Imperial College London

Multi-Input Cardiac Image Super-Resolution using Convolutional Neural Networks

Ozan Oktay, Wenjia Bai, Matthew Lee, Ricardo Guerrero, Konstantinos Kamnitsas, Jose Caballero, Antonio de Marvao, Stuart Cook, Declan O'Regan, and Daniel Rueckert

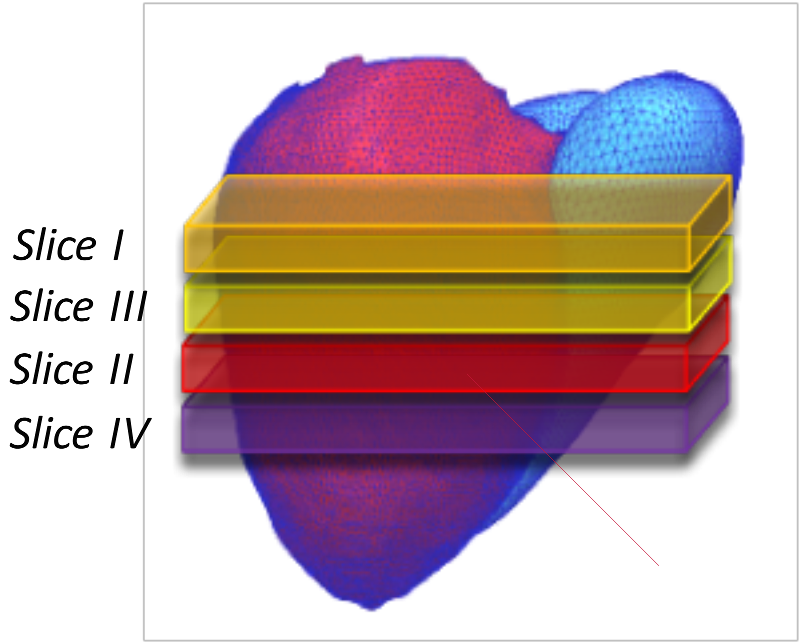
19th International Conference on Medical Image Computing and Computer Assisted Interventions (MICCAI 2016)
October 2016, Athens

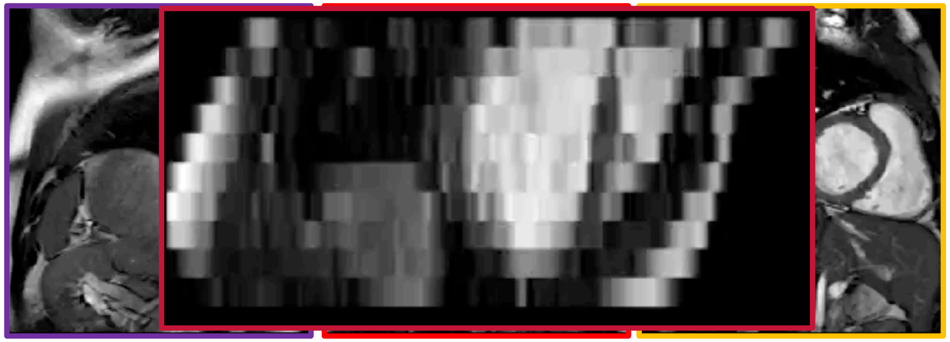
Clinical Motivation

<u>SAX Cardiac MR Image Acquisition</u>

- Large slice thickness (8-10 mm)
  - Due to constrains on SNR, acquisition and breath-hold time

- It hampers subsequent image analysis and quantitative measurements.

