Temporal HeartNet: Towards Human-Level Automatic Analysis of Fetal Cardiac Screening Video

Weilin Huang, **Christopher P. Bridge**, J. Alison Noble, and Andrew Zisserman

Department of Engineering Science, University of Oxford, UK

MICCAI (12^{th} September 2017)

Congenital Heart Disease (CHD)

- Range of structural heart defects present at birth
- Leading cause of infant mortality
  - 42% of infant deaths reported to WHO
- Heart examination during routine second trimester abnormality screening using **2D ultrasound**
- Multiple viewing planes
- Highly skilled

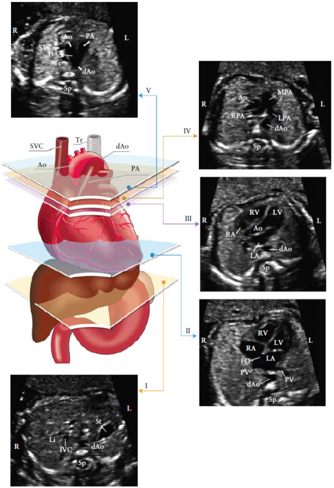


Figure: Source: ISUOG

Congenital Heart Disease (CHD)

- Range of structural heart defects present at birth
- Leading cause of infant mortality
  - 42% of infant deaths reported to WHO
- Heart examination during routine second trimester abnormality screening using **2D ultrasound**
- Multiple viewing planes
- Highly skilled

Aim

Can automated methods assist in analysing this video?

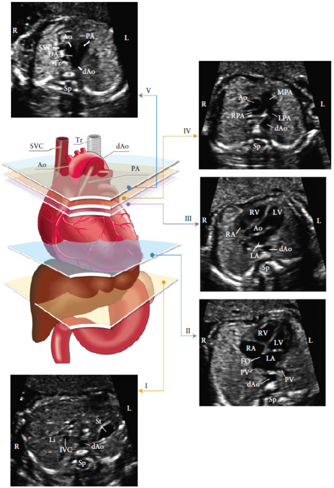
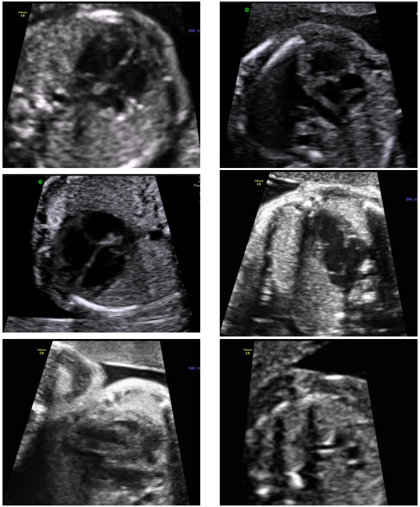


Figure: Source: ISUOG

Challenges

Challenges

- Unconstrained video stream
- Unpredictable probe/fetal movement
- Variable orientation
- Imaging artefacts (speckle, shadowing, enhancement)
- Low contrast and indistinct structures

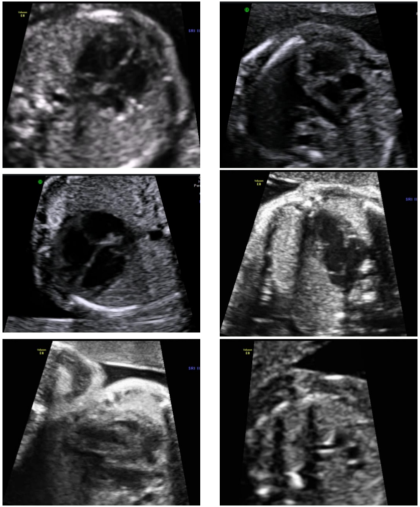


Challenges

Challenges

- Unconstrained video stream
- Unpredictable probe/fetal movement
- Variable orientation
- Imaging artefacts (speckle, shadowing, enhancement)
- Low contrast and indistinct structures

- **How can we make sense of this raw video data?**



Aims

- Interpret the video stream
- Estimate key low-level variables of interest:

