Deep learning for medical image reconstruction, segmentation and analysis

Daniel Rueckert
Biomedical Image Analysis Group
Department of Computing, Imperial College London, UK
Disclosures

- Co-founder – IXICO
- Adviser – HeartFlow, Circle Cardiovascular Imaging
- Grant funding from:
Deep learning for medical imaging: Opportunities

Value proposition

Level of diagnostic support

Reasoning

Perception

Computing

Diagnosis and Prediction

Screening, Diagnosis and Monitoring

Quantification of Imaging Biomarkers

Semantic Image Interpretation

Image Enhancement

Image Acquisition and Reconstruction

Accelerated imaging

Denoising, Super-resolution, Image fusion

Detection, Localisation, Segmentation

Analysis of shape and texture, Radiomics

Computer-aided decision support systems
MR image acquisition: Challenges

• Magnetic Resonance Imaging (MRI)
  – MRI acquisition is inherently a slow process
  – Slow acquisition is
    • ok for static objects (e.g. brain, bones, etc)
    • problematic for moving objects (e.g. heart, liver, fetus)
  – Options for MRI acquisition:
    • real-time MRI: fast, but 2D and relatively poor image quality
    • gated MRI: fine for period motion, e.g. respiration or cardiac motion but requires gating
      (ECG or navigators) leading to long acquisition times (30-90 min).
Example: Cardiac imaging
Cardiac MRI: Full acquisition is slow

- MRI acquisition is performed in k-space by sequentially traversing sampling trajectories.
Cardiac MRI: Full acquisition is slow

- MRI acquisition is performed in k-space by sequentially traversing sampling trajectories.
Cardiac MRI: Full acquisition is slow

- MRI acquisition is performed in k-space by sequentially traversing sampling trajectories.
Cardiac MRI: Full acquisition is slow

- MRI acquisition is performed in k-space by sequentially traversing sampling trajectories.

There is significant spatio-temporal redundancy.
Deep learning for image reconstruction

- **K-space**
- **Signal space**

**Full sampling** (slow)

\[ F^{-1} \{ \cdot \} \rightarrow X_f \]

Learning-based reconstruction

E.g. compressed sensing
Deep learning for image reconstruction
Deep learning for image reconstruction: How?
Deep learning for image reconstruction

Schlemper et al. IEEE TMI 2018
Deep learning for image reconstruction

Schlemper et al. IEEE TMI 2018
Deep learning for image reconstruction

\[ F(k) = \begin{cases} 
\tilde{x}_{\text{cnn}}(k) & \text{if } k \notin \Omega \\
\tilde{x}_{\text{cnn}}(k) + \lambda \tilde{x}_u(k) / (1 + \lambda) & \text{if } k \in \Omega 
\end{cases} \]

Schlemper et al. IEEE TMI 2018
Magnitude reconstruction (6-fold)

(a) 6x Undersampled
(b) CNN reconstruction
(c) Ground Truth

Schlemper et al. IEEE TMI 2018
Magnitude reconstruction (11-fold)

(a) 11x Undersampled  (b) CNN reconstruction  (c) Ground Truth

Schlemper et al. IEEE TMI 2018
Replacing Deep Cascade of CNNs with Recurrent Neural Networks

(a) Traditional Optimisation

\[ x_{rec}^{(i)} \rightarrow \text{argmin}_z \rightarrow \text{DC} \rightarrow z^{(i)} \]

Replacing Deep Cascade of CNNs with Recurrent Neural Networks

RNNs over both iterations and time:
- Embed the *iterative optimisation* process in a learning setting
- Exploit *spatio-temporal* redundancies

Replacing Deep Cascade of CNNs with Recurrent Neural Networks

Exploiting \( k-t \) Correlations

\( k-t \) NEXT: Exploiting spatio-temporal redundancies in **complementary** domains

**k-t NEtwork with X-f Transform:**
- Exploit \( k-t \) correlations in \( x-f \) domain with CNNs
- Alternate between both \( x-f \) and **image domains** learning complementary information

