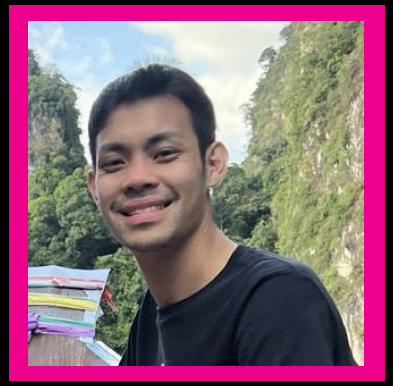
MEDINFO23
8 – 12 JULY 2023 | SYDNEY, AUSTRALIA

之江实验室
ZHEJIANG LAB
RITSUMEIKAN UNIVERSITY
浙江大學
ZHEJIANG UNIVERSITY



MSPA-DLA++: A multi-scale phase attention deep layer aggregation for lesion detection in multi-phase CT images

Titinunt Kitrungrotsakul

Reserach Scientist

Zhejiang lab, China

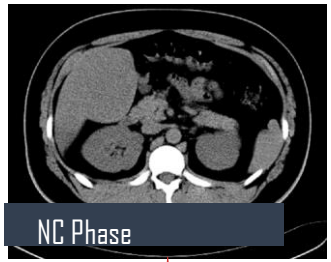
AIDH IMIA

Reserach motivation

- Number of deaths from liver cancer worldwide.
  - Detection
  - Segmentation
  - etc.
- Why medical imaging methods are important for healthcare?
  - High accuray
  - Reduce task and cost of medical staff
- How to utilizing medical data to bring out their potential?