Aims

- Interpret the video stream
- Estimate key low-level variables of interest:
  1. Heart Visibility, $h_t \in \{0, 1\}$
  2. Heart Centre Position, $\mathbf{x}_t \in \mathbb{R}^2$
  3. View Label, $v_t \in \{4C, LVOT, 3V\}$
     - Four chamber
     - Left Ventricular Outflow Tract
     - Three vessels
  4. Heart Orientation, $\theta_t \in [0, 2\pi)$
  5. Heart Radius, $r_t \in \mathbb{R}^+$

Aims

- Interpret the video stream
- Estimate key low-level variables of interest:
  1. Heart Visibility, $h_t \in \{0, 1\}$
  2. Heart Centre Position, $\mathbf{x}_t \in \mathbb{R}^2$
  3. View Label, $v_t \in \{4C, LVOT, 3V\}$
     - Four chamber
     - Left Ventricular Outflow Tract
     - Three vessels
  4. Heart Orientation, $\theta_t \in [0, 2\pi)$
  5. Heart Radius, $r_t \in \mathbb{R}^+$

Summary

Find a basic 'global coordinate system' for each frame

Aims

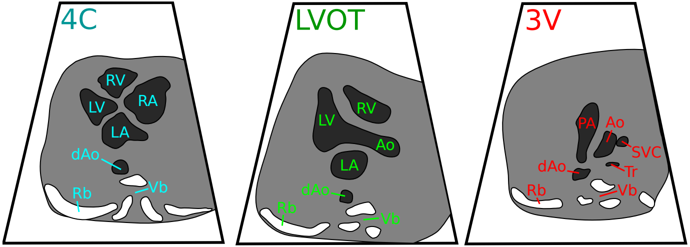
- Interpret the video stream
- Estimate key low-level variables of interest:
  1. Heart Visibility, $h_t \in \{0, 1\}$
  2. Heart Centre Position, $\mathbf{x}_t \in \mathbb{R}^2$
  3. View Label, $v_t \in \{4C, LVOT, 3V\}$
     - ⋆ Four chamber
     - ⋆ Left Ventricular Outflow Tract
     - ⋆ Three vessels
  4. Heart Orientation, $\theta_t \in [0, 2\pi)$
  5. Heart Radius, $r_t \in \mathbb{R}^+$

Summary

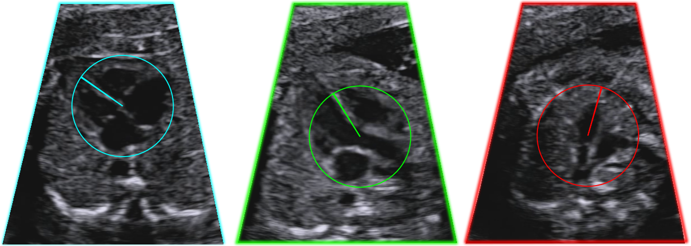
Find a basic 'global coordinate system' for each frame

- Prior work based on particle-filtering model (Bridge et al. 2017)

Viewing Plane Definitions



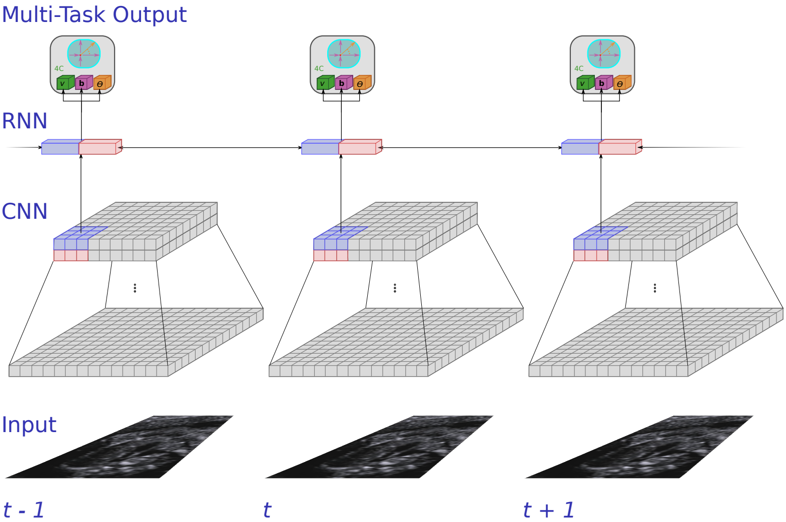
LV/RV left/right ventricle, **LA/RA** left/right atrium, (**d**)**Ao** (descending) aorta, **PA** pulmonary artery, **SVC** superior vena cava, **Tr** trachea, **Vb** vertebra, **Rb** ribs