---

Qin et al. MICCAI, 2019
Exploiting $k$-$t$ Correlations

$k$-$t$ NEXT: Exploiting spatio-temporal redundancies in **complementary** domains

**k-t NEtwork with X-f Transform:**
- Exploit $k$-$t$ correlations in **x-f domain** with CNNs
- Alternate between both **x-f** and **image domains** learning complementary information

---

Qin et al. MICCAI, 2019
Exploiting \( k-t \) Correlations

\( k-t \) NEXT: Exploiting spatio-temporal redundancies in \textit{complementary} domains

\textbf{\( k-t \) NEtwork with X-f Transform:}
- Exploit \( k-t \) correlations in \textit{x-f domain} with CNNs
- Alternate between both \textit{x-f} and \textit{image domains} learning complementary information

\[
\rho^{(n)} = DC(\hat{\rho}_{\text{rec}}^{(n-1)}) + xf - \text{CNN}(\rho_{\text{rec}}^{(n-1)} - \hat{\rho}_{\text{rec}}^{(n-1)})
\]

\[
\sigma_{\text{rec}}^{(n)} = \text{CRNN}\left(\mathcal{F}_f\rho^{(n)};\mathbf{v}^{(0)}\right), \quad \rho_{\text{rec}}^{(n)} = \mathcal{F}_f\sigma_{\text{rec}}^{(n)}
\]

Qin et al. MICCAI, 2019
Exploiting $k$-$t$ Correlations

<table>
<thead>
<tr>
<th></th>
<th>$k$-$t$ FOCUS</th>
<th>CRNN-MRI</th>
<th>$k$-$t$ NEXT</th>
</tr>
</thead>
<tbody>
<tr>
<td>9x</td>
<td>29.52/0.951</td>
<td>32.45/0.969</td>
<td>34.23/0.979</td>
</tr>
<tr>
<td>12x</td>
<td>28.14/0.937</td>
<td>31.30/0.962</td>
<td>33.28/0.975</td>
</tr>
</tbody>
</table>

Qin et al. MICCAI, 2019
Exploiting motion for extremely undersampled dynamic MRI reconstruction

Fully sampled ‘x1’
✗ High time-cost
✓ Accurate reconstruction
Exploiting motion for extremely undersampled dynamic MRI reconstruction

- Fully sampled ‘x1’
- ✗ High time-cost
- ✓ Accurate reconstruction
DC-CNN: Data-Consistent CNN

Zero-filled reconstruction → De-aliasing network → DC: Data consistency → Dynamic MRI Reconstruction

iterate

image space image space k-space image space

Schlemper et. al (2018)
Data-consistent motion-augmented cine (DC-MAC)

Exploiting motion for extremely undersampled dynamic MRI reconstruction
G. Seegoolam et al. MICCAI 2019, ISMRM 2020
ME-CNN: Motion-Exploiting CNN

Exploiting motion for extremely undersampled dynamic MRI reconstruction
G. Seegoolam et al. MICCAI 2019, ISMRM 2020
Data-consistent motion-augmented-cine (DC-MAC) example (x16 acceleration rate)

Exploiting motion for extremely undersampled dynamic MRI reconstruction
G. Seegoolam et al. MICCAI 2019, ISMRM 2020
ME-CNN vs DC-CNN example (x16 acceleration rate)

Exploiting motion for extremely undersampled dynamic MRI reconstruction
G. Seegoolam et al. MICCAI 2019, ISMRM 2020

1 Schlemper et. al (2018)
ME-CNN vs DC-CNN example  
(x51 acceleration rate)

Exploiting motion for extremely undersampled dynamic MRI reconstruction
G. Seegoolam et al. MICCAI 2019, ISMRM 2020

1 Schlemper et. al (2018)
Wide-range comparison of SOTA approaches including
- various data consistency layers
- implicit and explicit coil combination
- loss functions
- model ensembling

Paper:

Code:
https://github.com/khammernik/sigmanet
**Σ-net: Systematic Evaluation of Iterative Deep Neural Networks for Fast Parallel MR Image Reconstruction**

**Architectures**
- *Sensitivity Networks*: Require explicit coil information
- *Parallel Coil Networks*: Learn implicit coil combination
Sensitivity networks required a *style-transfer layer* to match the intensities of the target root-sum-of-squares (RSS) reconstruction. This has NO practical relevance!

