Regularization of CNN with Shape Priors

Seminar Machine Learning

Neelu Madan

Universität Duisburg-Essen
Department of Computer Science and Applied Cognitive Science

26. June 2018

Outline

1. Regularization
2. GridNet
3. Anatomically Contrained Neural Networks (ACNN)
4. Anatomically Contrained Neural Networks (ACNN)
5. References

What is regularization?

- Ensures that model to generalizes well to the real world data.
- Reduces Overfitting.
- Techniques:
  - Dropout
  - Weight penalty L1 and L2
  - Dataset augmentation
  - Early stopping
- There is no best generic regularization technique.

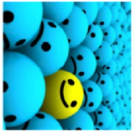
CNN Regularization

- Training data size is big.
- Number of features are very large.
- L2/L1 regularization or dropout like other neural networks.
- **L1/L2 regularization**: Add an extra term to the loss function.
  - L2 norm: Penalize the square weights ($|w|^2$).
  - L1 norm: Penalize the absolute values ($|w|$).

UNet: Regularization

- Training data size is very small.
- Data augmentation
  - Rotate
  - Scale
  - Flip
- Makes network invariant to certain transformations.

Image



Rotate

Scale

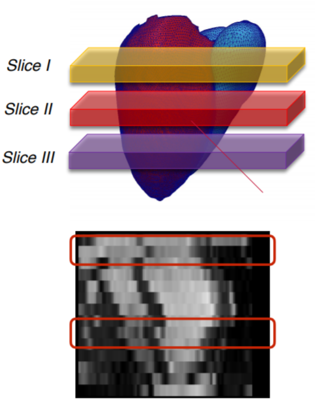
Flip

Shape priors

- A very specific type of "regularization" technique.
- Prior information related with the geometry of the object .
  - Boundries and edge polarity
  - Topology specification
  - Distance prior between the region
  - Atlas Prior (Most suitable for medical images)

.

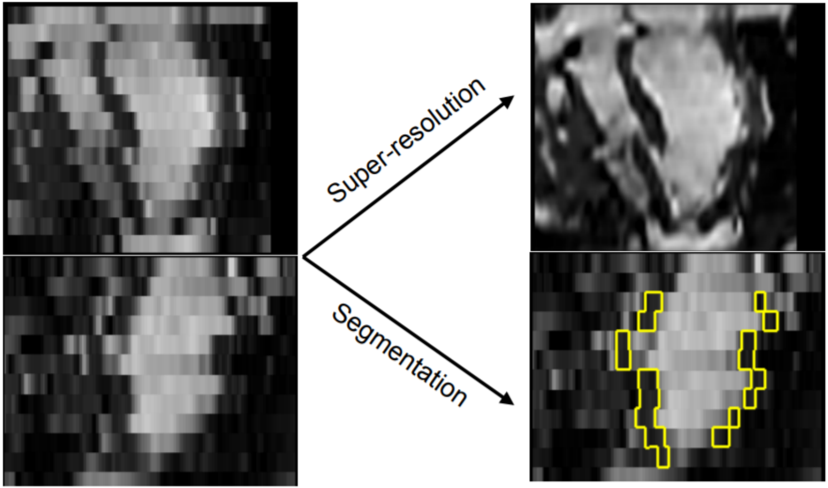
Shape priors



Cardiac MRI: 2D multislice data

- Useful where images are corrupted and contain artefact.
- Application in field of Medical image analysis.
- Improves the final reliability and accuracy of the segmentation result.

Conventional CNNs without shape prior



CNNs with shape prior

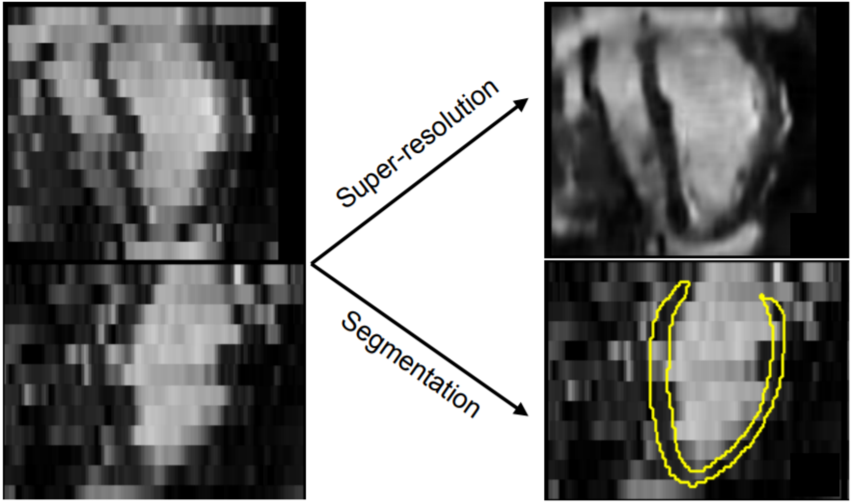
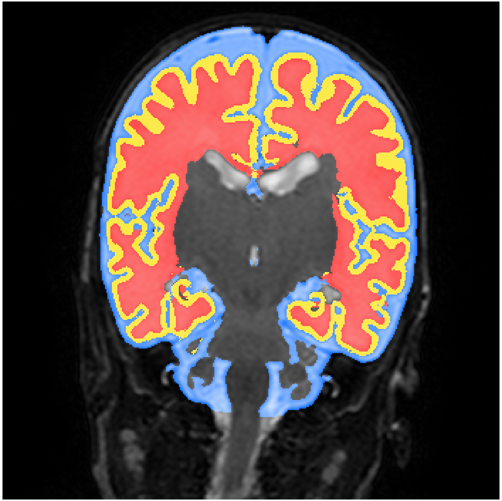