Model Overview

Components

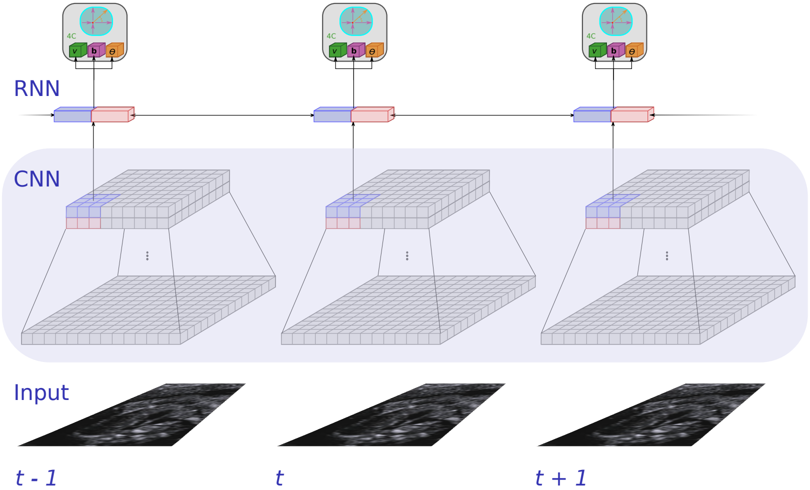
- Spatial CNN (VGG-16)
- Temporal RNN (BLSTM)
- Multi-task Layers:
  - View
  - Location
  - Orientation
  - Radius
- Two alternatives:
  - Circular anchors
  - IoU



Model Overview

Components

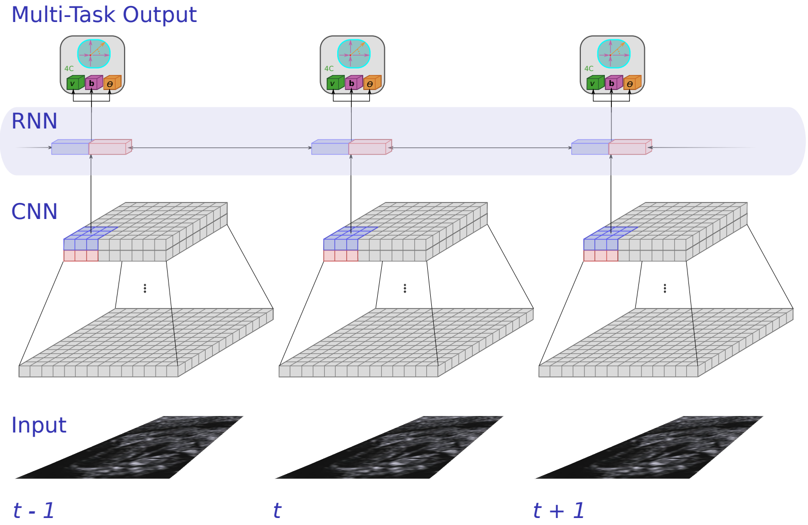
- Spatial CNN (VGG-16)
- Temporal RNN (BLSTM)
- Multi-task Layers:
  - View
  - Location
  - Orientation
  - Radius
- Two alternatives:
  - Circular anchors
  - IoU