Example reconstruction for a fat-saturated PDw image of the fastMRI validation set at R=8.
Image Enhancement vs Σ-net Reconstruction
Deep learning for image super-resolution

• Acquisition of cardiac MRI typically consists of 2D multi-slice data due to
  – constraints on SNR
  – breath-hold time
  – total acquisition time

• This leads to thick slice data (thickness 8-10 mm per slice)
Super-resolution via image-to-image translation
Super-resolution via image-to-image translation

O. Oktay et al. MICCAI 2016
Super-resolution via image-to-image translation

O. Oktay et al. MICCAI 2016
Deep learning for image super-resolution

O. Oktay et al. MICCAI 2016
Deep learning for image super-resolution

What we want

3D HR Image

PSF kernel and patient motion

Sinc Filter

Down-sample

Sub-sampling

What we have

Input

2D LAX Images

Outpt

Image Super Resolution Model

2D SAX Images

What we want

3D HR Image

PSF kernel and patient motion

Sinc Filter

Down-sample

Sub-sampling

What we have

Input

2D LAX Images

Output

Image Super Resolution Model

2D SAX Images
Deep learning for image super-resolution

O. Oktay et al. MICCAI 2016
Deep learning for image segmentation

- Convolution + RELU
- Max pooling
- Transposed convolution
- Softmax
- Skip layers

Bai et al., JCMR 2018
Deep learning for image segmentation

Lavdas et al. 2017, Medical Physics

DeepMedic: K. Kamnitsas et al. Medical Image Analysis, 2017

Bai et al., JCMR 2018
Deep learning for image segmentation

• Fully connected networks (Long et al., 2015)
• Manual annotations of 4,872 subjects (QMUL/Oxford) with 93,128 pixelwise annotated 2D images slices
• Divided into training/validation/test: 3,972/300/600
Deep learning for image segmentation

SA, basal
SA, mid-ventricular
SA, apical
LA, 2 chamber
LA, 4 chamber

Bai et al., JCMR 2018
### Evaluation of segmentation accuracy

##### Comparison to expert observers

#### (a) Absolute difference

<table>
<thead>
<tr>
<th></th>
<th>Auto vs Man (n = 600)</th>
<th>O1 vs O2 (n = 50)</th>
<th>O2 vs O3 (n = 50)</th>
<th>O3 vs O1 (n = 50)</th>
</tr>
</thead>
<tbody>
<tr>
<td>LVEDV (mL)</td>
<td>6.1±5.3</td>
<td>6.1±4.4</td>
<td>8.8±4.8</td>
<td>4.8±3.1</td>
</tr>
<tr>
<td>LVESV (mL)</td>
<td>5.3±4.9</td>
<td>4.1±4.2</td>
<td>6.7±3.1</td>
<td>2.0±1.8</td>
</tr>
<tr>
<td>LVM (gram)</td>
<td>6.9±5.5</td>
<td>4.2±3.2</td>
<td>6.7±3.1</td>
<td>3.0±2.4</td>
</tr>
<tr>
<td>RVEDV (mL)</td>
<td>8.5±7.1</td>
<td>11.1±4.5</td>
<td>8.7±5.8</td>
<td>5.7±3.0</td>
</tr>
<tr>
<td>RVESV (mL)</td>
<td>7.2±6.8</td>
<td>5.7±3.0</td>
<td>11.7±6.9</td>
<td>5.7±3.0</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