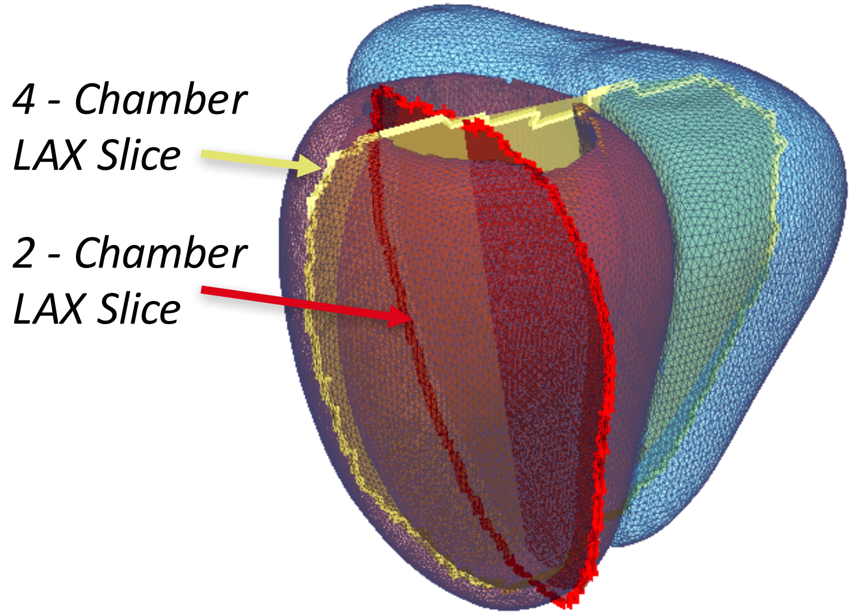


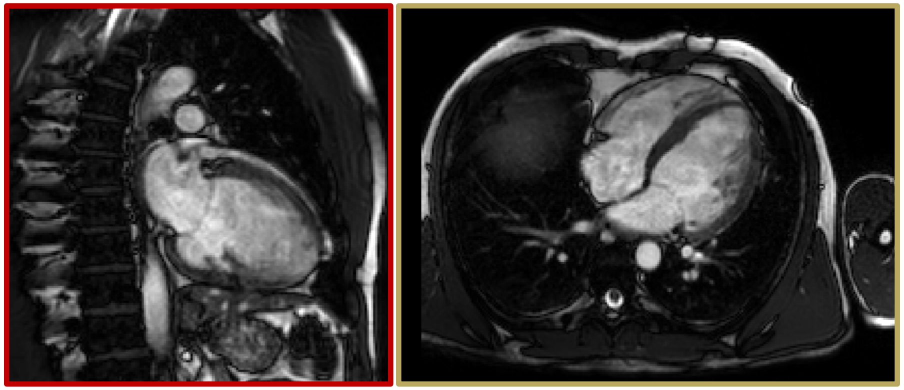


Clinical Motivation

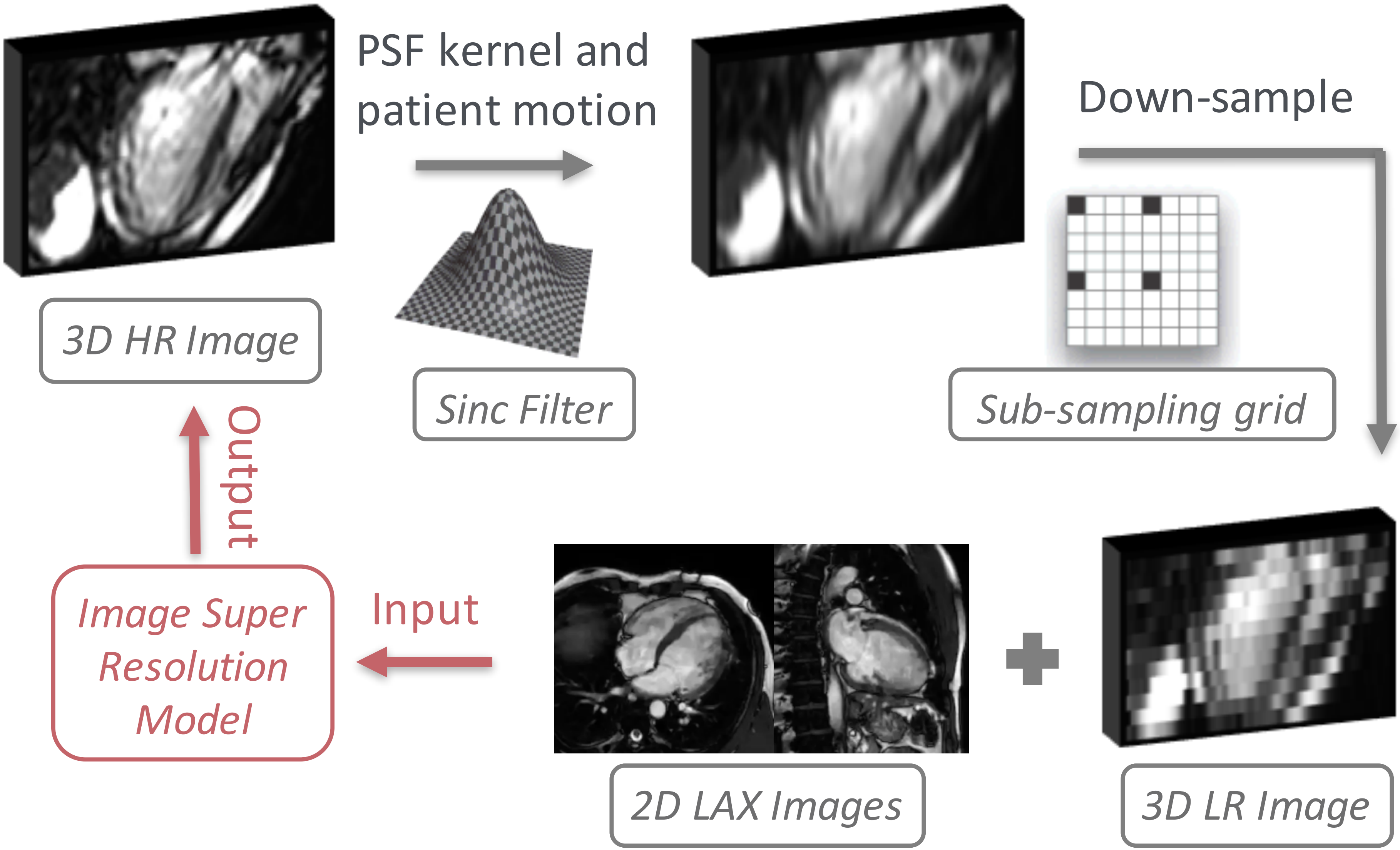
SAX Cardiac MR Image Acquisition

- Large slice thickness (8-10 mm)
  - Due to constrains on SNR, acquisition and breath-hold time
- It hampers subsequent image analysis and quantitative measurements.
- LAX image acquisitions are performed to complement SAX images