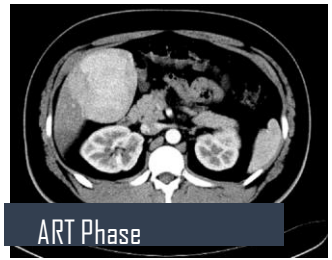
Multi-phase CT image

Before injection

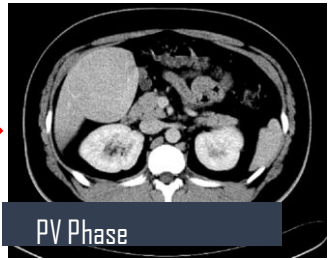


Contrast Injection

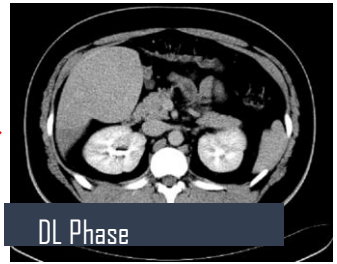
30–40 s



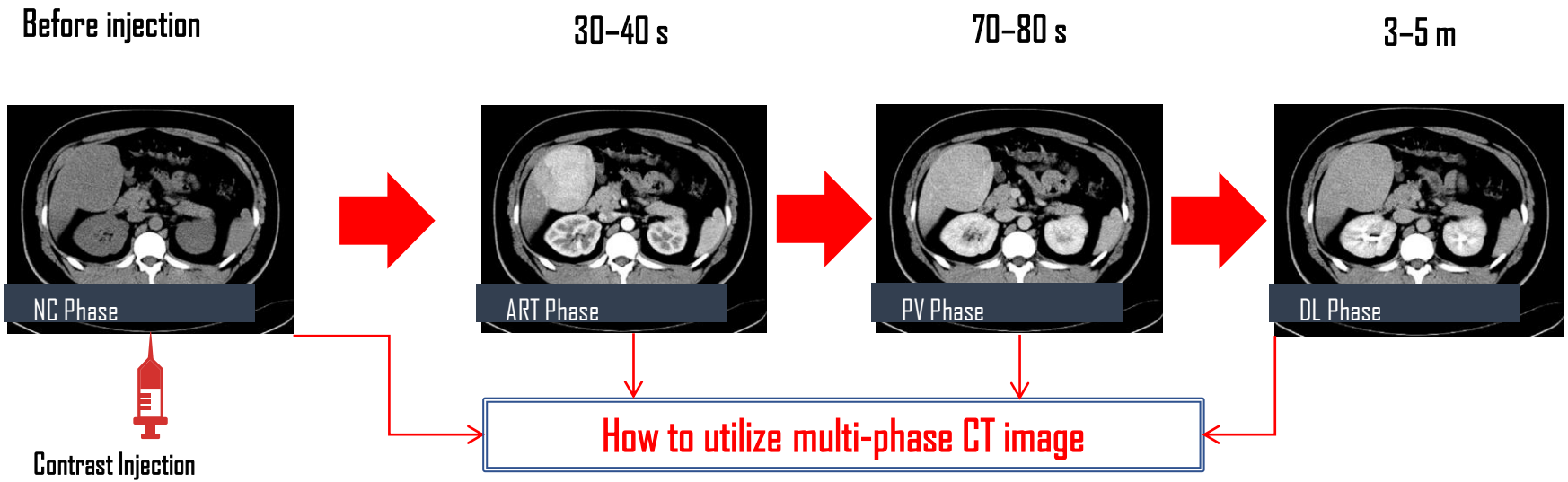
70–80 s



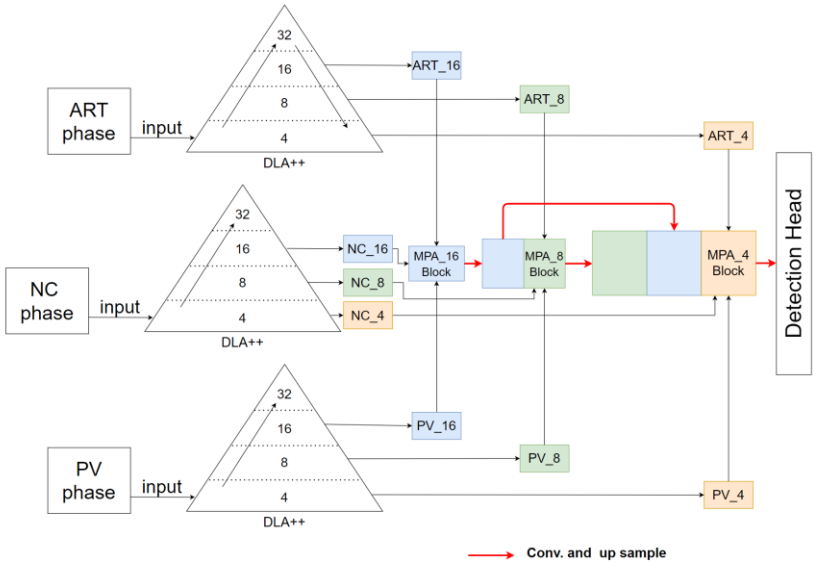
3–5 m

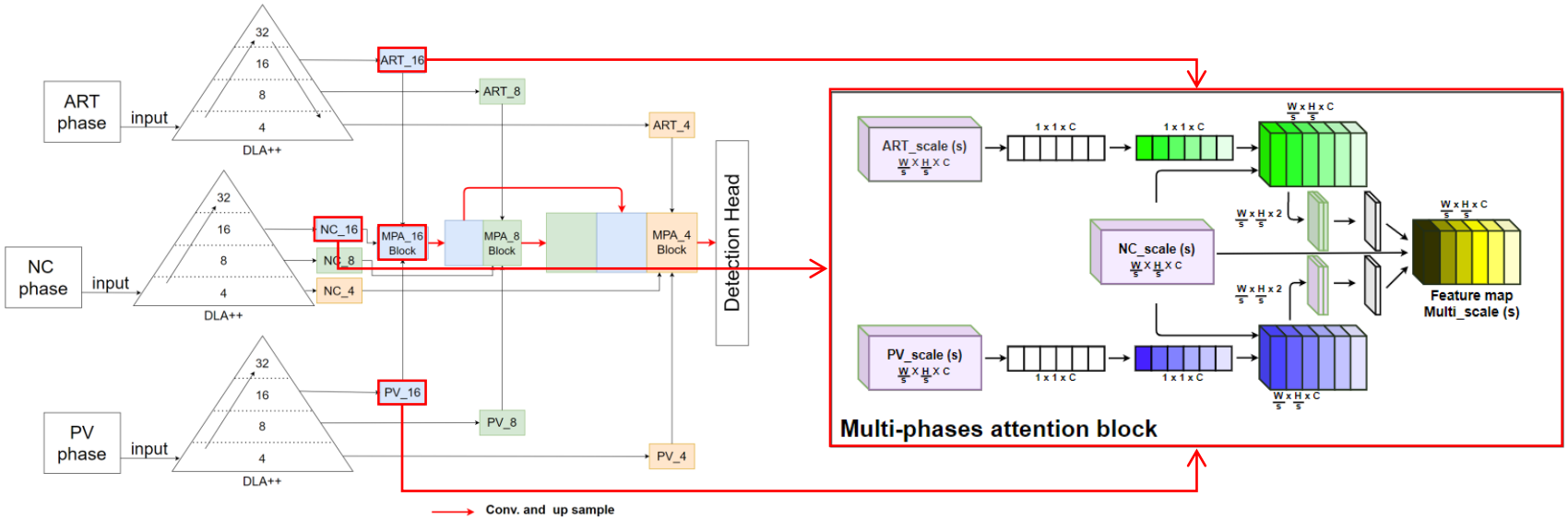


Multi-phase CT image



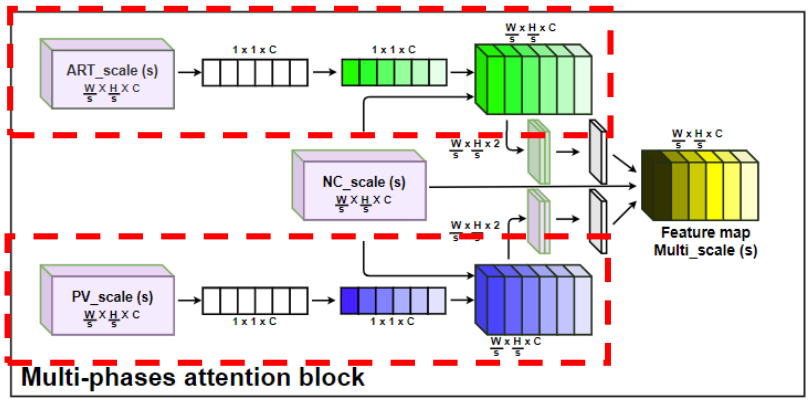
Architecture of MSPA-DLA++




ART phase
input
32
16
8
4
DLA++
ART_16
ART_8
ART_4
NC phase
input
32
16
8
4
DLA++
NC_16
NC_8
NC_4
MPA_16 Block
MPA_8 Block
MPA_4 Block
PV phase
input
32
16
8
4
DLA++
PV_16
PV_8
PV_4
Detection Head
Conv. and up sample
ART_scale (s)
NC_scale (s)
PV_scale (s)
1 x 1 x C
W/s x H/s x C
W/s x H/s x 2
Feature map Multi_scale (s)
Multi-phases attention block

Channel attention features

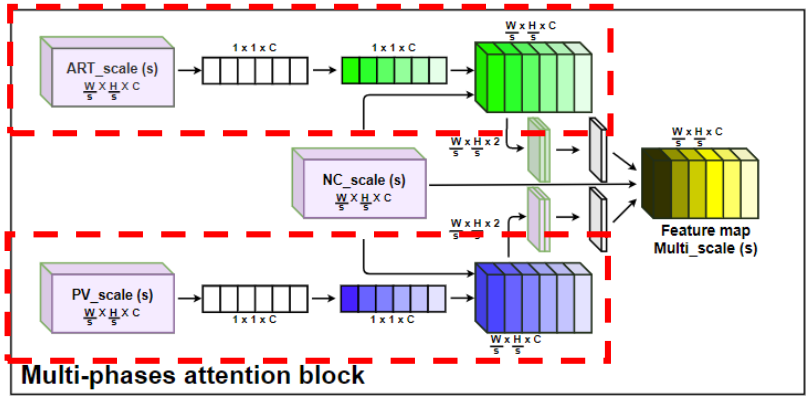
$$F_{x1,x2}^{C,s} = F_{x1}^{s} \otimes \sigma(MLP(AvgPool(F_{x2}^{s})))$$



Channel attention features

$$F_{x1,x2}^{C,s} = F_{x1}^{s} \otimes \sigma(MLP(AvgPool(F_{x2}^{s})))$$

Why only ART and PV on channel attention features extraction proess?

