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

Ozan Oktay[1], Wenjia Bai[1], Matthew Lee[1], Ricardo Guerrero[1], Konstantinos Kamnitsas[1], Jose Caballero[3], Antonio de Marvao[2], Stuart Cook[2], Declan O'Regan[2], and Daniel Rueckert[1]

[1] Biomedical Image Analysis Group, Imperial College London, UK
[2] Institute of Clinical Science, Imperial College London, UK [3] Magic Pony Technology, London, UK

## Motivation

3D cardiac MR imaging requires long acquisition time and breath-hold. For this reason, the clinical routine is still dominated by multi-slice 2D imaging, which hamper the visualization of anatomy and quantitative measurements as relatively thick slices are acquired. As a solution, we propose a novel image super-resolution (SR) approach that is based on a residual convolutional neural network model. It reconstructs high resolution 3D volumes from 2D image stacks for more accurate image analysis. Our model allows the use of multiple input data acquired from different viewing planes for improved performance.

## Qualitative Assessment

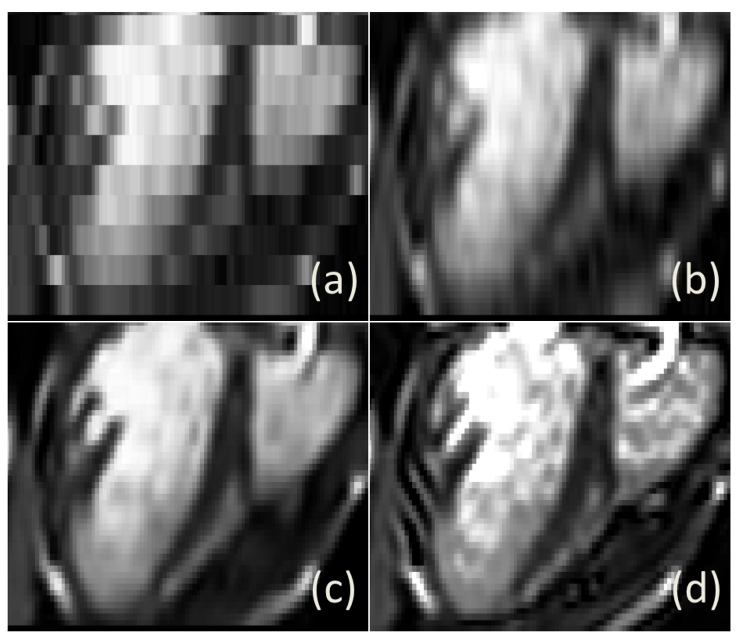


**Figure 3:** LR image (a), cubic-spline interpolation (b), our model (c), and ground-truth HR image (d).

## Residual Learning in SR

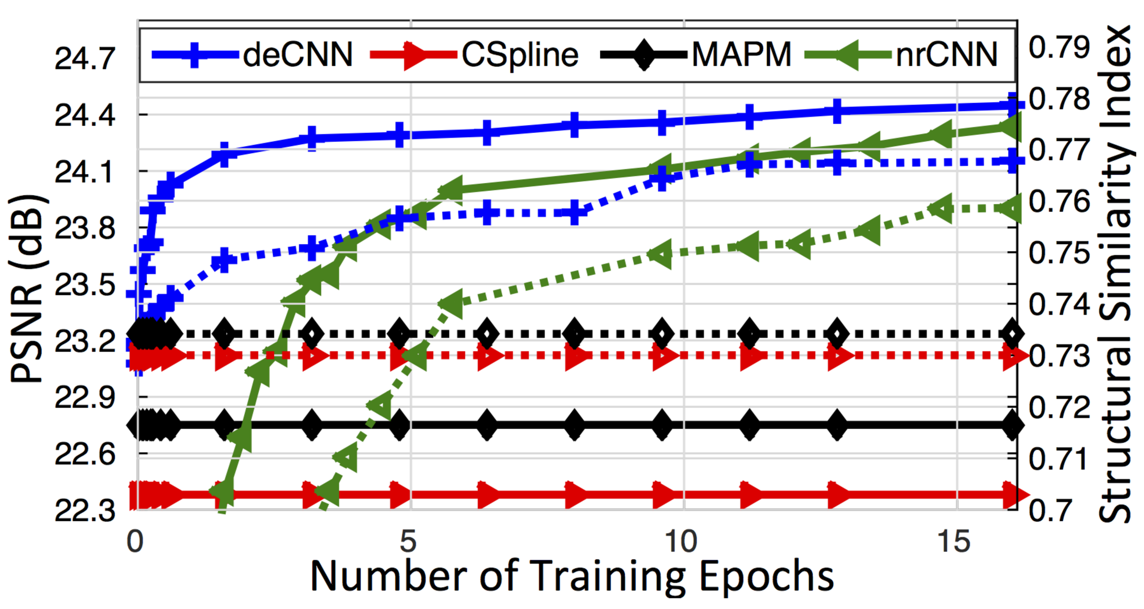


**Figure 4:** CNN based methods can learn better HR models, and it can be seen that the model without the residual learning (nrCNN) underperforms and requires a large number of training iterations.