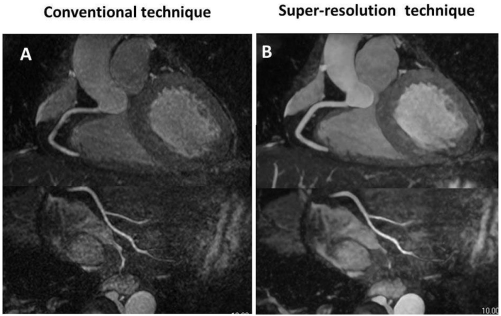
Figure: Using prior/expert knowledge about the shape result into anatomically more meaningful and smooth image.

Applications

a) Segmenation



b) Image Enhancement



GridNet

- Generalization of U-Net.
- Incorporate shape prior (S_i) regularization: via alignment of S_i on top of input.
- Specific to MRI cardiac segmentation.
- Extract both global and local contexts.
  - Global context: Distinguishes among the organs.
  - Local context: For segmentation.

GridNet: Comparison with U-Net

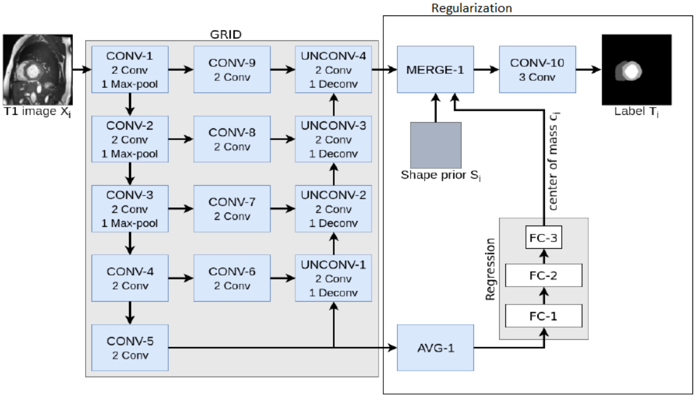


Fig. : Our network architecture.

- Additional
  CONV-6 to 9: Feature extraction at different resolution.
  CONV-5: Estimate CoM.
- 8 millions params in U-Net as compare to 32 millions in U-Net.

GridNet: Architecture

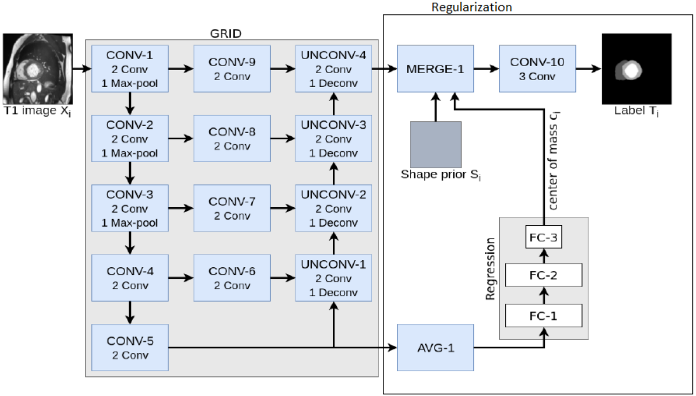


Fig. : Our network architecture.

Input:

- 256 x 256 MR image (X_i)
- Shape prior (S_i) of the corresponding slice.

