Localization of organs and structures

Jan Kybic

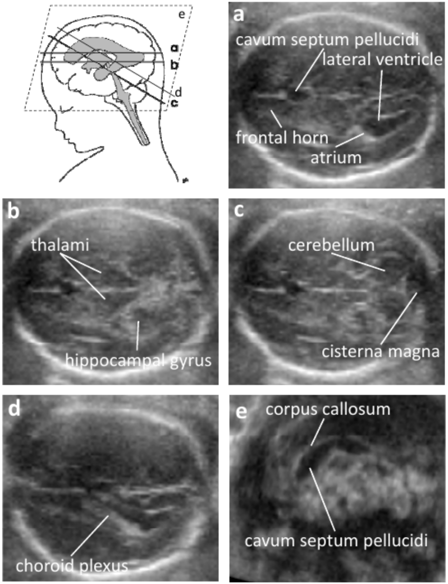
2020

Sofka et al: Automatic Detection and Measurement of Structures in Fetal Head Ultrasound Volumes Using Sequential Estimation and Integrated Detection Network (IDN)

Key points:

- Motivation - fetus development examination
- Find organs or structures in 3D (ultrasound)
- Probabilistic hierarchical sequential detection
- Probabilistic model learnt from data - relative position and appearance

Standard planes



Measurements to be performed

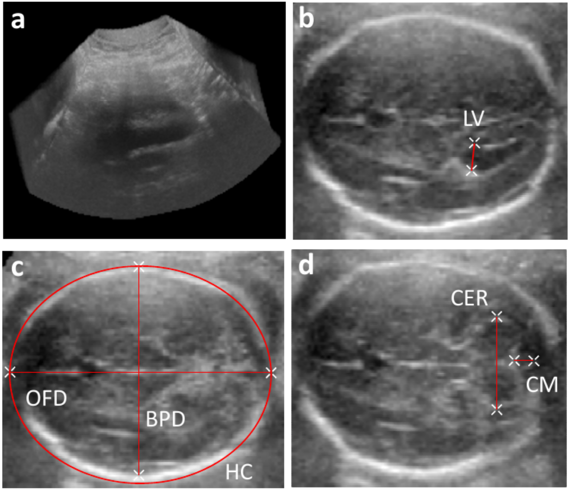
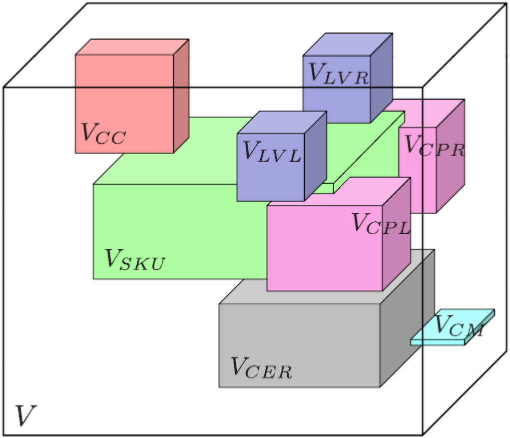


Fig. 2. Using the input volume (a), Automatic Fetal Head and Brain (AFHB) system provides the following measurements: (b) Lateral Ventricles (LV), (c) Cerebellum (CER) and Cisterna Magna (CM), and (d) Occipitofrontal Diameter (OFD), Biparietal Diameter (BPD), and Head Circumference (HC). In addition, the system provides automatic detection of median plane for visualization of Corpus Callosum (CC) and a plane for visualization of

Observations



Notation

detected by a multi-object detection system as follows. The state (pose) of the modeled object s is denoted as $\boldsymbol{\theta}_s$ and the sequence of multiple objects as $\boldsymbol{\theta}_{0:s} = \{\boldsymbol{\theta}_0, \boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_s\}$. In our case, $\boldsymbol{\theta}_s = \{\mathbf{p}, \mathbf{r}, \mathbf{s}\}$ denotes the position $\mathbf{p}$, orientation $\mathbf{r}$, and size $\mathbf{s}$ of the object s. The set of observations for object s are obtained from the image neighborhood V_s. The

