UNIVERSITY OF OXFORD

Deep Neural Networks in Computer Vision and Biomedical Image Analysis

Weidi Xie

Supervised by :

Professor Alison Noble

Professor Andrew Zisserman

Outlines

1. Microscopy Cell Counting and Detection (§5, 2014-2015)

2. Layer Recurrent Neural Networks (§3, 2015-2016)

3. Image Analysis in 3D Fetal Neurosonography (§6, 2016-2017)

4. Cardiac Magnetic Resonance Imaging Analysis (§7, 2016-2017)

5. Comparator Networks (§4, 2016-2017)

Publications (Accepted)

— Conferences:

[1] **Weidi Xie**, J. Alison Noble, and Andrew Zisserman, “Microscopy Cell Counting with Fully Convolutional Regression Networks,” in MICCAI 1st Deep Learning Workshop, Munich, 2015.

[2] Ana Namburete, **Weidi Xie** and J. Alison Noble, “Robust Regression of Brain Maturation from 3D Fetal Neurosonography using CRNs”, in *MICCAI Workshop on Fetal and InFant Image analysis* (FIFI 2017), Best Paper Award.

[3] Davis M. Vigneault, **Weidi Xie**, David A. Bluemke and J. Alison Noble, “Feature Tracking Cardiac Magnetic Resonance via Deep Learning and Spline Optimization”, in *Functional Imaging and Modelling of the Heart* (FIMH 2017), Best Poster Award.

(New) [4] Qiong Cao, Li Shen, **Weidi Xie**, Omkar M. Parkhi and Andrew Zisserman, “VGGFace2: A Dataset for Recognizing Faces Across Pose and Age”, in IEEE Conference on Automatic Face and Gesture Recognition (F&G 2018).

— Journals:

[5] **Weidi Xie**, J. Alison Noble, and Andrew Zisserman, “Microscopy Cell Counting And Detection with Fully Convolutional Regression Networks,” in *Computer Methods in Biomechanics and Biomedical Engineering : Imaging & Visualization*.

(New) [6] **Weidi Xie**[*], Ana Namburete[*], Mohammad Yaqub, Andrew Zisserman and J. Alison Noble, “Fully-Automated Standardized Reorientation of 3D Fetal Neurosonography Images using Multi-Task FCNs”, in *Medical Image Analysis (MedIA).*

(New) [7] Ruobing Huang, **Weidi Xie** and J. Alison Noble, “VP-Nets : Efficient Automatic Localization of Key Brain Structures in 3D Fetal Neurosonography”, in *Medical Image Analysis (MedIA).*

Potential Publications (Under Review)

— Conferences:

[1] **Weidi Xie**, Li Shen, Andrew Zisserman, “Comparator Networks”, submitted to ECCV 2018 (under review).

[2] Ana Namburete, **Weidi Xie** and J. Alison Noble, “AffineNet: Spatial Alignment of Volumetric Images”, submitted to MICCAI 2018 (under review).

[3] Mohammad MA, **Weidi Xie** and J. Alison Noble, “Can Dilated Convolutions Capture Ultrasound Video Dynamics?”, submitted to MICCAI 2018 (under review).

— Journals:

[4] **Weidi Xie***, Davis M. Vigneault*, Carolyn Ho, David A. Bluemke and J. Alison Noble, “Ω-Net: Fully Automatic, Multi-View Cardiac MR Detection, Orientation Alignment, and Segmentation with Deep Neural Networks”, submitted to *Medical Image Analysis* (under second review). https://arxiv.org/pdf/1711.01094.

— Technical Report:

[5] **Weidi Xie**, J. Alison Noble and Andrew Zisserman, “Layer Recurrent Neural Networks”, https://openreview.net/ forum?id=rJJRDvcex

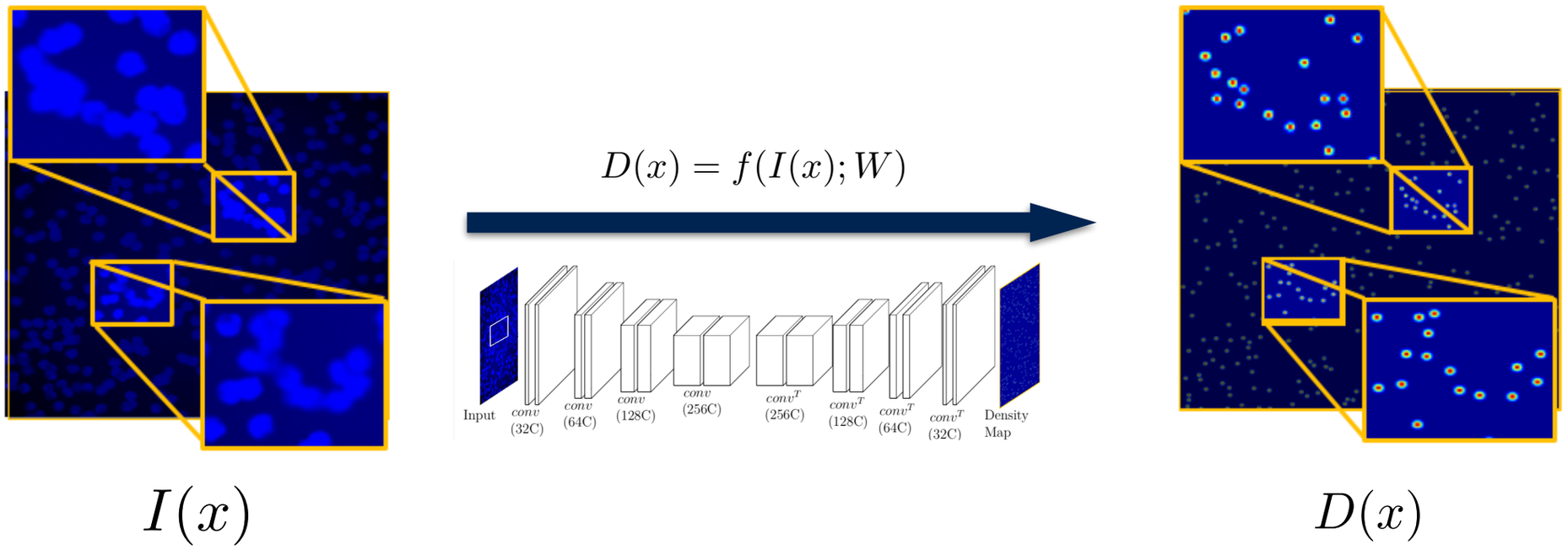
Microscopy Cell Counting and Detection

— Objectives:

- Count the number of cells in an image or a region.

— Approach:

- Find a mapping $f(\cdot)$ that maps a input image $I(x)$ to a density map $D(x)$.
- Groundtruth density maps is defined as number of cells per pixel.
- Integral over the predicted density map gives the cell counts.
- Local maxima detection servers as the cell detection.



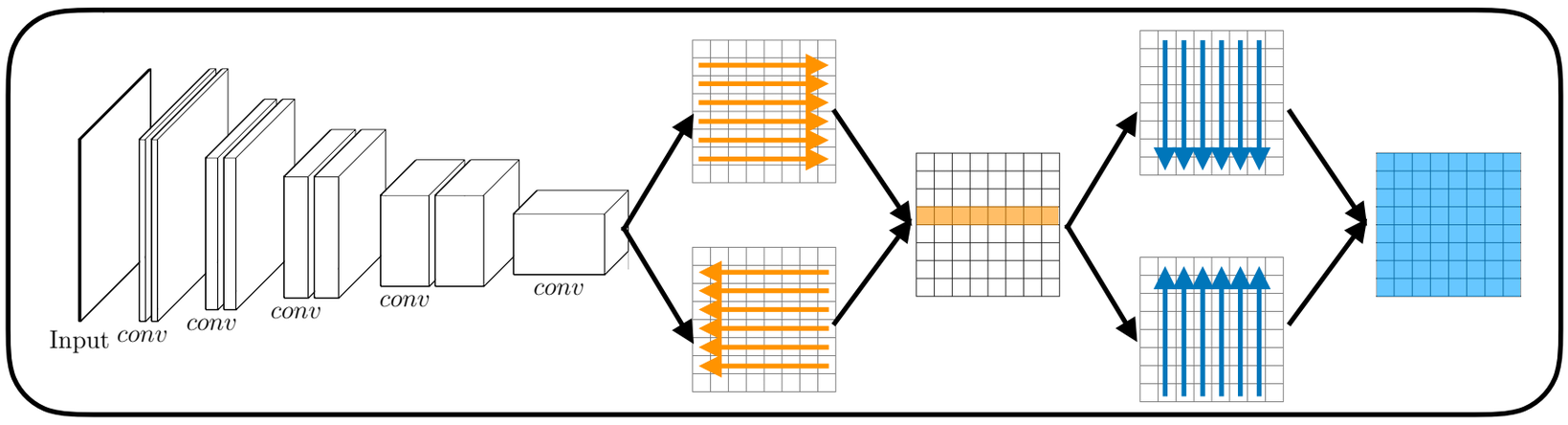
Layer Recurrent Neural Networks (LRNNs)

— Objectives:

- Semantic segmentation in natural images.
- Learning multi-scale contextual information.

— Approach:

- CNNs interleaved with spatial RNNs.
- Capture long-range dependencies by within-layer recurrence.
- Initialize as identity function, and followed by end-to-end finetuning.



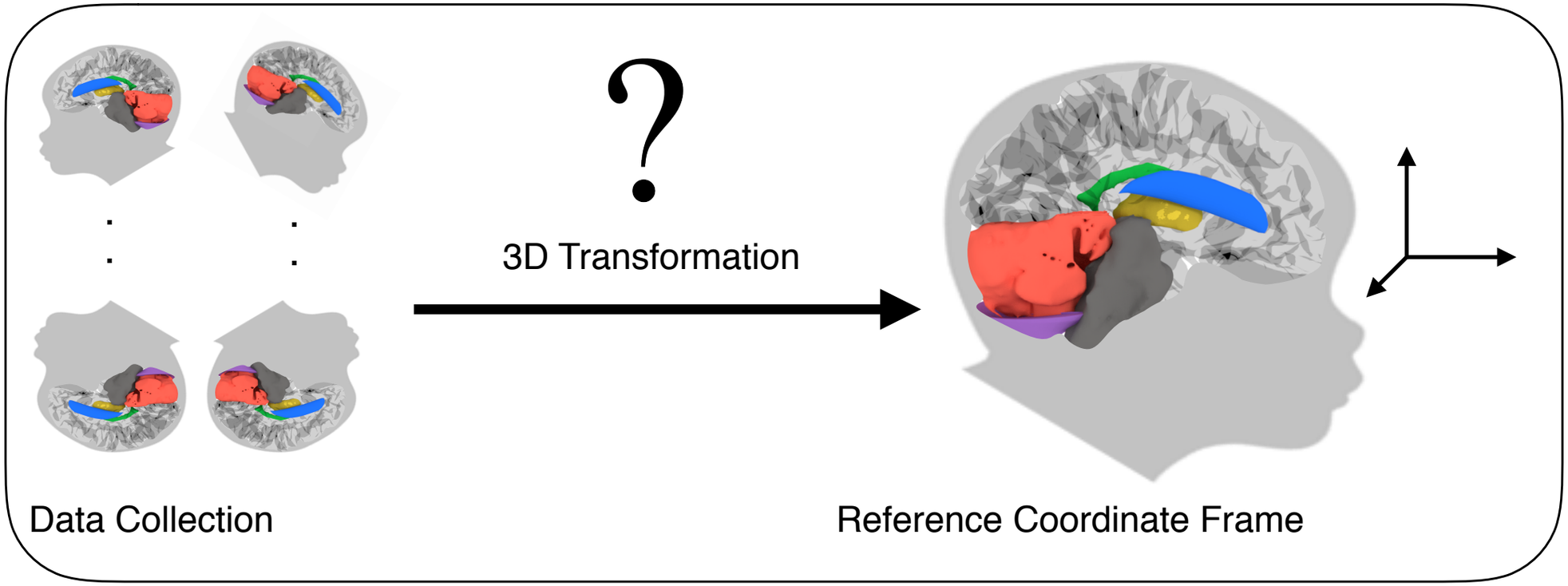
3D Fetal Neurosonography

— Objectives:

- Monitoring the fetal brain growth, i.e. important structures.

— Approach:

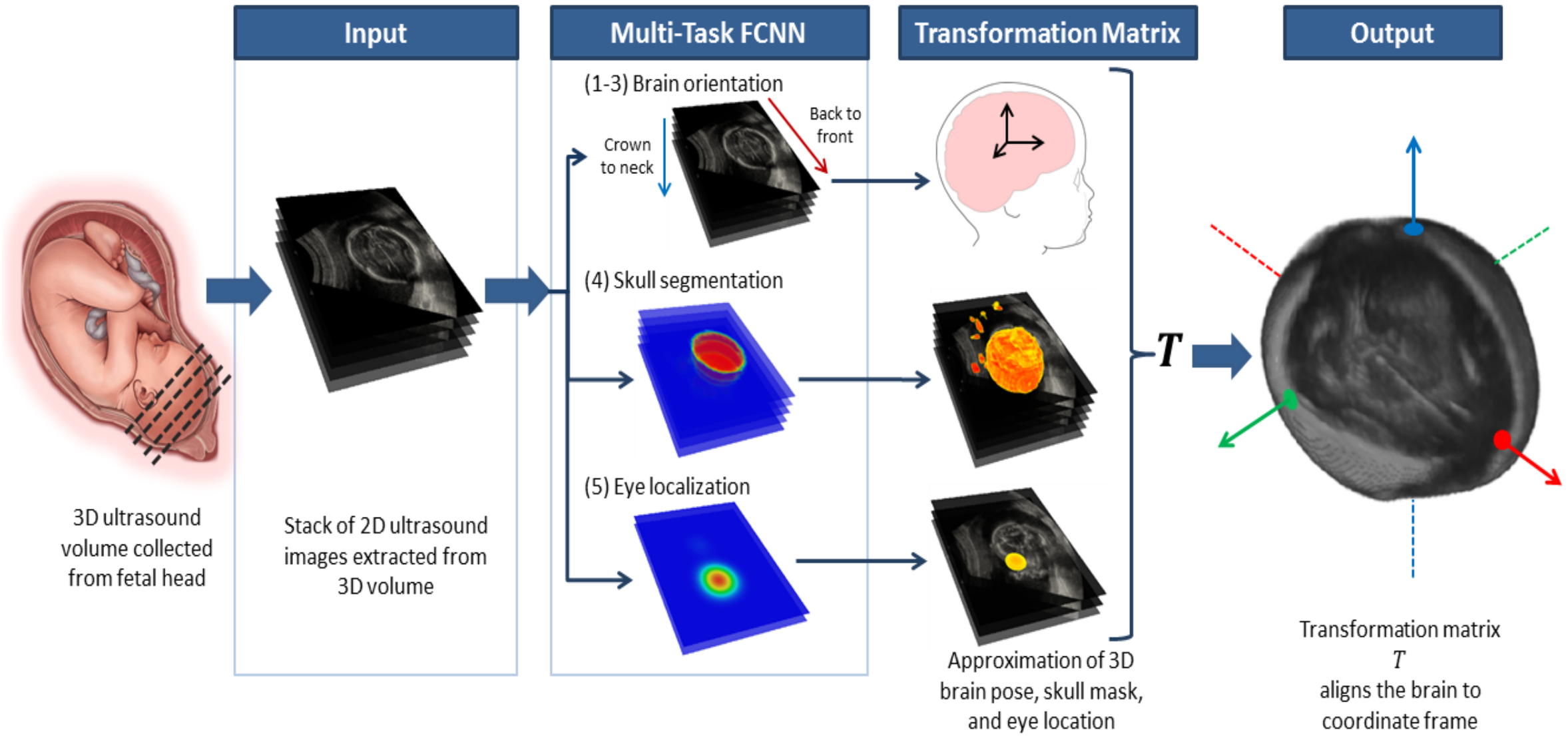
- Transform the 3D ultrasound volumes to a reference coordinate frame.



3D Fetal Neurosonography-1

— Proposed Solution

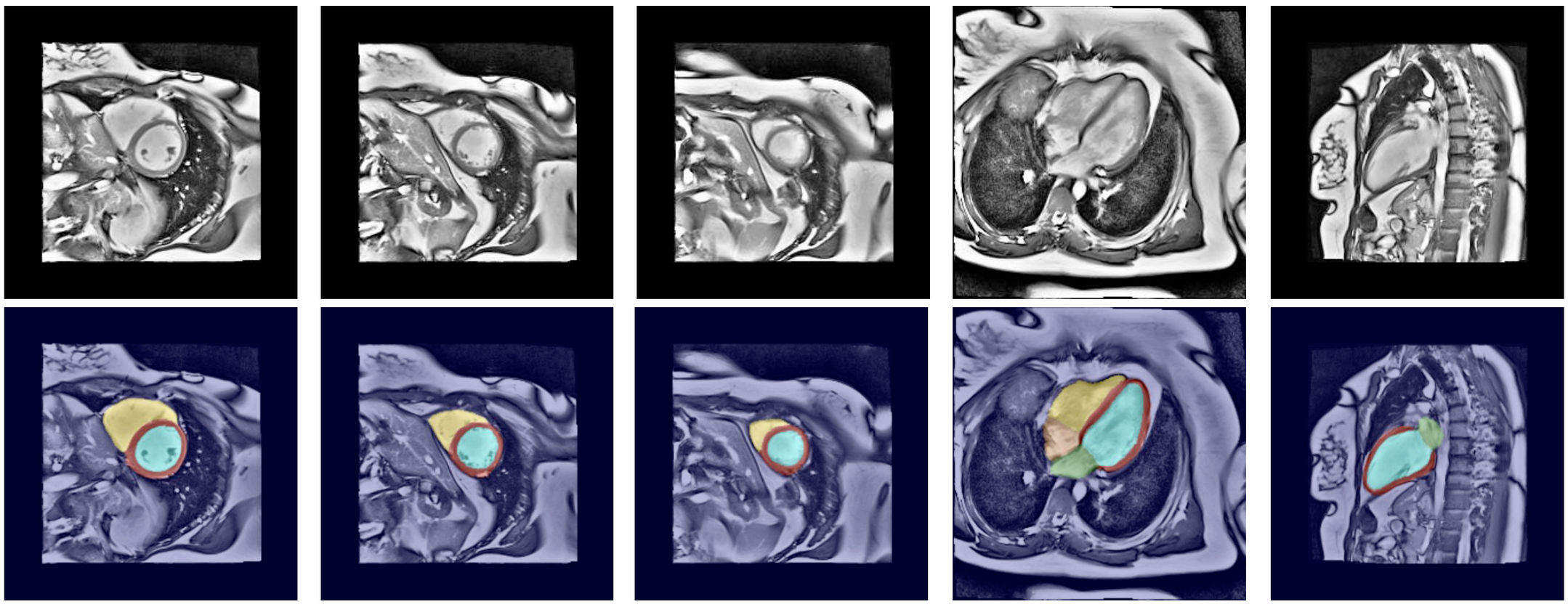
- Multi-task Convolutional Neural Networks.



Cardiac Magnetic Resonance Imaging

— Objectives:

- Evaluating the anatomy and function of the heart chambers.
- Multiview, multi-structure segmentation.