Output:

- CoM (c_i) (bottom right)
- label field (T_i) (top right)

GridNet: Architecture

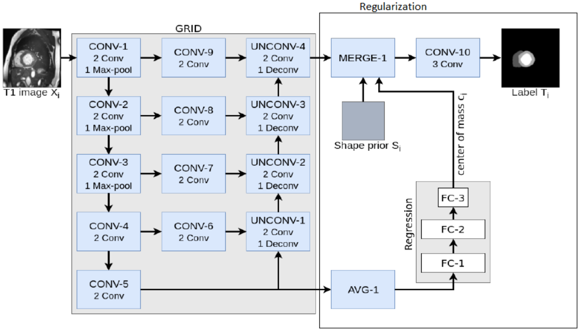


Fig. : Our network architecture.

Segmentation network (GRID) :

- **Column 1(from CONV-1 to CONV-5)**: Global context increases as we move down.
- **Column 2** : 4 convolution layers to compute the features at various resolutions.
- **Column 3**: UNCONV-4 layer contains entire information, used to segment the image.

GridNet: Architecture

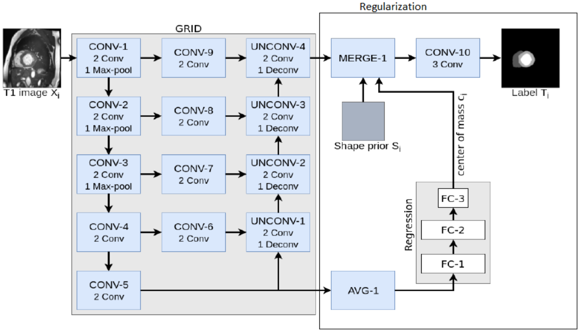


Fig. : Our network architecture.

Regularization :

- CoM (c_i) is the output of regression module (from FC-3).
- "Merge-1" realigns shape prior based on predicted CoM.

GridNet with shape prior: Loss

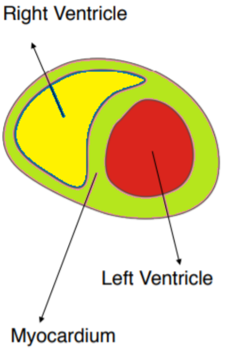
Standard loss for segmentation:

Cross-Entropy

Regularization:

- Contour loss.
- "Shape regularization loss" as difference between CoMs

GridNet: Results



- Regularization improved the results significantly especially close to boundaries.
- Outperforms popular segmentation networks like U-Net, ConvDeconv.
- Measures:
  - Clinical metric:
    - Correlation coefficient for cavity volume
    - Ejection fraction(EF) of LV and RV
  - Geometric metric:
    - Dice coefficient
    - Hausdorff distance.

GridNet: Results

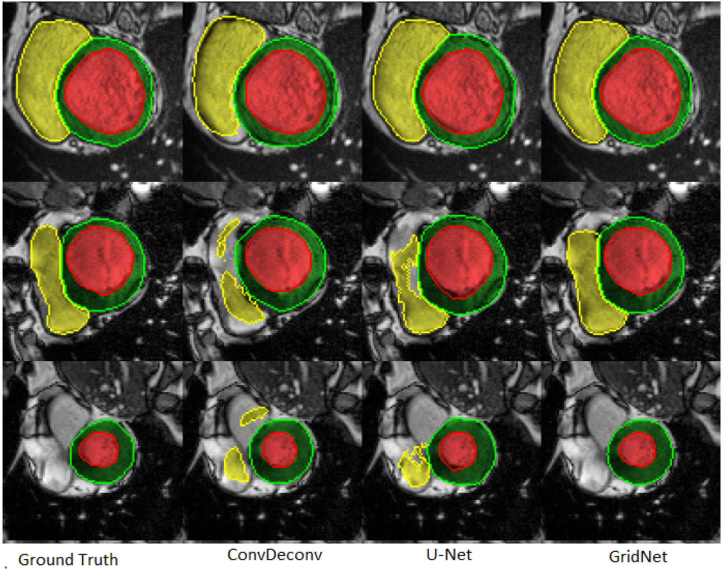
Table : Results for the validation dataset.

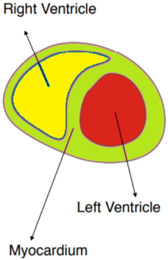
	Dice LV		Dice RV		Dice MYO	
	ED	ES	ED	ES	ED	ES
ConvDeconv	0.92	0.87	0.82	0.64	0.76	0.81
UNet	**0.96**	0.92	0.88	0.79	0.78	0.76
GridNet	**0.96**	**0.94**	**0.94**	**0.87**	**0.89**	**0.90**

	HD LV (mm)		HD RV (mm)		HD MYO (mm)	
	ED	ES	ED	ES	ED	ES
ConvDeconv	8.77	10.34	22.59	28.45	13.92	11.64
UNet	6.17	8.29	20.51	21.20	15.25	17.92
GridNet	**5.96**	**6.57**	**13.48**	**16.66**	8.68	**8.99**