- observations conditionally independent, likelihood $f(V_s|\boldsymbol{\theta}_s)$
- initial distribution $f(\theta_0)$
- transition distribution $f(\boldsymbol{\theta}_s|\boldsymbol{\theta}_{0:s-1})$

Prediction and update

- prediction

$$f(\boldsymbol{\theta}_{0:s}|V_{0:s-1}) = f(\boldsymbol{\theta}_s|\boldsymbol{\theta}_{0:s-1})f(\boldsymbol{\theta}_{0:s-1}|V_{0:s-1}).$$

- update

$$f(\boldsymbol{\theta}_{0:s}|V_{0:s}) = \frac{f(V_s|\boldsymbol{\theta}_s)f(\boldsymbol{\theta}_{0:s}|V_{0:s-1})}{f(V_s|V_{0:s-1})},$$

Particle filtering/importance sampling

- represent $f(\boldsymbol{\theta}_{0:s}|V_{0:s})$ by a set of weighted particles $\{\boldsymbol{\theta}^j_{0:s}, w^j_s\}^m_{j=1}$
- particles generated by a *proposal distribution* $p(\boldsymbol{\theta}_{0:s}|V_{0:s})$

$$p(\boldsymbol{\theta}_{0:s}|V_{0:s}) = p(\boldsymbol{\theta}_{0:s-1}|V_{0:s-1})p(\boldsymbol{\theta}_s|\boldsymbol{\theta}_{0:s-1}, V_{0:s}).$$

Particle filtering/importance sampling

- represent $f(\boldsymbol{\theta}_{0:s}|V_{0:s})$ by a set of weighted particles $\{\boldsymbol{\theta}_{0:s}^j, w_s^j\}_{j=1}^m$
- particles generated by a *proposal distribution* $p(\boldsymbol{\theta}_{0:s}|V_{0:s})$

$$p(\boldsymbol{\theta}_{0:s}|V_{0:s}) = p(\boldsymbol{\theta}_{0:s-1}|V_{0:s-1})p(\boldsymbol{\theta}_s|\boldsymbol{\theta}_{0:s-1}, V_{0:s}).$$

- adjust weights

$$\tilde{w}_s^j = \frac{f(V_{0:s}|\boldsymbol{\theta}_{0:s}^j)f(\boldsymbol{\theta}_{0:s}^j)}{p(\boldsymbol{\theta}_{0:s}^j|V_{0:s})} \qquad = \tilde{w}_{s-1}^j \frac{f(V_s|\boldsymbol{\theta}_s^j)f(\boldsymbol{\theta}_s^j|\boldsymbol{\theta}_{0:s-1}^j)}{p(\boldsymbol{\theta}_s^j|\boldsymbol{\theta}_{0:s-1}^j, V_{0:s})}.$$

$$w_s^j = \tilde{w}_s^j / \sum_{i=1}^m \tilde{w}_s^i.$$

$$\tilde{w}_s^j = \tilde{w}_{s-1}^j f(V_s|\boldsymbol{\theta}_s^j).$$

Object detection by importance sampling

1) Obtain m samples from the proposal distribution, $\boldsymbol{\theta}_s^j \sim p(\boldsymbol{\theta}_s^j|\boldsymbol{\theta}_{0:s-1}^j)$.
2) Reweight each sample according to the importance ratio

$$\tilde{w}_s^j = \tilde{w}_{s-1}^j f(V_s|\boldsymbol{\theta}_s^j). \tag{10}$$

   Normalize the importance weights.
3) Resample the particles using their importance weights to obtain more particles in the peaks of the distribution. Finally, compute the approximation of $f(\boldsymbol{\theta}_{0:s}|V_{0:s})$:

$$f(\boldsymbol{\theta}_{0:s}|V_{0:s}) \approx \sum_{j=1}^{m} w_s^j \delta(\boldsymbol{\theta}_{0:s} - \boldsymbol{\theta}_{0:s}^j), \tag{11}$$