LV_BP : Left Ventricle Blood Pool

LA_BP : Left Atrium Blood Pool

LV_MY : Left Ventricle Myocardium

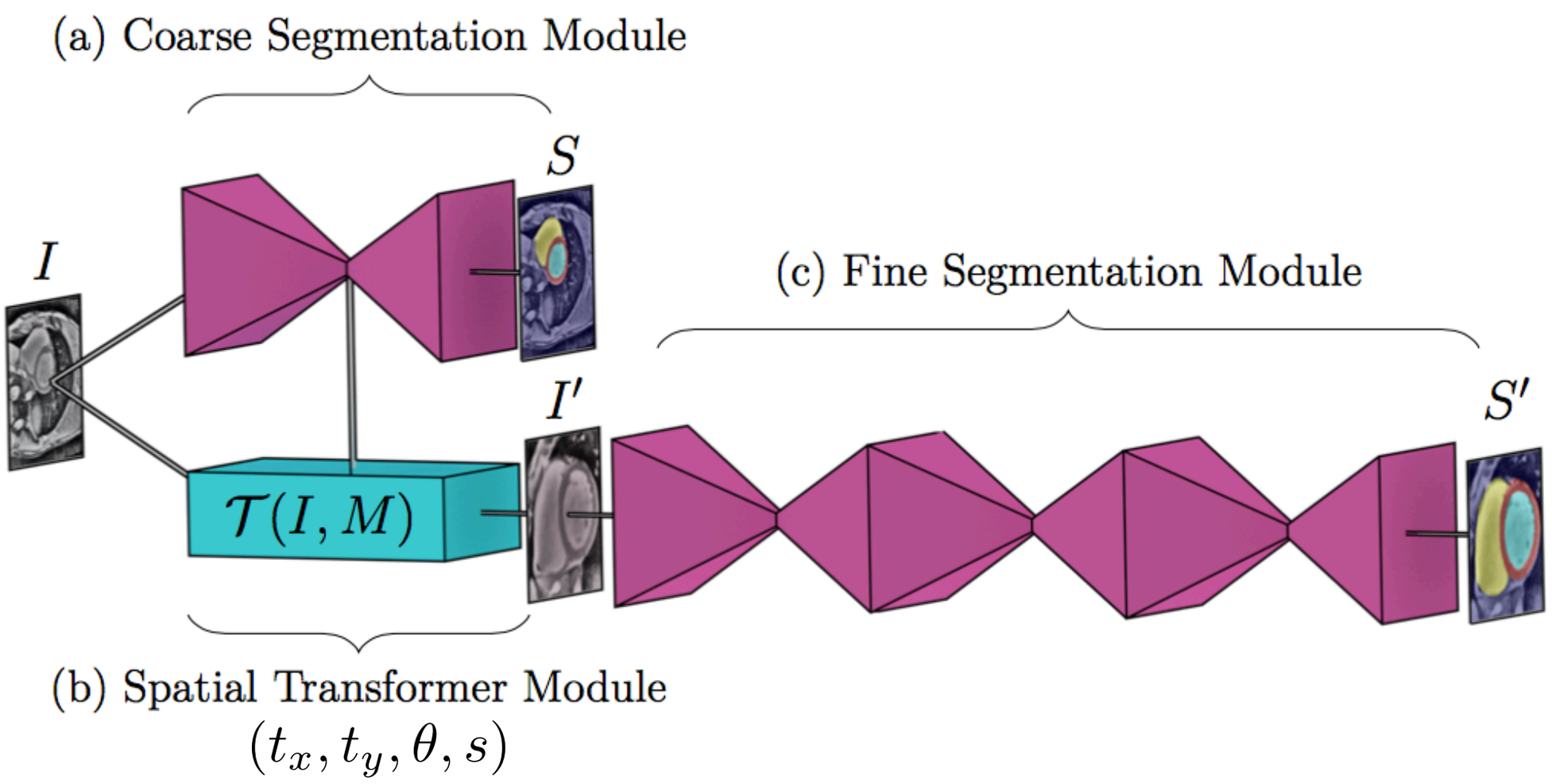
RV_BP : Right Ventricle Blood Pool

RA_BP : Right Atrium Blood Pool

Cardiac Magnetic Resonance Imaging

— Approach:

- Heart localization—> transformation into a canonical orientation—> semantic segmentation.
- All in one end-to-end trainable model.

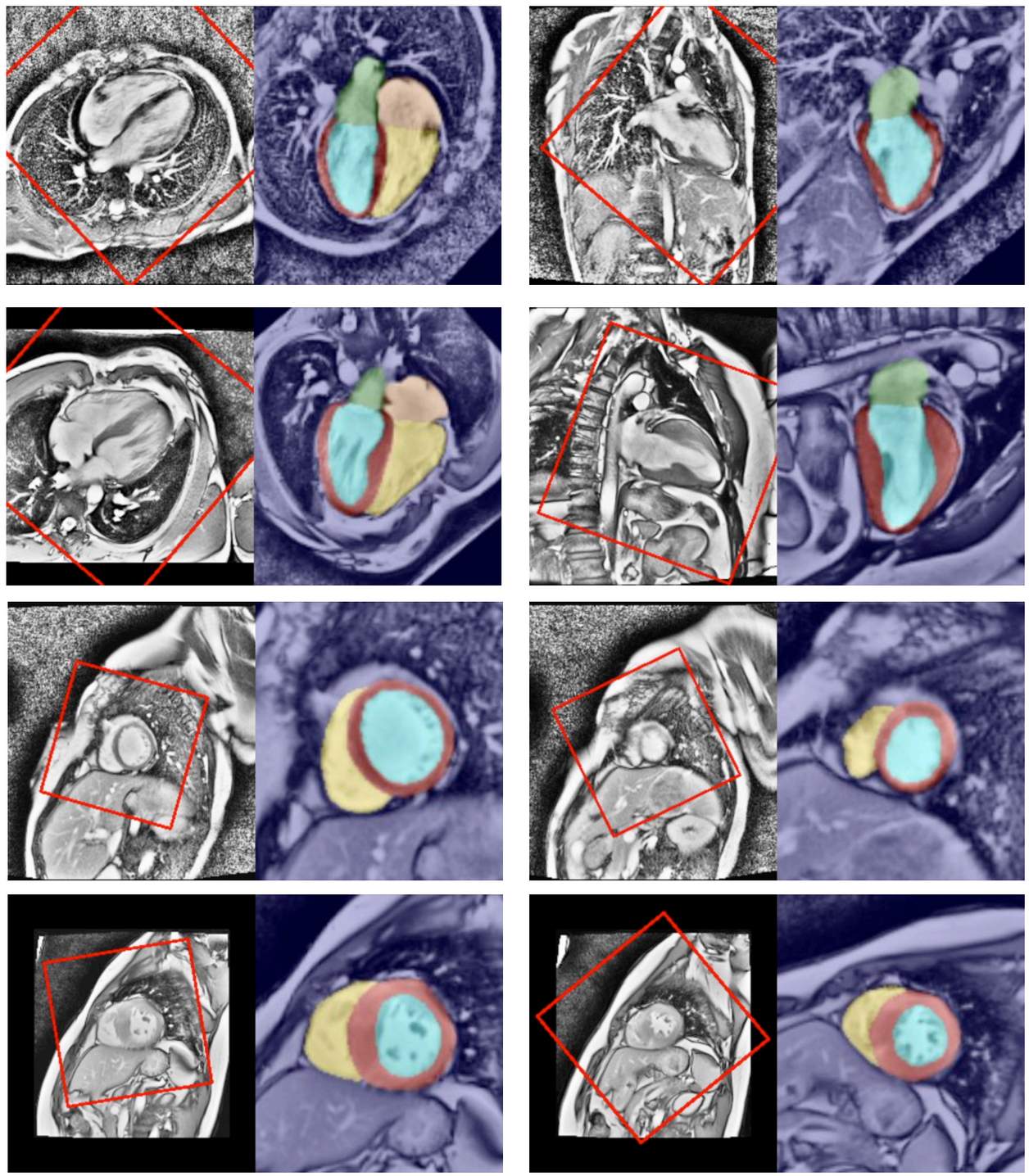


Cardiac Magnetic Resonance Imaging

— Results on in-door dataset

- Model can always transform input frame to canonical views, with orientation, scale normalized.
- Segmentation on canonical views are robust for different planes (short-axis, four-chamber, etc.).

Red boxes refer to the sampling location.



Cardiac Magnetic Resonance Imaging

— Extended evaluation:

- Train from scratch on 2017 MICCAI Automated Cardiac Diagnosis Challenge (ACDC)
- Multiple diseases, i.e. normal, myocardial infarction, dilated cardiomyopathy, hypertrophic cardiomyopathy, abnormal right ventricle.

proposed →

Structure	LV Bloodpool	RV Bloodpool	LV Myocardium
	Jaccard Index (IoU)		
Ω-Net	**0.912**	**0.852**	0.803
Isensee et al. (2018)	0.896	0.832	**0.826**
Isensee et al. (2017)	0.869	0.784	0.775
	Dice Coefficient		
Ω-Net	**0.954**	**0.920**	0.891
Isensee et al. (2018)	0.945	0.908	**0.905**
Isensee et al. (2017)	0.930	0.879	0.873

Table 3: Segmentation accuracy on the 2017 MICCIA ACDC dataset. Segmentation accuracy is reported as Dice coefficient in the ACDC challenge, but as IoU elsewhere in this work; therefore, both are reported here. (Note that Dice $= 2 * \text{IoU}/(1 + \text{IoU})$). Results are reported for the Network B variant of the Ω-Net; for the results by Isensee et al. (2017) published in STACOM; and for the same group's unpublished `arxiv.org` revision Isensee et al. (2018). Boldface formatting indicates the best performing model for each foreground class.

Comparator Networks

— Objectives:

- Given a pair of sets, specify whether the two sets belong to the same person.
- Each set may contain a variable number of faces, generalisation of conventional image based face verification.

— Conventional Solution:

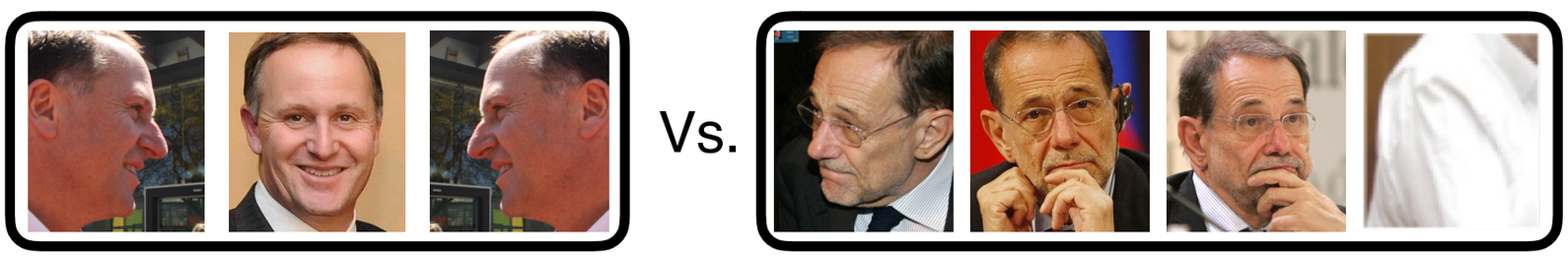
- Networks are trained with standard classifications on single images.
- Image features are aggregated to template features, e.g. by averaging over all image vectors.
- Compute similarity.

$$T_1 = \frac{1}{3}(V_1 + V_2 + V_3)$$

$$T_2 = \frac{1}{2}(V_1 + V_2)$$

Challenges

Consider the template-based verification example:

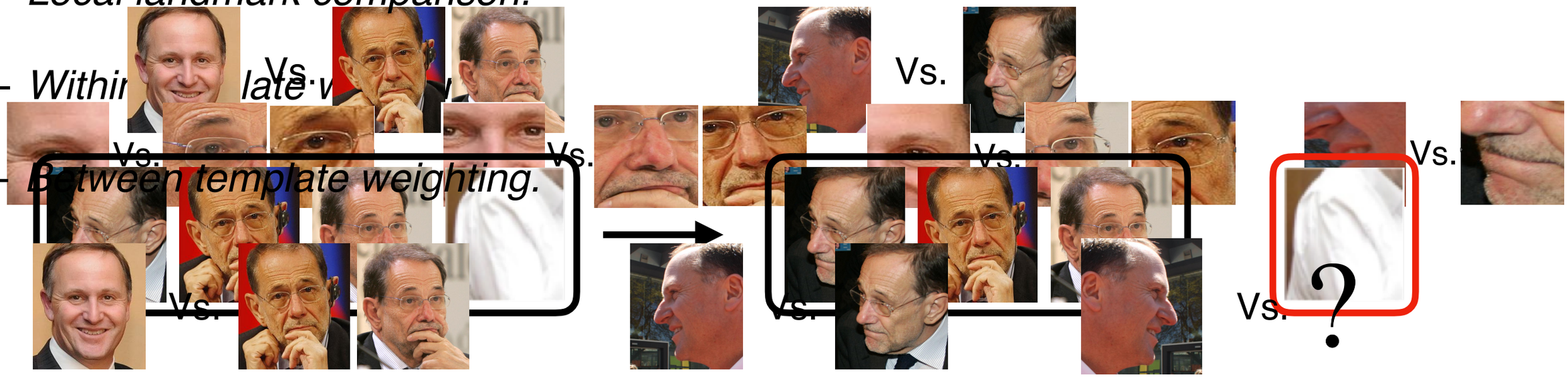


— *Viewpoint conditioned similarity.*

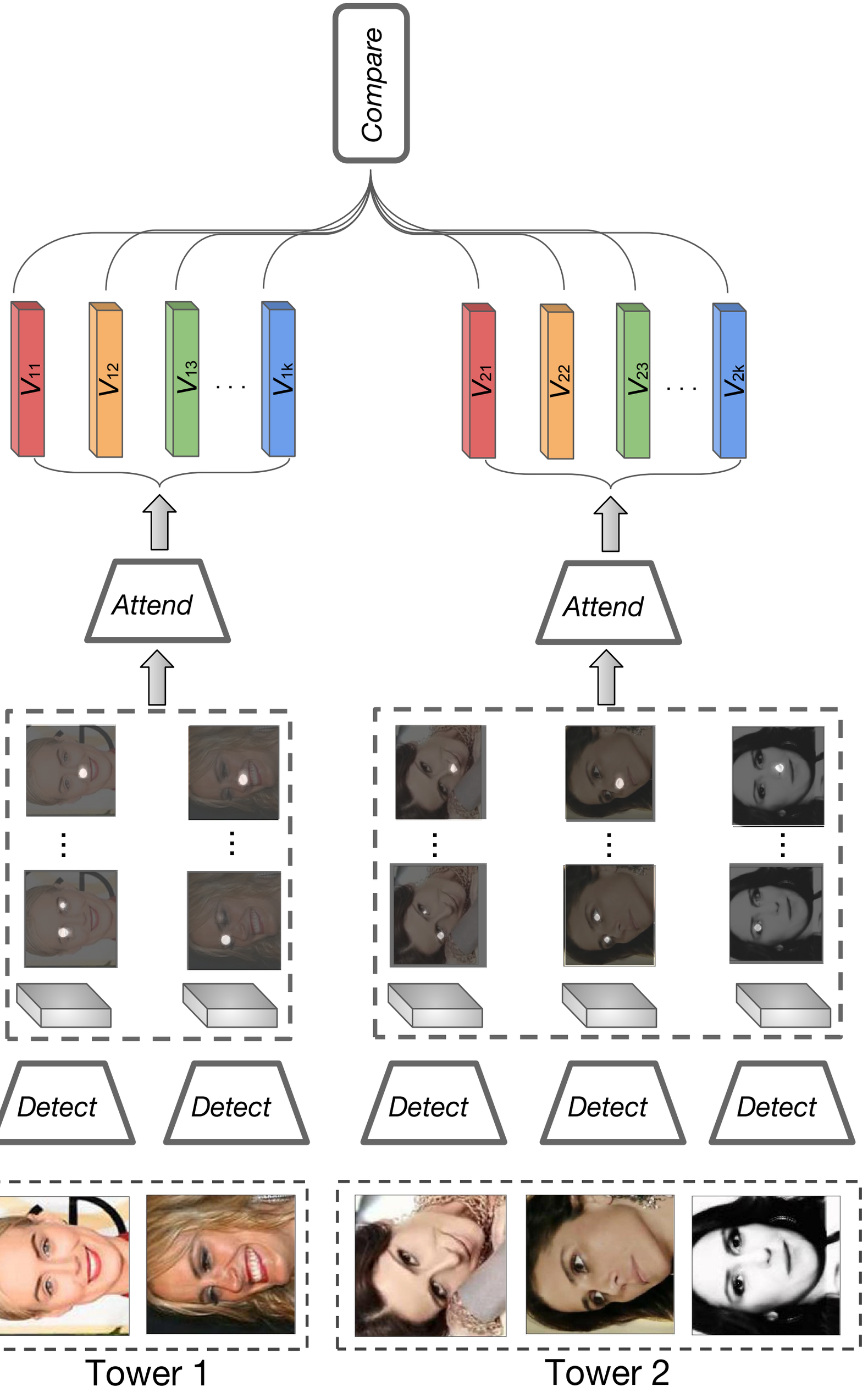
— *Local landmark comparison.*

— *Within template weighting.*

— *Between template weighting.*



Comparator Networks

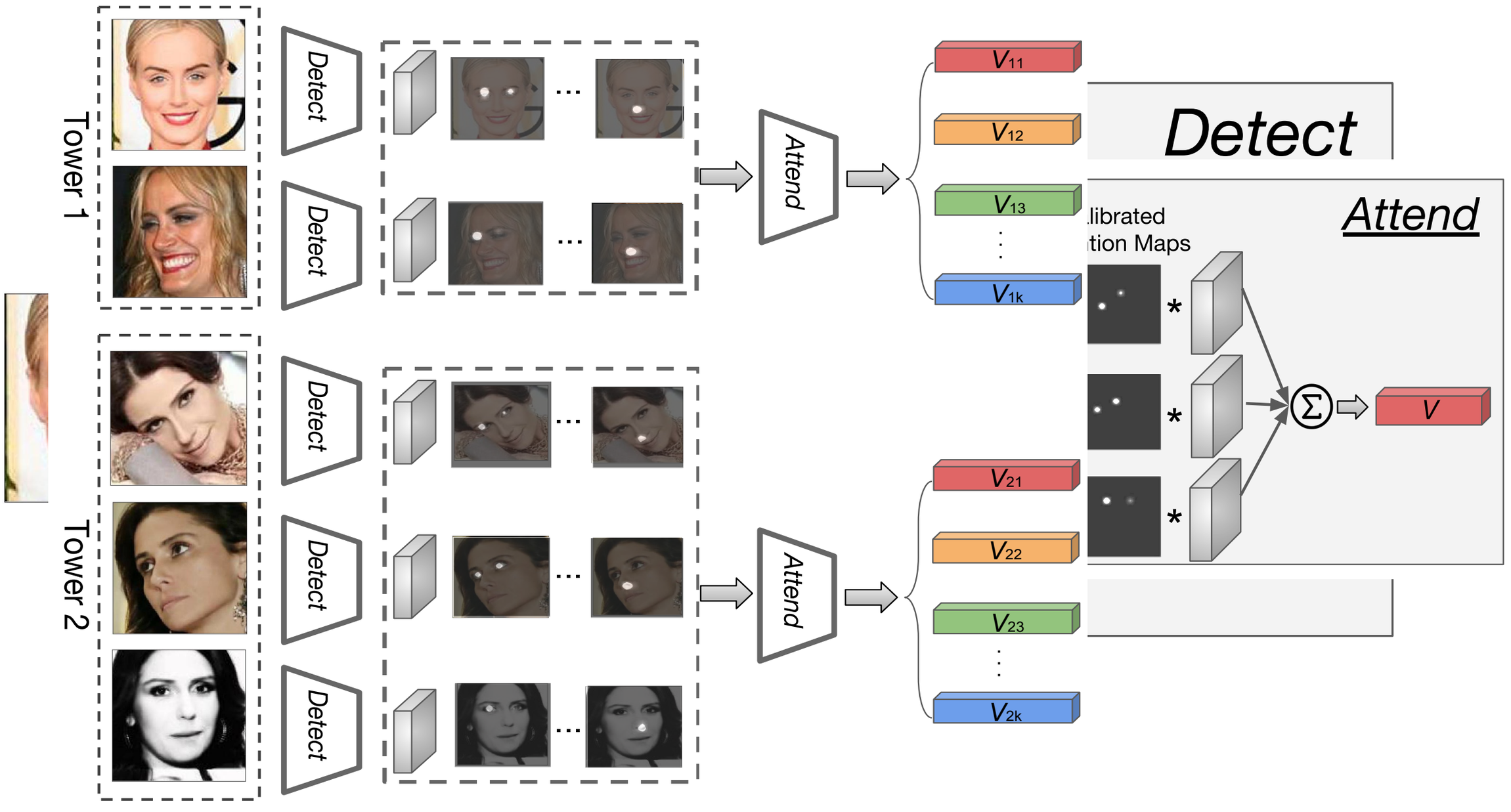


Comparator Networks (*Detect, Attend, Compare*)