Model Overview

Components

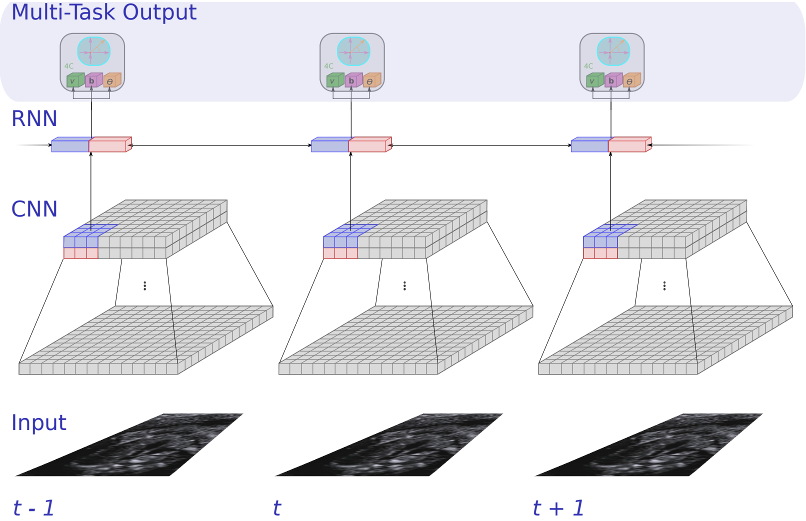
- Spatial CNN (VGG-16)
- Temporal RNN (BLSTM)
- Multi-task Layers:
  - View
  - Location
  - Orientation
  - Radius
- Two alternatives:
  - Circular anchors
  - IoU



Model Overview

Components

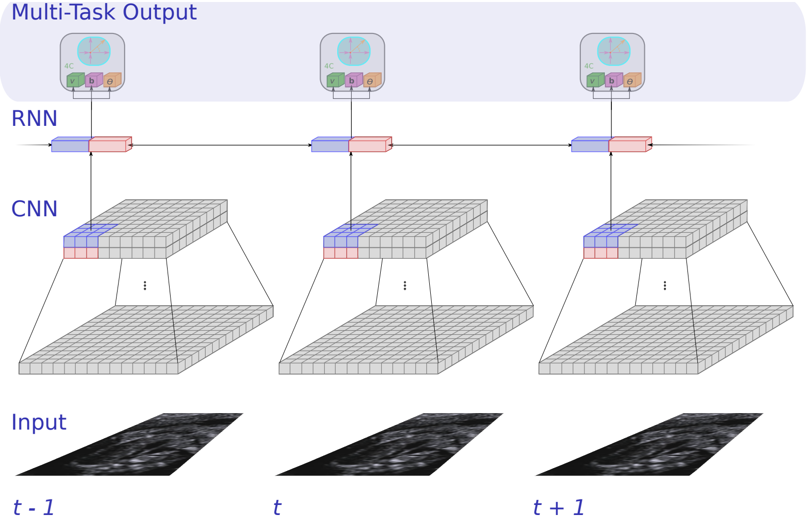
- Spatial CNN (VGG-16)
- Temporal RNN (BLSTM)
- Multi-task Layers:
  - View
  - Location
  - Orientation
  - Radius
- Two alternatives:
  - Circular anchors
  - IoU



Model Overview

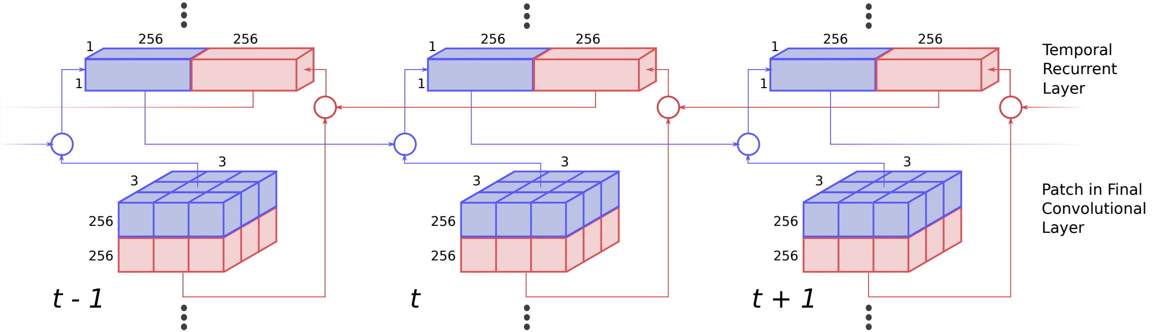
Components

- Spatial CNN (VGG-16)
- Temporal RNN (BLSTM)
- Multi-task Layers:
  - View
  - Location
  - Orientation
  - Radius
- Two alternatives:
  - Circular anchors
  - IoU