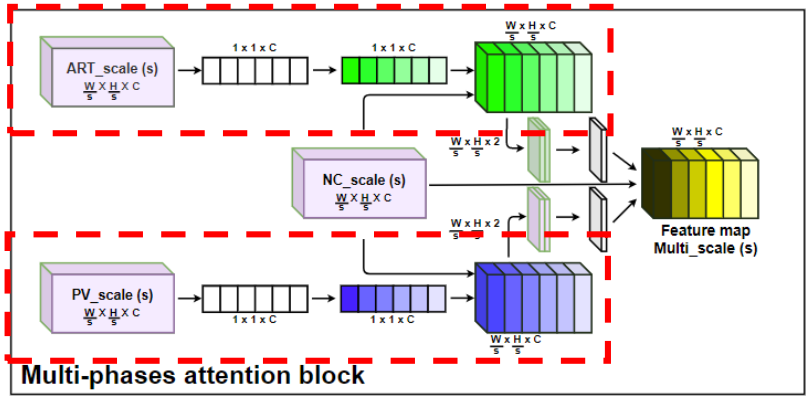


Channel attention features

$$F_{x1,x2}^{C,s} = F_{x1}^{s} \otimes \sigma(MLP(AvgPool(F_{x2}^{s})))$$

Why only ART and PV on channel attention features extraction proess?

1. NC provide unnecessary features (not important features).

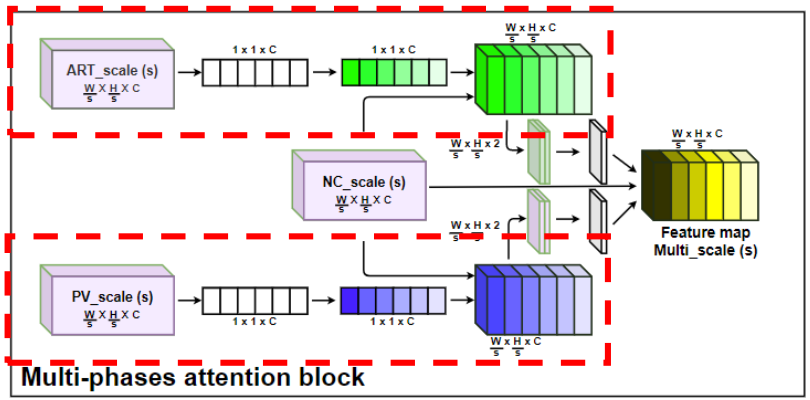


Channel attention features

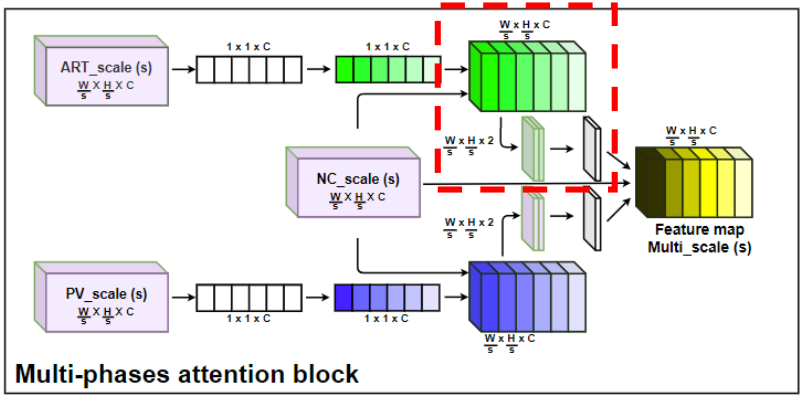
$$F_{x1,x2}^{C,s} = F_{x1}^{s} \otimes \sigma(MLP(AvgPool(F_{x2}^{s})))$$

Why only ART and PV on channel attention features extraction proess?

1. NC provide unnecessary features (not important features).
2. Based on our experiment w./wo. chennel attention from NC make not significant improvement.

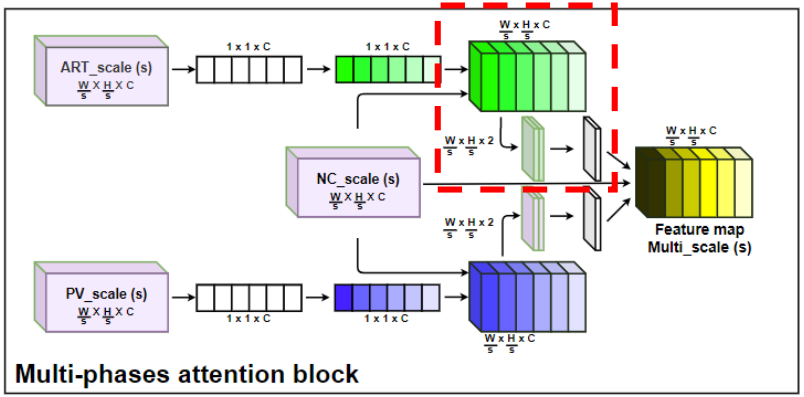


Spatial attention features



Spatial attention features

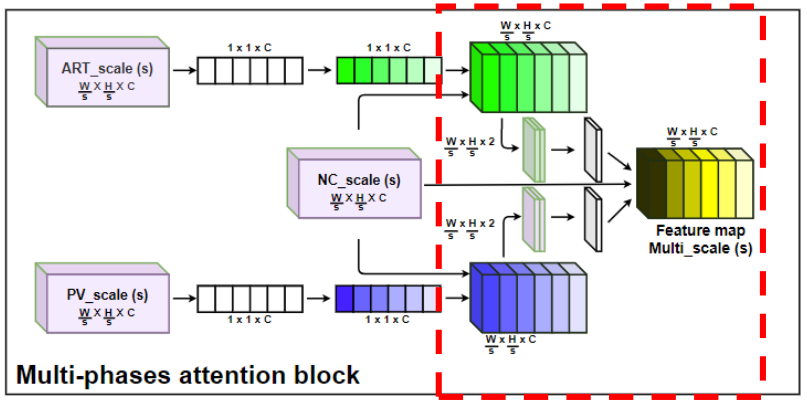
$$F_{x1,x2}^{S,s} = f^{7\times7}([AvgPool(F_{x1,x2}^{C,s}), MaxPool(F_{x1,x2}^{C,s})])$$



Spatial attention features

$$F_{x1,x2}^{S,s} = f^{7\times7}([AvgPool(F_{x1,x2}^{C,s}), MaxPool(F_{x1,x2}^{C,s})])$$