— *Viewpoint conditioned similarity.*

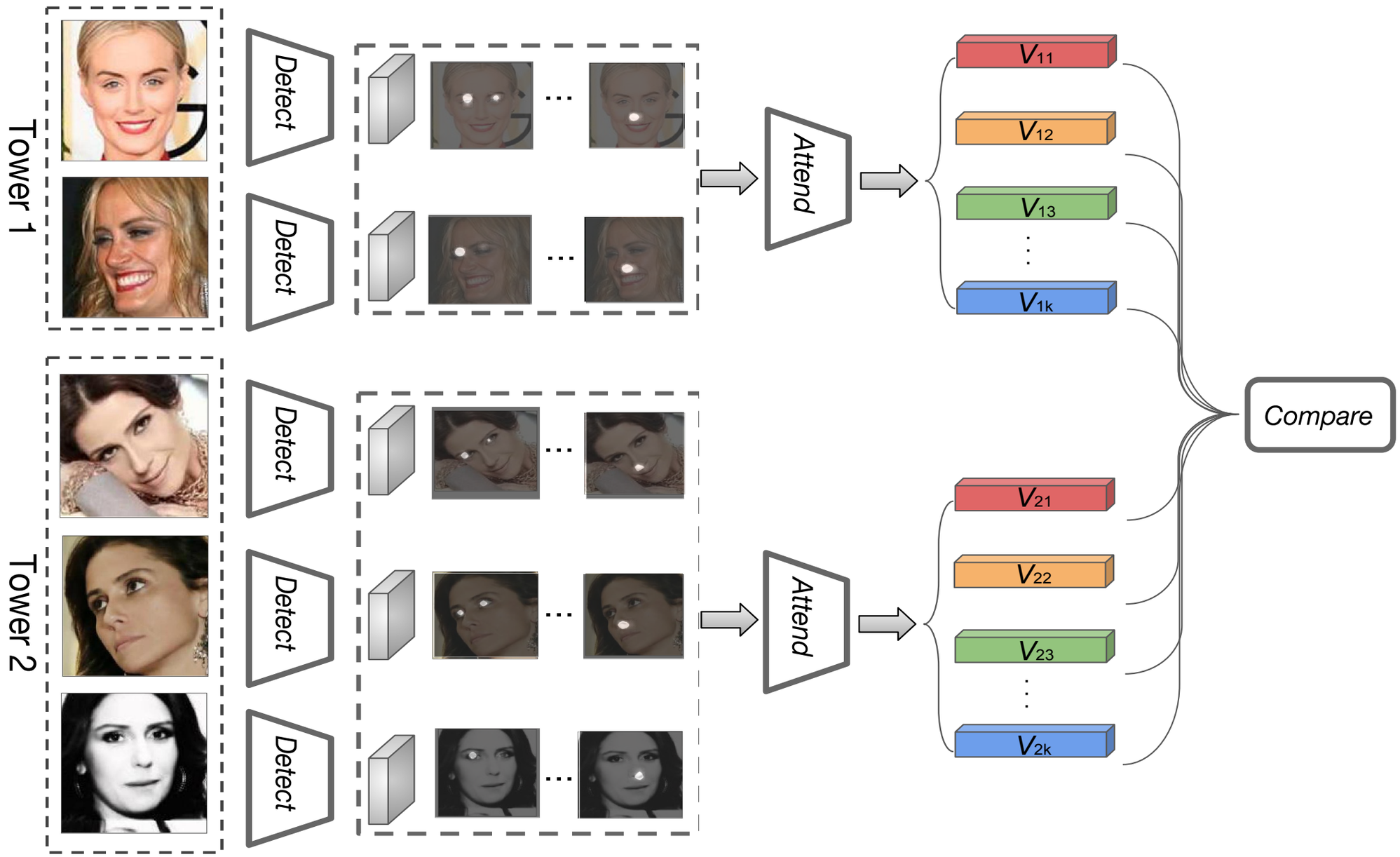
— *Local landmark comparison.*

— *Within template weighting.*



Comparator Networks (*Detect, Attend, Compare*)

- *Viewpoint conditioned similarity.*
- *Local landmark comparison.*
- *Within template weighting.*
- *Between template weighting.*

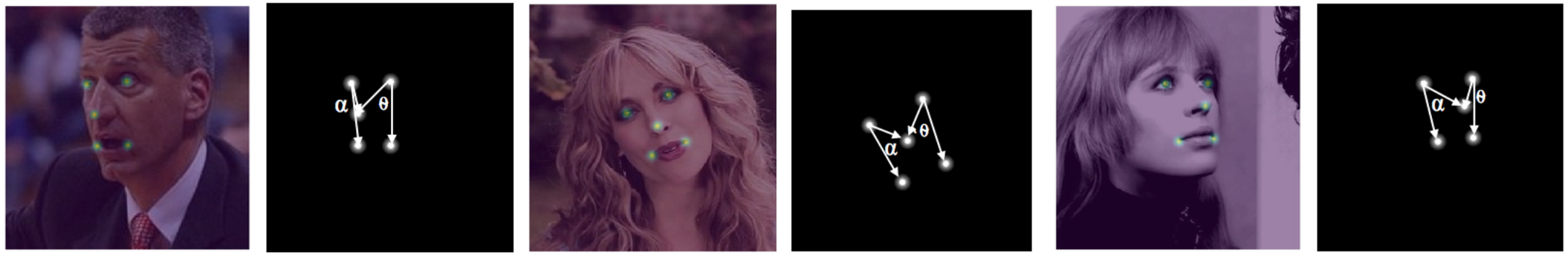


Training Comparator Network

Contributions:

— Learn discriminative, informative landmark detectors.

- **Diversity**: penalize mutual overlap of the landmark maps.
- **Keypoints**: provide extra landmark supervisions, use pseudo keypoints ground-truth from pretrained facial landmark detectors, and estimate the pose based angle ratio.



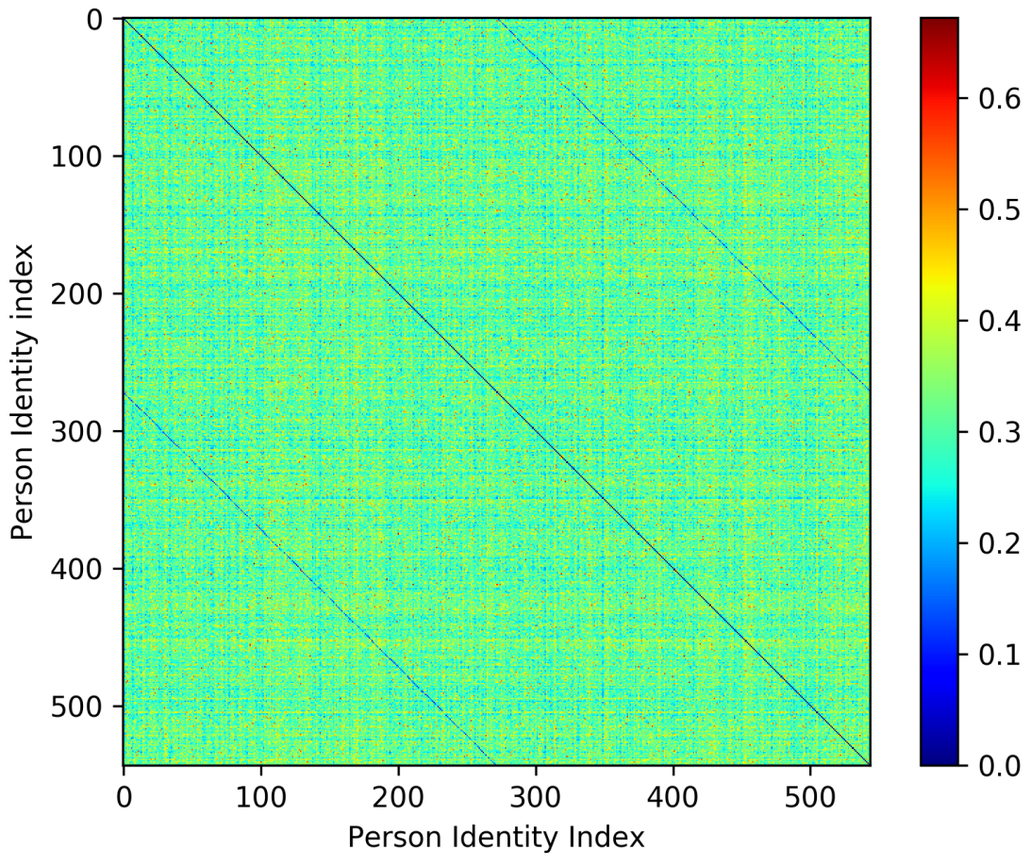
Comparator Networks : Sampling Strategy

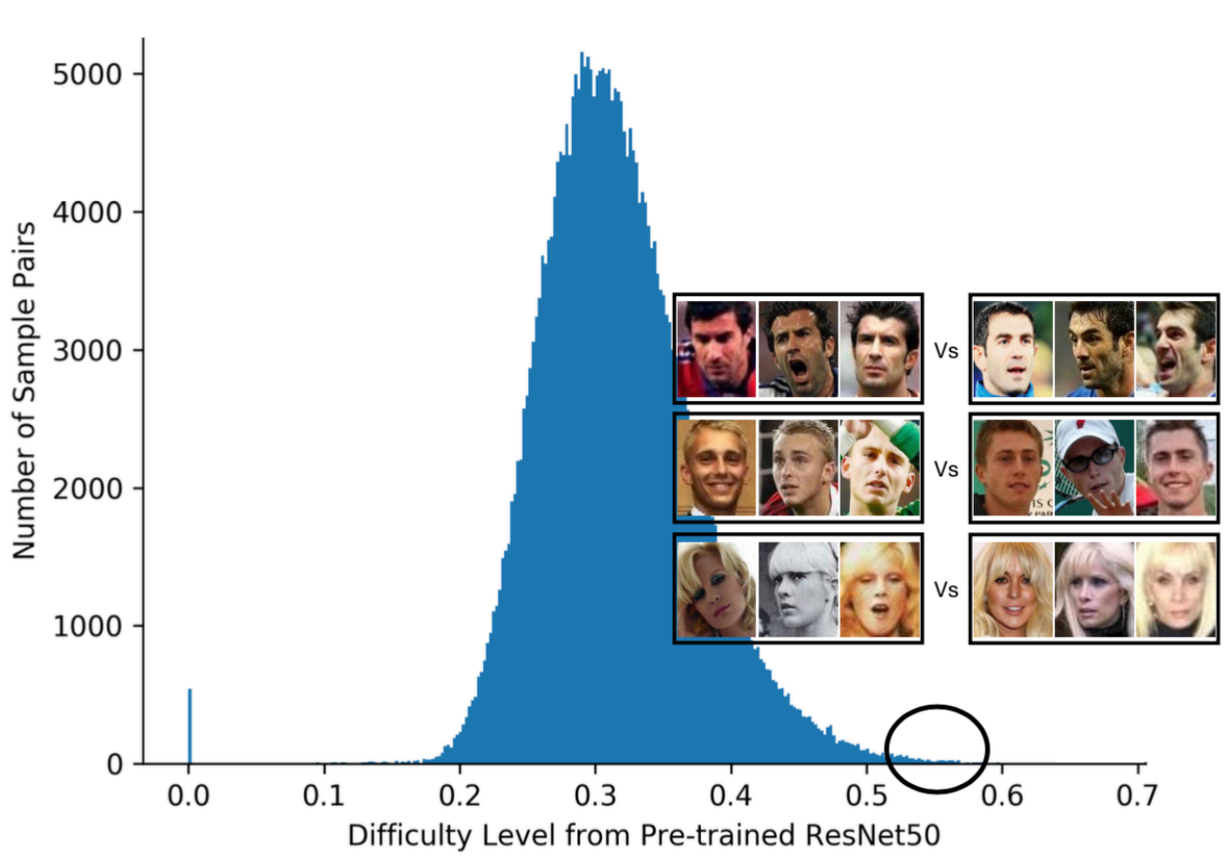
— Take inspiration from instance/image retrieval.

- Train a CNNs with standard image-wise classification. (fast training)
- Use the pre-trained CNNs to approximate template descriptors. (conventional approach)
- Explicitly control the template pair sampling process,

  e.g. sample 256 identity, 2 templates from each identity, construct the difficulty matrix:

$$d = |groundtruth - M_s| \qquad M_s \text{ refers to the verification score from pre-trained CNN}$$





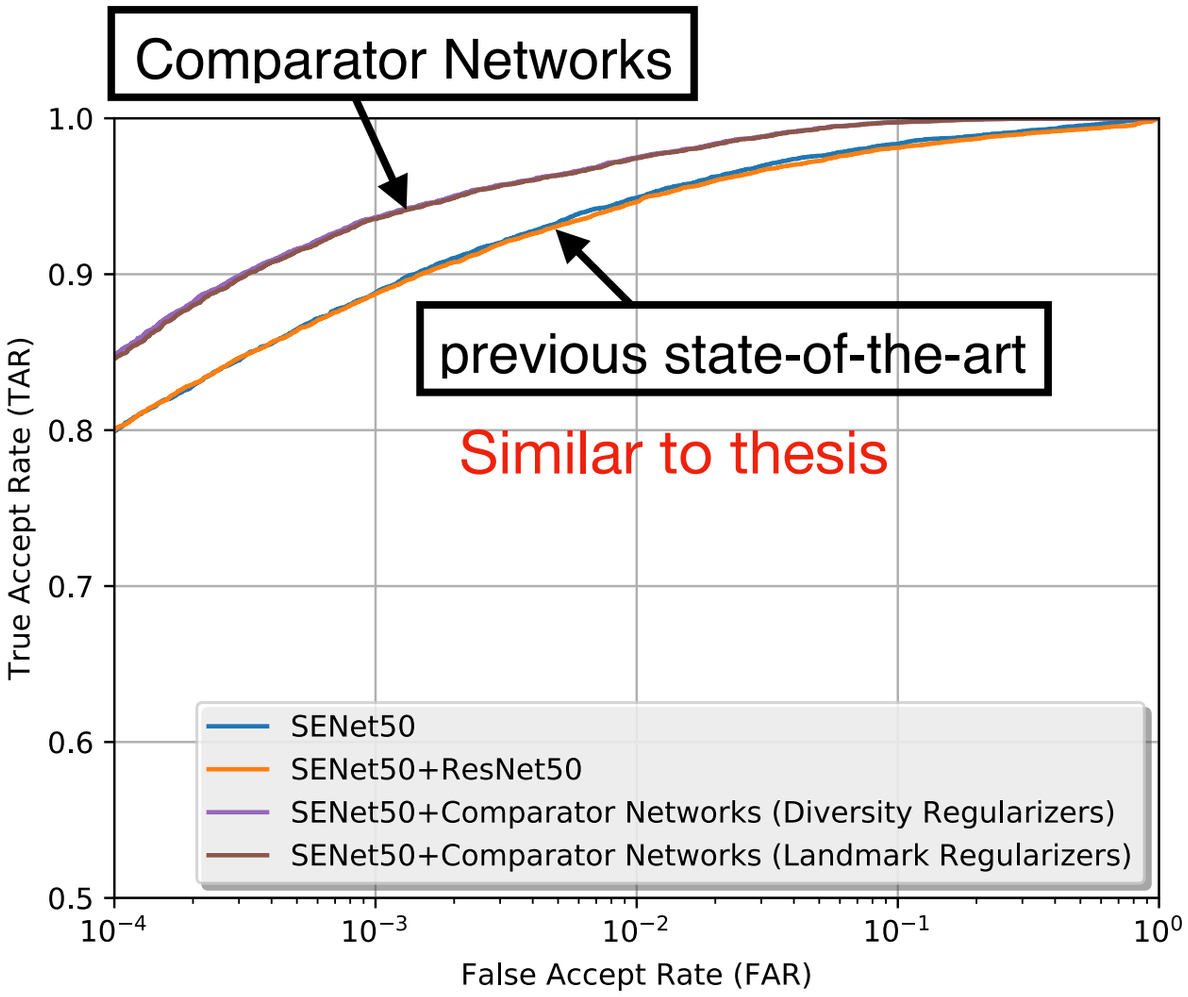
Experiments Results (1:1 Template Verification)

— IARPA Janus IJBB Benchmarks:

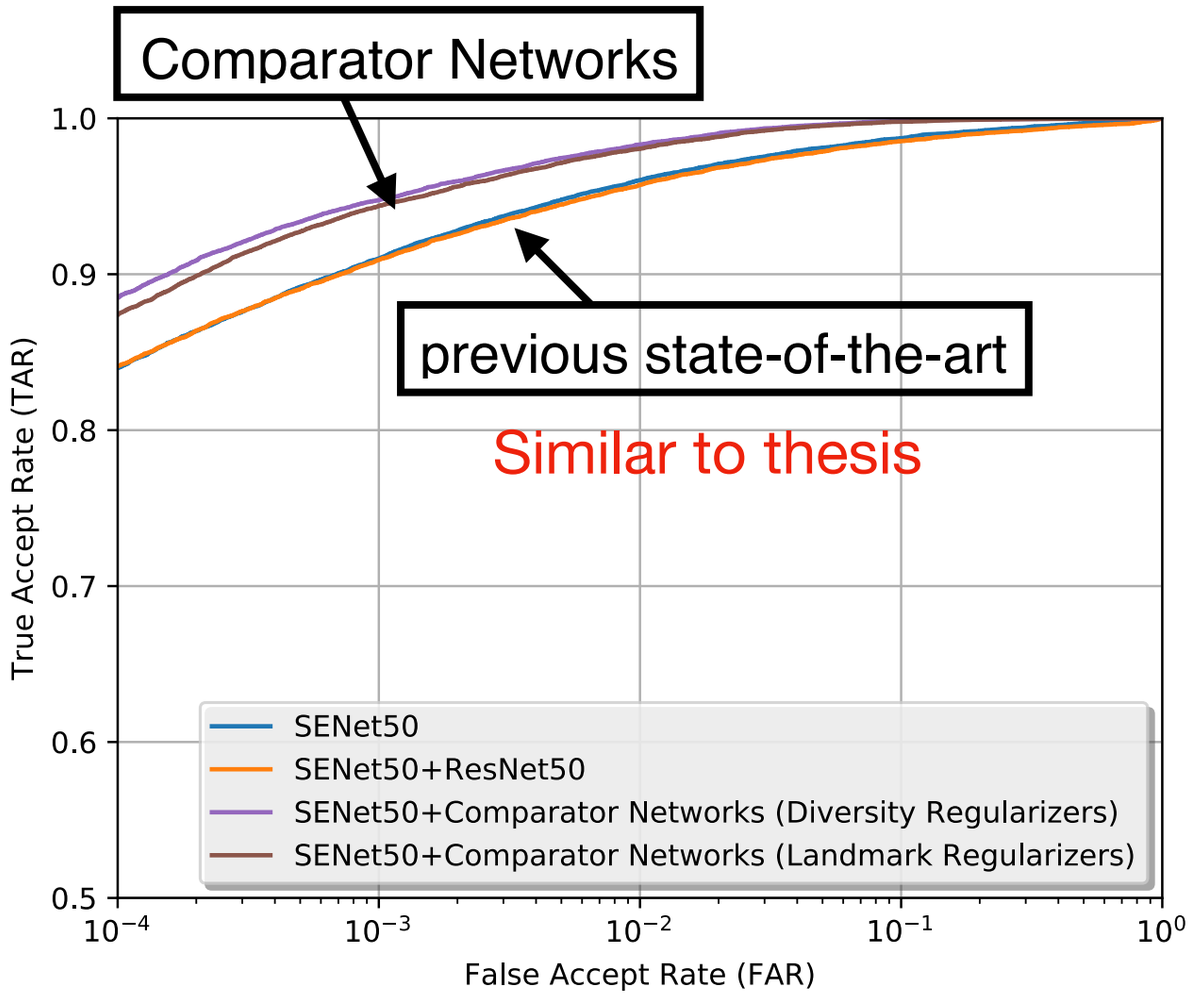
- It contains 1,845 subjects with 21.8K still images and 55K frames from 7,011 videos.

