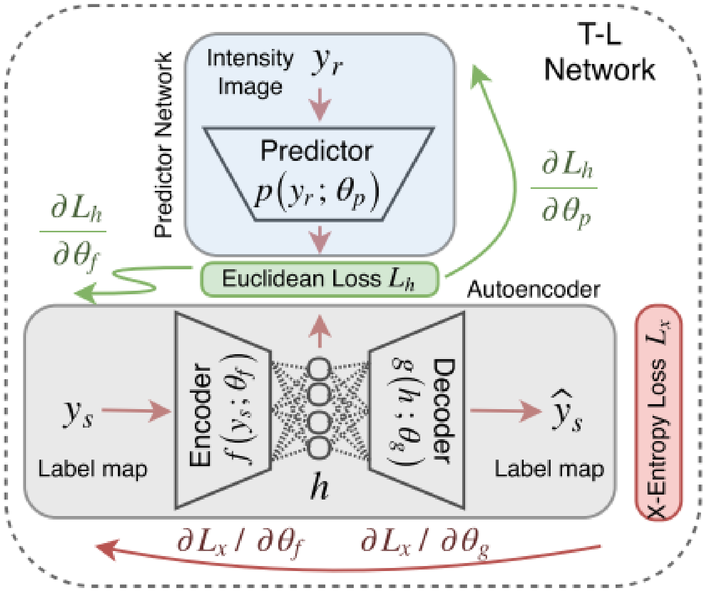


Fig. 3: Block diagram of the stacked convolutional autoencoder (AE) network (in grey), which is trained with segmentation labels. The AE model is coupled with a predictor network (in blue) to obtain a compact non-linear representation that can be extracted from both intensity and segmentation images. The whole model is named as T-L network.

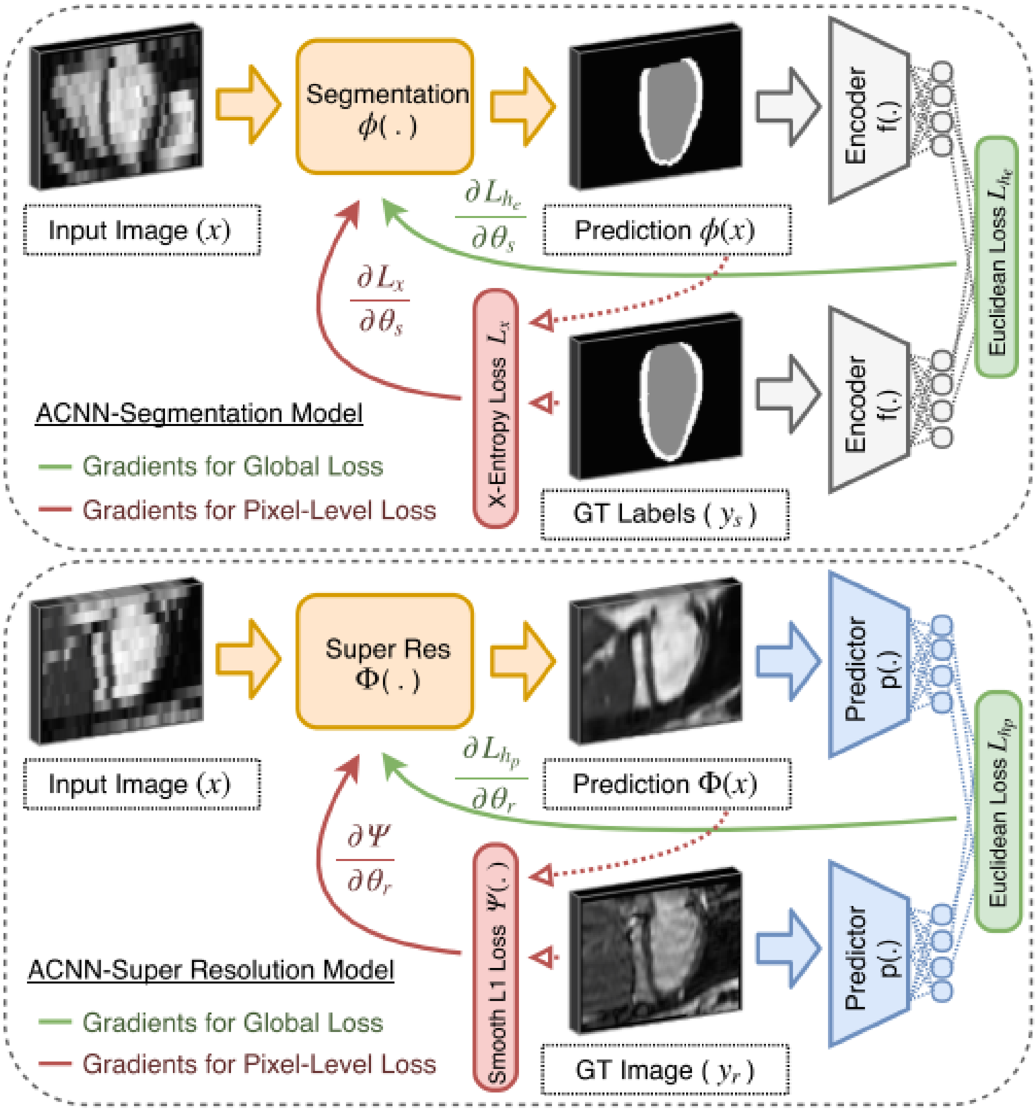


Fig. 4: Training scheme of the proposed anatomically constrained convolutional neural network (ACNN) for image segmentation and super-resolution tasks. The proposed T-L network is used as a regularisation model to enforce the model predictions to follow the distribution of the learnt low dimensional representations or priors.