Low and High Resolution Images

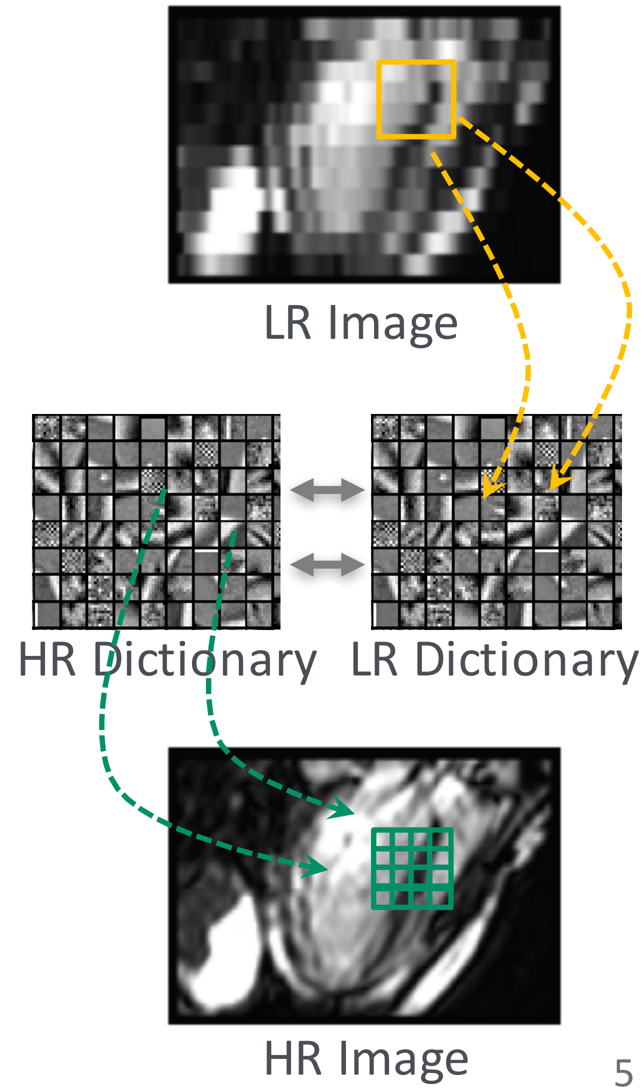


Related Work on Super Resolution

BioMedIA
Imperial College London

External Example & Model Based SR

I. Coupled Dictionary Learning and S. Coding [Yang et al TIP'12, Bhatia K. ISBI'14]



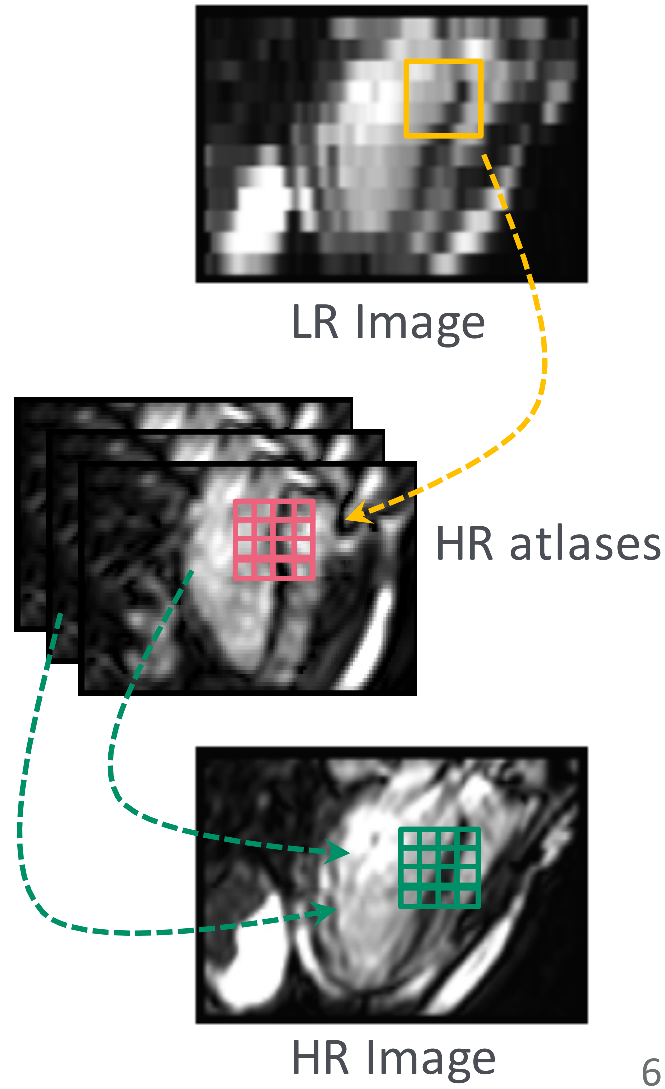
Related Work on Super Resolution

BioMedIA
Imperial College London

External Example & Model Based SR

I. Coupled Dictionary Learning and S. Coding
[Yang et al TIP’12, Bhatia K. ISBI’14]

II. Multi-Atlas Based SR Techniques
[Shi et al. MICCAI’13]



Related Work on Super Resolution

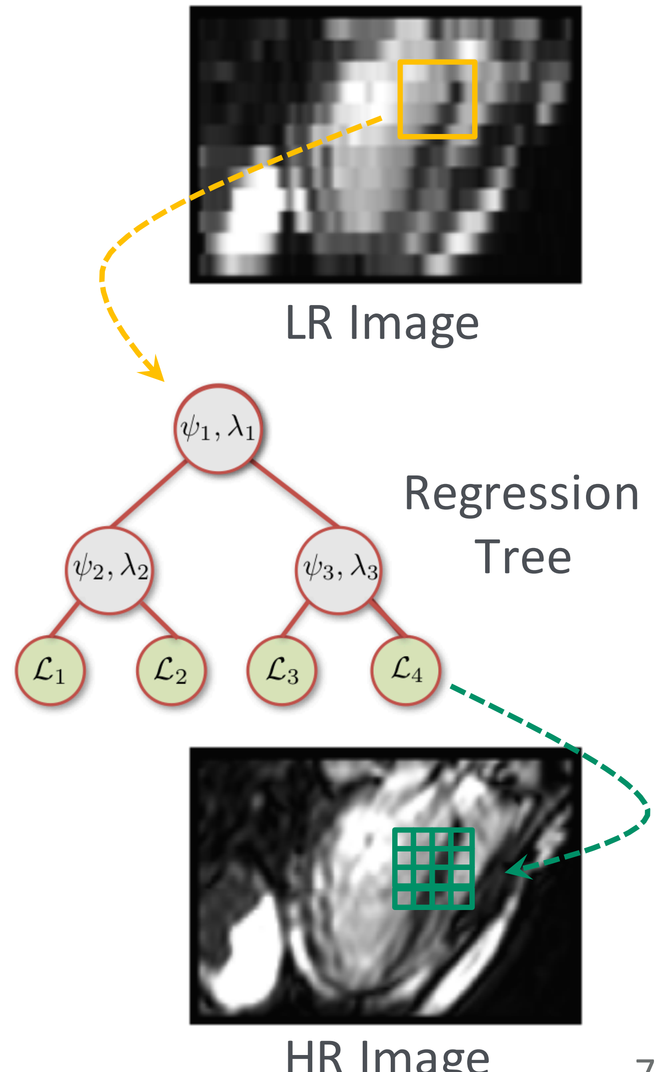
BioMedIA
Imperial College London

External Example & Model Based SR

I. Coupled Dictionary Learning and S. Coding
[Yang et al TIP'12, Bhatia K. ISBI'14]

II. Multi-Atlas Based SR Techniques
[Shi et al. MICCAI'13]

III. Decision Forest based Regression
[Alexander et al MICCAI'14,
Schulter S. CVPR'15]



Related Work on Super Resolution

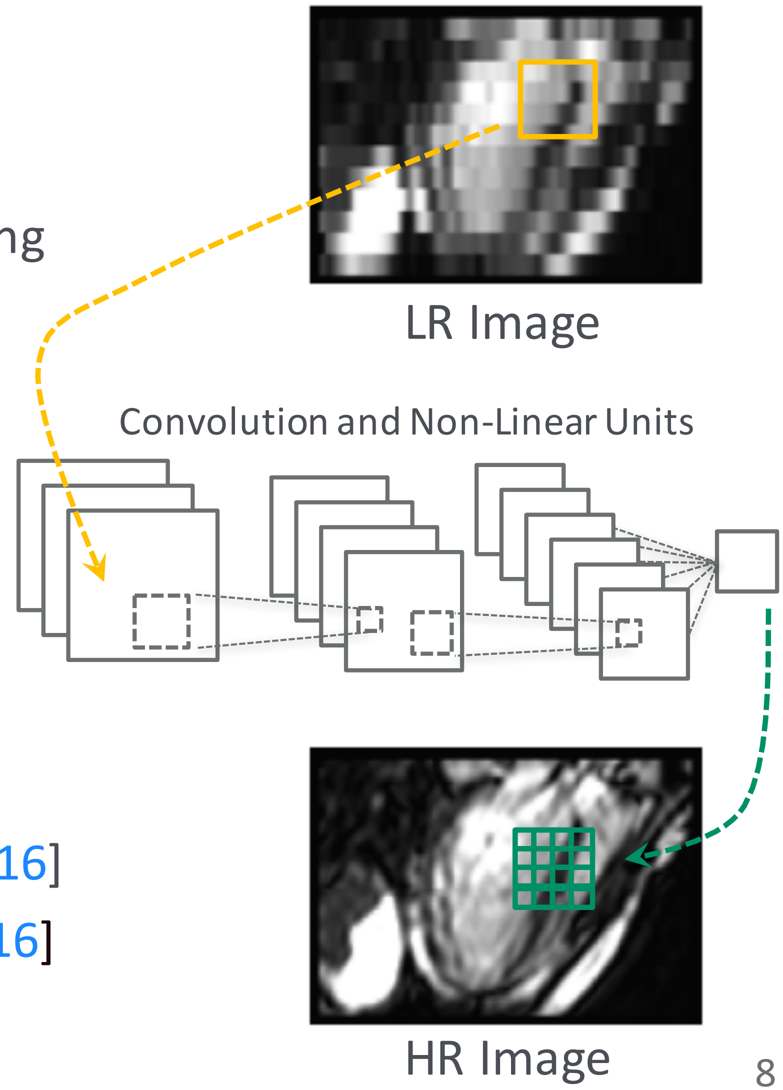