— IARPA Janus IJBC Benchmarks:

- It contains 3,531 subjects with 31.3K still images and 117.7K frames from 11,779 videos. In total, 23124 templates with 19557 genuine matches and 15639K impostor matches.



ROC curve for IJBB (Higher is better)



ROC curve for IJBC (Higher is better)

Visualization (Keypoints)

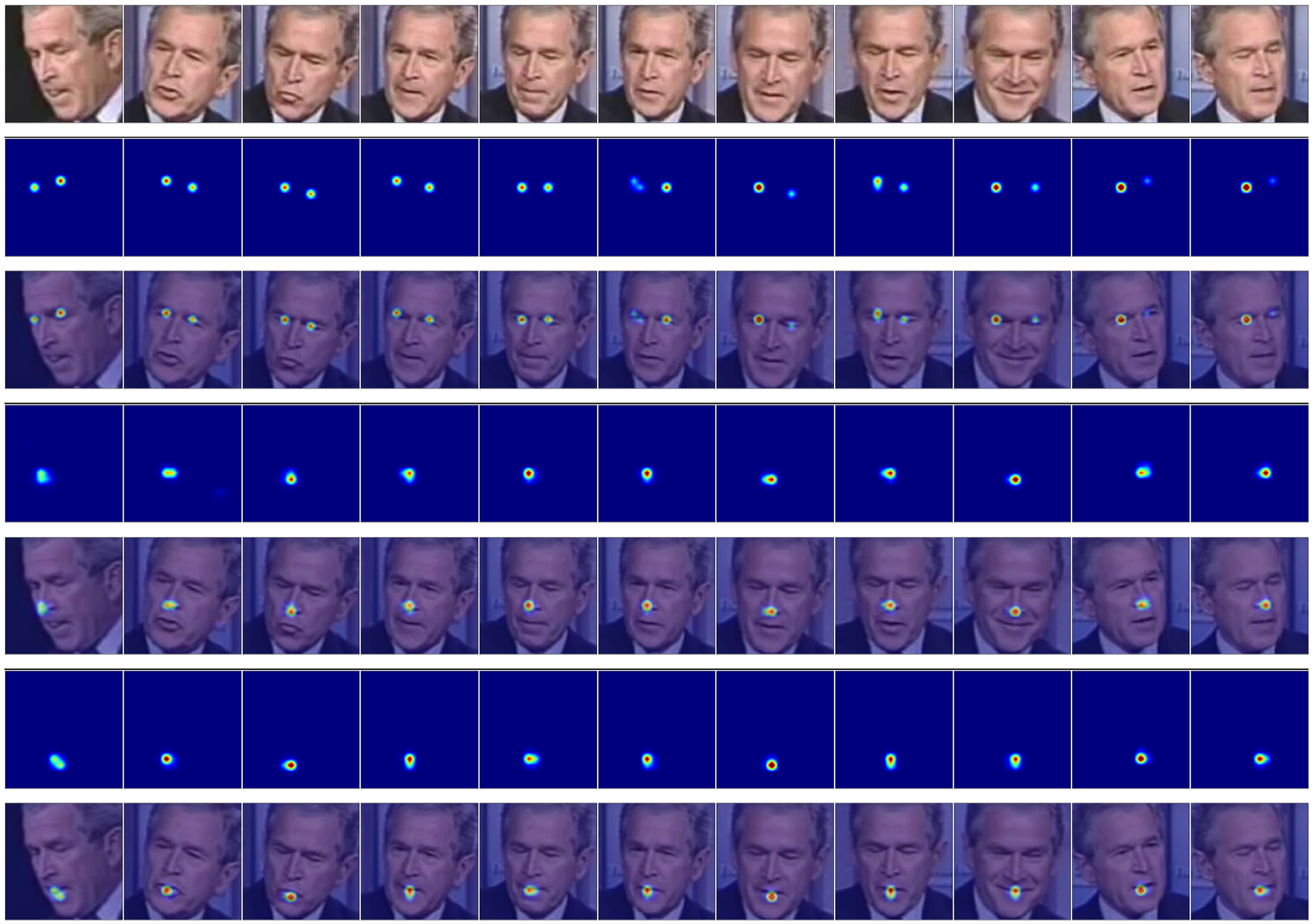


Fig. 1. Predicted facial landmark score maps after self-normalizing for three of the landmark detectors. Additional examples are given in the supplementary material. *1st row*: raw images in the template, faces in a variaty of poses are shown from left to right; *2nd,4th,6th row*: self-normalized landmark score maps (attention maps); *3rd, 5th, 7th row*: images overlayed with the attention maps.

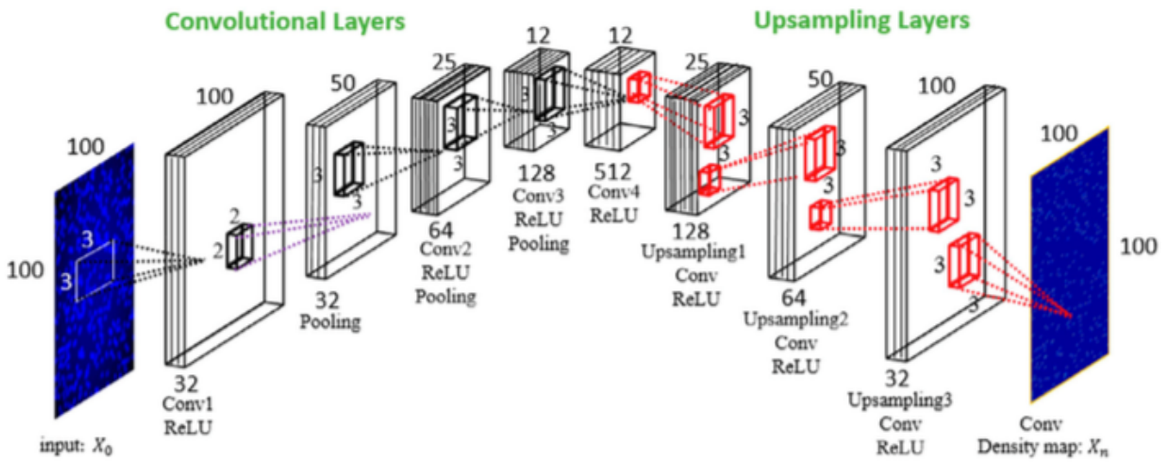
Visualization (Diversity)

Thank You For Your Attention !

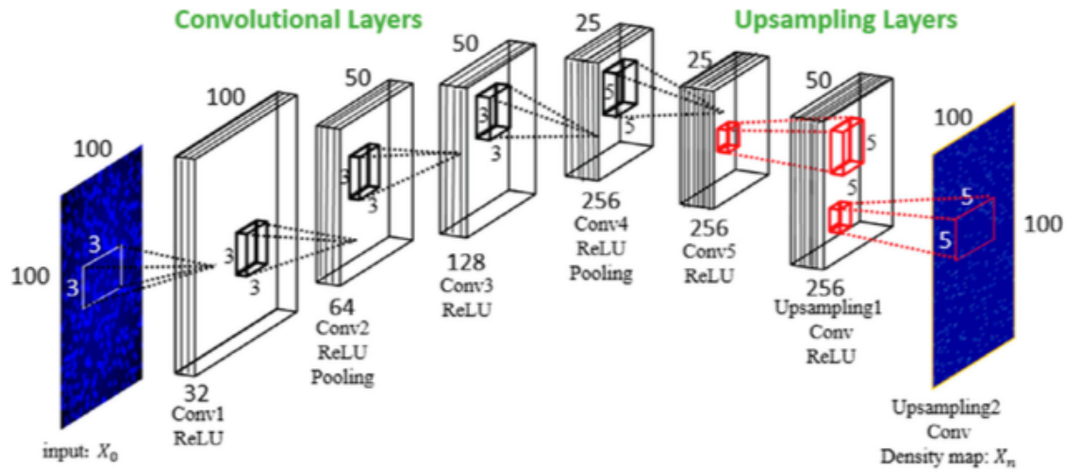
Appendix

Appendix: Cell Counting

— Architectures



Fully Convolutional Regression Networks A (**FCRN-A**)



Fully Convolutional Regression Networks B (**FCRN-B**)

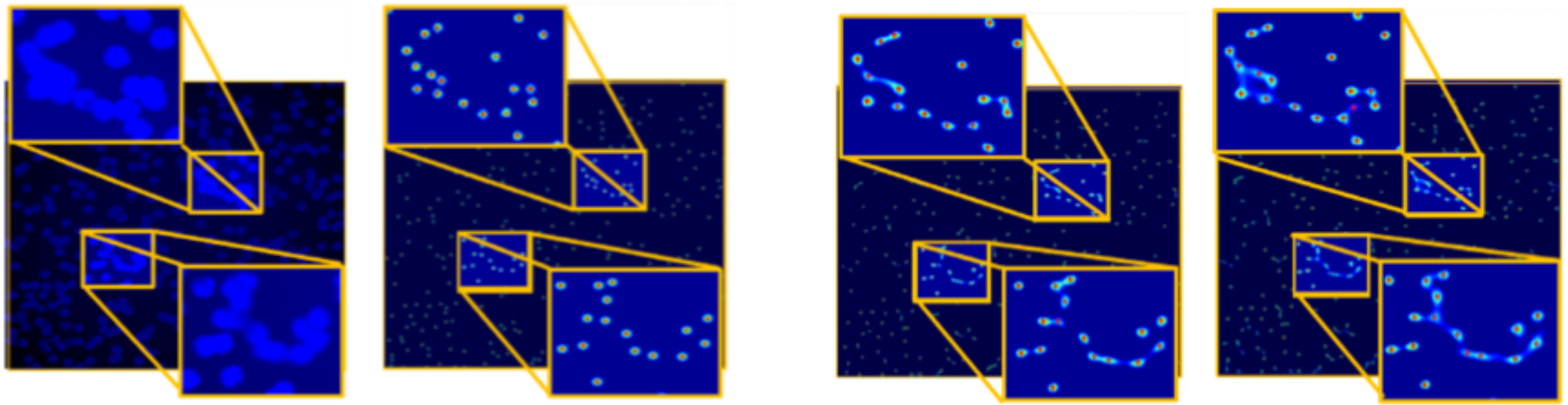
— Results

Method	174 ± 64 cells			
	$N = 8$	$N = 16$	$N = 32$	$N = 64$
Lempitsky and Zisserman (2010)	8.8 ± 1.5	6.4 ± 0.7	5.9 ± 0.5	N/A
Lempitsky and Zisserman (2010)	4.9 ± 0.7	3.8 ± 0.2	3.5 ± 0.2	N/A
Fiaschi et al. (2012)	3.4 ± 0.1	N/A	3.2 ± 0.1	N/A
Arteta et al. (2014)	4.5 ± 0.6	3.8 ± 0.3	3.5 ± 0.1	N/A
Proposed FCRN-A	3.9 ± 0.5	3.4 ± 0.2	2.9 ± 0.2	2.9 ± 0.2
Proposed FCRN-B	4.1 ± 0.5	3.7± 0.3	3.3 ± 0.2	3.2 ± 0.2

The columns correspond to the number of training images. Standard deviation corresponds to five different draws of training and validation sets

Appendix: Cell Counting

— Results

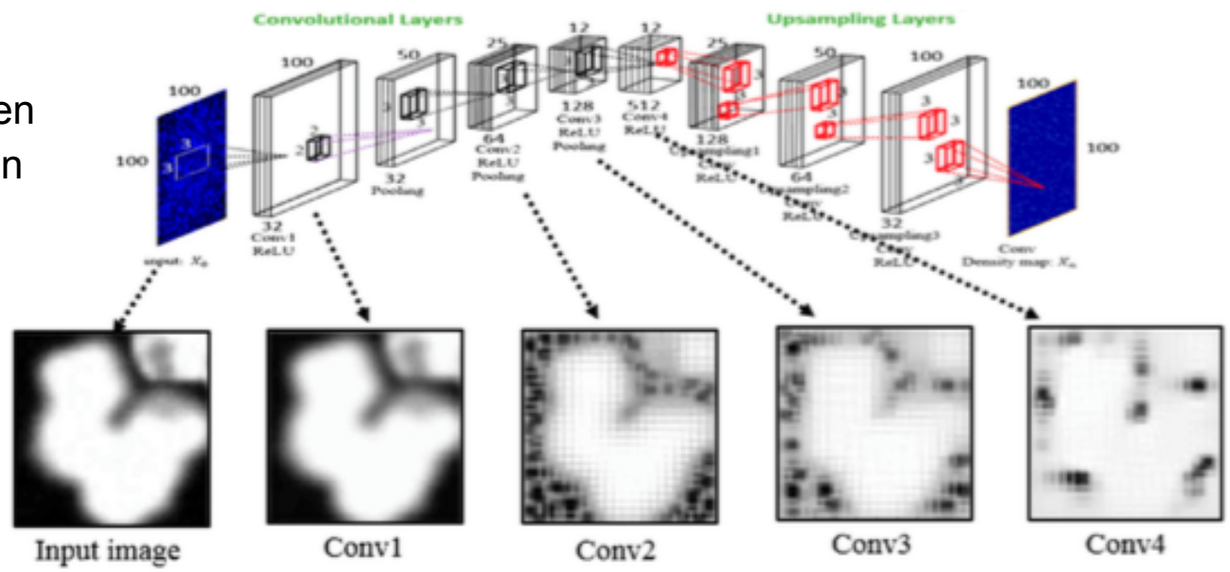


(a) Synthetic Image: $I(x)$ (b) Ground-truth Density Map (c) Density Map by FCRN-A (d) Density Map by FCRN-B

— Visualisation

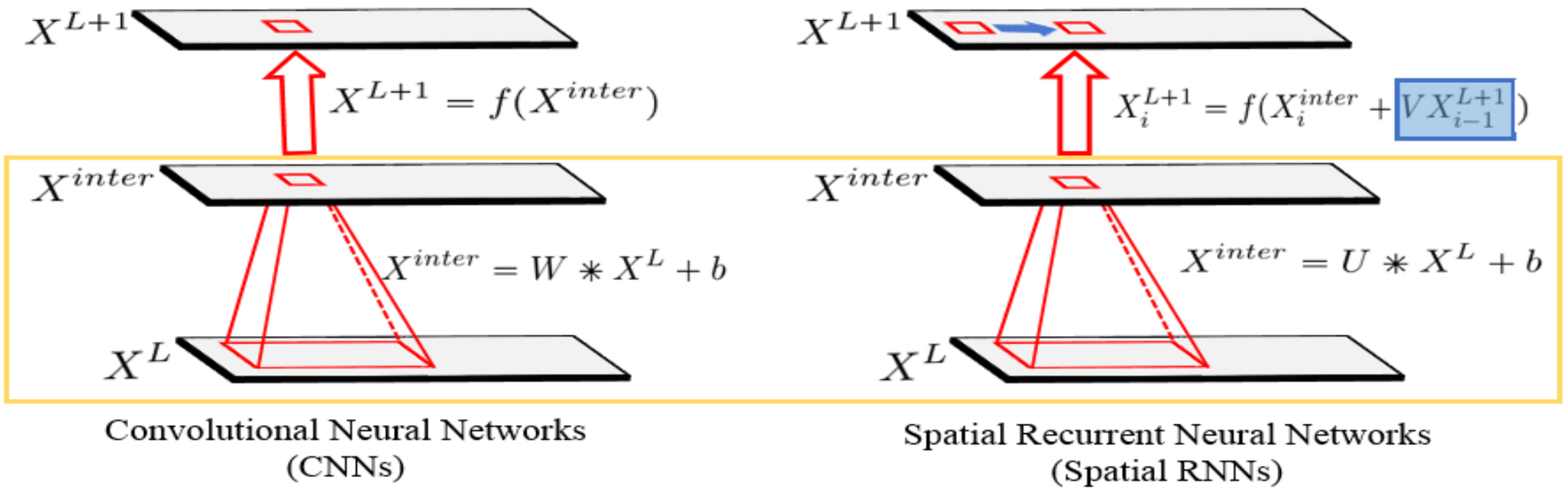
The problem can be formalised as a reconstruction problem. Given a representation function $F : \mathrm{R}^{H\times W\times C} \rightarrow \mathrm{R}^{d}$ and a representation for image x as $\varphi = \varphi(x)$, the reconstruction process aims to find another image $x' \in \mathrm{R}^{H\times W\times C}$ that minimises the objective:

$$L(\phi(x), \phi(x')) = \|\phi(x) - \phi(x')\|^2$$



Appendix: Layer Recurrent Neural Networks

— CNN vs RNN



Convolutional Neural Networks (CNNs)

Spatial Recurrent Neural Networks (Spatial RNNs)

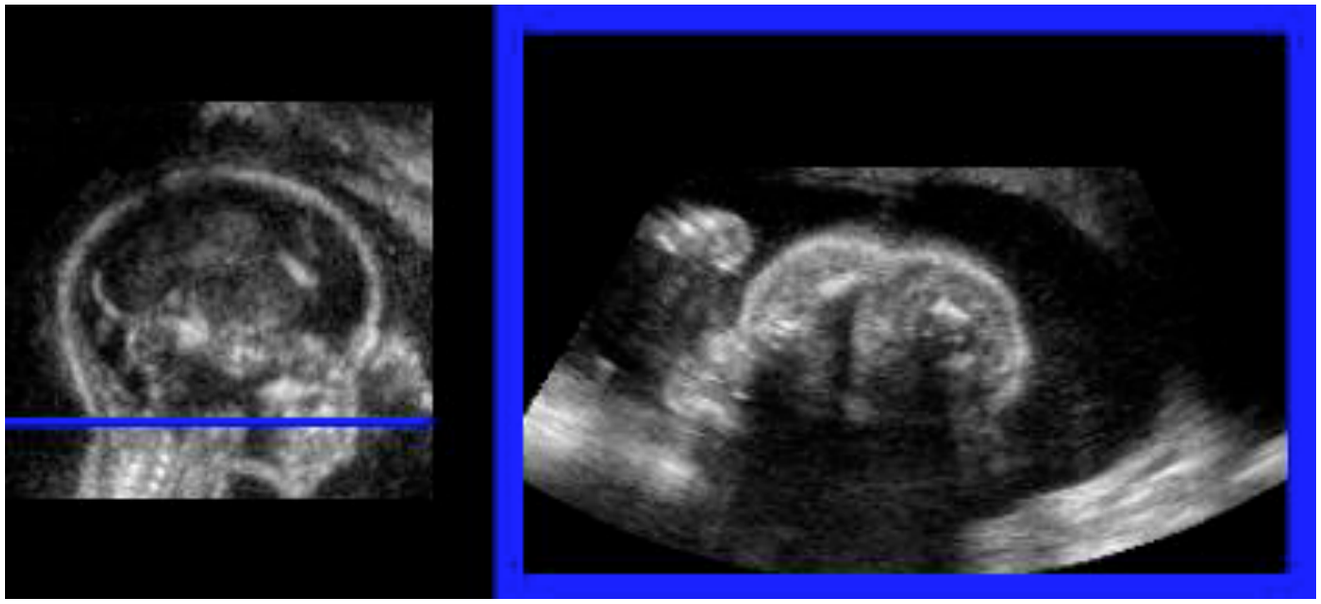
$$X^{inter} = U * X^L + b \quad \text{(Convolution)}$$