## Image SR & LV Tracking

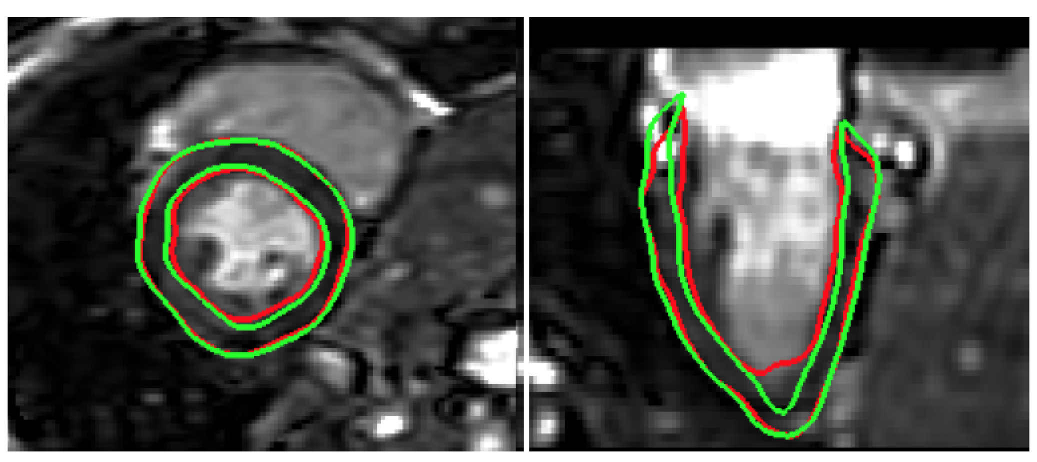

**Figure 5:** Left ventricular (LV) tracking results (end-systole phase) obtained with linearly interpolated (red) and super-resolved images (green).

## References

[1] Dong, C., Loy, C.C., He, K., Tang, X.: Image super-resolution using deep convolutional networks. IEEE PAMI 38(2), 295–307 (2016)

[2] Shi, W., et al.: Cardiac image super-resolution with global correspondence using multi-atlas patchmatch. In: MICCAI (2013)