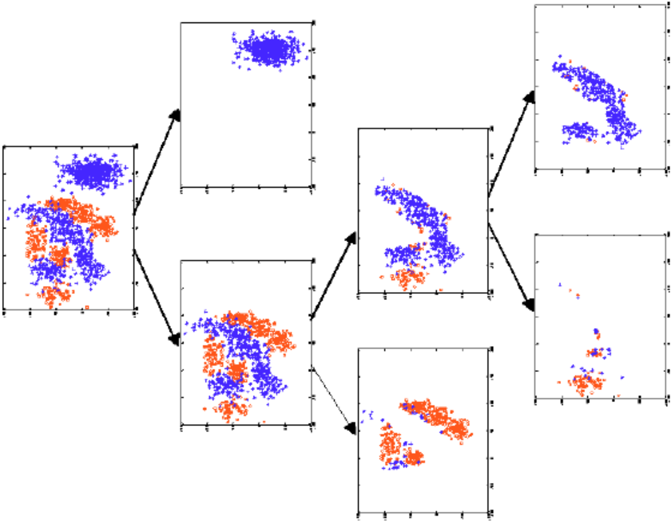
   where δ is the Dirac delta function.

Probabilistic boosting tree (PBT)

To train a tree of maximum depth of L:

- Given: A training set $S = \{(x_1, y_1, w_1), ..., (x_m, y_m, w_m); \quad x_i \in \chi, \quad y_i \in \{-1, +1\}, \sum_i w_i = 1$.
- Compute the empirical distribution $\hat{q}(y) = \sum_i w_i \delta(y_i = y)$.
- On training set S, train a strong classifier using a boosting algorithm with T weak classifiers but exit early if $\epsilon_t > \theta$, e.g. $\theta = 0.45$.
- If the current tree depth is L then exits.
- Initialize two empty sets S_{left} and S_{right}.
- For each sample (x_i, y_i), compute the probability $q(+1|x_i)$ and $q(-1|x_i)$ using the learned strong classifier.
- If $q(+1|x_i) - \frac{1}{2} > \varepsilon$ then $(x_i, y_i, 1) \rightarrow S_{right}$
  else If $q(-1|x_i) - \frac{1}{2} > \varepsilon$ then $(x_i, y_i, 1) \rightarrow S_{left}$
  else $(x_i, y_i, q(+1|x_i)) \rightarrow S_{right}$ and $(x_i, y_i, q(-1|x_i)) \rightarrow S_{left}$.
- Normalize all the weights of the samples in S_{left}.
- Repeat the procedure recursively.
- Normalize all the weights of the samples in S_{right}.
- Repeat the procedure recursively.

PBT illustration



Observation model

- Optimize instance parameters

$$\hat{\boldsymbol{\theta}}_s = \arg\max_{\boldsymbol{\theta}_s} P(V_s|\boldsymbol{\theta}_s).$$

- Probabilistic boosting tree (PBT), $y_s = 1$ iff object $s$is in V_s

$$\hat{\boldsymbol{\theta}}_s = \arg\max_{\boldsymbol{\theta}_s} P(y_s = +1|\boldsymbol{\theta}_s, V_s)$$

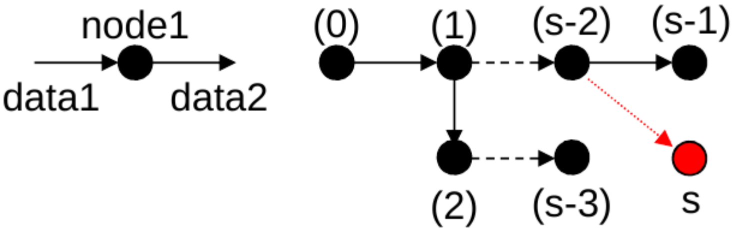
Transition model

context. In this paper, we use a pairwise dependency

$$f(\boldsymbol{\theta}_s|\boldsymbol{\theta}_{0:s-1}) = f(\boldsymbol{\theta}_s|\boldsymbol{\theta}_j), \quad j \in \{0, 1, \ldots, s-1\}. \qquad (14)$$

We model $f(\boldsymbol{\theta}_s|\boldsymbol{\theta}_j)$ as a Gaussian distribution estimated from the training data. The statistical model captures spatial