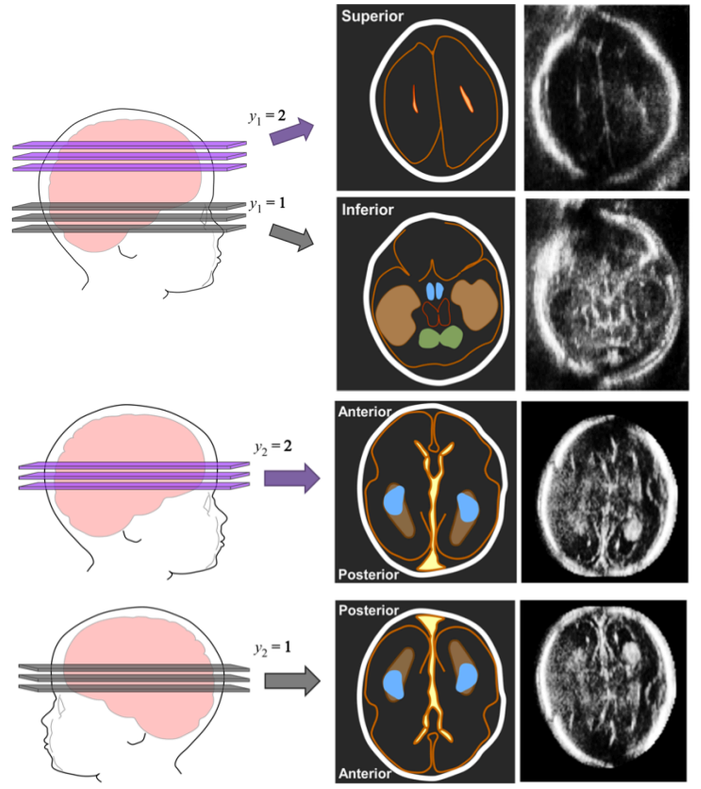
$$X_i^{L+1} = f(X_i^{inter}) \quad \text{(i = 1, zero initial states)}$$

$$X_i^{L+1} = f(X_i^{inter} + VX_{i-1}^{L+1}) \quad (i > 1)$$

Appendix: 3D Fetal Neurosonography

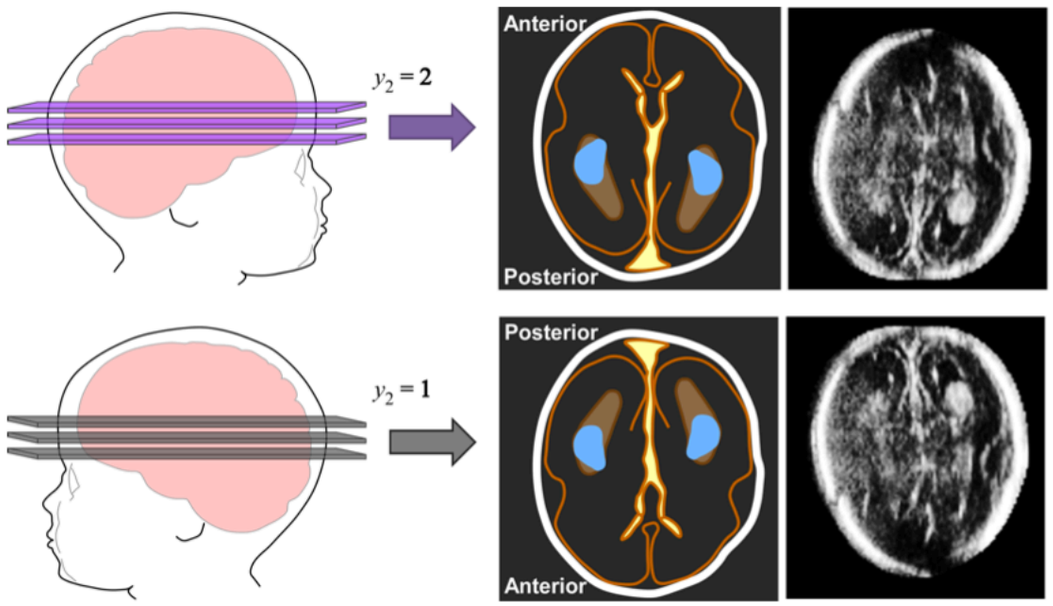
— Observation

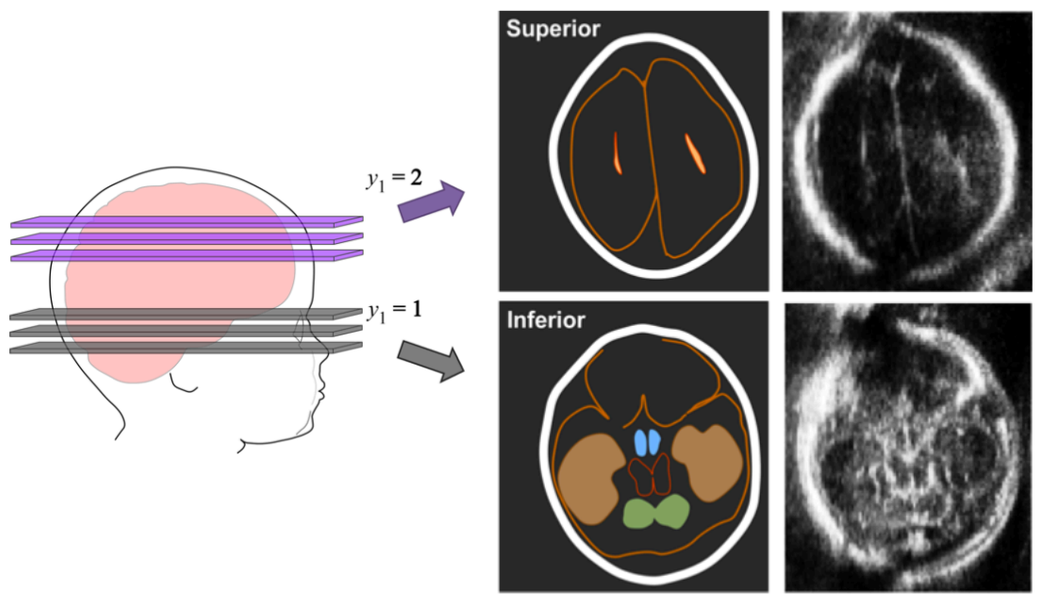


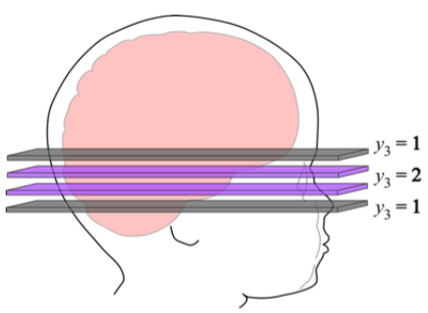


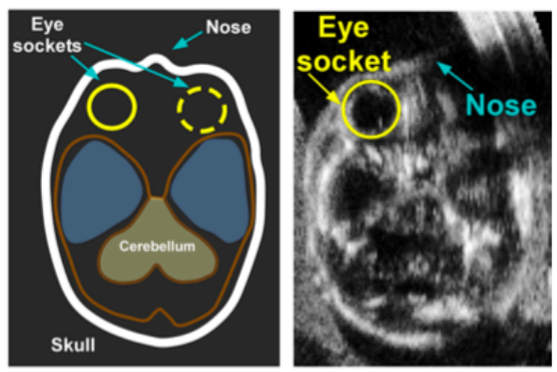
Appendix: 3D Fetal Neurosonography

— Slice annotation:

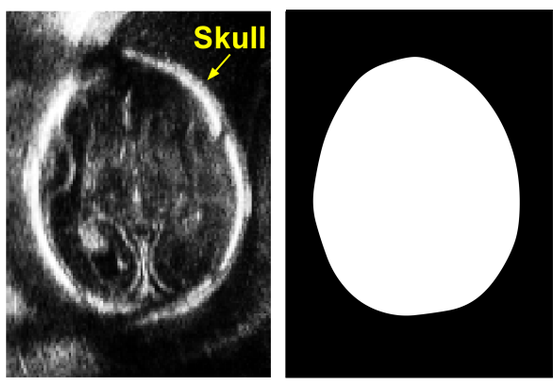




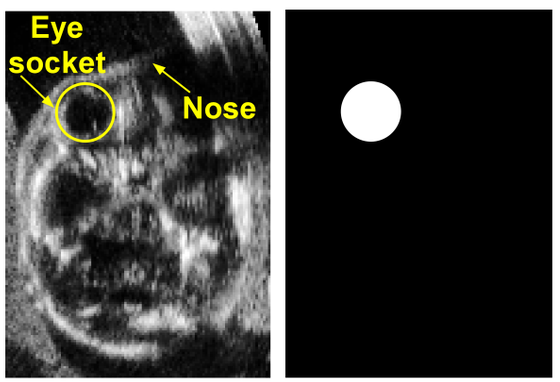




$y_3 = 2$



$\mathbf{Y}_5$

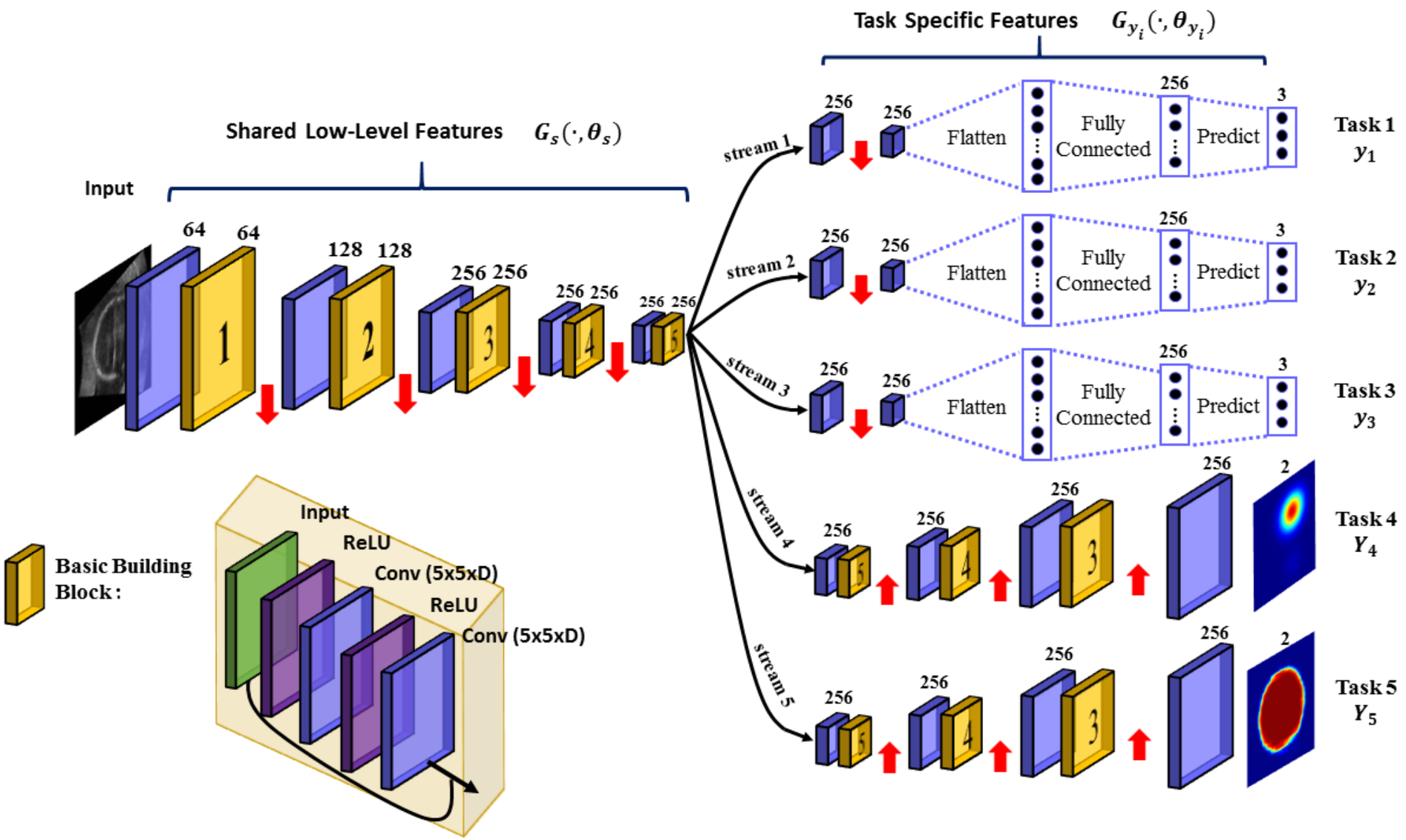


$(y_3 = 2)$

$\mathbf{Y}_4$

Appendix: 3D Fetal Neurosonography

— Models



Appendix: 3D Fetal Neurosonography

— Results

Network		Kernel Size	Classification (3-way)			Classification (2-way)			No. Params
			y_1	y_2	y_3	y_1	y_2	y_3	
3-pool	A	3 × 3	90.4 ± 1.0	90.2 ± 1.0	87.9 ± 0.6	94.7 ± 0.5	63.8 ± 5.3	83.2 ± 0.4	13.7 M
	B	5 × 5	90.6 ± 1.0	90.7 ± 0.9	89.5 ± 0.8	94.9 ± 0.7	66.7 ± 3.3	85.4 ± 1.3	25.1 M
	C	7 × 7	91.4 ± 1.4	91.3 ± 1.3	89.1 ± 1.1	95.0 ± 1.0	68.6 ± 5.3	84.6 ± 2.1	40.0 M
4-pool	D	3 × 3	91.0 ± 0.5	90.8 ± 0.9	88.5 ± 1.0	94.4 ± 0.5	68.9 ± 2.2	83.5 ± 1.6	11.9 M
	E	**5 × 5**	**92.1 ± 0.7**	**91.9 ± 0.7**	**90.3 ± 0.6**	**95.8 ± 0.4**	**70.8 ± 2.0**	**86.4 ± 1.5**	**29.6 M**
	F	7 × 7	91.3 ± 1.1	91.3 ± 0.7	89.7 ± 1.2	95.5 ± 1.1	66.3 ± 2.2	85.7 ± 2.2	55.4 M

Table 1: **Slice classification accuracy**
Mean accuracy (± standard deviation) of the slice-wise classification computed over the five-fold cross-validation sets. Axial slice label prediction accuracy on 2D images for tasks y_1, y_2, y_3. The accuracy for 3-way (where $y_k \in \{-1, 0, +1\}$) and 2-way (excluding $y_k = 0$) classification is reported for all six network architectures. Network E (in bold) outperformed the others in all classification tasks.

Network		Kernel Size	Segmentation		Localization (mm)		No. Params
			$\mathbf{Y}_4$ (eye)	$\mathbf{Y}_5$ (skull)	d_4 (eye)	d_5 (skull)	
3-pool	A	3 × 3	0.47 ± 0.13	0.71 ± 0.22	2.50 ± 2.96	1.82 ± 1.71	13.7 M
	B	5 × 5	0.53 ± 0.13	0.79 ± 0.21	2.11 ± 2.01	1.50 ± 1.68	25.1 M
	C	7 × 7	0.52 ± 0.14	0.80 ± 0.20	2.49 ± 3.02	1.49 ± 1.64	40.0 M
4-pool	D	3 × 3	0.52 ± 0.15	0.79 ± 0.20	2.49 ± 2.95	1.50 ± 1.68	11.9 M
	E	**5 × 5**	**0.55 ± 0.12**	**0.83 ± 0.18**	**1.92 ± 1.36**	**1.19 ± 1.26**	**29.6 M**
	F	7 × 7	0.55 ± 0.13	0.82 ± 0.19	2.14 ± 2.17	1.24 ± 1.25	55.4 M

Table 2: **Slice segmentation and object localization accuracy**
Mean accuracy (± standard deviation) of the slice-wise segmentation and object localization computed over the five-fold cross-validation sets. Jaccard index ($J_k = 0.0$: no overlap; $J_k = 1.0$: perfect overlap) and centerpoint distance (d_k in mm, between GT and predicted segmentations) shown for the eye and skull segmentation tasks, $\mathbf{Y}_4$ and $\mathbf{Y}_5$, respectively.

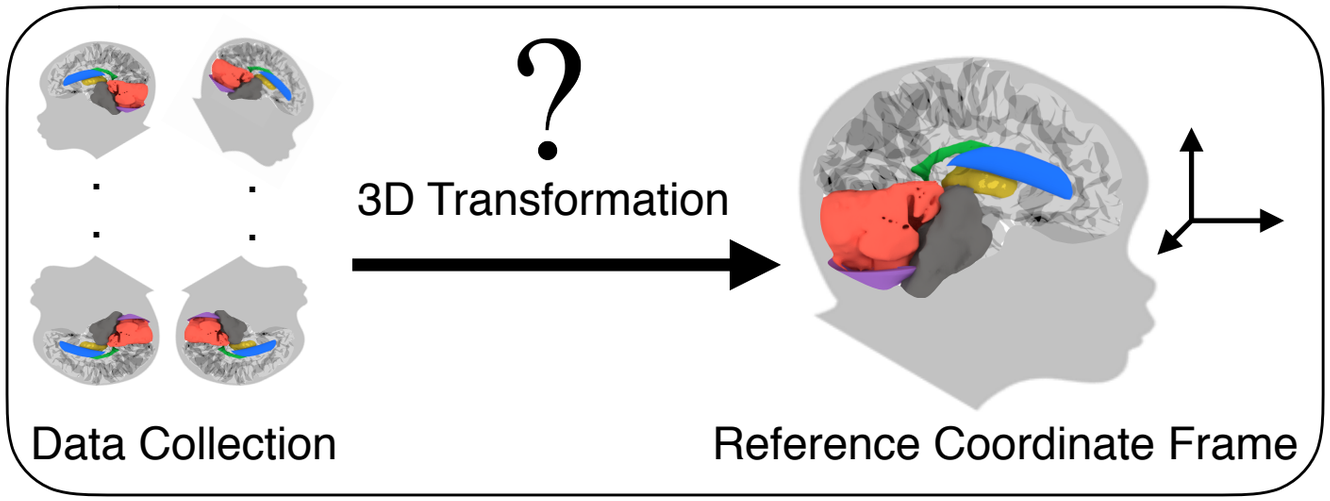
3D Fetal Neurosonography

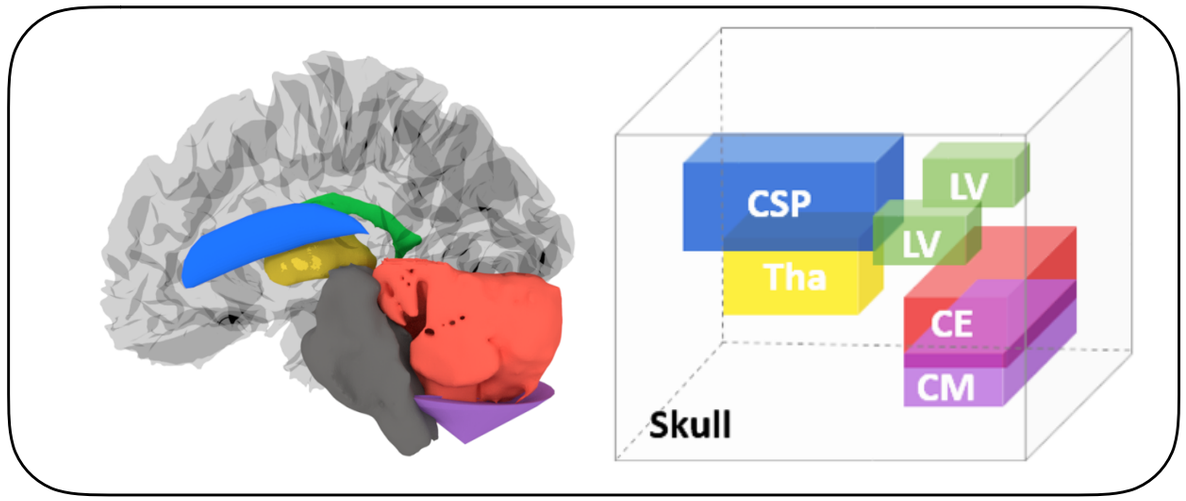
— Objectives:

- Monitoring the fetal brain growth, i.e. important structures.