	Corr EF LV	Corr EF RV	Corr MYO ED	Corr LV vol	Corr RV vol
ConvDeconv	0.988	0.764	0.927	0.990	0.957
UNet	0.991	0.824	0.921	0.995	0.821
GridNet	**0.992**	0.898	**0.975**	**0.997**	**0.978**

GridNet: Results





Anatomically Contrained Neural Networks

- PCA based statistical model unlike aligment in GridNet
  - Learns latent represenation of anatomy via auto-encoders.
  - Final prediction follows the learnt space.
- Components :

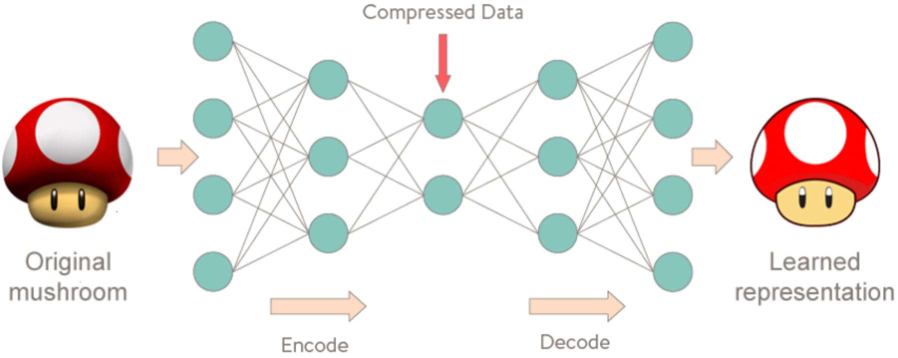
  - For Segmenation:

  Segmenation Network + TL Network (for shape regularization)

  - For Image Enhancement:

  SR Network + TL Network(for shape regularization)

AutoEncoder: Overview



- NN that aim to learn intermediate representation.
- Output could be reconstructed from this intermediate representation.
- Used for dimensionality reduction.
- Integrated as regularization model., s.t. prediction (y) will be anatomically meaningful and more accurate

AutoEncoder: Loss

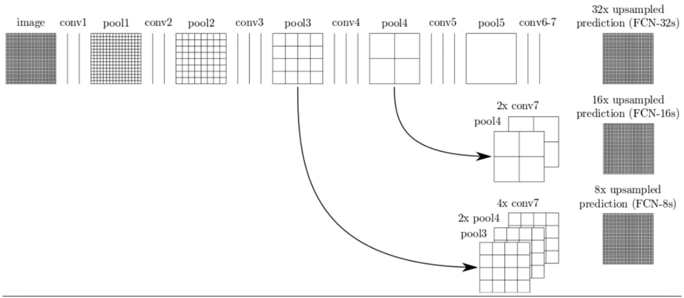
$f(y_s)$: Encoder component of AE.
$g(f(y_s))$: Decoder component of AE.

Loss: $L_x(y_s, g(f(y_s)))$

L_x is penalising $g(f(y_s))$ being dissimilar from y_s

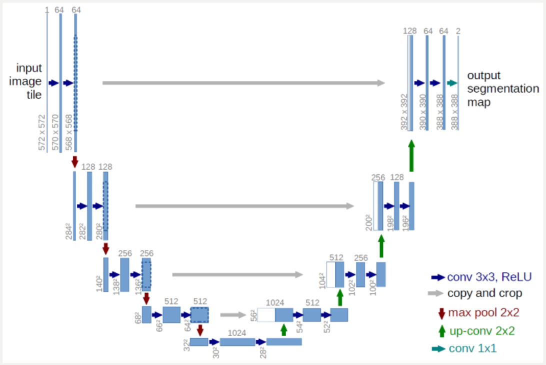
Segmentation networks

- We can use any segmentation network.
- Aim: To get per pixel classification.
- Popular segmentation networks we discussed:
  FCN:



Segmentation networks

U-Net:



Anatomically Contrained Neural Networks

- PCA based statistical model unlike aligment in GridNet
  - Learns latent represenation of anatomy via auto-encoders.
  - Final prediction follows the learnt space.
- Components :

  - For Segmenation:

  Segmenation Network + TL Network (for shape regularization)

  - For Image Enhancement:

  SR Network + TL Network(for shape regularization)

