94.4**±1.53	**95.8**±1.22	**95.2**±1.01	**96.8**±0.42	**95.3**±0.84	**6.5**



Column (a) shows original CXR images with pneumonia. Column (b) shows Grad-CAM visualization of saliency maps of LBT-PCDARTS. Column (c) shows the overlay of saliency maps on original images.

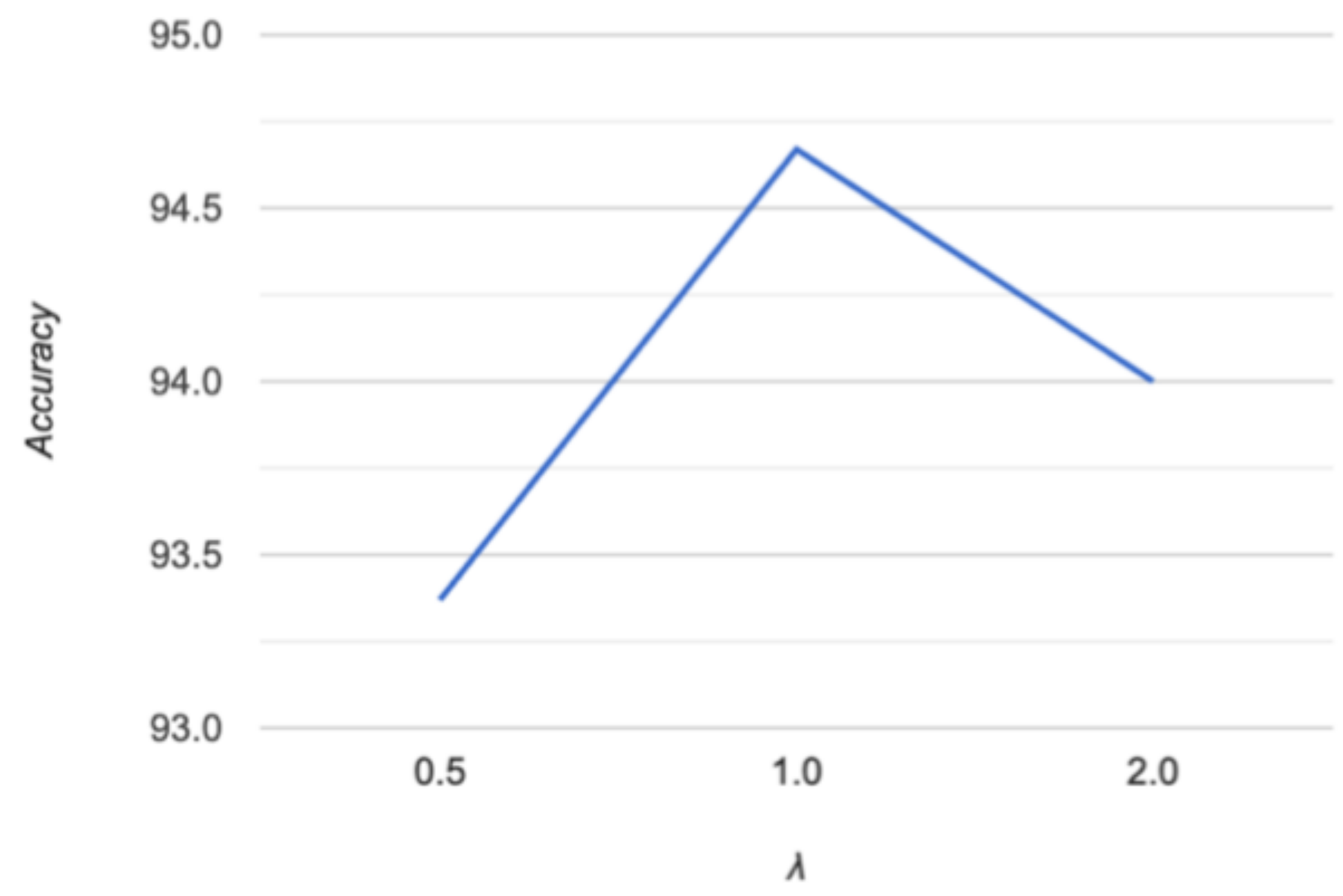
Ablation Studies

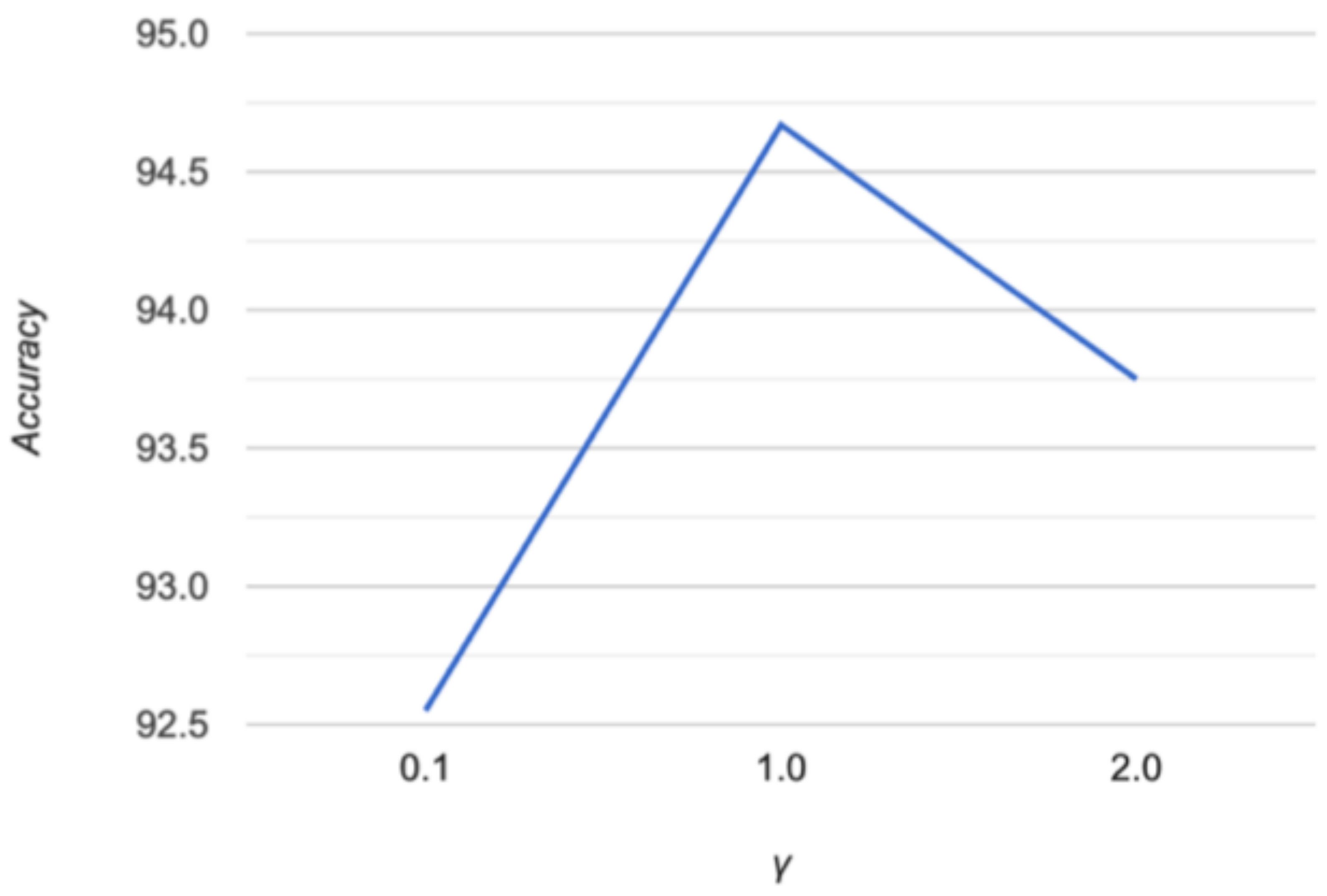
Ablation setting 1: the teacher updates its architecture by minimizing the validation loss of the student only, without considering the validation loss of itself. <u>Accuracy: 92.85%</u>

Ablation setting 2: in the second stage of LBT, only the pseudo labeled dataset is used to train the student. The training data of the student, labeled by humans, is not used. <u>Accuracy: 93.75%</u>

Ablation setting 3: We investigate how the teacher's validation performance changes with the tradeoff parameter λ.

Ablation setting 4: We investigate how the teacher's validation performance changes with the tradeoff parameter γ.





Thank you! Any questions?