— Approach:

- Transform the 3D ultrasound volumes to a reference coordinate frame.
- Localize the important structures, e.g. lateral ventricles (green), thalami (yellow), cerebellum (red), etc. (not covered by the thesis)





CSP: Cavum Septi Pellucidi, **Tha**: Thalami,

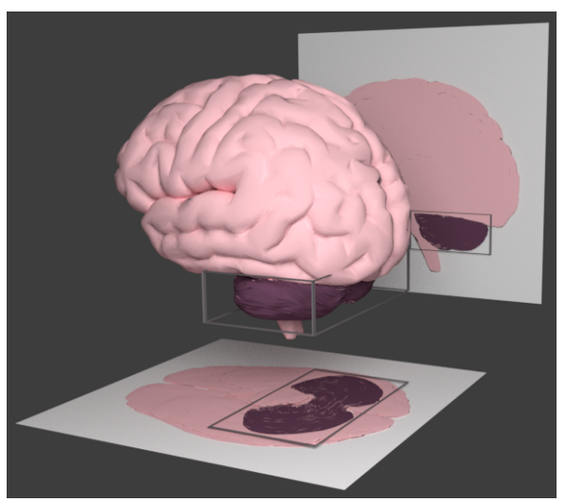
LV: Lateral Ventricles, **CE**: Cerebellum,

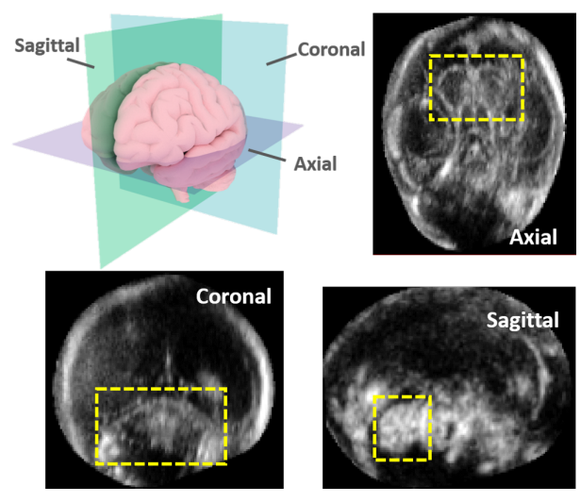
CM: Cisterna Magna

3D Fetal Neurosonography-2

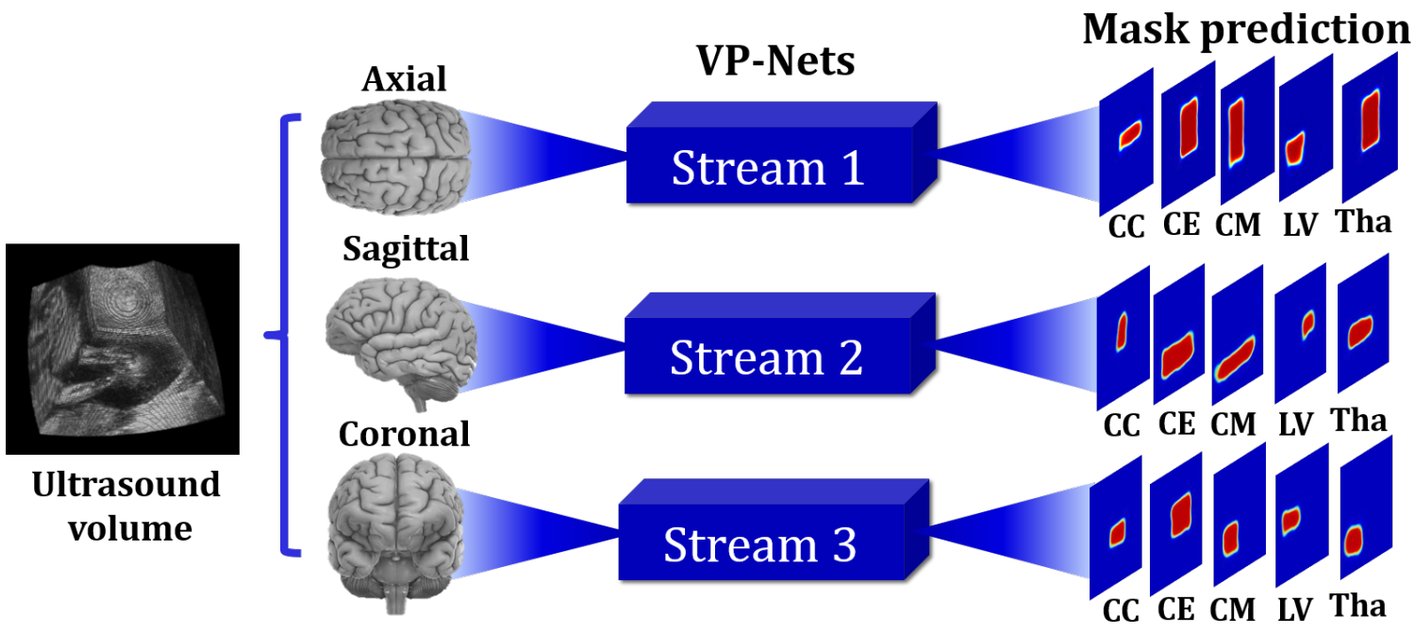
— Proposed Solution

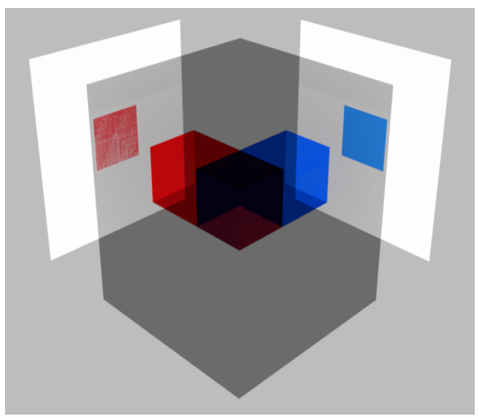
- Observations: learning from silhouettes





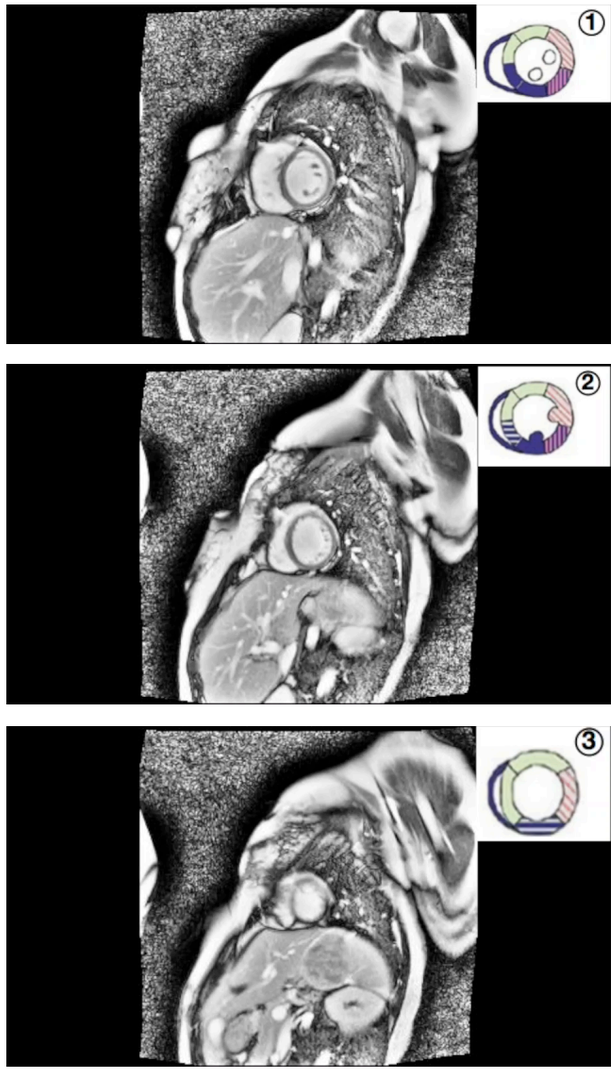
- Architectures

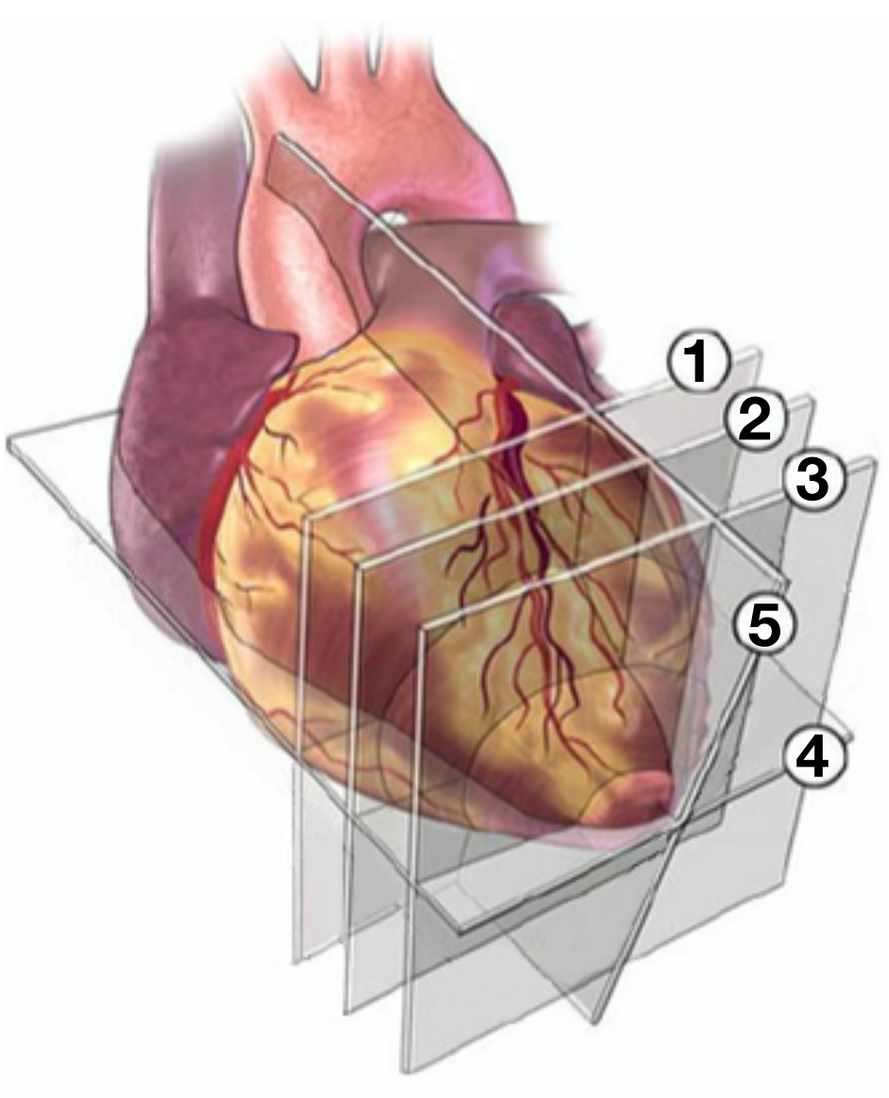


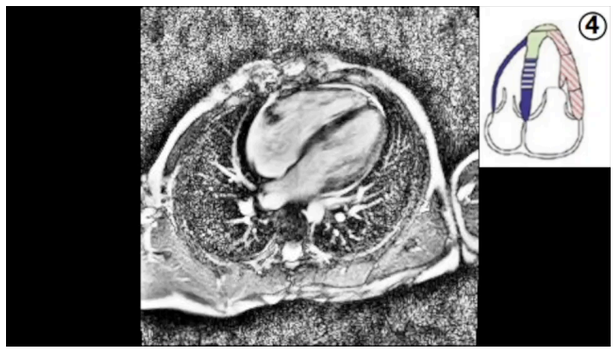


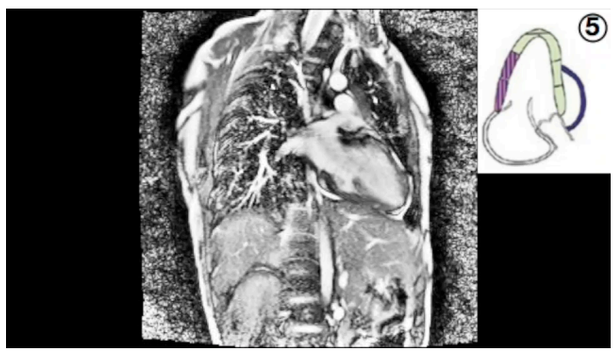
Appendix: Cardiac MRI

— Data Acquisition





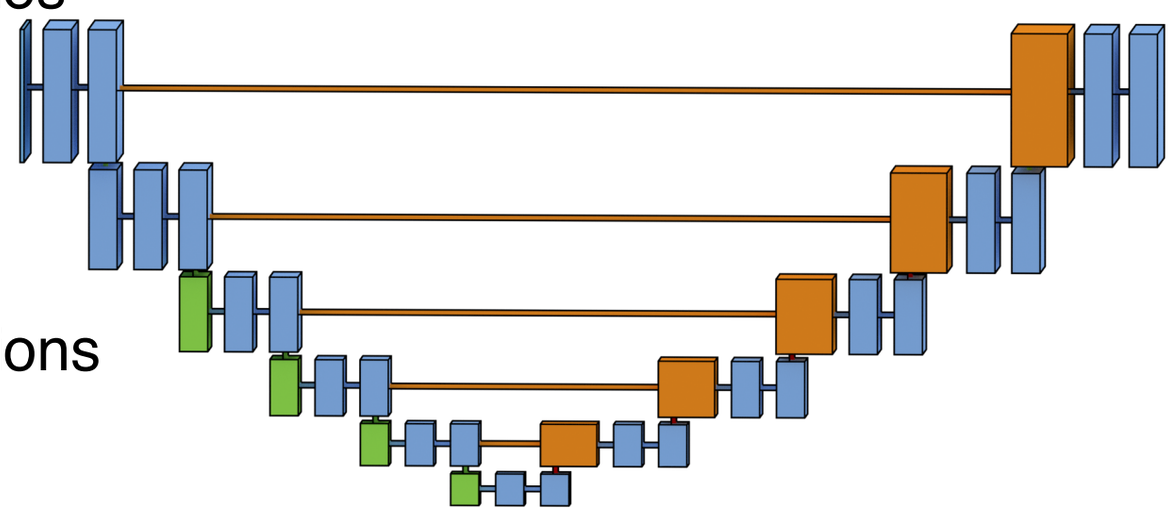




Appendix: Cardiac MRI

— Modules

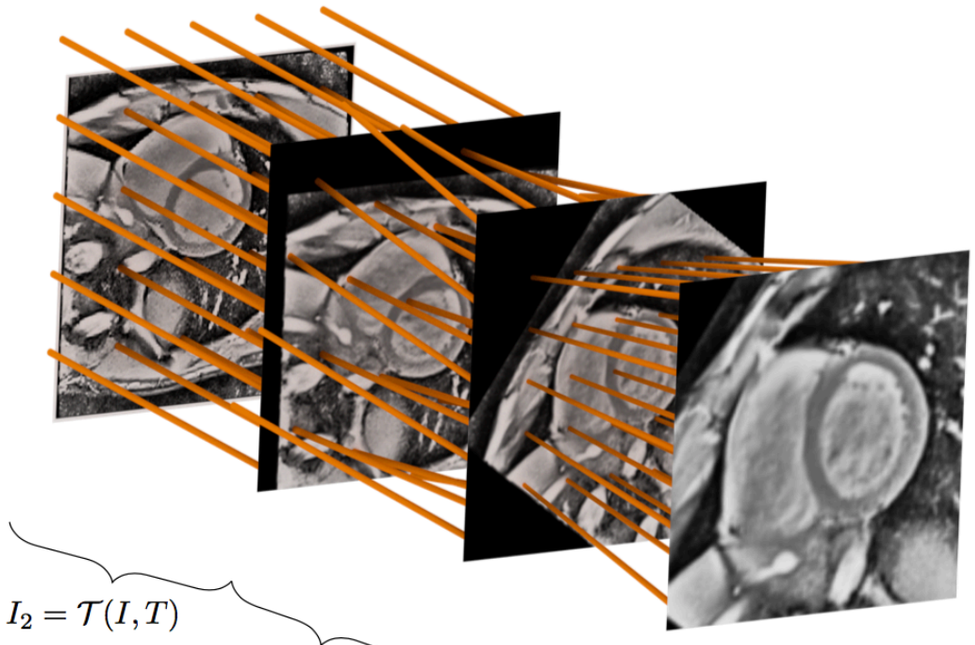
(a) Coarse & Fine-grained Segmentation Modules



- Max-pooling layers after every two convolutional layers.
- Concatenation is used to fuse representations of different levels.

convolution max pooling concatenation

(b) Spatial Transformer Module

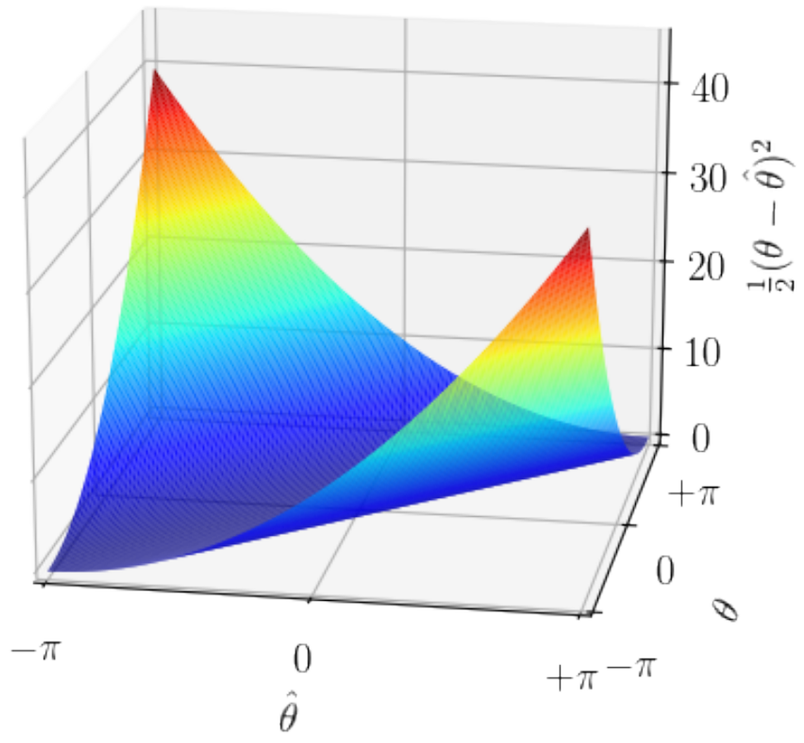


$$I'_{h',w',c} = \sum_{h=1}^{H} \sum_{w=1}^{W} I_{h,w,c} \cdot \max\left(0, 1 - |\alpha_v G'_{1,h',w'} + \beta_v - h|\right) \cdot \max\left(0, 1 - |\alpha_u G'_{2,h',w'} + \beta_u - w|\right)$$

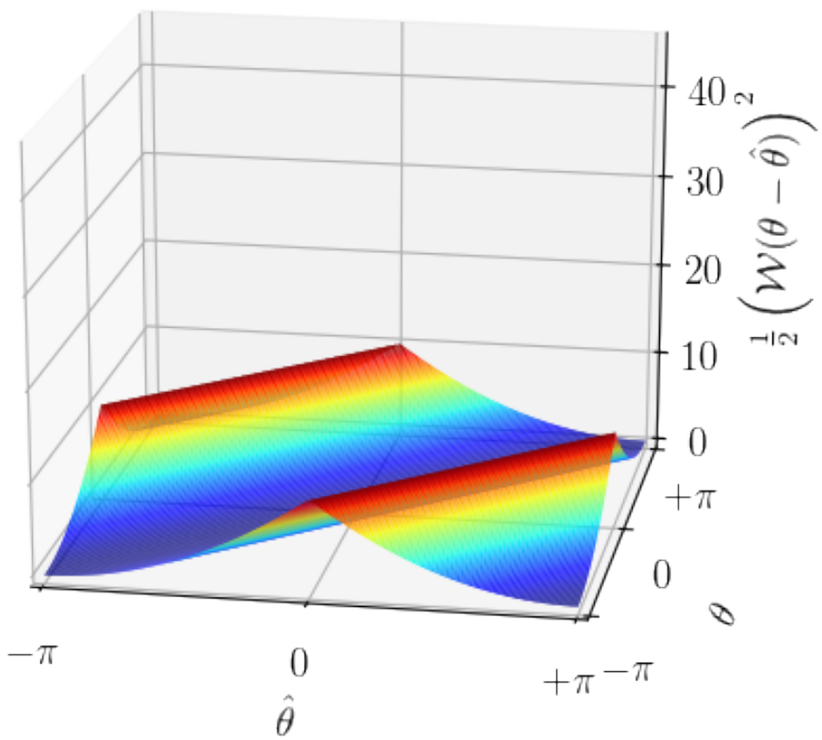
$$\alpha_v = +\frac{H-1}{2}, \quad \beta_v = -\frac{H+1}{2}, \quad \alpha_u = +\frac{W-1}{2}, \text{and} \quad \beta_u = -\frac{W+1}{2}.$$