External Example & Model Based SR

I. Coupled Dictionary Learning and S. Coding
[Yang et al TIP'12, Bhatia K. ISBI'14]

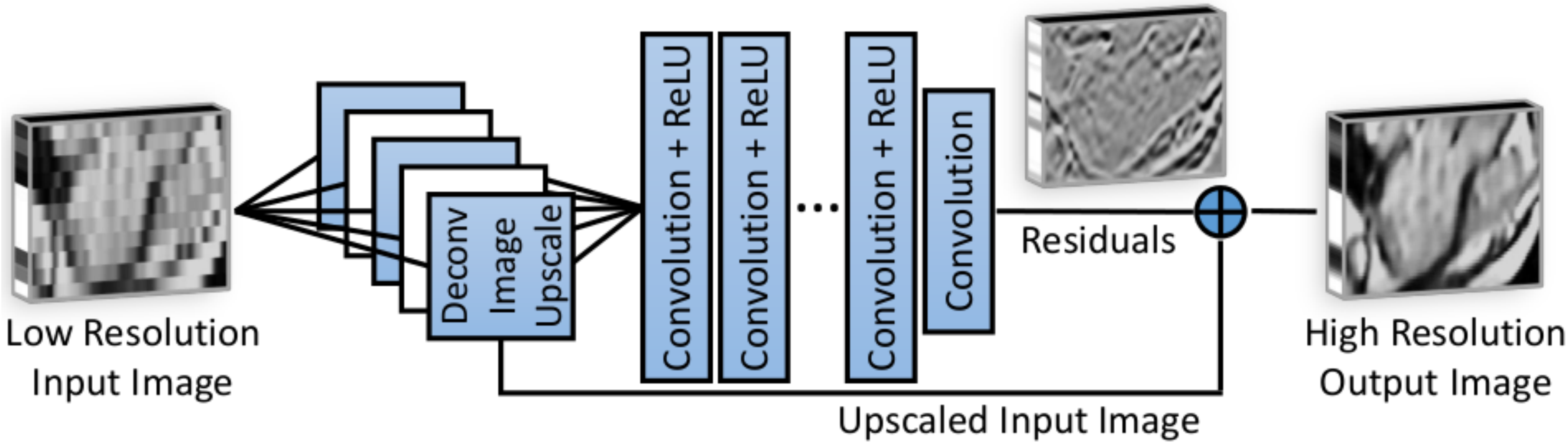
II. Multi-Atlas Based SR Techniques
[Shi et al. MICCAI'13]

III. Decision Forest based Regression
[Alexander et al MICCAI'14,
Schulter S. CVPR'15]

IV. Neural Network based Regression
   i. CNNs
   [Dong et al. ECCV'14, Shi et al. CVPR'16]
   ii. CNNs + GANs [Ledig et al Arxiv Sept'16]



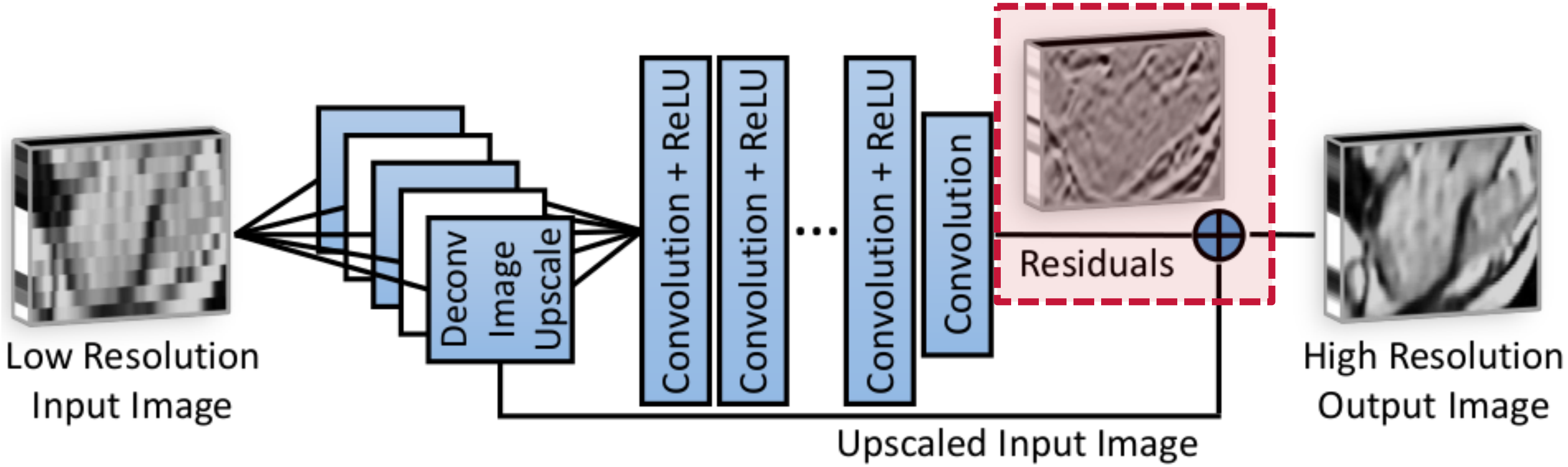
Proposed 3D-SR Model (Single-Image)



Components of the model

- 3D Convolution and Deconvolution (inverse convolution) Kernels
- Rectified Linear Units (ReLUs)
- Regression Based Cost Function (Smooth L1-Norm)
- Input (2D Stack-LR) and Output (3D-HR) Images

Proposed 3D-SR Model (Single-Image)

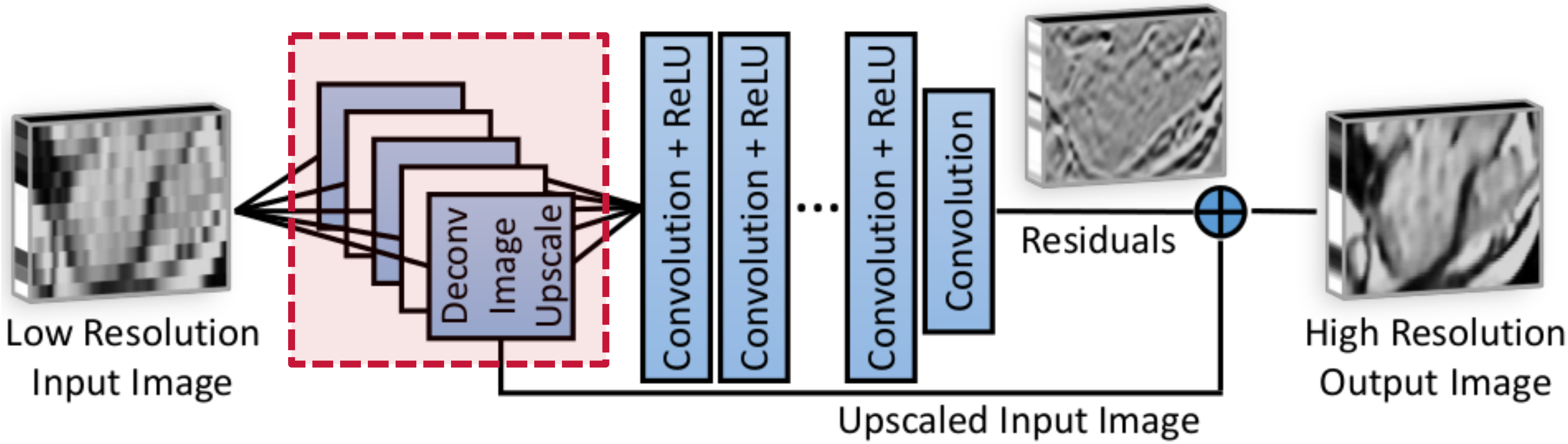


Proposed improvements on SR-CNN model:

I. Residual Learning

- An easier regression problem to solve
- Robust and faster model convergence

Proposed 3D-SR Model (Single-Image)