TL Network

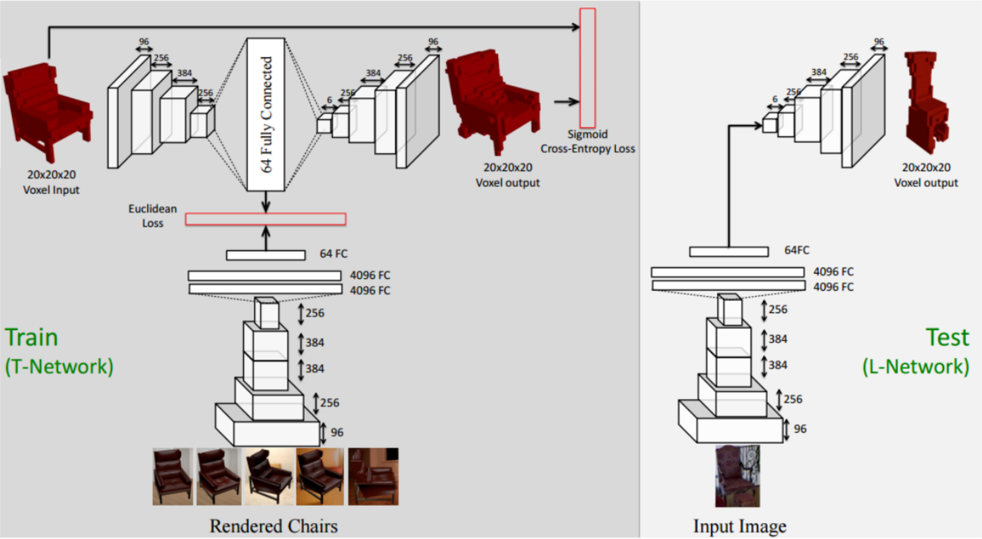


Figure: (a) T-Network (training time) : Takes 2 inputs 3D voxel map and 2D image (b) L-Network (test time) : Removes encoder part and uses image as input.

TL network

- Requirements:
  a. Generative: Should be able to generate voxel in 3D from latent space.
  b. Predictable: Should be able to predict the 3D represenation from 2D image of the object.

- **T Network**

  Components:

  - Auto-encoder: 3D voxel map -> latent space -> 3D voxel map
  - CNN

Total Loss = Reconstruction loss + Regression loss

  - Reconstruction loss: For voxel output
  - Regression loss: For 64D embeddings

- **L Network**: Remove encoder part only use image as input.

TL networks:Training

Three stage training:

1. Stage 1: Train the auto-encoder part of the network independently.
2. Stage 2: Train ConvNet to regress 64D representation generated by auto-encoder.
3. Stage 3: Fine tune the network jointly with both the losses.

Shape prediction using TL-Network vs PCA

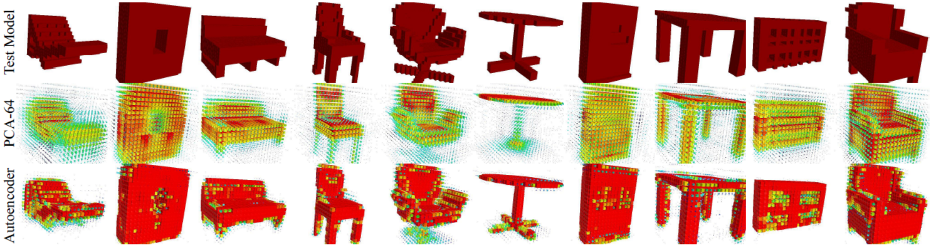
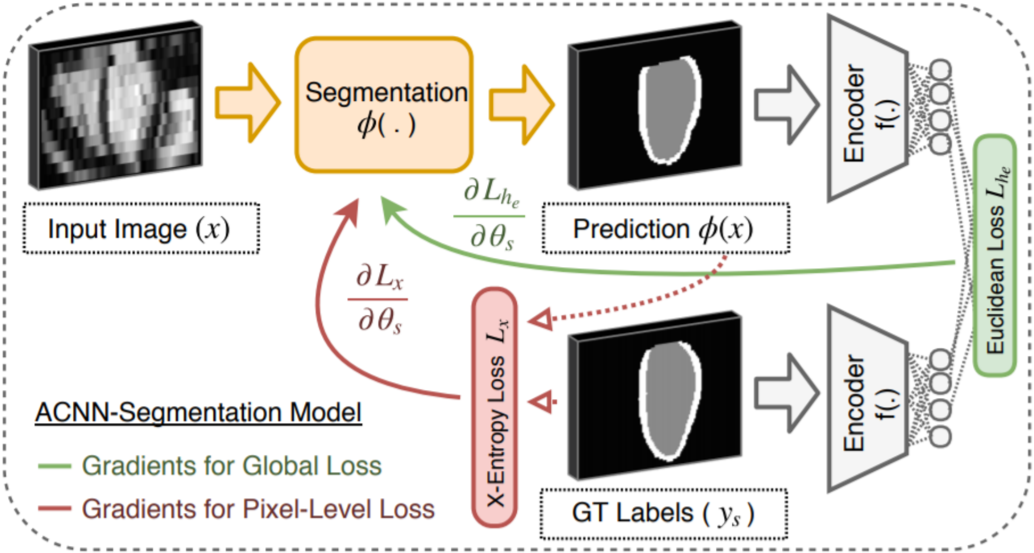


Figure: Predicted voxel are colored by level of prediction confidence i.e. from red to blue.