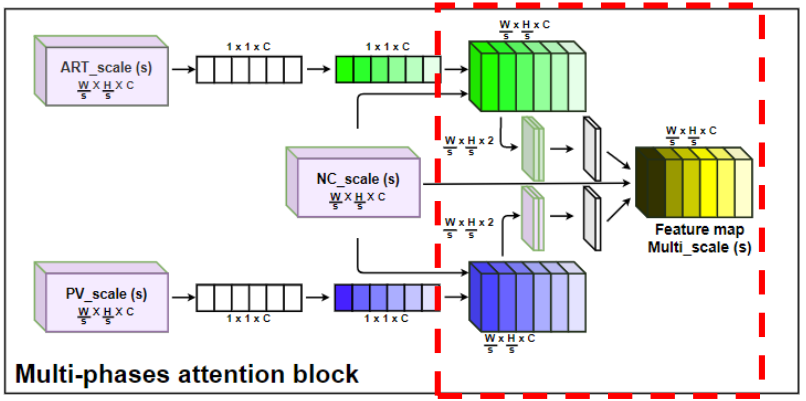
$$\hat{F}^{s} = F_{NC}^{s} \otimes \sigma(F_{NC,ART}^{S,s} + F_{NC,PV}^{S,s})$$



Spatial attention features

$$F^{S,s}_{x1,x2} = f^{7\times7}([AvgPool(F^{C,s}_{x1,x2}), MaxPool(F^{C,s}_{x1,x2})])$$

$$\hat{F}^{s} = F^{s}_{NC} \otimes \sigma(F^{S,s}_{NC,ART} + F^{S,s}_{NC,PV})$$

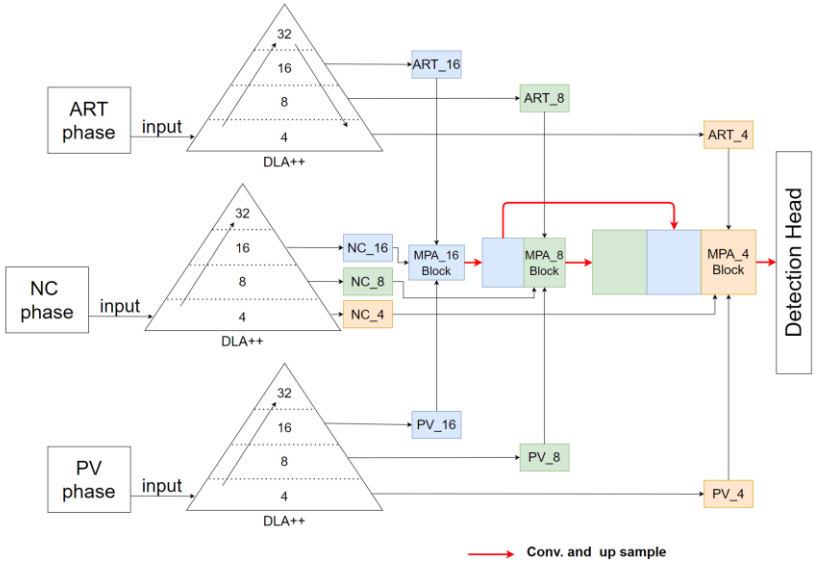


Multi-scale features

$$F^{S,s}_{x1,x2} = f^{7\times7}([AvgPool(F^{C,s}_{x1,x2}), MaxPool(F^{C,s}_{x1,x2})])$$

$$\hat{F}^{s} = F^{s}_{NC} \otimes \sigma(F^{S,s}_{NC,ART} + F^{S,s}_{NC,PV})$$

$$\hat{\hat{F}}^{2} = H(\hat{F}^{4}_{NC}, H(\hat{F}^{8}_{NC}, \hat{F}^{16}_{NC}), H(H(\hat{F}^{16}_{NC})))$$

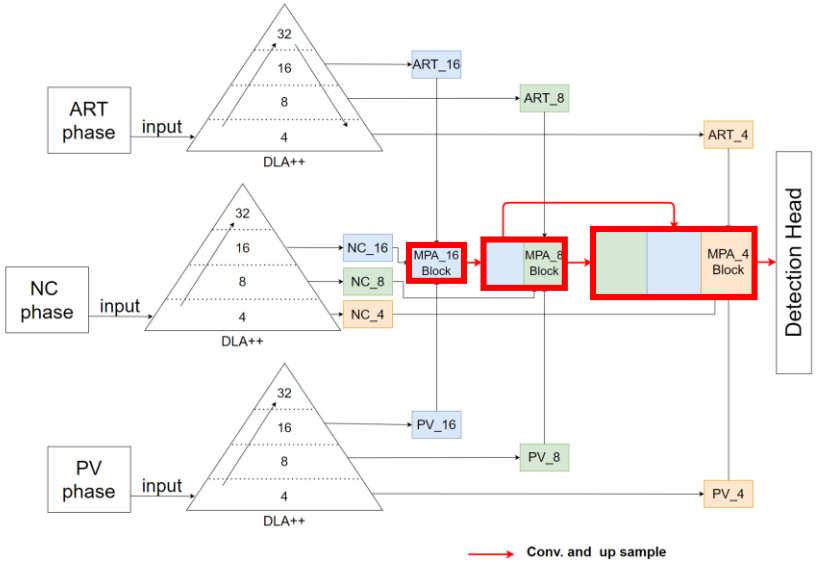


Multi-scale features

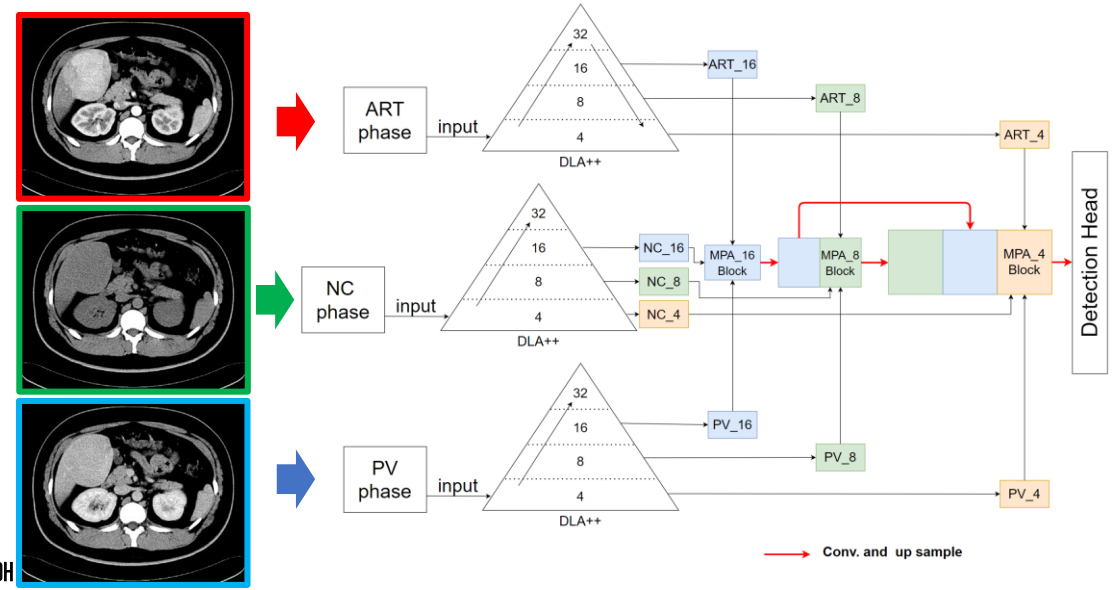
$$F^{S,s}_{x1,x2} = f^{7\times7}([AvgPool(F^{C,s}_{x1,x2}), MaxPool(F^{C,s}_{x1,x2})])$$