Integrated detection network (IDN)



Order to maximize posterior probability (greedily)

$$S[s,(j)] = \int_{\substack{\boldsymbol{\theta}_s \in \Omega(\tilde{\boldsymbol{\theta}}_s) \\ \boldsymbol{\theta}_{(0:s-1)} \in \Omega(\tilde{\boldsymbol{\theta}}_{(0:s-1)})}} f(\boldsymbol{\theta}_{(0:s-1)} | V_{(0:s-1)}) f(\boldsymbol{\theta}_s | \boldsymbol{\theta}_{(j)}) f(V_s | \boldsymbol{\theta}_s) d\boldsymbol{\theta}_s d\boldsymbol{\theta}_{(0:s-1)}, \tag{15}$$

Anatomy specific transition models

the pose of the cerebellum *within* the cerebellar plane. We can write the transition model from (14) as

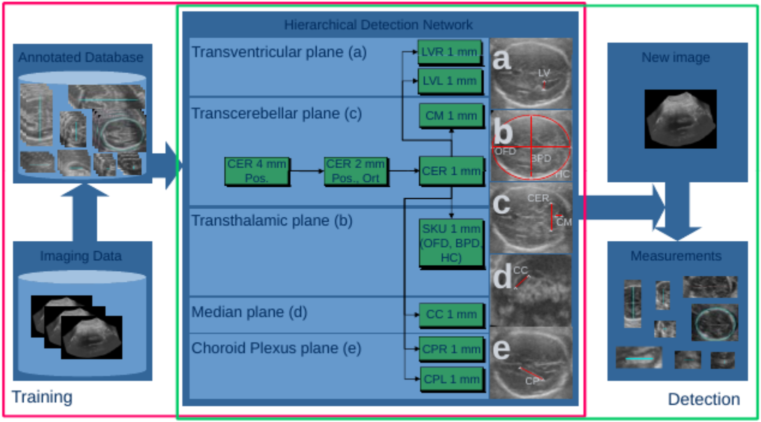
$$f(\boldsymbol{\theta}_s|\boldsymbol{\theta}_j) = f(\boldsymbol{\theta}_{CM}|\hat{\boldsymbol{\theta}}_{CER}) = \mathcal{N}(\mu_{CER-CM}, \sigma^2_{CER-CM}), \tag{16}$$

Dataset

- 2089 heads, 1982 for training, 107 for testing
- augmentation - X,Y,Z flipping

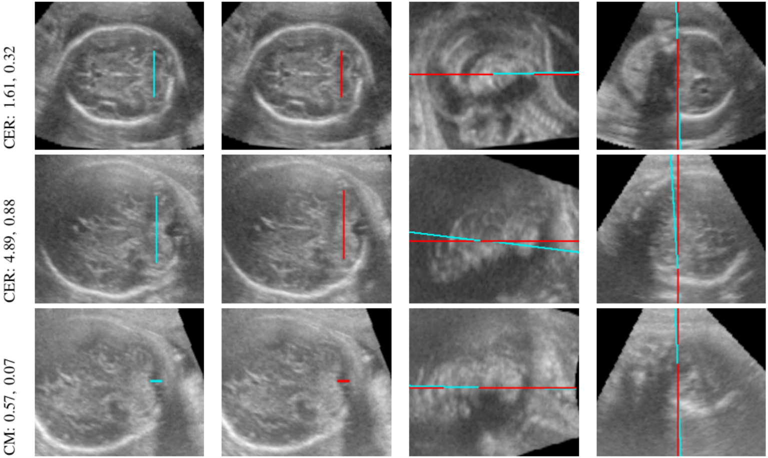
	Antares	S2000	Total
CER, CM, LV	884	1205	2089
HC, BPD, OFD	365	1206	1571
CC, CP	0	1193	1193
Total (all structures)	3747	9619	13366