Anatomically Contrained CNN: Segmenation framework



Anatomically Contrained CNN: Segmenation

Standard Loss for segmentation: Cross- Entropy per pixel

Issues?

Anatomically Contrained CNN: Segmenation

Standard Loss for segmentation: Cross- Entropy per pixel

Issues?

Individual pixel level class prediction

- No guarantee about the global consistency

Anatomically Contrained CNN: Segmenation

Standard Loss for segmentation: Cross- Entropy per pixel

Issues?

Individual pixel level class prediction

- No guarantee about the global consistency
- No idea of possible anatomical shapes.

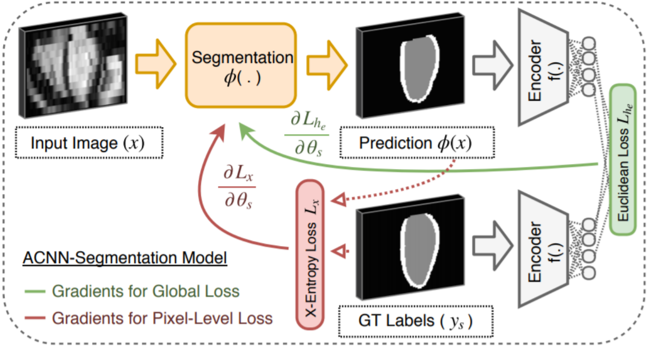
Anatomically Contrained CNN: Additional Regularization

L_{h_e} : Shape regularization loss
θ_f : Model parameter to be trained.
$\phi(x)$: Predicted output labels.
y : Ground truth labels.

$$L_{h_e} = ||f(\phi(x); \theta_f) - f(y; \theta_f)||^2$$

L_{h_e} : Ensures that **latent space of generated segmentations** are similar to that of **Ground Truth**.

Anatomically Contrained CNN: Total Loss

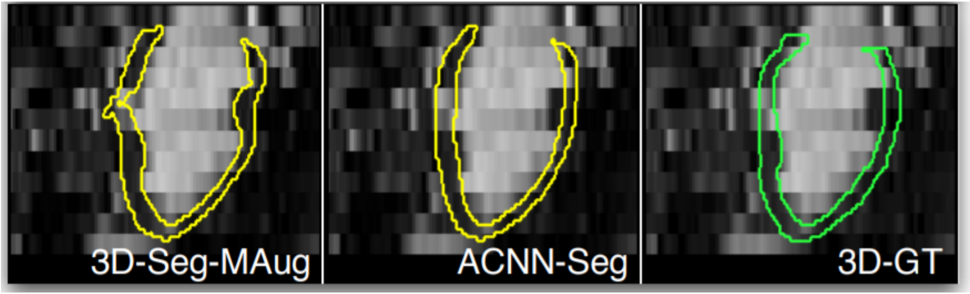


Segmenation Loss with shape regularization

L_x: Cross-entropy (L_x)