Appendix: Cardiac MRI

— Losses



L2 Loss



Wrapped Phase Loss

— Loss Functions

$$L_\Omega = \alpha_1 L_{S_U}$$ ← (a) Coarse Segmentation Loss

$$+ \alpha_2 (L_{t_x} + L_{t_y} + L_\theta + L_s)$$ ← (b) STN Matrix Loss

$$+ \alpha_3 (L_{I_t} + L_{I_\theta} + L_{I_s})$$ ← (b) STN Reconstruction Loss

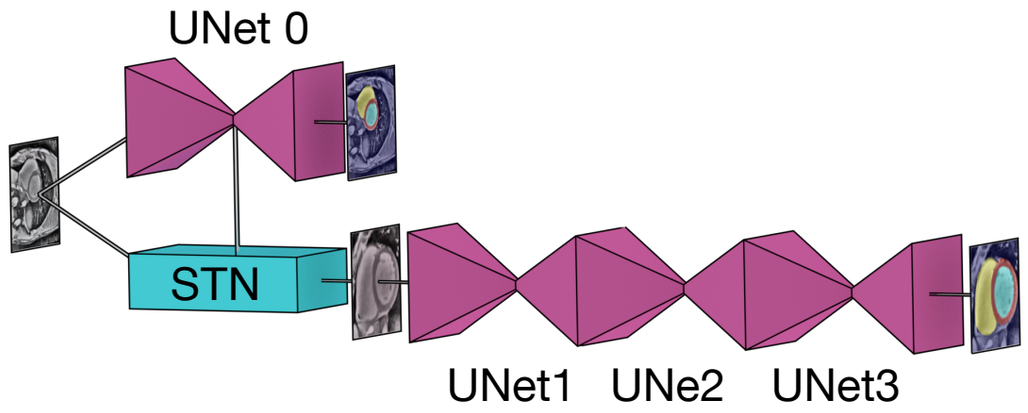
$$+ \alpha_4 \sum_{d=1}^{D} L_{S_{H,d}},$$ ← (c) Fine Segmentation Loss

Appendix: Cardiac MRI

— Architectures

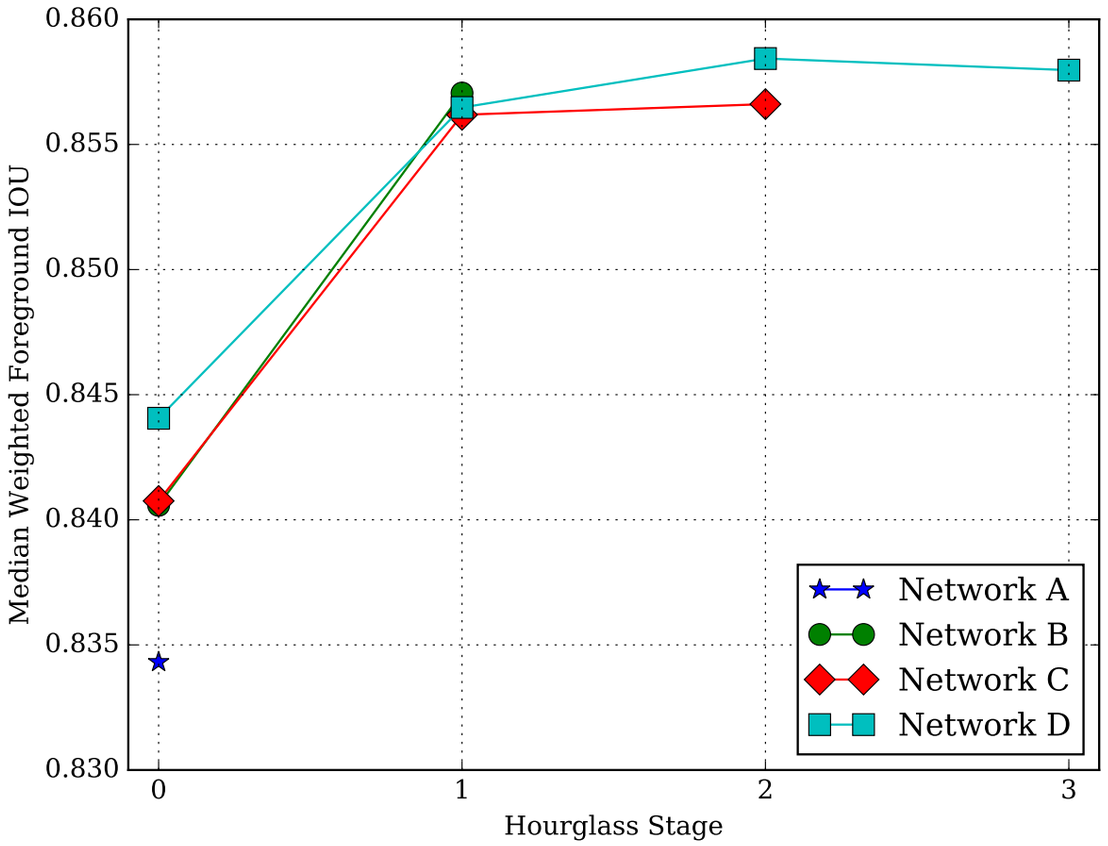
Name	UNet 0	UNet 1	UNet 2	UNet 3	Parameters (Millions)
Network A	128	–	–	–	7.0
Network B	64	64	–	–	3.5
Network C	64	64	64	–	4.5
Network D	64	64	64	64	5.5

* 128 or 64 channels in each convolutional layer



— Evaluations

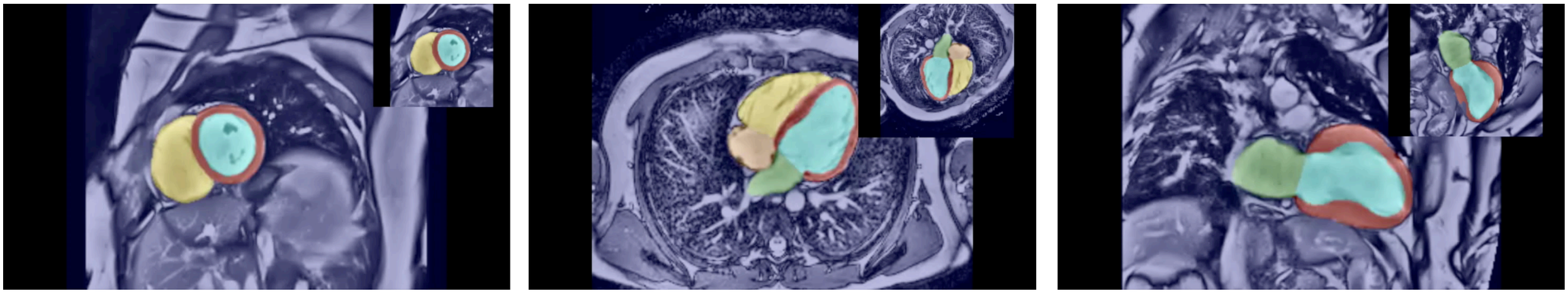
- Segmentation on canonical view improve performance by a large margin.
- Hourglass helps with the coarse segmentation results.



Appendix: Cardiac MRI

— Artificial Transformation

- Upper right shows the network prediction of inputs under artificial transformation.



Appendix: Cardiac MRI

— In-door data

- Data was collected at 10 medical centres from 2009 to 2010.
- 9 centres used 1.5T magnets, and one used 3T magnet.
- 42 subjects with overt hypertrophic cardiomyopathy (HCM),
- 21 healthy control subjects.
- Each Volume was cropped or padded to 256x256 pixels spatially, and 20 to 50 frames temporally.
- All reported results are based on three-fold (patient level split) cross-validation.

Appendix: Cardiac MRI

— External data

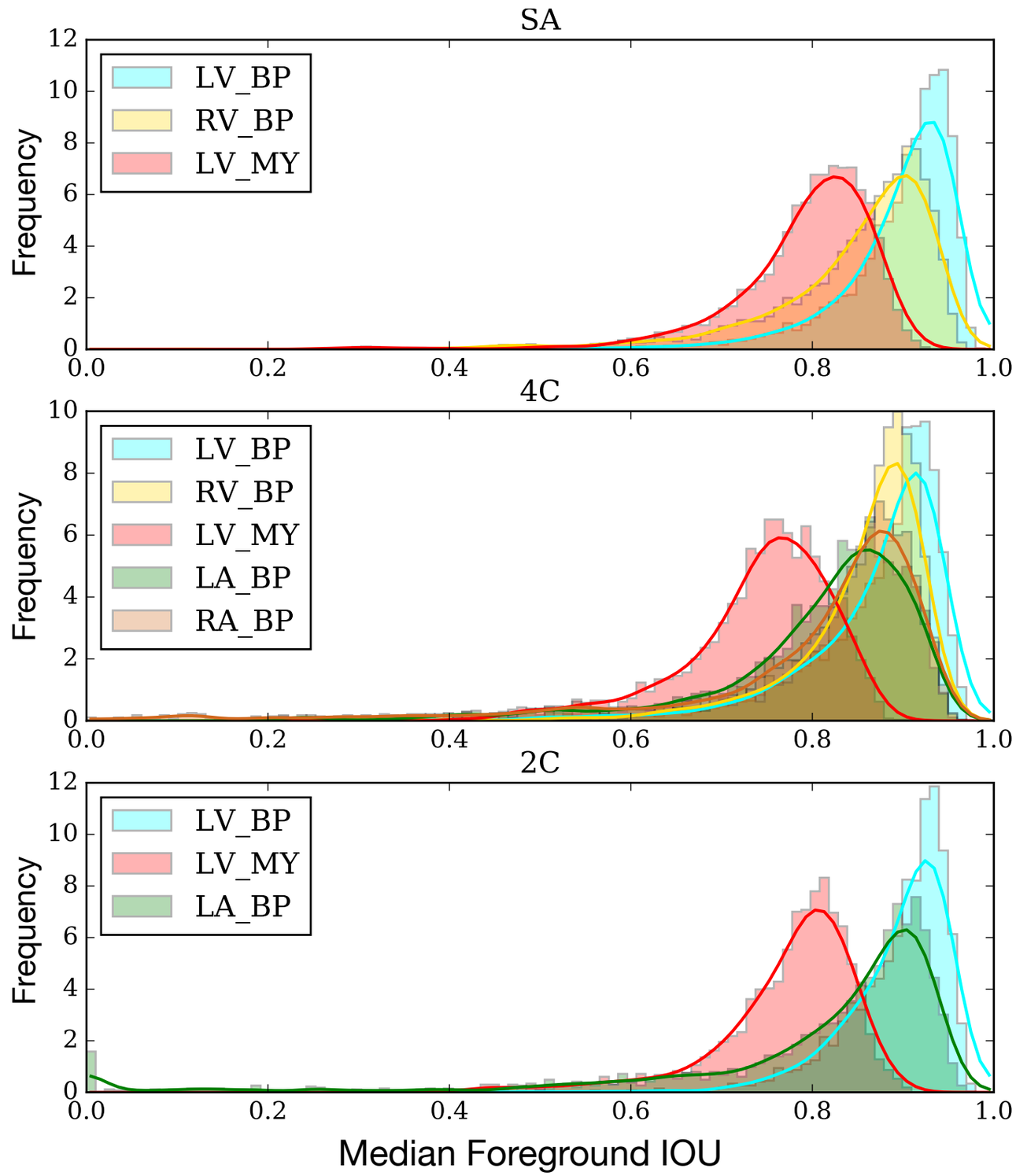
- 30 normal subjects - NOR
- 30 patients with previous myocardial infarction (ejection fraction of the left ventricle lower than 40% and several myocardial segments with abnormal contraction) - MINF
- 30 patients with dilated cardiomyopathy (diastolic left ventricular volume >100 mL/m^2 and an ejection fraction of the left ventricle lower than 40%) - DCM
- 30 patients with hypertrophic cardiomyopathy (left ventricular cardiac mass high than 110 g/m^2, several myocardial segments with a thickness higher than 15 mm in diastole and a normal ejecetion fraction) - HCM
- 30 patients with abnormal right ventricle (volume of the right ventricular cavity higher than 110 mL/m^2 or ejection fraction of the rigth ventricle lower than 40%) - RV

Appendix: Cardiac MRI

— Imagewise Histogram

- In all three clinical planes, performance is worst for the LV myocardium, best in LV blood pool.

- Segmentation error mainly concentrates at structure boundaries (Myocardium).



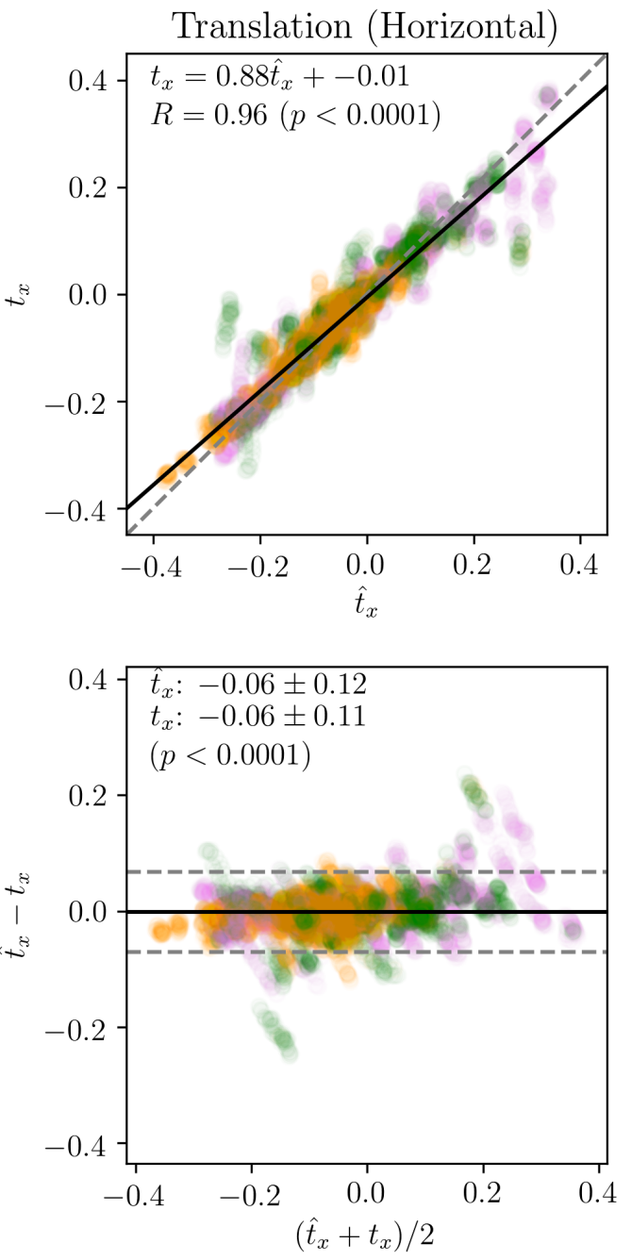
Appendix: Cardiac MRI

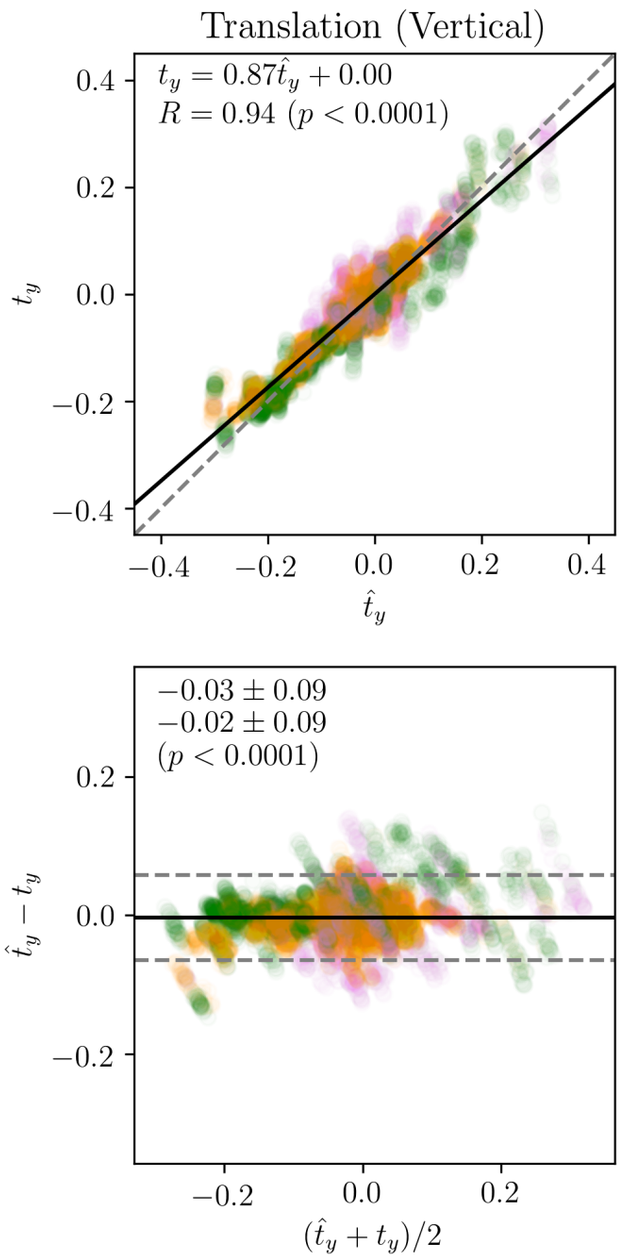
— Results Table

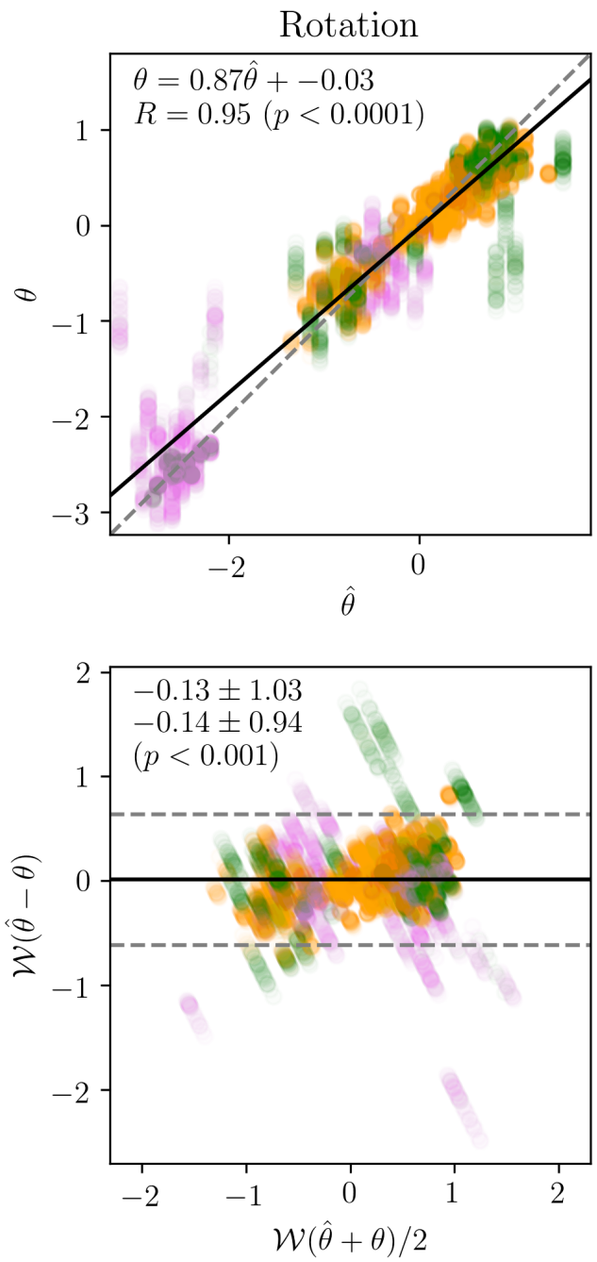
Name	View	U-Net 0	U-Net 1	U-Net 2	U-Net 3
Network A	All	0.834 [0.783, 0.871]	–	–	–
	SA	0.843 [0.789, 0.880]	–	–	–
	4C	0.819 [0.765, 0.855]	–	–	–
	2C	0.831 [0.788, 0.863]	–	–	–
Network B	All	0.841 [0.793, 0.876]	0.857 [0.819, 0.885]	–	–
	SA	0.848 [0.800, 0.884]	0.862 [0.820, 0.891]	–	–
	4C	0.831 [0.780, 0.861]	0.845 [0.812, 0.871]	–	–
	2C	0.832 [0.787, 0.864]	0.856 [0.822, 0.882]	–	–
Network C	All	0.841 [0.792, 0.875]	0.856 [0.816, 0.884]	0.857 [0.816, 0.885]	–
	SA	0.849 [0.797, 0.883]	0.862 [0.820, 0.890]	0.862 [0.819, 0.890]	–
	4C	0.830 [0.779, 0.861]	0.843 [0.804, 0.869]	0.844 [0.805, 0.869]	–
	2C	0.830 [0.793, 0.863]	0.855 [0.818, 0.883]	0.857 [0.819, 0.884]	–
Network D	All	0.844 [0.797, 0.877]	0.856 [0.819, 0.884]	0.858 [0.821, 0.886]	0.858 [0.821, 0.886]
	SA	0.851 [0.798, 0.886]	0.862 [0.821, 0.890]	0.863 [0.822, 0.892]	0.863 [0.822, 0.892]
	4C	0.832 [0.781, 0.861]	0.839 [0.805, 0.868]	0.843 [0.811, 0.870]	0.843 [0.811, 0.869]
	2C	0.842 [0.804, 0.869]	0.858 [0.828, 0.884]	0.860 [0.831, 0.886]	0.859 [0.830, 0.886]