$$\hat{F}^s = F^s_{NC} \otimes \sigma(F^{S,s}_{NC,ART} + F^{S,s}_{NC,PV})$$

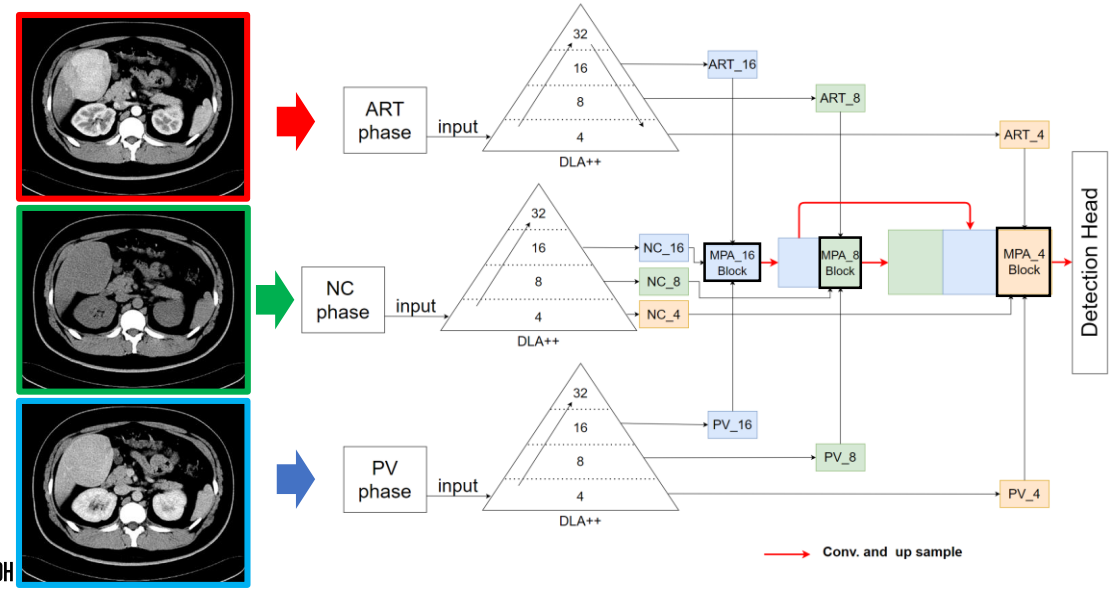
$$\hat{\hat{F}}^2 = H(\hat{F}^4_{NC}, H(\hat{F}^8_{NC}, \hat{F}^{16}_{NC}), H(H(\hat{F}^{16}_{NC})))$$