$$\text{Total Loss} = L_x + L_{h_e}$$

Anatomically Contrained CNN: Segmenation results

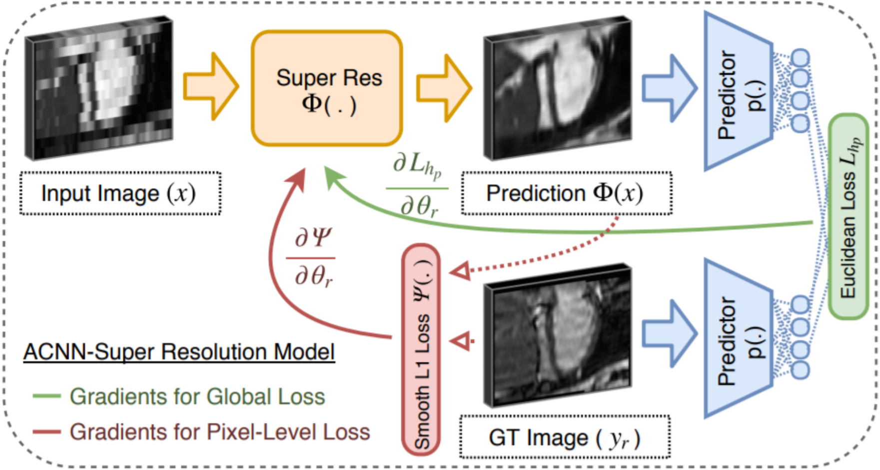


Anatomically Contrained CNN: Segmenation results

	Endocardium			Myocardium			Capacity
	Mean Dist. (mm)	Hausdorff Dist. (mm)	Dice Score (%)	Mean Dist. (mm)	Hausdorff Dist. (mm)	Dice Score (%)	# Trainable Parameters
2D-FCN	2.07±0.61	11.37±7.15	.908±.021	1.58±0.44	9.19±7.22	.727±.046	1.39×10^6
3D-Seg	1.77±0.84	10.28±8.25	.923±.019	1.48±0.51	10.15±10.58	.773±.038	1.60×10^6
3D-UNet	1.66±0.74	9.94±9.22	.923±.019	1.45±0.47	9.81±11.77	.764±.045	1.64×10^6
AE-Seg	1.75±0.58	8.42±3.64	.926±.019	1.51±0.29	8.52±2.72	.779±.033	1.68×10^6
3D-Seg-MAug	1.59±0.74	8.52±8.13	.928±.019	1.37±0.41	9.41±9.17	.785±.041	1.60×10^6
AE-Seg-M	1.59±0.48	**7.52±3.78**	.927±.017	1.32±0.26	**7.12±2.79**	.791±.036	1.91×10^6
ACNN-Seg	**1.37±0.42**	7.89±3.83	**.939±.017**	**1.14±0.22**	7.31±3.59	**.811±.027**	1.60×10^6
p-values	$p \ll 0.001$	$p \approx 0.890$	$p \ll 0.001$	$p \ll 0.001$	$p \approx 0.071$	$p \ll 0.001$	-

Figure: Dataset of 2D cardiac MR images and segmentation accuracy measured in terms of "Dice Metric" and "Surface to Surface distances"

Anatomically Contrained CNN: Super-resolution framework



Anatomically Contrained CNN: Loss

- Standard loss for SR Network: Smooth l_1 loss

Anatomically Contrained CNN: Loss

- **Standard loss for SR Network**: Smooth l_1 loss

- **Shape regularization loss**:

$$L_{h_p} = ||p(\phi(x);\theta_p) - p(y_r;\theta_p)||_2^2$$

Anatomically Contrained CNN: Loss

- **Standard loss for SR Network**: Smooth l_1 loss

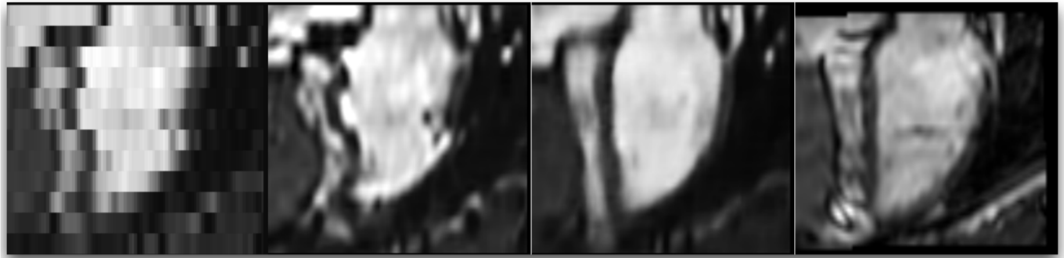
- **Shape regularization loss**:

$$L_{h_p} = ||p(\phi(x);\theta_p) - p(y_r;\theta_p)||_2^2$$

- **Segmenation Loss using shape regularization**:

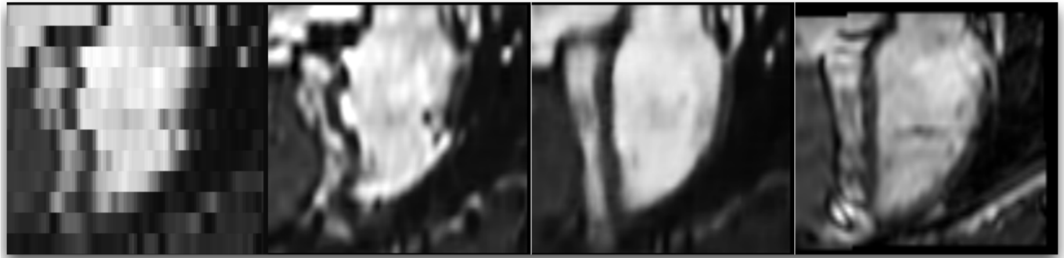
$$\text{Total Loss} = \Psi_{l_1} + L_{h_p}$$

Anatomically Contrained CNN: Super-resolution results



Original LR image | Baseline SR approach | Anatomically constrained SR model | Ground-truth HR image

Anatomically Contrained CNN: Super-resolution results



Original LR image

Baseline SR approach

Anatomically constrained SR model

Ground-truth HR image

Summary

- Prior knowledge about the anatomy provided guidance and robustness to models.
- Discussed two major models that used shape priors regularization:
  - GridNet
  - Anatomically constrained NN
- Future Research:
  - Human pose estimation.
  - Landmark localization on partially occluded images.