#### (b) Relative difference

<table>
<thead>
<tr>
<th></th>
<th>Auto vs Man (n = 600)</th>
<th>O1 vs O2 (n = 50)</th>
<th>O2 vs O3 (n = 50)</th>
<th>O3 vs O1 (n = 50)</th>
</tr>
</thead>
<tbody>
<tr>
<td>LVEDV (%)</td>
<td>4.1±3.5</td>
<td>4.2±3.2</td>
<td>6.3±3.3</td>
<td>3.4±2.2</td>
</tr>
<tr>
<td>LVESV (%)</td>
<td>9.5±9.5</td>
<td>6.8±7.5</td>
<td>12.5±8.5</td>
<td>11.7±5.1</td>
</tr>
<tr>
<td>LVM (%)</td>
<td>8.3±7.6</td>
<td>4.4±3.3</td>
<td>6.0±3.7</td>
<td>6.7±4.6</td>
</tr>
<tr>
<td>RVEDV (%)</td>
<td>5.6±4.6</td>
<td>8.0±5.0</td>
<td>4.2±3.1</td>
<td>5.7±3.6</td>
</tr>
<tr>
<td>RVESV (%)</td>
<td>11.8±12.2</td>
<td>30.6±15.5</td>
<td>10.9±8.3</td>
<td>16.9±9.2</td>
</tr>
</tbody>
</table>

---

**Computer performs as well as different expert observers**

---

W. Bai et al., JCMR 2018
Deep learning for joint image segmentation and motion tracking: Motion-Seg Net

Challenges:
• Segmentation: only ED/ES frames are annotated
• Motion field: no ground truth are available

Beyond supervised learning:
Self-supervised motion estimation + weakly-supervised segmentation

Qin et al. MICCAI, 2018
Deep learning for joint image segmentation and motion tracking: Motion-Seg Net

Qin et al. MICCAI, 2018
Deep learning for joint image segmentation and motion tracking: Motion-Seg Net

• Loss function:
  \[ \mathcal{L} = \mathcal{L}_m + \lambda_1 \mathcal{L}_s + \lambda_2 \mathcal{L}_w, \]

• Motion estimation loss:
  \[ \mathcal{L}_m = \frac{1}{T} \sum_{k=1}^{T} \| I_t - I'_{t+k} \|^2 + \alpha \mathcal{H}(\delta_{x,y} \Delta_{t+k}) + \beta \mathcal{H}(\Delta_{t}) \]

• Segmentation loss on labeled data:
  \[ \mathcal{L}_s = - \sum_{l \in L} y_l \log(f(x_l; \Theta)) \]

• Warped segmentation loss on unlabeled data:
  \[ \mathcal{L}_w = - \sum_{n \in U} y_l \log(f_w(x_n; \Theta)) \]

Qin et al., MICCAI 2018
Cardiac MR image analysis: Is the problem solved?

- Currently, UK Biobank is in the process of acquiring comprehensive multi-modal images of different organ systems from 100,000 participants using a highly standardised protocol, including magnetic resonance (MR) scans of the heart, brain and abdomen, ultrasound scans of carotid arteries, whole body dual-energy X-ray absorptiometry (DXA) scan of bones and joints, retinal photographs and optical coherence tomography (OCT) images. These will add additional imaging phenotypes for understanding the determinants of diseases.

- Deriving quantitative imaging phenotypes at this scale forms a major challenge. Recently, an image analysis pipeline has been developed for UK Biobank brain MR images, which generates about 4,350 imaging phenotypes of brain structure and function for 10,000 subjects. The derived brain imaging phenotypes, along with the breadth of life-style and health information collected by UK Biobank, provide a valuable resource for studying the influence of ageing, progression of neuropathology and identifying early-stage image-based biomarkers for diseases.

- An initial genome-wide association study (GWAS) has been performed, which identified 148 clusters of associations between single nucleotide polymorphisms (SNPs) and brain imaging phenotypes that replicate at $p < 0.05$, providing insights into the genetic architecture relevant to the brain.