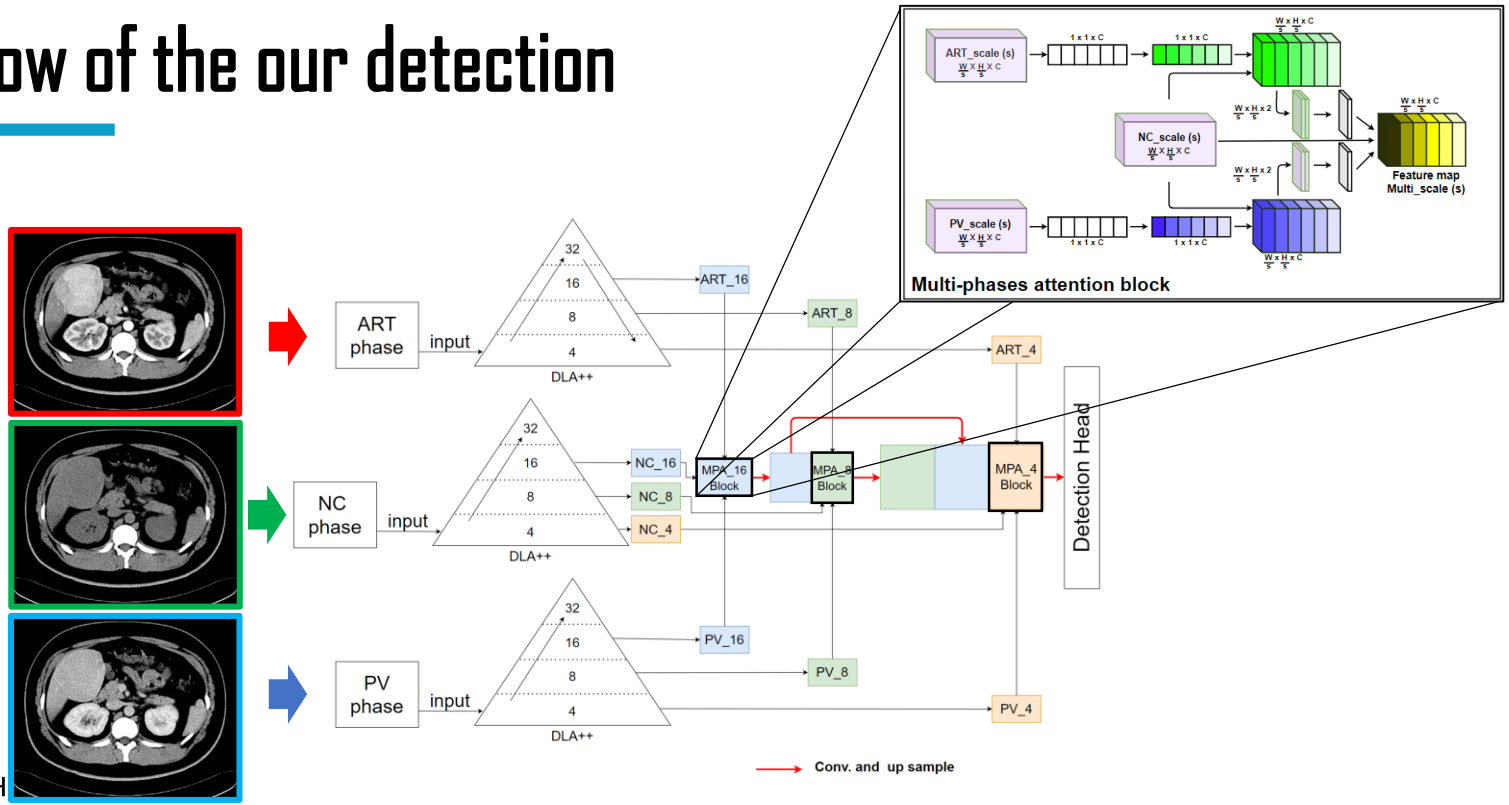
Flow of the our detection



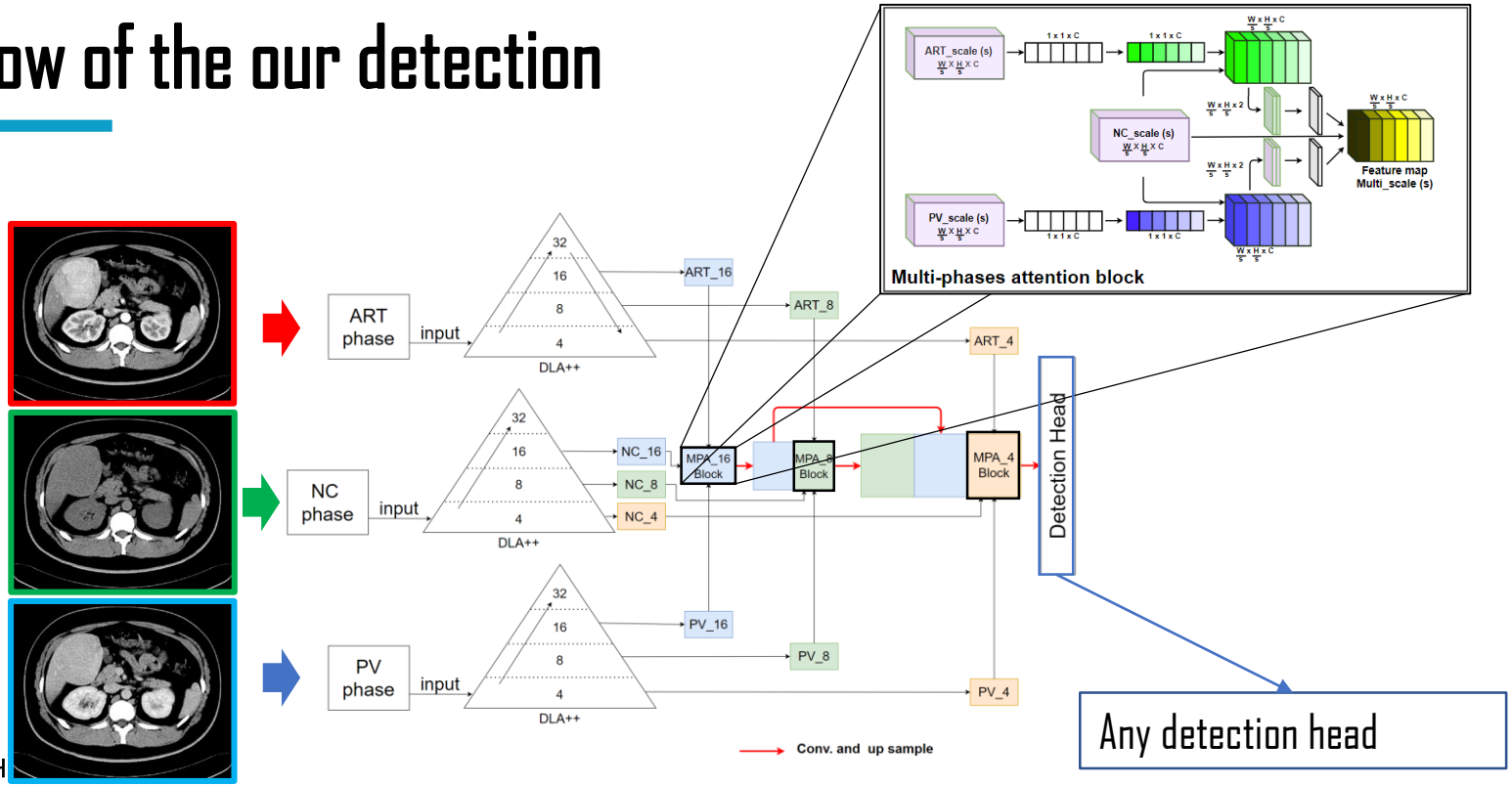
Flow of the our detection



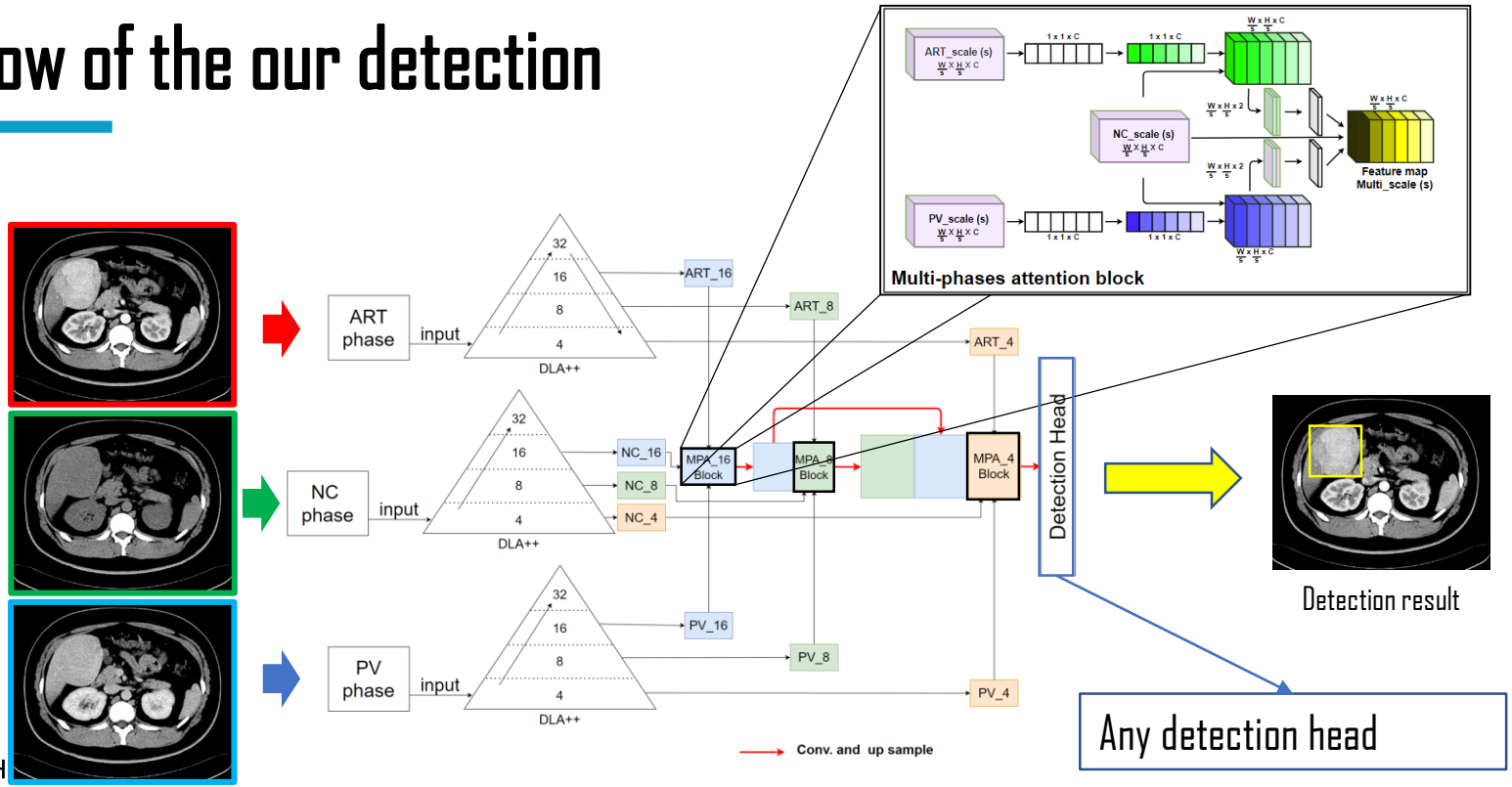
Flow of the our detection



Flow of the our detection



Flow of the our detection



Comparison of detection result

Approach	Method	Sir Run Run Shaw Hospital			LiTS2017	
		AP@0.1	AP@0.5	AP@0.75	AP@0.5	AP@0.75
Anchor-based	SSD [3]	51.48± 1.78	47.42±2.57	32.33±8.30	45.33±4.75	29.90±10.17
	Faster RCNN [2]	76.62±1.43	66.38±4.10	41.82±10.97	58.24±3.91	35.74±9.93
	GSSD [10]	63.57± 2.09	57.81±1.99	46.25±4.88	50.21±4.81	34.52±10.04
	DeepLung [29]	80.33±4.47	68.01±9.91	47.44±9.19	70.29±4.79	45.41±7.89
	PixelLink [12]	83.08± 4.32	69.94±9.25	-	-	-
	Liang et al [11]	88.45± 2.46	77.06 ±5.85	-	-	-
	3DCE, 9 slices [30]	55.97±4.44	53.47±9.94	42.25±9.79	60.72±5.35	36.00±9.26
	3DCE, 27 slices [30]	81.84±1.22	68.09±1.70	47.60±4.78	69.22±4.90	38.72±8.91