Network structure

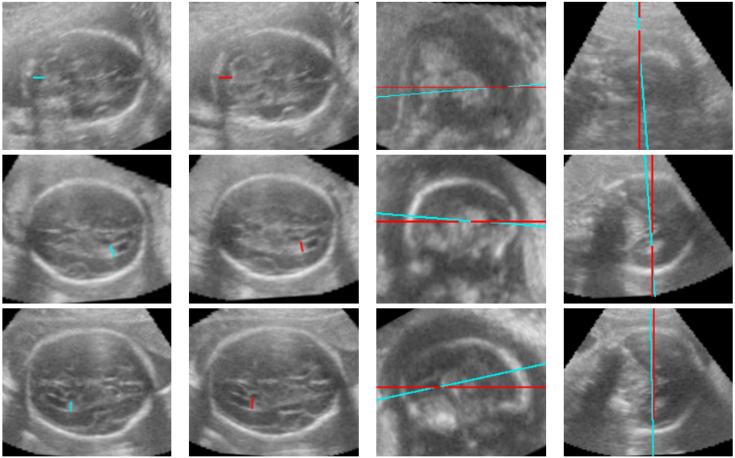


- 8 structures + 2 resolution levels for CER, detectors for position, position+orientation, position+orientation+size (not at low resolution)
- 2 level classification ("bootstrap") $\rightarrow$ 54 classifiers

Example results



Example results (2)

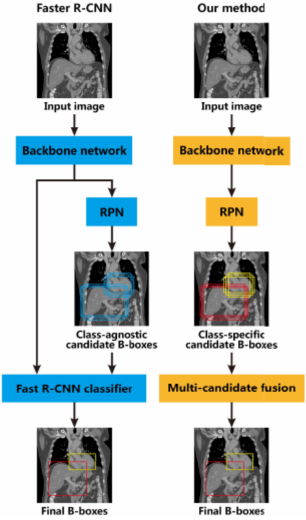


Xu: Efficient Multiple Organ Localization in CT Image Using 3D Region Proposal Network. IEEETMI 2019

Key points:

- MultipleOrgan localization (finding a bounding box)
- CNN
- Region proposals in 3D - simplified and streamlined
- high resolution feature maps

Flowchart



simplify, only 1 organ

Datasets

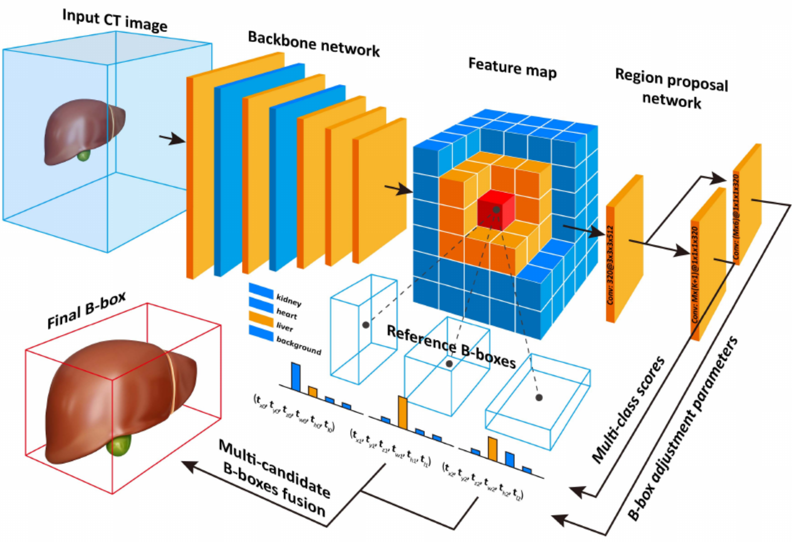
Dataset	Subset	Image number	Slice size	Slice number	In-plane resolution[mm]	Slice thickness[mm]
Abdomen	Training	118	512×512	74-987	0.56-1.00	0.70-5.00
	Validation	13	512×512	122-846	0.68-1.00	0.70-3.00
	Testing	70	512×512	42-1026	0.60-0.98	0.45-5.00
Head	Training	80	512×512	109-202	0.62-1.27	3.00
	Validation	9	512×512	130-151	0.88-1.56	3.00
	Testing	30	256×256 512×512	118-208	0.90-1.95	3.00

Abdomen - 11 organs (heart, lungs, liver, spleen...). **Head** - 12 anatomical structures (eyes, optic nerve, inner ear, oral cavity...)