- Here we present cardiac and aortic structural and functional imaging phenotypes for 26,893 subjects and demonstrate association studies enabled by these imaging phenotypes at a population level. The phenotypes were derived from UK Biobank CMR images using an automated machine learning-based analysis pipeline, built upon previously proposed cardiac and aortic image segmentation methods using convolutional neural networks. The pipeline evaluates comprehensive imaging phenotypes for the heart and aorta, including global phenotypes of the four cardiac chambers and two aortic sections: the left ventricle (LV), right ventricle (RV), left atrium (LA), right atrium (RA), ascending aorta (AAo) and descending aorta (DAo), as well as regional phenotypes of the LV myocardial wall thickness and strain. We report associations of the cardiac and aortic imaging phenotypes with sex, age and traditional cardiovascular risk factors. We then conducted a first large, population-based phenome-wide association study to relate the cardiac and aortic phenotypes to non-imaging phenotypes. We discovered a wide range of highly significant associations with life style, early-life factors, mental health and cognitive function of the participants.
Cardiac MR image analysis: Is the problem solved?

• Acquisition of cardiac MRI typically consists of 2D multi-slice data due to
  – constraints on SNR
  – breath-hold time
  – total acquisition time
• This leads to thick slice data (thickness 8-10 mm per slice)
• Motion between slices can lead to artefacts
Cardiac MR image analysis:
Is the problem solved?
Conventional CNNs: What we want

- Super-resolution
- Segmentation
CNNs: No explicit use of prior knowledge

- Standard Loss for **segmentation**: Cross-Entropy loss

\[
L_x = - \sum_{i \in S} \sum_{c=1}^{C} \log \left( \frac{e^{f(c,i)}}{\sum_{j} e^{f(j,i)}} \right)
\]

- Standard loss for **super-resolution**: L2 or L1 loss

\[
\sum_{i \in S} \left\| \Phi(x_i, \theta_r) - y_i \right\|^2
\]
Anatomically constrained CNNs

Low-resolution input

Segmentation & Super Resolution Network

High-resolution output

O. Oktay et al. IEEE TMI 2017
T-L networks for representing priors

Figure 6.3: 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.

6.2.2 Convolutional Autoencoder Model and ACNN-Seg

An autoencoder (AE)\footnote{267} is a neural network that aims to learn an intermediate representation from which the original input can be reconstructed. Internally, it has a hidden layer $h$ whose activations represent the input image, often referred as codes. To avoid the AE to directly copy its output, the AE are often designed to be undercomplete so that the size of the code is less than the input dimension as shown in Fig. 6.3. Learning an AE forces the network to capture the most salient features of the training data. The learning procedure minimises a loss function $L_x(y_s, g(f(y_s; \theta_f)))$, where $L_x$ is penalising $g(f(y_s; \theta_f))$ being dissimilar from $y_s$. The functions $g$ and $f$ are usually defined as follows:

\[
L_x(y_s, g(f(y_s; \theta_f))) = \frac{1}{2} \sum_{i,j} (y_{si} - g(f(y_{si}; \theta_f)))^2
\]

where $y_{si}$ is the $i$-th pixel of the $j$-th slice of the input image. The loss function $L_h$ of the predictor network $p(y_r; \theta_p)$ is defined as:

\[
L_h = \frac{1}{2} \sum_{i,j} (y_{ri} - p(y_{ri}; \theta_p))^2
\]

where $y_{ri}$ is the $i$-th pixel of the $j$-th slice of the intensity image $y_r$.

The gradients of the loss functions with respect to the parameters of the networks are computed using backpropagation.