	3DCE, 9 slices multi-phases [30]	92.41± 1.22	83.68±3.78	59.48±5.35	-	-
Anchor-free	CornerNet [13]	84.79±2.92	72.48±4.76	49.54±7.33	64.56±5.01	42.73±9.13
	ExtremeNet [14]	88.30±4.13	75.52±3.30	56.91±5.93	66.87±5.77	45.98±6.90
	CenterNet [16]	90.32±1.55	81.07±3.89	53.29±6.01	73.55±4.21	45.31±7.27
	MSPA-DLA++ (single phase)	-	-	-	**74.00±4.23**	**48.25±6.32**
	MSPA-DLA++	**94.02±0.98**	**84.26±3.02**	**60.08±5.16**	-	-

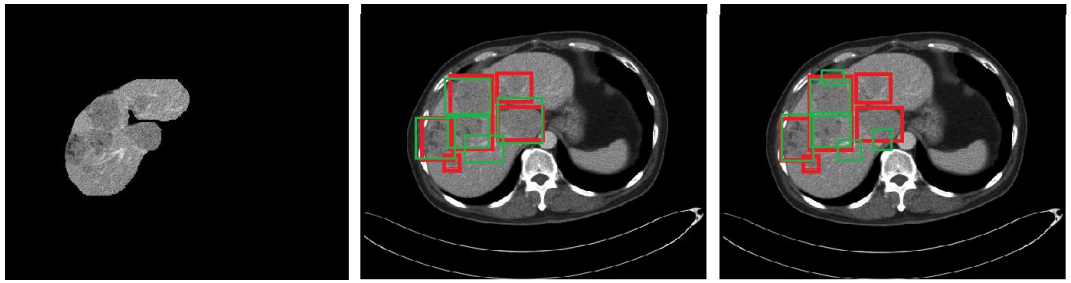
The effectiveness of DLA++

Model	Backbone	Phase	AP@0.1	AP@0.5
S1	DLA [16]	Single phase	90.32±1.55	81.07±3.89
S2	DLA++	Single-phase	91.27±1.32	81.90±3.65
M1	DLA	Multi-phase	91.29±1.33	81.79±3.75
M2	DLA++	Multi-phase	91.81±1.40	82.07±3.51

Ablation study

Model	Backbone	Phase	Multi-scales	PA	AP@0.1	AP@0.5
S2	DLA++ [31]	-	-	-	91.27±1.32	81.90±3.65
M2	DLA++	✓	-	-	91.81±1.40	82.07±3.51
M3	G-DLA++	✓	✓	-	92.30±1.15	82.56±3.38
M4	MA-DLA++	✓	-	✓	93.01±1.03	82.72±3.40
M5	MSPA-DLA++	✓	✓	✓	94.02±0.98	84.26±3.02