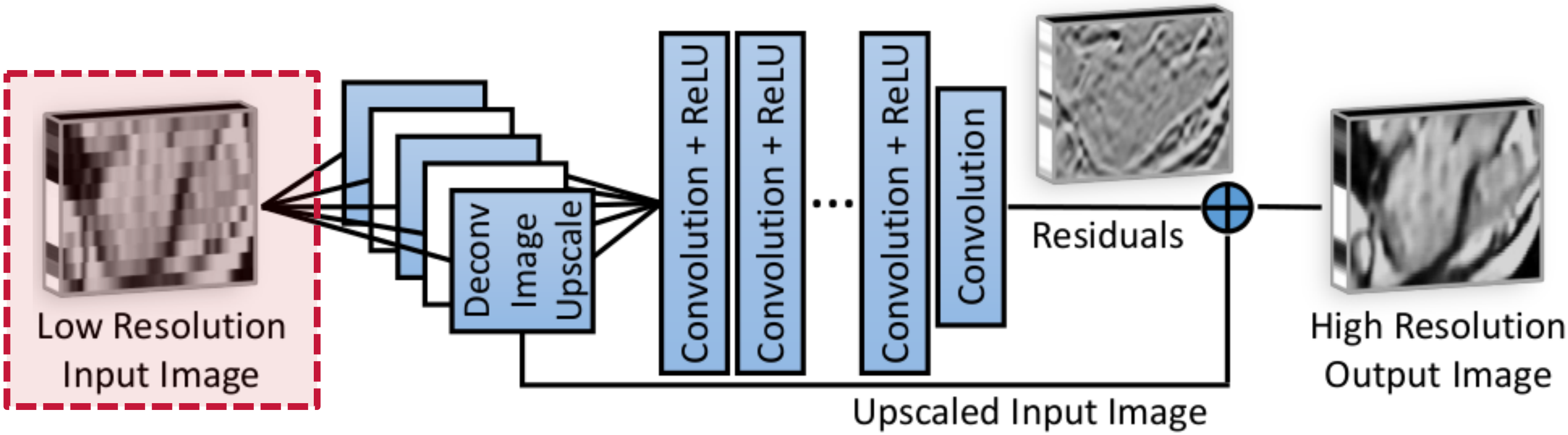


Proposed improvements on SR-CNN model:

II. Learning Upsampling Layers

- End-to-end training of convolution and upsampling kernels

Proposed 3D-SR Model (Single-Image)



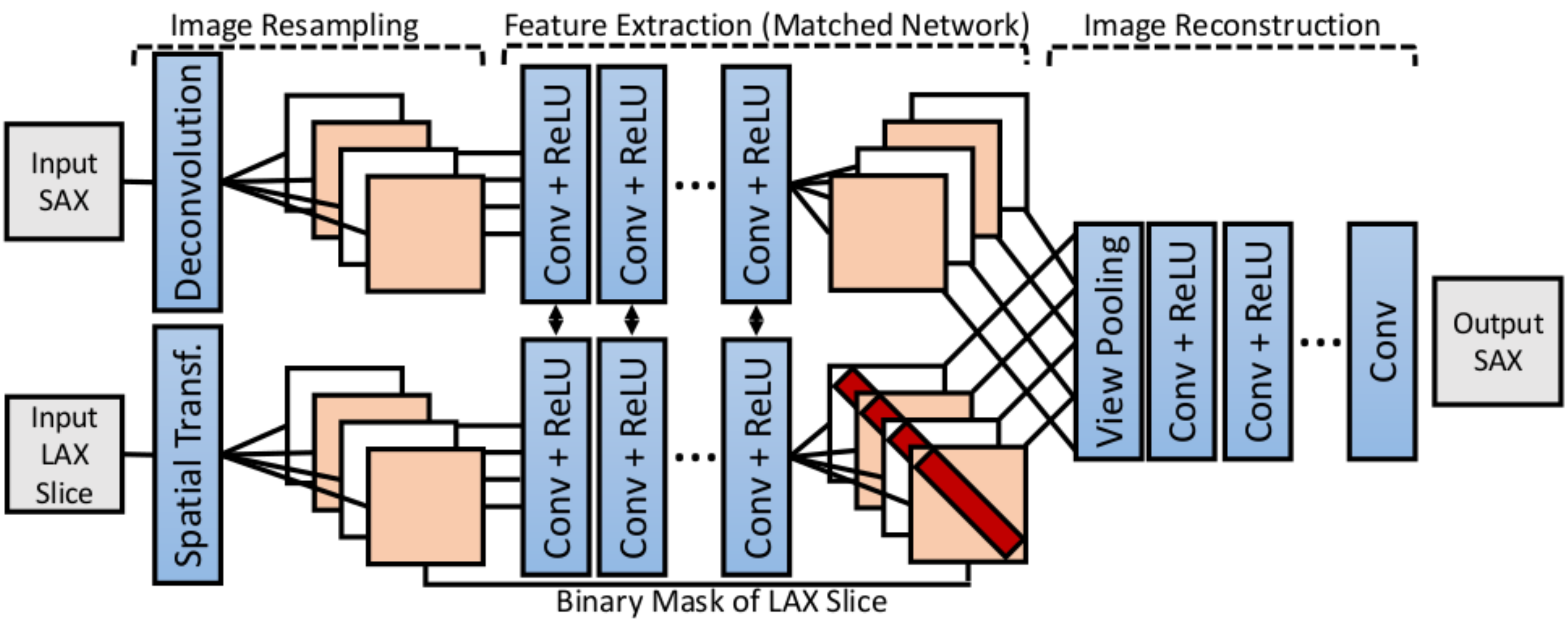
Proposed improvements on SR-CNN model:

III. Multi-Input model extension

- Constrains the regression task with more input data
- In cardiac imaging usually multiple image stacks are acquired.

Proposed 3D-SR Model (Multi-Image)

BioMedIA
Imperial College London



- Siamese model is used to combine information from multiple stacks
- The learned kernels can be easily integrated in this multi-model.

Method Evaluation Strategy

I. Image Quality Analysis

- Peak-to-Signal-Noise Ratio (PSNR) (Images from 300 Subjects)
- Structural Similarity Index Measure (SSIM) [Wang et al. IEEE TIP'04]

II. Subsequent Image Analysis (SR is used for pre-processing)

- Cardiac Image Segmentation (Images from 18 Subjects)
- Cardiac Motion Tracking (Images from 10 Subjects)

III. Our method is compared against:

- Linear, C-Spline, MAPM [Shi MICCAI'13], CNN [Dong TPAMI'15]

Image Quality Assessment

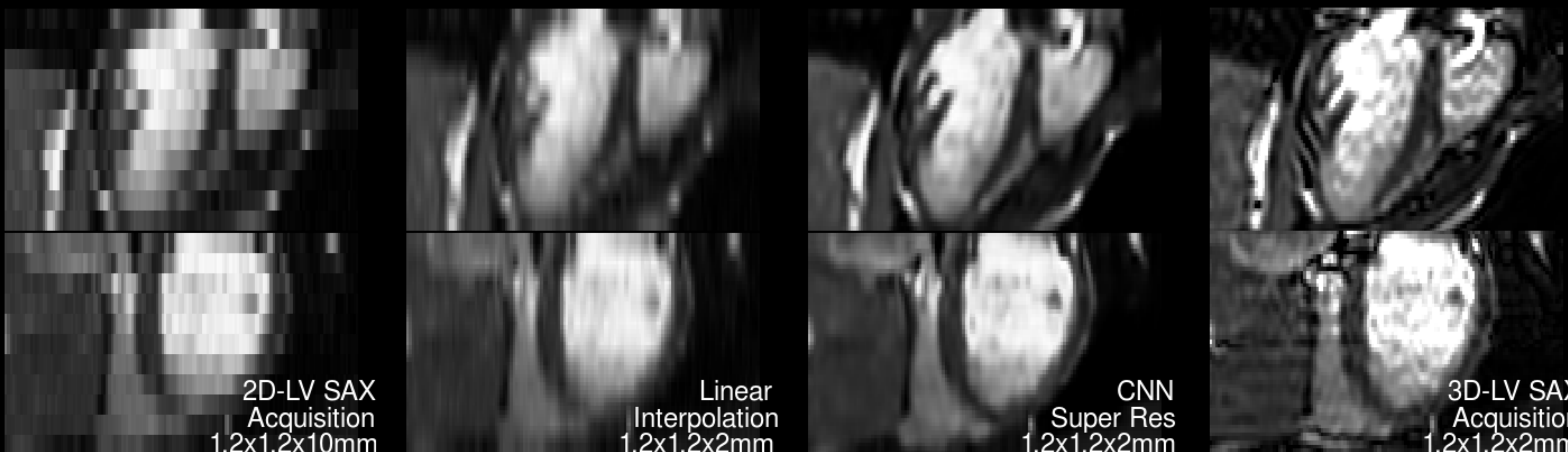
BioMedIA
Imperial College London

Table 1: Quantitative comparison of different image upsampling methods.