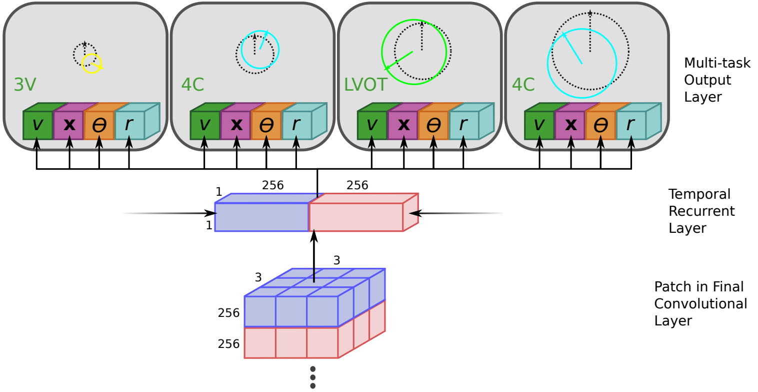
Bi-directional LSTM RNN

- Recurrent 512-D representation for an image region
- LSTM (long short-term memory) cells learn long-term dependencies
- Parameters shared between image regions
- Bi-directional: Two separate 256-D recurrent vectors: one forwards, one backwards



Multi-Task Prediction Architecture 1: ‘Circular Anchors’

- Independently predict offsets from four *circular anchors* with radii $\{80, 120, 160, 240\}$ (Ren et al. 2015, “Faster R-CNN”)



Training – Circular Anchor Architecture

- Loss functions:
  - L_{cls} **Classification** (v): Softmax
  - L_{loc} **Localisation** $(\mathbf{x}, \theta, r)$: Smooth-l_1 loss
  - **Total:**

$$L = L_{cls} + \lambda_1 L_{loc}$$

Training – Circular Anchor Architecture

- Loss functions:
  - L_{cls} **Classification** (v): Softmax
  - L_{loc} **Localisation** $(\mathbf{x}, \theta, r)$: Smooth-l_1 loss
  - **Total:**

$$L = L_{cls} + \lambda_1 L_{loc}$$

- Training samples:
  - **Positives:** Anchors with ground-truth IoU overlap > 0.7
  - **Negatives:** Anchors with ground-truth IoU overlap < 0.5
  - **Excluded:** Anchors with $0.5 <$ IoU < 0.7

Training – Circular Anchor Architecture

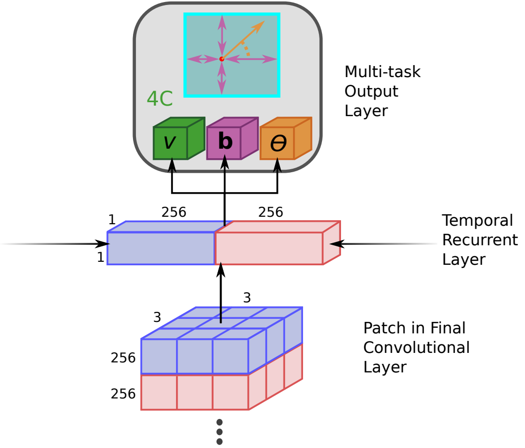
- Loss functions:
  - L_{cls} **Classification** (v): Softmax
  - L_{loc} **Localisation** $(\mathbf{x}, \theta, r)$: Smooth-l_1 loss
  - **Total:**

$$L = L_{cls} + \lambda_1 L_{loc}$$

- Training samples:
  - **Positives:** Anchors with ground-truth IoU overlap > 0.7
  - **Negatives:** Anchors with ground-truth IoU overlap < 0.5
  - **Excluded:** Anchors with $0.5 <$ IoU < 0.7
- Location, orientation and radius gradients only applied for positive anchors