O. Oktay et al. IEEE TMI 2017
Anatomically constrained CNN: Segmentation framework

ACNN-Segmentation Model
- Gradients for Global Loss
- Gradients for Pixel-Level Loss

X-Entropy Loss $L_x$

Prediction $\phi(x)$

Encoder $f(.)$

GT Labels ($y_s$)

Cross-Entropy Loss $L_{he}$

Euclidean Loss $L_{ee}$

O. Oktay et al. IEEE TMI 2017
Anatomically constrained CNN: Segmentation results

Figure 6.7: Segmentation results on two different 2D stack cardiac MR images. The proposed ACNN model is insensitive to slice misalignments as it is anatomically constrained and it makes less errors in basal and apical slices compared to the 2D-FCN approach. The results generated from low resolution image is better correlated with the HR ground-truth annotations (green).

Performance in basal and apical parts of the heart as shown in Fig. 6.7. Previous slice by slice segmentation approaches validated their methods on LR annotations; however, we see that the produced label maps are far from the true underlying ventricular geometry and it can be a limiting factor for the analysis of ventricle morphology. Similar results were obtained in clinical studies [66], which however required HR image acquisition techniques.

(II) The results also show that introduction of shape priors in segmentation models can be useful to tackle false-positive detections and motion-artefacts. As can be seen in the bottom row of Fig. 6.7, without the learnt shape priors, label map predictions are more prone to imaging artefacts. Indeed, it is the main reason why we observe such a large difference in terms of Hausdorff distance. For endocardium labels, on the other hand, the difference in dice score metric is observed to be less due to the larger size of the LV blood pool compared to the myocardium.

Lastly (III), we observe a performance difference between the cascaded AE based segmentation (AE-Seg [217]) and the proposed ACNN-Seg models: the segmentations generated with the former model are strongly regularised due to the second stage AE. It results in reduced Hausdorff distance with marginal statistical significance, but the model overlooks fine details of the myocardium surface since the segmentations are generated only from the coarse level feature-maps. More importantly, cascaded approaches add additional computational complexity...
Figure 6.5: Training scheme of the proposed anatomically constrained convolutional neural network (ACNN) for image super-resolution task. The predictor part of the proposed T-L network is used as a regularisation model to enforce the model predictions to follow the distribution of the learned low dimensional representations or priors.

Training details are further discussed in Section 6.3.2. It is important to note that the T-L regulariser model is used only at training time but not during inference; in other words, the fully convolutional (FCN) segmentation and super-resolution models can still be used for applications using different image sizes. In this paper, the proposed SR model is referred to as ACNN-SR and its training scheme is shown in Figure 6.5.
Anatomically constrained CNN: Super-resolution results

Original LR image  Baseline SR approach  Anatomically constrained SR model  Ground-truth HR image
Traditional medical imaging

- Serial process with no interaction between different components of imaging pipeline
- Limited ability for adjustment of upstream imaging pipeline based on downstream requirements
- Stages of imaging pipeline not optimized for clinical endpoint
AI-enabled medical imaging

- Close coupling of acquisition, reconstruction, analysis and interpretation
- Feedback and interaction between components of imaging pipeline
- Optimization of whole imaging pipeline with respect to clinical endpoint
AI-enabled medical imaging

Do we need images at all?
AI-enabled medical imaging: Example

Ground truth

J. Schlemper et al.
MICCAI 2018
Acknowledgements

Ghalib Bello
Tim Dawes
Antonio de Marvao
Declan O'Regan
Stuart Cook
Hideaki Suzuki
Paul M. Matthews
Aaron M. Lee
Nay Aung
Elena Lukaschuk
Mihir M. Sanghvi
Filip Zemrak
Kenneth Fung
Jose Miguel Paiva
Valentina Carapella
Young Jin Kim
Steffen E. Petersen

Stefan K. Piechnik
Stefan Neubauer
Aurelien Bustin
Giulia Ginami
Gastao Cruz
Teresa Correia
Tevfik F. Ismail
Imran Rashid
Radhouene Neji
Claudia Prieto
Rene M. Botnar
Jo Hajnal
Alberto Gomez
Veronika Zimmer
Jacqueline Matthew
David Lloyd
Reza Razavi

This research has been conducted mainly using the UKBB Resource under Application Number 2946. The initial stage of the research was conducted using the UKBB Resource under Application Number 18545.