Method	PSNR (dB)	SSIM
Linear	20.83±1.10	.70±.03
CSpline	22.38±1.13	.73±.03
MAPM	22.75±1.22	.73±.03
sh-CNN	23.67±1.18	.74±.02
CNN	24.12±1.18	.76±.02
de-CNN	**24.45±1.20**	**.77±.02**

- MAPM: Multi-Atlas Patch Match [Shi et al MICCAI'13]
- sh-CNN: 4 - Layer Network without Deconvolution Layer [Dong TPAMI'15]
- CNN: 7 - Layer Network without Deconvolution Layer
- de-CNN: 7 - Layer Network with Deconvolution Layer

Image Quality Assessment

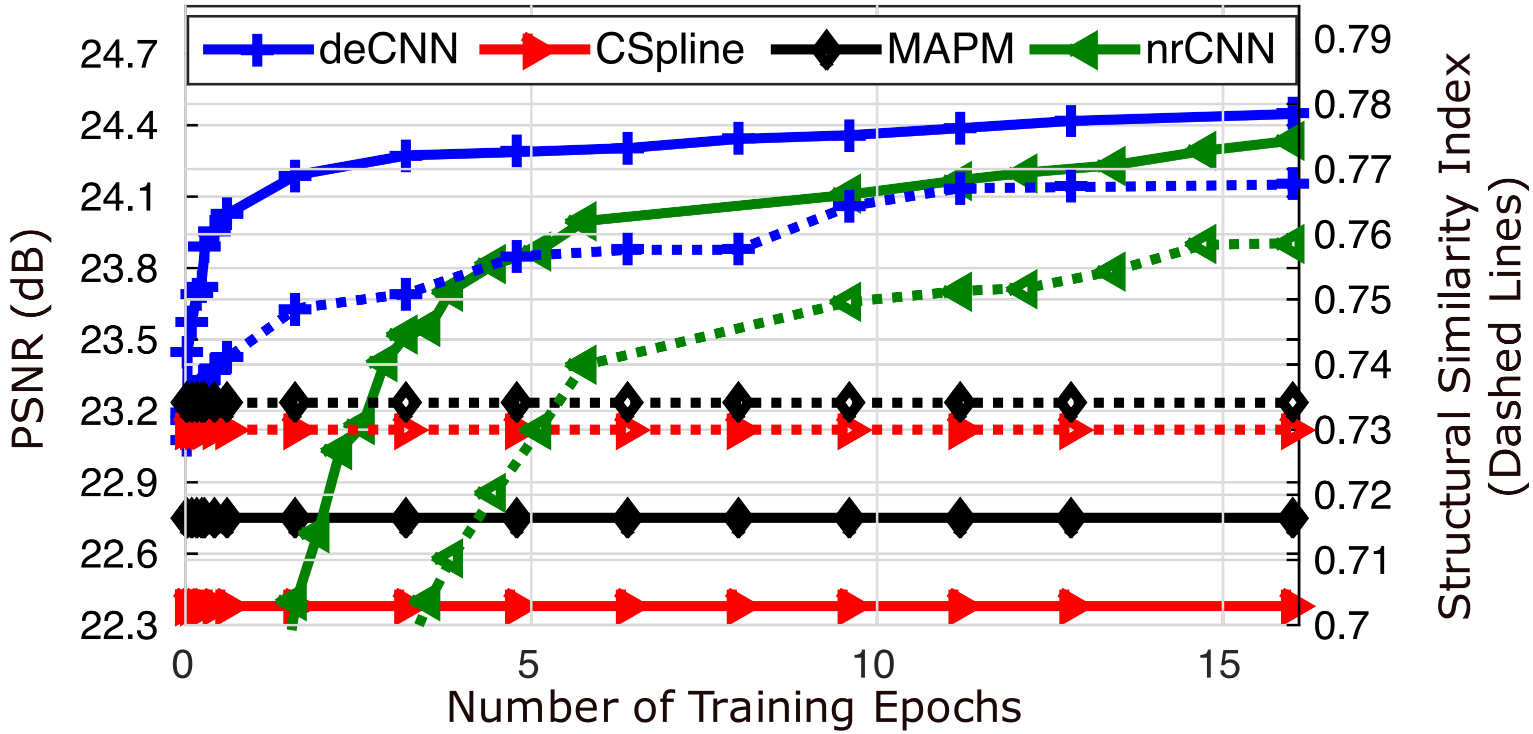


Low Resolution Input Image | Linear Interpolation | The Proposed Method | High Resolution Ground-truth

Upsampling x5

Inference Time: 6-8 Seconds for image size (140x140x10)

Image Quality Assessment



- nr-CNN: 7 - Layer Network without Residual Learning.
- de-CNN: 7 - Layer Network with Residual Learning