## Image Super Resolution Problem

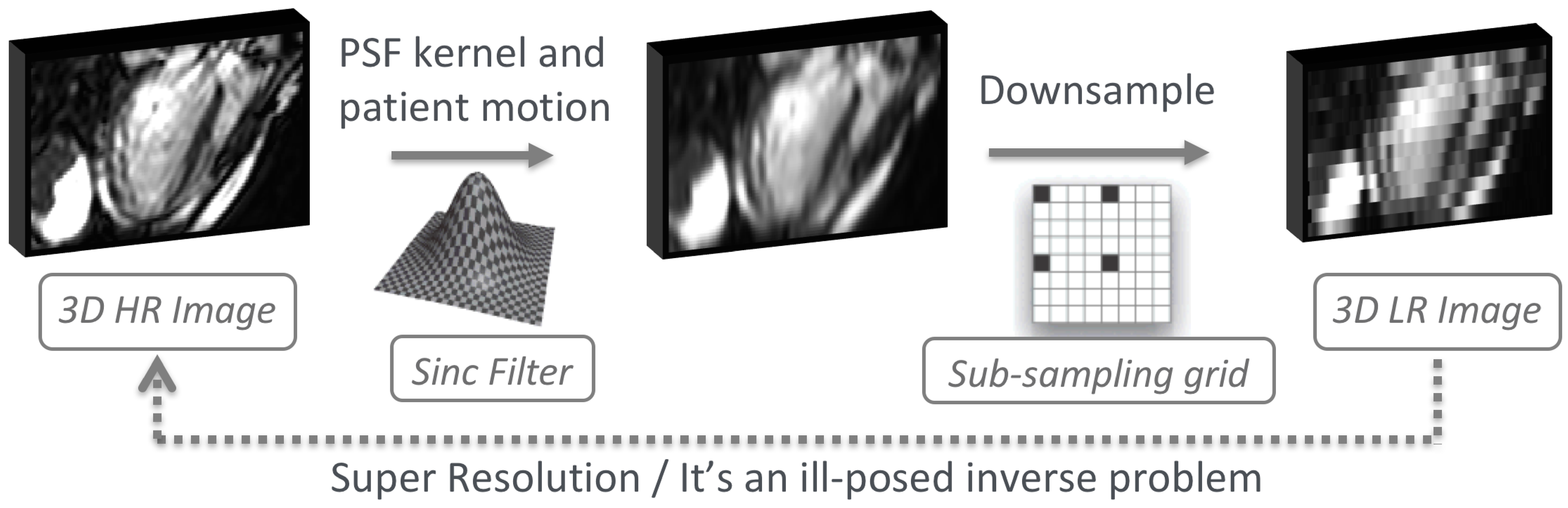


**Figure 1:** Low and high resolution image model

## Proposed Super Resolution Single Input Model

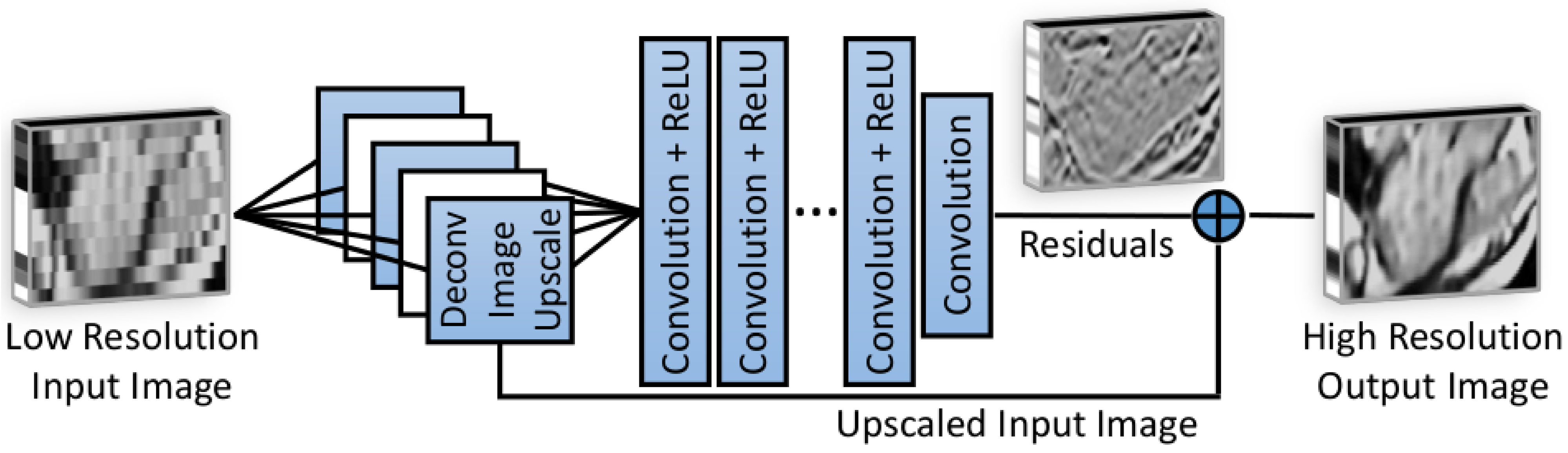


**Figure 2:** The proposed single input residual super-resolution network model

**We extend the SR model in [1] with:**

(I) Residual learning of 3D kernels

(II) End-to-end training of upsampling layer

(III) Multi-input model extension (SAX-LAX)

**The proposed model training details:**

(I) 930 LR-HR training image pairs

(II) Batch-normalization [Ioffe and Szegedy'15]

(III) Smooth $\ell_1$ norm is used as a cost function

## Quantitative Results

**Table 1:** Quantitative comparison of results obtained with different image upsampling methods on 150 testing images

| | 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 |

⇒ Image qualities are evaluated based on peak signal-to-noise ratio (PSNR) and structural similarity index measure (SSIM).

⇒ The proposed method achieves better image quality compared to multi-atlas patch match (MAPM) [2], shallow network with 4 layers (sh-CNN) [1], and network without a deconvolution layer (CNN).

**Table 2:** Quantitative comparison of multi-input network models (A separate testing dataset is used 150 SAX/LAX image pairs)

| | PSNR (dB) | SSIM | $p$ - values |
|---|---|---|---|
| de-CNN (SAX) | 24.76±0.48 | .807±.009 | 0.005 |
| Siamese (SAX/4CH) | 25.13±0.48 | .814±.013 | 0.016 |
| MC (SAX/4CH) | 25.15±0.47 | .814±.012 | 0.017 |
| MC (SAX/2/4CH) | **25.26±0.37** | **.818±.012** | - |

⇒ Multi-input models improve high-resolution image synthesis results. Siamese and multi-channel (MC) input models are compared. 2/4 chamber images (2/4 CH) are processed together with short axis (SAX) stacks.

## Image SR Based Left Ventricle Segmentation

**Table 3:** As a subsequent image analysis, 18 SAX upsampled images are automatically segmented using a state-of-the-art multi-atlas patch fusion method. Those segmentations are compared with the manual annotations performed on HR images. Segmentation results for CSpline ($p = .007$), MAPM ($p = .009$), and deCNN are compared in terms of mean and Hausdorff distances (MYO) and LV cavity volume differences.

| | Linear | CSpline | MAPM | de-CNN | High Res |
|---|---|---|---|---|---|
| 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 |