Multi-Task Prediction Architecture 2: ‘IoU’

- Regress top, bottom, left, and right edges of bounding box (**b**) with IoU loss
- Orientation is regressed separately



Training – IoU Architecture

- Loss functions:
  - L_{cls} **Classification** (v): Softmax
  - L_{loc} **Localisation** $(\mathbf{b})$: IoU (intersection over union)
  - L_{ori} **Orientation** (θ): Cosine loss

$$L_{\text{ori}} = 1 - \cos\left(\hat{\theta} - \theta\right)$$

  - **Total:**

$$L = L_{cls} + \lambda_1 L_{loc} + \lambda_2 L_{ori}$$

Training – IoU Architecture

- Loss functions:
  - L_{cls} **Classification** (v): Softmax
  - L_{loc} **Localisation** $(\mathbf{b})$: IoU (intersection over union)
  - L_{ori} **Orientation** (θ): Cosine loss

$$L_{\text{ori}} = 1 - \cos\left(\hat{\theta} - \theta\right)$$

  - **Total:**

$$L = L_{cls} + \lambda_1 L_{loc} + \lambda_2 L_{ori}$$

- Training samples:
  - **Positives:** Regions within $0.7\times$ annotated heart radius
  - **Negatives:** Regions outside annotated heart radius
  - **Excluded:** Regions between 0.7 and 1.0 annotated heart radius

Training – IoU Architecture

- Loss functions:
  - L_{cls} **Classification** (v): Softmax
  - L_{loc} **Localisation** $(\mathbf{b})$: IoU (intersection over union)
  - L_{ori} **Orientation** (θ): Cosine loss

$$L_{\text{ori}} = 1 - \cos\left(\hat{\theta} - \theta\right)$$

  - **Total:**

$$L = L_{cls} + \lambda_1 L_{loc} + \lambda_2 L_{ori}$$

- Training samples:
  - **Positives:** Regions within $0.7\times$ annotated heart radius
  - **Negatives:** Regions outside annotated heart radius
  - **Excluded:** Regions between 0.7 and 1.0 annotated heart radius
- Location, orientation and radius gradients only applied for positive regions

Experiments

- Database of 91 videos from 12 healthy subjects
- Multiple views and range of gestational ages (20–35 weeks), orientations, magnifications
- Leave-one-subject-out cross-validation
- Pre-trained VGG-16 or random initialisation

Example Output

Output

Ground Truth

Results

Method	Class Error *or* Outside $0.25\hat{r}$ (%)*	Class Error *or* IoU < 0.25 (%)	Orient. Error†
Circular Anchor	28.8	30.3	0.074
IoU Loss	26.8	28.7	0.084
RNN + Circular Anchor	<u>21.6</u>	<u>27.7</u>	<u>0.072</u>

* Estimated inter-rater variation: 26%, intra-rater variation: 15%

† Orientation Error $= \frac{1}{2}\left(1 - \cos\left(\theta - \hat{\theta}\right)\right)$

Conclusions

- Deep architecture for predicting basic information in each frame of fetal cardiac screening video
- Three parts:
  - CNN
  - Spatially-localised RNN
  - Multi-task output
- Approaching human-level accuracy on highly ambiguous problem
- IoU architecture gives better localisation than circular anchors
- RNN significantly improves results

Thank You

- Christos Ioannou, John Radcliffe Hospital, Oxford
- EPSRC 'Seebibyte' Programme Grant (EP/M013774/1)
- EPSRC Doctoral Training Award

Weilin Huang: whuang@robots.ox.ac.uk
Chris Bridge: christopher.bridge@eng.ox.ac.uk
Alison Noble: alison.noble@eng.ox.ac.uk
Andrew Zisserman: az@robots.ox.ac.uk