Experiments with Multi-Image Model

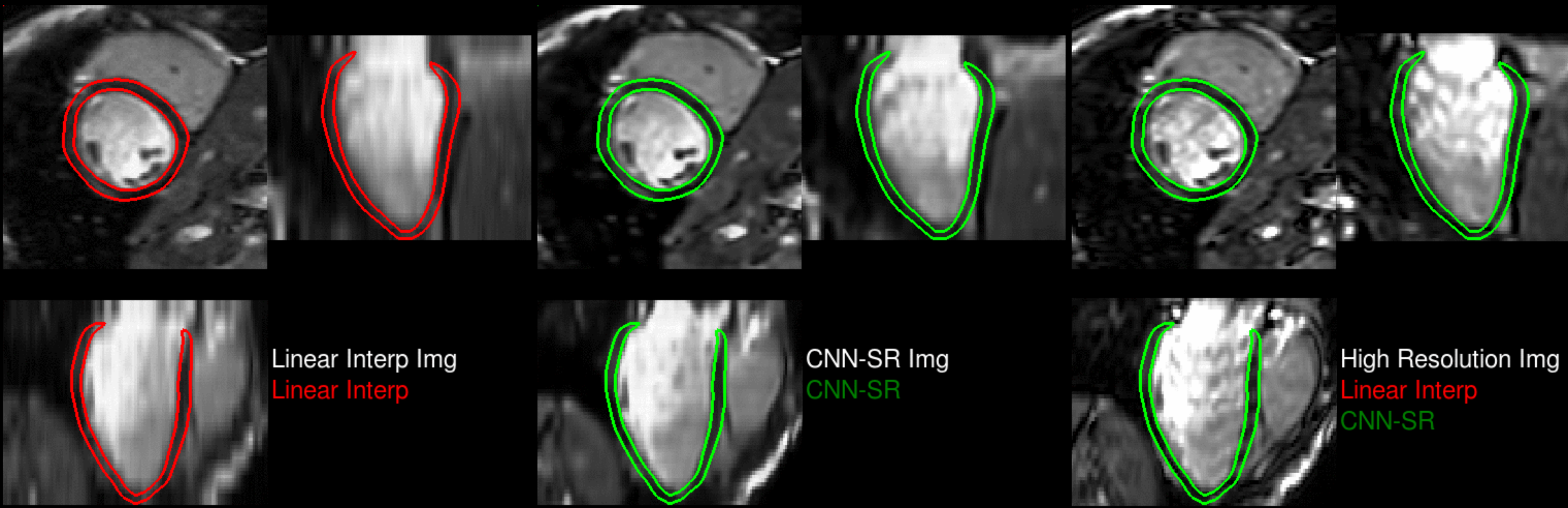
BioMedIA Imperial College London

Table 2: Image quality results obtained with three different models: single-image de-CNN, Siamese, and multi-channel (MC) CNN.

Method	PSNR (dB)	SSIM
de-CNN(SAX)	24.76±0.48	.807±.009
Siamese(SAX/4CH)	25.13±0.48	.814±.013
MC(SAX/4CH)	25.15±0.47	.814±.012
MC(SAX/2/4CH)	**25.26±0.37**	**.818±.012**

- MC (SAX/4CH): Multi-Channel input – SAX and 4 Chamber LAX Images
- MC (SAX/2/4CH): Multi-Channel input – SAX and 2/4 Chamber LAX Images

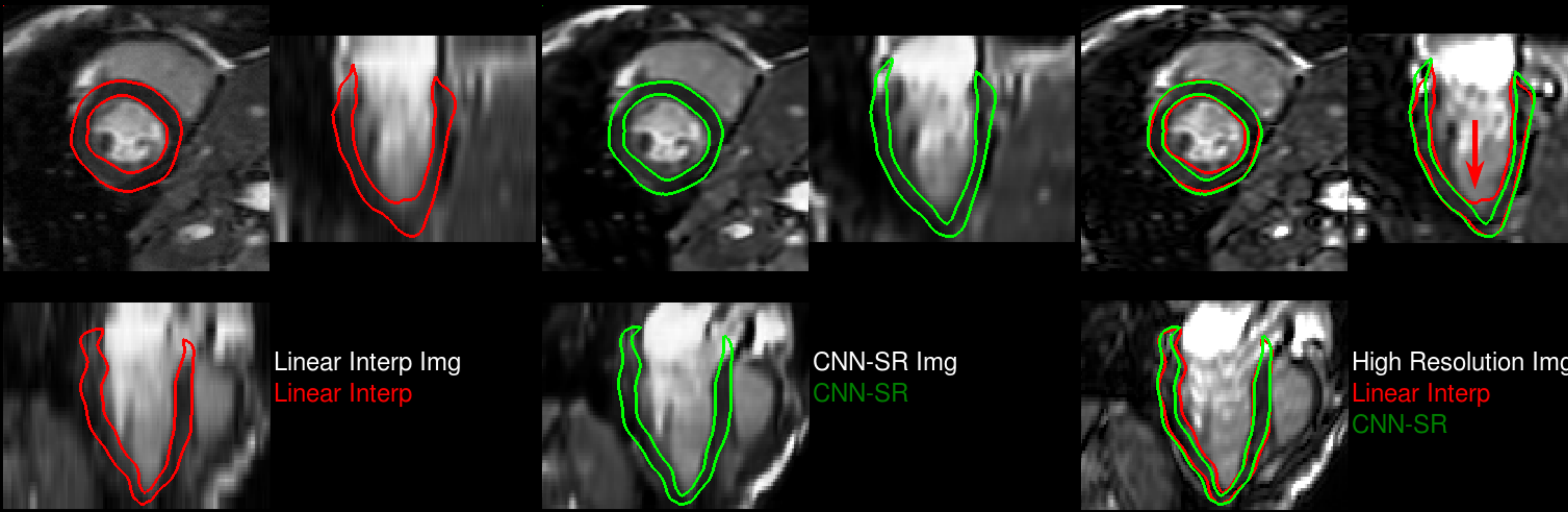
Motion Tracking Experiments (SR is used as a preprocessing method)



Surface to Surface Distance (Linear vs HR) 5.50 mm

Surface to Surface Distance (Proposed vs HR) 4.73 mm

Motion Tracking Experiments (SR is used as a preprocessing method)



LV Segmentation Experiments (SR is used as a preprocessing method)

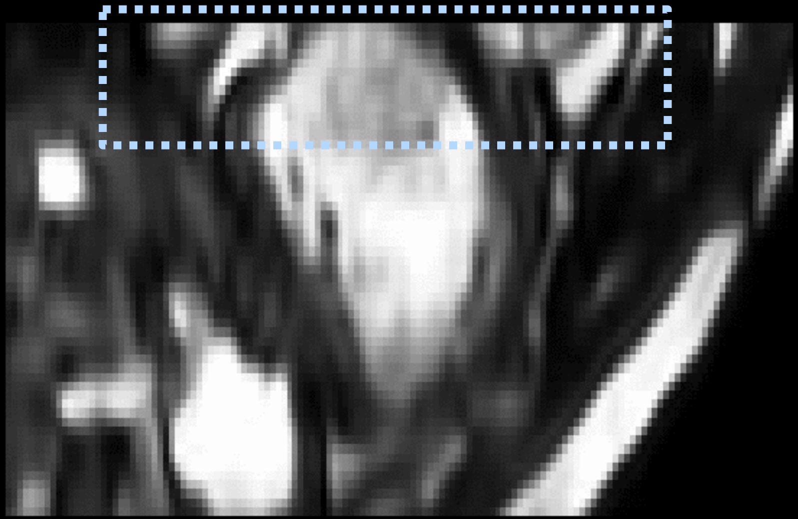
BioMedIA Imperial College London

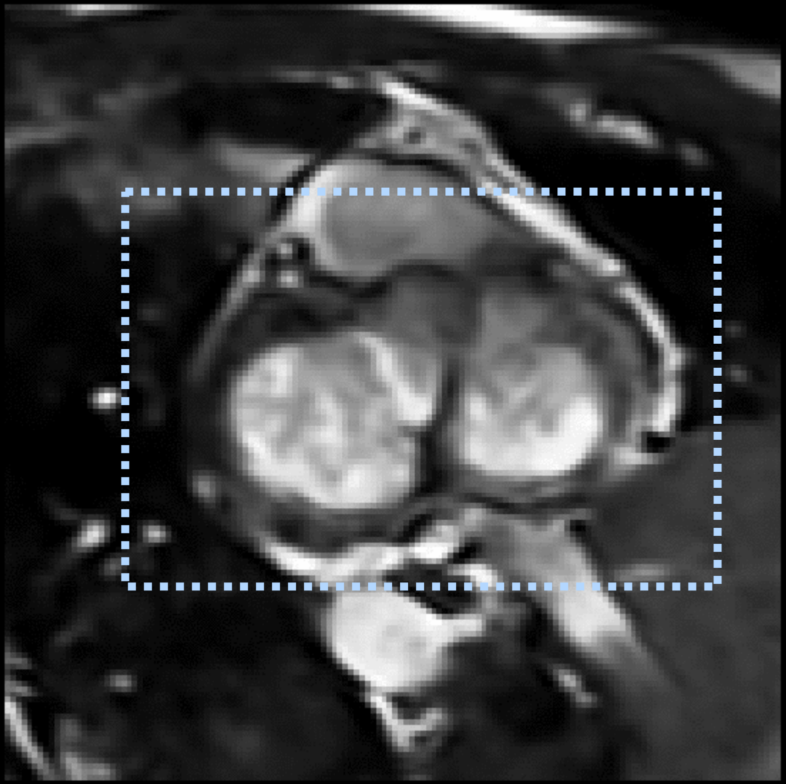
Table 3: Segmentation results for different upsampling methods, CSpline ($p = .007$) and MAPM ($p = .009$). They are compared in terms of mean and Hausdorff distances (MYO) and LV cavity volume differences (w.r.t. manual annotations).