Preprocessing

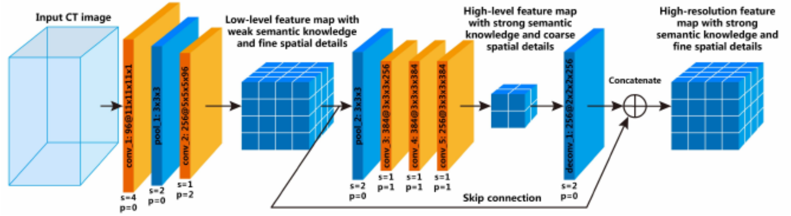
- resampling to uniform resolution,
- cropping,
- intensity rescaling
- augmentation - translation, slice subset

Flowchart


Input CT image
Backbone network
Feature map
Region proposal network
Conv: 320@3x3x3x512
Conv: Mx(K+1)@1x1x1x320
Conv: (Mx6)@1x1x1x320
kidney
heart
liver
background
Final B-box
Reference B-boxes
$(t_{x0}, t_{y0}, t_{z0}, t_{w0}, t_{h0}, t_{l0})$
$(t_{x1}, t_{y1}, t_{z1}, t_{w1}, t_{h1}, t_{l1})$
$(t_{x2}, t_{y2}, t_{z2}, t_{w2}, t_{h2}, t_{l2})$
Multi-candidate B-boxes fusion
Multi-class scores
B-box adjustment parameters

Backbone network

- based on AlexNet, combine high and low resolution features



Region proposal network

- $3 \times 3 \times 3$ convolution, two $1 \times 1 \times 1$ convolution
- **input:**
  - $W \times H \times L$ spatial cells,
  - each cell Mreference bounding boxes
  - each box has $K+1$class scores and 6 adjustment parameters
    $(t_x, t_y, t_z, t_w, t_h, t_l)$
- **goal:** predict class scores and refine box positions

$$\begin{cases} x = x_r + t_x w_r \\ y = y_r + t_y h_r \\ z = z_r + t_z l_r \end{cases} \qquad \begin{cases} w = w_r e^{t_w} \\ h = h_r e^{t_h} \\ l = l_r e^{t_l} \end{cases}$$

- **reference boxes:** size 30,60,120,240mm in each dimension (for abdomen) - $M = 4^3 = 64$ boxes

Multiple prediction strategy

Candidate fusion

- multiple candidate boxes per organ
- keep if class score $p_i > T_1$ and ranks in top $T_2\%$ of the candidates
- average positions

$$\boldsymbol{B} = \frac{\sum_i p_i \boldsymbol{B}_i}{\sum_i p_i}$$

Training

- Assign labels to boxes if IoU is maximum or if IoU $> T_f$, background if max IoU $< T\_b$ otherwise ignore.
- Compute target adjustment parameters.
- **Classification loss** focuses on hard examples ($\gamma = 0, 2$)

$$L_{cls} = -\sum_i [u_i \geq 0] \frac{1}{N_{u_i}} (1 - p_i)^\gamma \log p_i$$

- **Regression loss**

$$L_{reg} = \frac{1}{N_{total}} \sum_i [u_i > 0] \; smooth_{L_1}(\boldsymbol{t}_i - \boldsymbol{t}_i^*) \tag{5}$$

in which

$$smooth_{L_1}(x) = \begin{cases} 0.5x^2, & \text{if } |x| < 1, \\ |x| - 0.5, & \text{otherwise}, \end{cases} \tag{6}$$