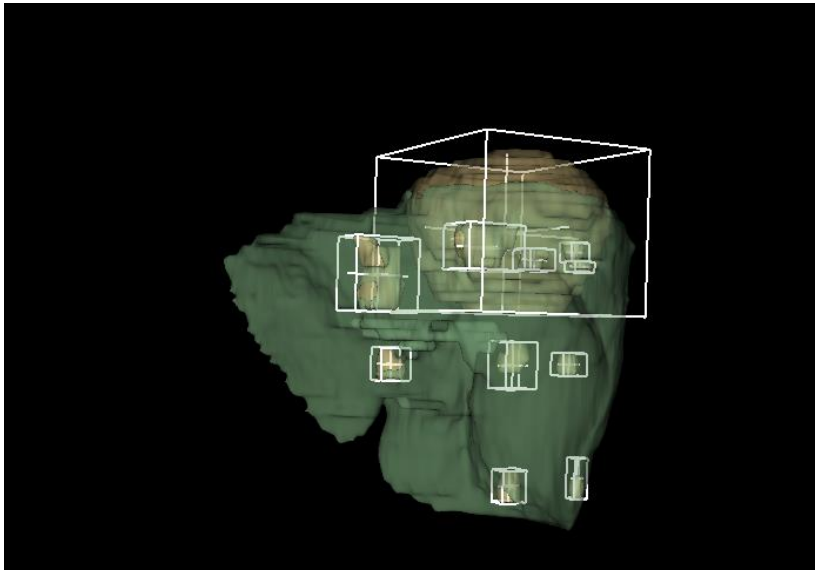
Visualization results

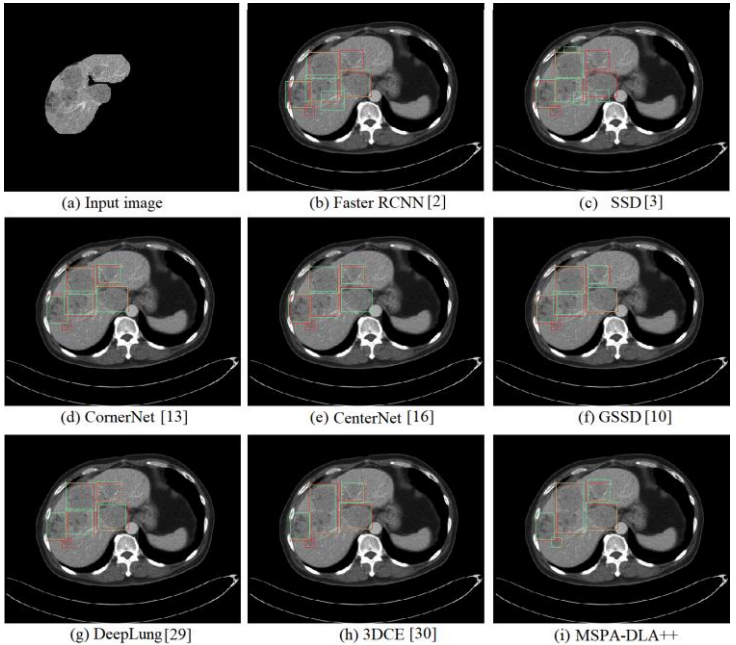


(a) Input image (b) Faster RCNN (c) SSD

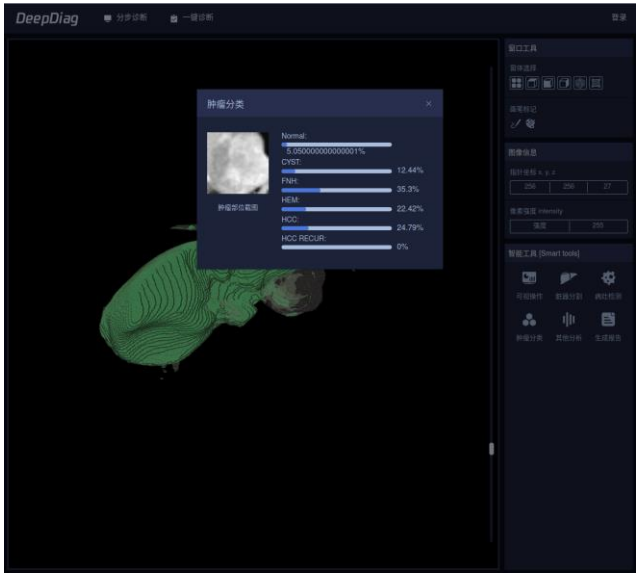
(d) ExtremeNet (e) CenterNet (f) RDLA++

Visualization results





Additional works

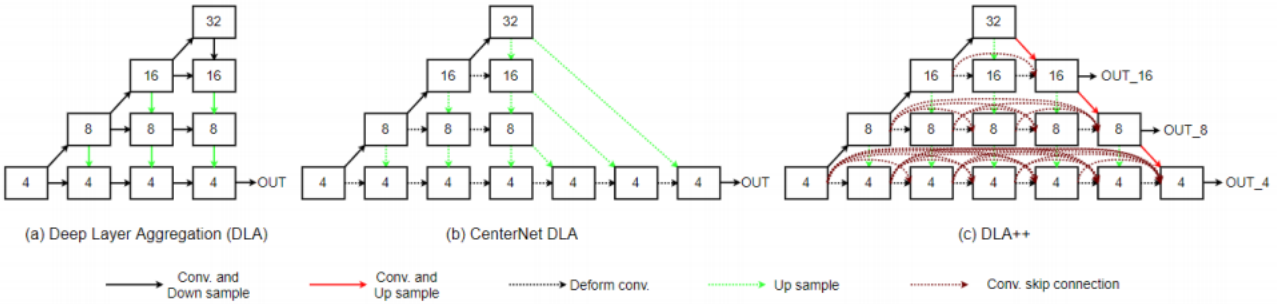


MEDINFO23

8 – 12 JULY 2023 | SYDNEY, AUSTRALIA

Thank you

DLA++



$$x_j^i = \begin{cases} H([x_j^{i-1}]), & \text{j} = 0 \\ H([U(x_{j-1}^{i+1}), [x_k^i]_{k=0}^{j-1}]), & 0 < \text{j} < J - ((i \times 2) + 1) \\ H([U(x_{j-2}^{i+1}), [x_k^i]_{k=0}^{j-1}]), & \text{otherwise} \end{cases}$$