		Linear	CSpline	MAPM	de-CNN	High Res
Exp (c)	LV Vol Diff (ml)	11.72±6.96	10.80±6.46	9.55±5.42	**9.09±5.36**	8.24±5.47
	Mean Dist (mm)	1.49±0.30	1.45±0.29	1.40±0.29	**1.38±0.29**	1.38±0.28
	Haus Dist (mm)	7.74±1.73	7.29±1.63	6.83±1.61	**6.67±1.77**	6.70±1.85

- Multi-Atlas patch based label fusion [Coupe NeuroImage'11] is used to segment images (20 Atlases)

Difference Between Trained and Fixed Deconvolution Kernels





Take Home Messages

BioMedIA
Imperial College London

I. SR as a preprocessing step / Could it replace standard interpolation techniques ?

II. Importance of learning upsampling filters and residual connections in SR models.

III. Models could be trained with combined images and stacks acquired from different directions.

IV. Future work
   a. Other imaging modalities or applications (DTI or MR Image Reconstruction)
   b. Perceptual loss function: Could it be applicable to medical images ?

Imperial College London

Multi-Input Cardiac Image Super-Resolution using Convolutional Neural Networks

Acknowledgments: EPSRC Engineering and Physical Sciences Research Council BioMedIA Imperial College London

Poster Session 1 – Cardiac Image Analysis (CARD) – PS1.40

Additional Details about the SR-CNN Model

Model Training Strategy

BioMedIA
Imperial College London

I. Batch Normalization [Ioffe and Szegedy ICML'15]

- Faster Model Convergence.
- Reduces the dependency of model on filter coefficient initialization.

II. Data Augmentation

- Training data, LR-HR pairs, are generated from 3D-HR Images based on the following model [Shi et al. MICCAI'13]:
- Trained with cine cardiac HR - MR images acquired from 930 healthy adult subjects.

$$\boldsymbol{x} = \boldsymbol{DBSMy} + \boldsymbol{\eta}$$

III. Smooth L1-Norm Function

- Improves the convergence when outliers are observed in training data.

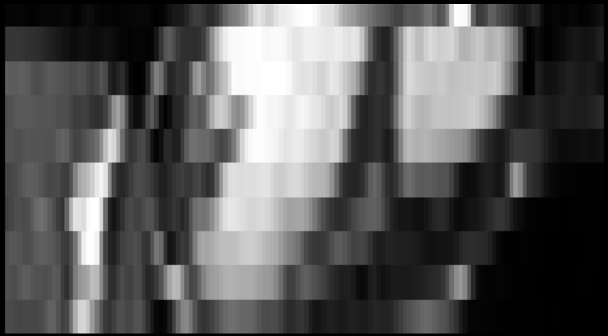
Number of Feature Maps / Atlases

Table 1: Quantitative comparison of different image upsampling methods.

Exp (a)	PSNR (dB)	SSIM	# Filters/Atlases
Linear	20.83±1.10	.70±.03	–
CSpline	22.38±1.13	.73±.03	–
MAPM	22.75±1.22	.73±.03	350
sh-CNN	23.67±1.18	.74±.02	64,64,32,1
CNN	24.12±1.18	.76±.02	64,64,32,16,8,4,1
de-CNN	**24.45±1.20**	**.77±.02**	64,64,32,16,8,4,1

- MAPM: Multi-Atlas Patch Match [Shi et al MICCAI'13]
- sh-CNN: 4 - Layer Network without Deconvolution Layer [Dong TPAMI'15]
- CNN: 7 - Layer Network without Deconvolution Layer
- de-CNN: 7 - Layer Network with Deconvolution Layer

SR-CNN (9-5-5) - ImageNet



Low Resolution
Input Image

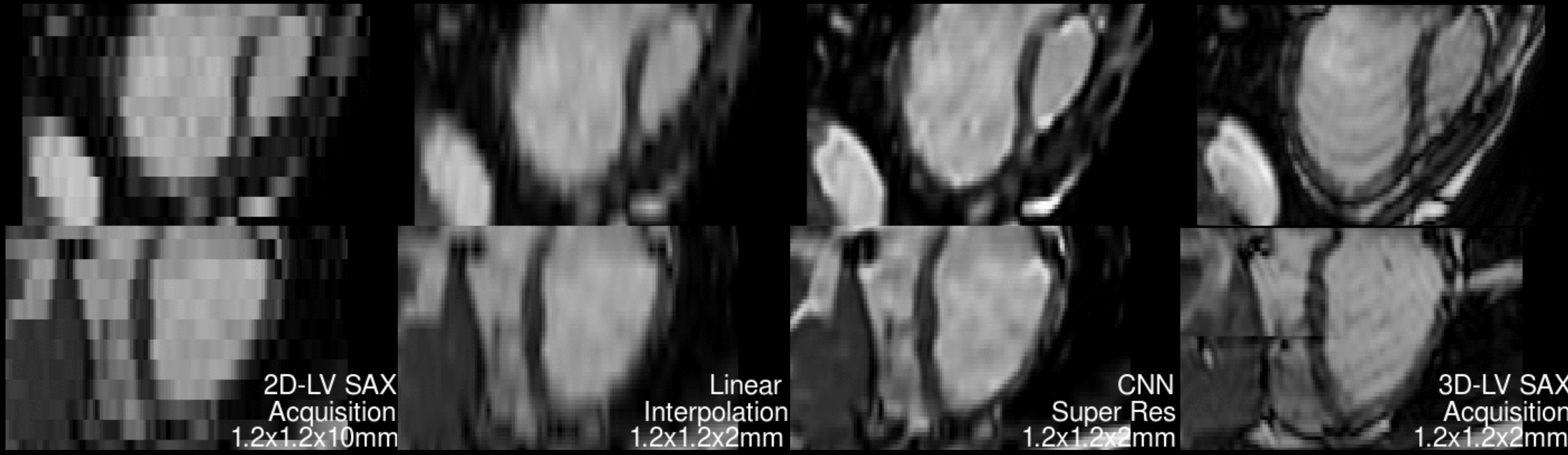
Cubic Spline
Interpolation

SR-CNN (9-5-5) Output
Image

Upsampling x4

A 3-Layer model is trained with ImageNet Dataset

Image Quality Assessment



Low Resolution Input Image	Linear Interpolation	SR-CNN Output Image	High Resolution Ground-truth

Upsampling x5

Inference Time: 6-8 Seconds for image size (140x140x10)