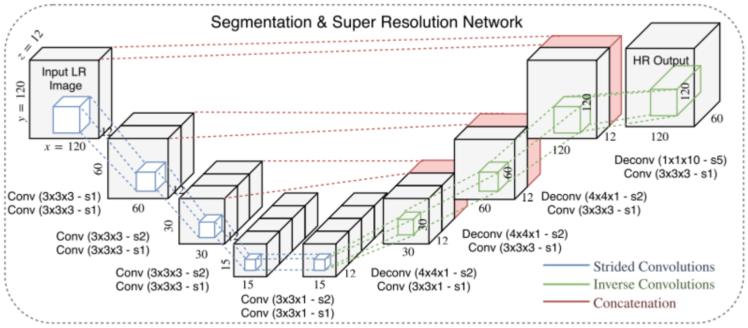


Fig. 2: Block diagram of the baseline segmentation (Seg) and super-resolution (SR) models which are combined with the proposed T-L regularisation block (shown in Fig. 3) to build the ACNN-Seg/SR frameworks. In SR, the illustrated model extracts SR features in low-resolution (LR) space, which increases computational efficiency. In segmentation, the model achieves sub-pixel accuracy for given LR input image. The skip connections between the layers are shown in red.

$$L_{h_p} = \| \, p\left(\Phi(\boldsymbol{x}); \boldsymbol{\theta}_p\right) - p\left(\boldsymbol{y}_r; \boldsymbol{\theta}_p\right) \|_2^2$$

$$\min_{\boldsymbol{\theta}_r} \left( \Psi_{\ell_1}\left(\Phi(\boldsymbol{x}\,; \boldsymbol{\theta}_r) - \boldsymbol{y}_r\right) + \lambda_1 \cdot L_{h_p} + \frac{\lambda_2}{2} ||\boldsymbol{w}||_2^2 \right) \tag{2}$$

$$\Psi_{\ell_1}(k) = \{0.5\, k^2 \text{ if } |k| < 1\,,\ |k| - 0.5 \text{ otherwise}\}$$

$$\min_{\boldsymbol{\theta}_r} \sum_{i \in \mathcal{S}} \Psi_{\ell_1}\left(\Phi(\boldsymbol{x}_i\,; \boldsymbol{\theta}_r) - \boldsymbol{y}_i\right)$$

TABLE III: Average inference time (Inf-T) of the SR models per input LR image (120x120x12) using a GPU (GTX-1080). ACNN-SR and SR-CNN [34] models are given the same number of filters and capacity. MOS [26] results, received from the clinicians (R1 and R2), are reported separately.

| | SSIM [47] | MOS-R1 | MOS-R2 | Inf-T |
|---|---|---|---|---|
| Linear | .777±.043 | 2.71±0.82 | 2.60±.91 | - |
| B-Spline | .779±.053 | 2.77±0.89 | 2.64±.84 | - |
| SR-CNN [34] | .783±.046 | 3.59±1.05 | 3.85±.70 | .29 s |
| 3D-UNet [12] | .784±.045 | 3.55±0.92 | 3.99±.71 | .07 s |
| ACNN-SR | **.796±.041** | **4.36±0.62** | **4.25±.68** | **.06** s |
| p-values | $p \ll 0.001$ | $p < 0.001$ | $p < 0.01$ | - |

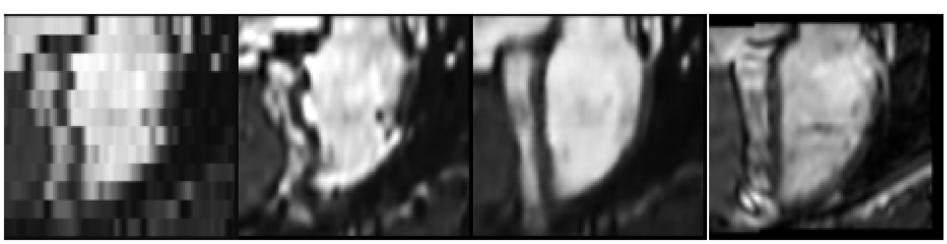

Fig. 8: Image super-resolution (SR) results. From left to right, input low resolution MR image, baseline SR approach [34] (no global loss), the proposed anatomically constrained SR model, and the ground-truth high resolution acquisition.