- learning rate 10^{-4}, 1000 epochs, best model by IoU, random initialization

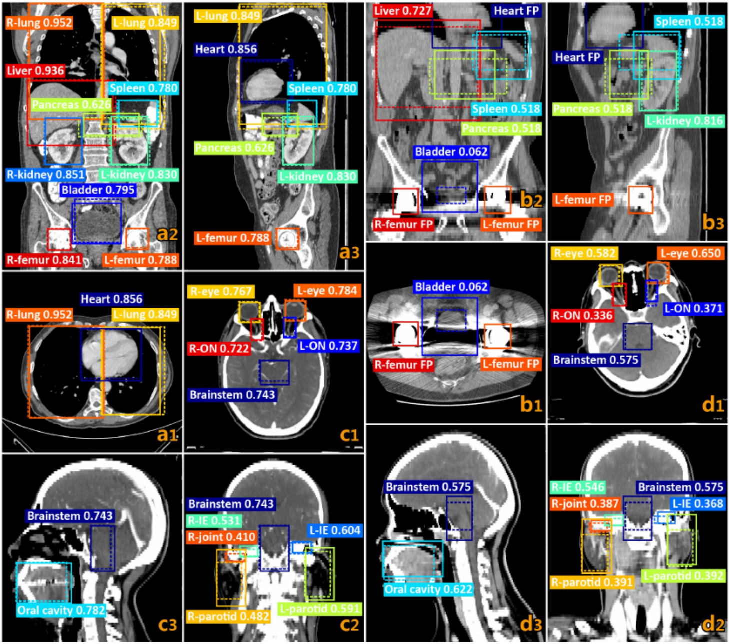
Quantitative results

DETECTION AND LOCALIZATION RESULTS OF 11 BODY ORGANS ON THE ABDOMINAL CLINICAL CT DATASET

Organs	IoU [%]		Wall dist. [mm]	Centroid dist. [mm]
	Mean	Worst		
Left lung	84.84	72.73	5.09(3.83)	7.55(2.64)
Right lung	86.88	72.31	4.87(4.93)	7.91(4.80)
Heart	80.52	66.46	4.07(4.63)	6.53(4.05)
Liver	77.83	58.62	8.46(9.36)	12.26(6.69)
Spleen	70.01	42.17	6.28(6.65)	9.95(5.84)
Pancreas	58.56	16.89	9.23(7.95)	13.25(6.48)
Left kidney	75.29	48.04	4.31(4.18)	6.19(3.75)
Right kidney	76.46	48.35	3.89(3.47)	5.59(2.86)
Bladder	58.23	6.24	7.32(6.53)	10.23(4.58)
Left femoral head	77.26	43.86	2.10(1.89)	3.25(1.79)
Right femoral head	79.77	61.72	1.85(1.62)	3.01(1.53)
Global	73.01	6.24	5.38(6.25)	7.94(5.73)

Precision[%]	Recall[%]	AP[%]	Time[s]
97.91	98.71	98.24	0.29

Examples


R-lung 0.952
L-lung 0.849
Liver 0.936
Spleen 0.780
Pancreas 0.626
R-kidney 0.851
L-kidney 0.830
Bladder 0.795
R-femur 0.841
L-femur 0.788
a2
L-lung 0.849
Heart 0.856
Spleen 0.780
Pancreas 0.626
L-kidney 0.830
L-femur 0.788
a3
Liver 0.727
Heart FP
Spleen 0.518
Pancreas 0.518
Bladder 0.062
R-femur FP
L-femur FP
b2
Spleen 0.518
Heart FP
Pancreas 0.518
L-kidney 0.816
L-femur FP
b3
Heart 0.856
R-lung 0.952
L-lung 0.849
a1
R-eye 0.767
L-eye 0.784
R-ON 0.722
L-ON 0.737
Brainstem 0.743
c1
Bladder 0.062
R-femur FP
L-femur FP
b1
R-eye 0.582
L-eye 0.650
R-ON 0.336
L-ON 0.371
Brainstem 0.575
d1
Brainstem 0.743
Oral cavity 0.782
c3
Brainstem 0.743
R-IE 0.531
R-joint 0.410
L-IE 0.604
R-parotid 0.482
L-parotid 0.591
c2
Brainstem 0.575
Oral cavity 0.622
d3
R-IE 0.546
Brainstem 0.575
R-joint 0.387
L-IE 0.368
R-parotid 0.391
L-parotid 0.392
d2