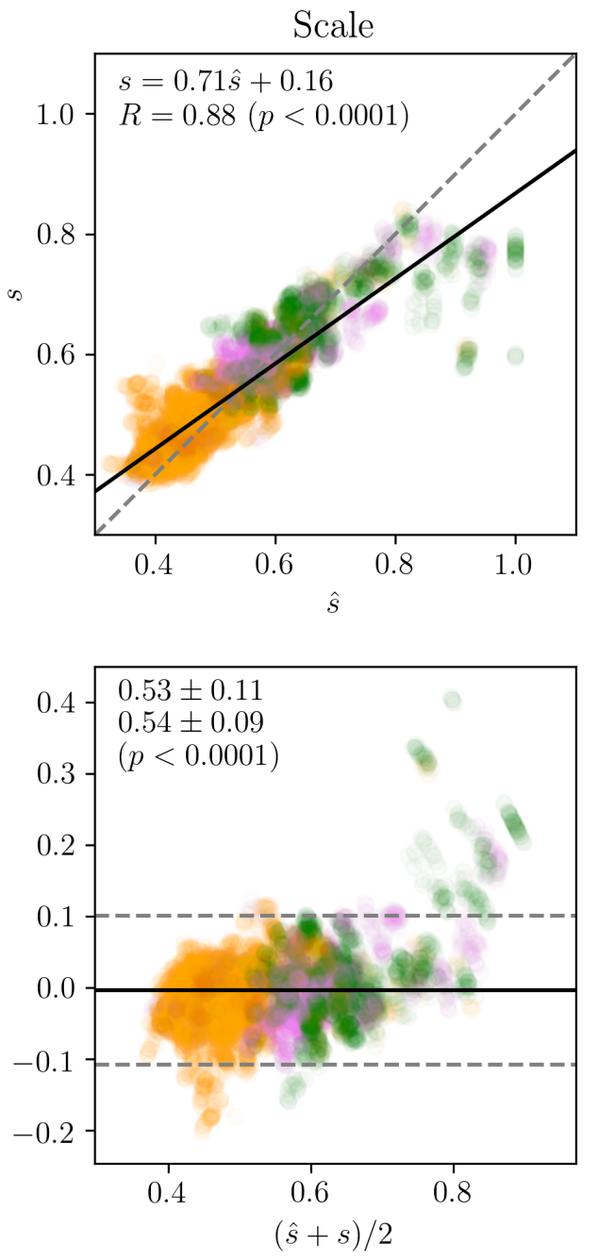
Appendix: Cardiac MRI

— Matrix Results









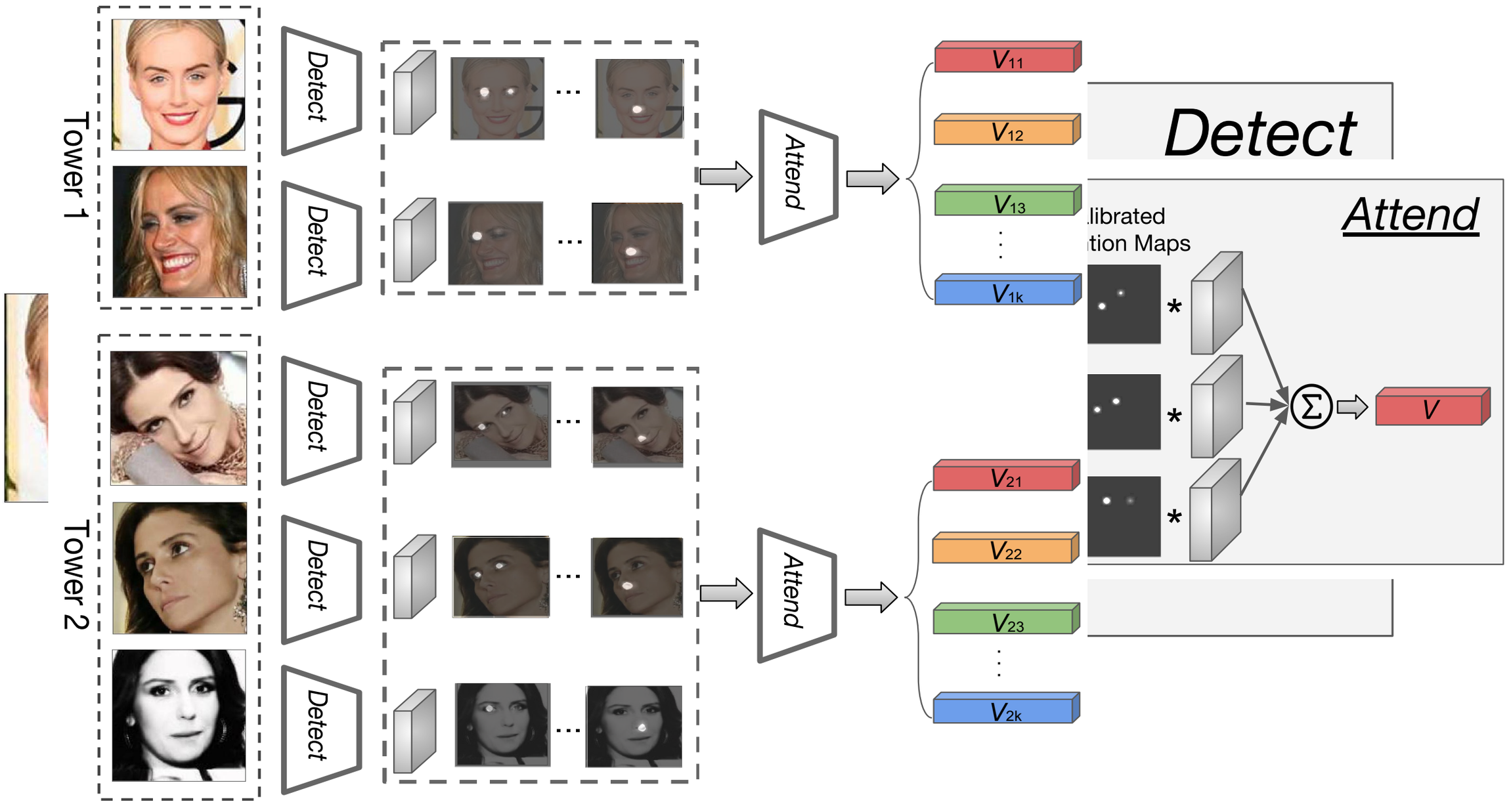
Orange : Short Axis, Purple : VLA (4 Chamber), Green : HLA (2 Chamber)

Comparator Networks (*Detect, Attend, Compare*)

— *Viewpoint conditioned similarity.*

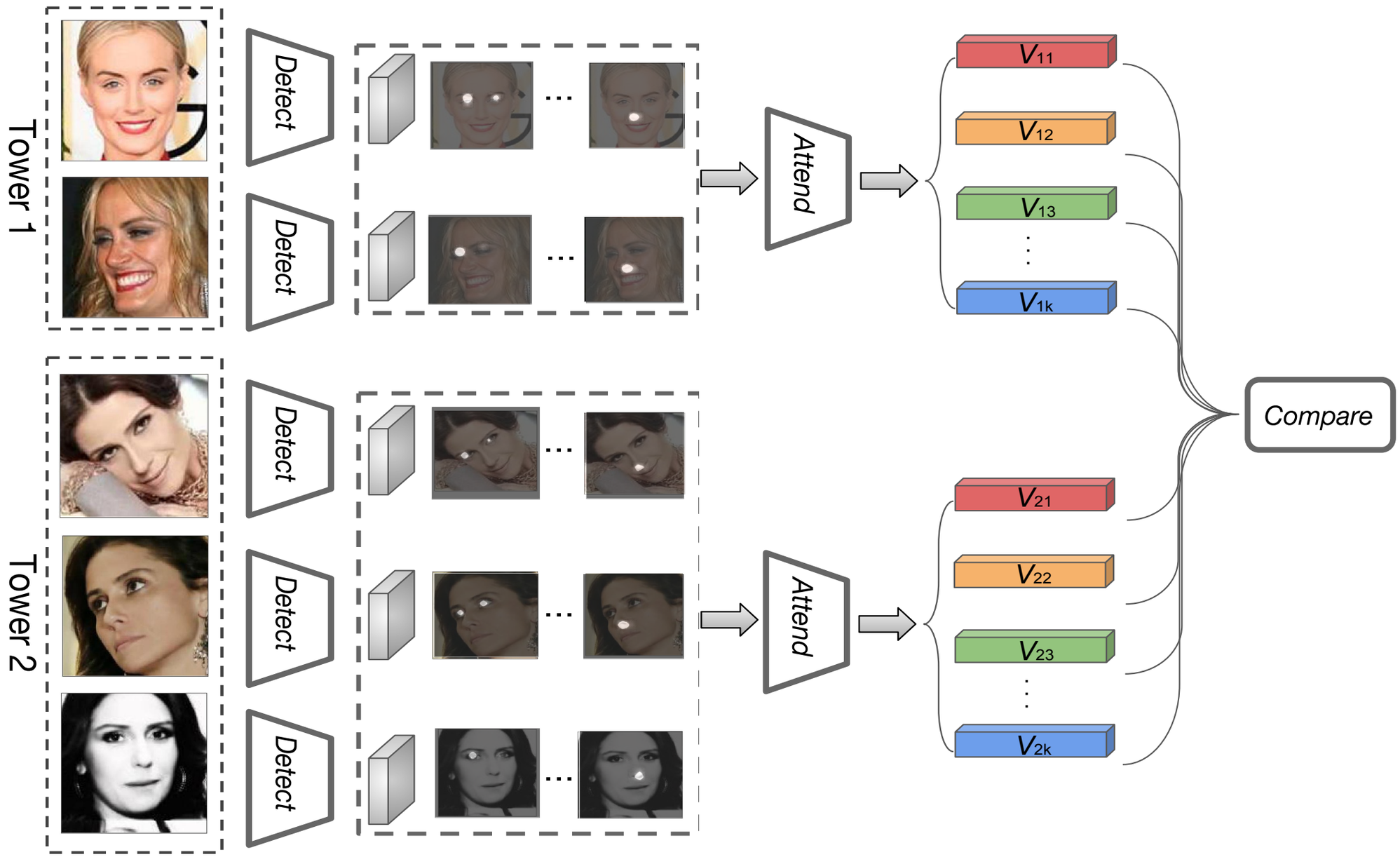
— *Local landmark comparison.*

— *Within template weighting.*

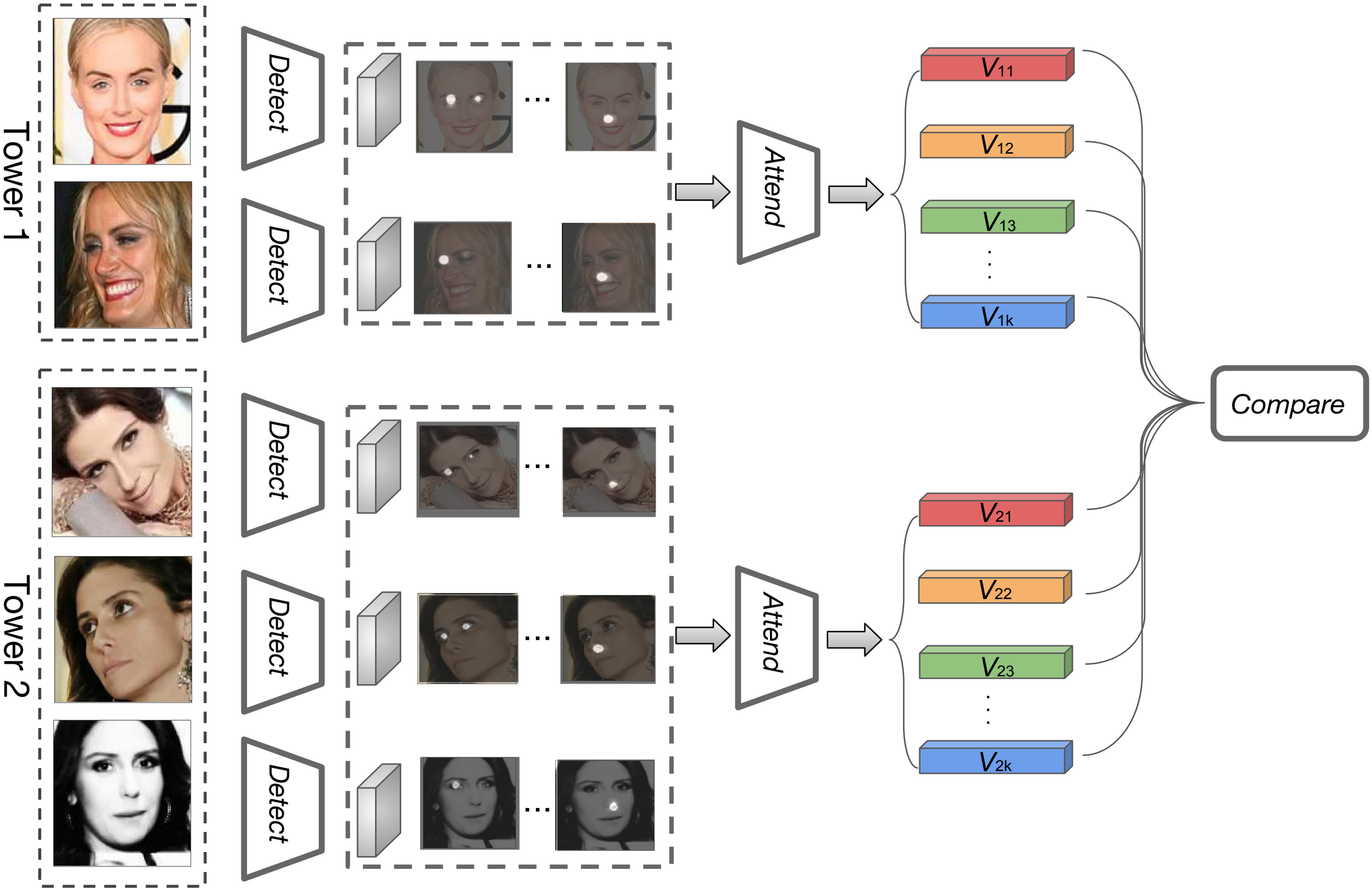


Comparator Networks (*Detect, Attend, Compare*)

- *Viewpoint conditioned similarity.*
- *Local landmark comparison.*
- *Within template weighting.*
- *Between template weighting.*



Appendix : Comparator Networks



Training Comparator Network

Challenges:

— Untrackable template pair combinations,

- VGGFace2 dataset contains *9131* identity, each identity contains *362* images in average, the possible template pairs will approximately scale up to :

$$\left(\binom{9131}{1} \times 362^{K} \right)^{2} \approx 1.88 \times 10^{23}$$ K=3 in our case

- Take inspiration from instance of image retrieval, i.e. use a simple model to choose hard samples.

Experiments Results (1:1 Template Verification)

— IARPA Janus IJBB Benchmarks:

Model	1:1 Verification TAR			
	FAR=$1E-4$	FAR=$1E-3$	FAR=$1E-2$	FAR=$1E-1$
Whitelam *et al.* [13]	0.540	0.700	0.840	−−
Navaneeth *et al.* [36]	0.685	0.830	0.925	0.978
ResNet50 [5]	0.784	0.878	0.938	0.975
SENet50 [5]	0.800	0.888	0.949	0.984
DCNs(Kpts)	0.823	0.921	0.966	0.991
DCNs(Divs)	0.835	0.923	0.971	0.995
ResNet50+SENet50	0.800	0.887	0.946	0.981
ResNet50+DCN(Kpts)	**0.850**	0.927	0.970	0.992
ResNet50+DCN(Divs)	0.841	0.930	0.972	0.995
SENet50+DCN(Kpts)	0.846	0.935	0.974	0.997
SENet50+DCN(Divs)	0.849	**0.937**	**0.975**	**0.997**

Table 1. Evaluation on 1:1 verification protocol on IJB-B dataset. (Higher is better)
Note that the result of Navaneeth *et al.* [36] was on the Janus CS3 dataset.
DCN(Divs) : Deep Comparator Network trained with Diversity Regularizer
DCN(Kpts): Deep Comparator Network trained with Keypoints Regularizer.

Experiments Results (1:1 Template Verification)

— IARPA Janus IJBC Benchmarks:

Model	1:1 Verification TAR			
	FAR=$1E-4$	FAR=$1E-3$	FAR=$1E-2$	FAR=$1E-1$
GOTS-1 [14]	0.160	0.320	0.620	0.800
FaceNet [14]	0.490	0.660	0.820	0.920
VGG-CNN [14]	0.600	0.750	0.860	0.950
ResNet50 [5]	0.825	0.900	0.950	0.980
SENet50 [5]	0.840	0.910	0.960	0.987
DCNs(Kpts)	0.851	0.921	0.969	0.992
DCNs(Divs)	0.862	0.930	0.972	0.994
ResNet50+SENet50	0.841	0.909	0.957	0.985
ResNet50+DCN(Kpts)	0.867	0.940	0.979	0.997
ResNet50+DCN(Divs)	0.880	0.944	0.981	0.998
SENet50+DCN(Kpts)	0.874	0.944	0.981	0.998
SENet50+DCN(Divs)	**0.885**	**0.947**	**0.983**	**0.998**

Table 2. Evaluation on 1:1 verification protocol on IJB-C dataset. (Higher is better) Results of GOTS-1, FaceNet, VGG-CNN are read from ROC curve in [14].

Experiments Results (1:1 Template Verification)