References I

Christian Osendorfer, Hubert Soyer and Patrick van der Smagt (2014). *Image Super-Resolution with Fast Approximate Convolutional Sparse Coding*. URL: http://arxiv.org/pdf/1609.04802.

OKtay, O et al. (2018). “Anatomically Constrained Neural Networks (ACNNs): Application to Cardiac Image Enhancement and Segmentation”. In: *IEEE transactions on medical imaging, ISSN: 1558-254X, DOI:10.1109/TMI.2017.2743464*.

Girdhar, Rohit et al. (2016). “Learning a Predictable and Generative Vector Representation for Objects”. In: *European Conference on Computer Vision. Springer, pp. 484–499.*

al., Zotti Clement et (2017). *"GridNet with automatic shape prior registration for automatic MRI cardiac segmentation"*. URL: https://arxiv.org/pdf/1705.08943.pdf.

Alain, Guillaume and Yoshua Bengio (2014). “What regularized auto-encoders learn from the data-generating distribution”. In: *The Journal of Machine Learning Research 15.1, pp. 3563–3593.*

References II

- Masci, Jonathan et al. (2011). “Stacked convolutional auto-encoders for hierarchical feature extraction”. In: *Artificial Neural Networks and Machine Learning–ICANN 2011, pp. 52–59.*
- Zeng, X. (2013). “Deep learning shape priors for object segmentation”. In: *Proc. IEEE CVPR, Jun. 2013, pp. 1870–1877.*
- J. Long, E. Shelhamer and T. Darrell (2015). “Fully convolutional networks for semantic segmentation”. In: *Proc. CVPR, 2015, pp. 3431–3440.*
- Courville, Aaron (2016). *Deep learning*. MIT Press.

Thank You!

APPENDIX A

GridNet

i: Corresponds to slice.

l: Classes (4 classes: RV, LV, MYO, Back)

v: Pixel location

$T_{i,l,v}$: True probability that for slice i pixel v is in class l

$\hat{T}_{i,l,v}$: Output of our model for slice i pixel v is in class l

$C_i, \hat{C}_i$: Countour extracted from T_i, $\hat{T}_i$

γ_T, γ_C: Constants

Loss functions:

$L_T = -\gamma_T \sum_{l=1}^{4} \sum_{v} T_{i,l,v} \ln \hat{T}_{l,v}$ (Cross-entropy of predicted labels)

$L_C = -\gamma_C \sum_{l=1}^{4} \sum_{v} C_{i,l,v} \ln \hat{C}_{l,v}$ (Cross-entropy of predicted contour)

$L_c = \gamma_c ||c_{i,w} - \hat{c}_i||^2$ (Euclidean distance between predicted CoM $c_{i,w}$ and true CoM)

$L_w = \gamma_w ||w||^2$ (Prior loss)

Total Loss:

$$L = \sum_i (-\gamma_T \sum_{l=1}^{4} \sum_{v} T_{i,l,v} \ln \hat{T}_{i,l,v} - \gamma_C \sum_{l=1}^{4} \sum_{v} C_{i,l,v} \ln \hat{C}_{l,v} + \gamma_c ||c_{i,w} - \hat{c}_i||^2) + \gamma_w ||w||^2$$

APPENDIX B

Anatomically Contrained CNN

Standard Loss for segmentation: Cross- Entropy

C : Set of classes.

$$L_x = -\sum_{c=1}^{C}\sum_{i\in S}\log(\frac{e^{f_{(c,i)}}}{\sum_j e^{f_{(j,i)}}})$$

Segmenation Loss with shape regularization:

L_x: Cross-entropy

θ_s: All the trainable parameters of segmentation model

λ_1 , λ_2: Weight of 'shape regularization loss' and 'weight decay term'

$$Loss = min\theta_s(L_x(\phi(x;\theta_s),y) + \lambda_1.L_{h_e} + \frac{\lambda_2}{2}||w||_2^2)$$

Super Resolution Loss using shape regularization:

Ψ_{l_1}: Smooth l_1 loss defined as { $0.5k^2$ if |k|<1 , |k| - 0.5 otherwise }

θ_r : All the trainable parameters in SR model.

λ_1 , λ_2: Weight of 'shape regularization loss' and 'weight decay term'

$$Loss = min\theta_s(\Psi_{l_1}(\phi(x;\theta_r) - y_r) + \lambda_1.L_{h_p} + \frac{\lambda_2}{2}||w||_2^2)$$