Learning Anatomical Image Representations for Cardiac Imaging

Ozan Oktay, PhD
Research Scientist @HeartFlow Inc. (CA/USA)
Research Fellow @Imperial College London

Deep Learning for Medical Imaging School
Lyon, April 16th, 2019
Research Background

- Siemens Corporate Research (NJ, USA) (2012-13)
  - Visiting Researcher (Thesis Project)

- Imperial College London (BioMedIA, London, UK)
  - Research Associate (2017 - 2018)
  - Honorary Research Fellow (2018 - Present)

- HeartFlow Inc. (CA, USA)
  - Imaging Research Scientist - Scientific Lead (2018 - Present)

- Specialisation:
  - Design and Development of Innovative Machine Learning Solutions for Improved Healthcare Services
Agenda

Machine Learning in Medical Image Analysis

Representation Learning in Medical Image Analysis

Inverse Problems and Image Quality Assessment

Building Clinical Solutions (HeartFlow Inc.)
Machine Learning in Medical Image Analysis

Stanford algorithm can diagnose pneumonia better than radiologists

Stanford researchers have developed a deep learning algorithm that evaluates chest X-rays for signs of disease. In just over a month of development, their algorithm outperformed expert radiologists at diagnosing pneumonia.

Deep Neural Networks Improve Radiologists’ Performance in Breast Cancer Screening

Machine Learning and AI: Are they the same?

1. What is the difference between AI and Machine Learning?

2. Do neural networks generalise well to unseen data points (e.g. scanner types)? What do these models actually learn?

3. Why do neural networks require large amount of data?

4. Can we fully rely on these algorithms in clinical practice?
Where do we stand?

Machine Learning (ML) models are good at:

- Automating well-defined and constrained (low variation) tasks.
- Reproducible results and fast predictions \(\rightarrow\) Scale up to thousands of images
- Can display performance close to the average annotator for some tasks \(\rightarrow\) Label quality

They are a good candidate for assistive clinical workflows as long as they are monitored.
Segmentation of Cardiac Images

**Challenges**
- Manual delineation of large number of images can be time-consuming.
- Reproducibility and large inter-observer variability are common issues.

**Objectives**
- Machine learning algorithm can automate and produce average annotator performance on some tasks.
- It can scale up to thousands of images and assist clinical workflow.
CNN Segmentation Model

Training dataset: 5000 Cases from the UKBB Dataset.
8 different annotators extracted manual segmentations for these images

[W. Bai et al. JCMR 2018]
Auto-Generated Image Segmentations

**Fig. 3** Illustration of the segmentation results for short-axis and long-axis images. The top row shows the automated segmentation, whereas the bottom row shows the manual segmentation. The automated method segments all the time frames. However, only end-diastolic (ED) and end-systolic (ES) frames are shown, as manual analysis only annotates ED and ES frames. The cardiac chambers are represented by different colours.  

- **a** short-axis.  
- **b** long-axis (2 chamber view).  
- **c** long-axis (4 chamber view)
## Variability Between the Annotators and Neural Network Model

<table>
<thead>
<tr>
<th>(a) Dice metric</th>
<th>Auto vs Manual</th>
<th>O1 vs O2</th>
<th>O2 vs O3</th>
<th>O3 vs O1</th>
</tr>
</thead>
<tbody>
<tr>
<td>(n = 600)</td>
<td>(n = 50)</td>
<td>(n = 50)</td>
<td>(n = 50)</td>
<td></td>
</tr>
<tr>
<td>LV cavity</td>
<td>0.94 (0.04)</td>
<td>0.94 (0.04)</td>
<td>0.92 (0.04)</td>
<td>0.93 (0.04)</td>
</tr>
<tr>
<td>LV myocardium</td>
<td>0.88 (0.03)</td>
<td>0.88 (0.02)</td>
<td>0.87 (0.03)</td>
<td>0.88 (0.02)</td>
</tr>
<tr>
<td>RV cavity</td>
<td>0.90 (0.05)</td>
<td>0.87 (0.06)</td>
<td>0.88 (0.05)</td>
<td>0.89 (0.05)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>(b) Mean contour distance (mm)</th>
<th>LV cavity</th>
<th>LV myocardium</th>
<th>RV cavity</th>
</tr>
</thead>
<tbody>
<tr>
<td>(n = 600)</td>
<td>1.04 (0.35)</td>
<td>1.14 (0.40)</td>
<td>1.78 (0.70)</td>
</tr>
<tr>
<td>(n = 50)</td>
<td>1.00 (0.25)</td>
<td>1.16 (0.34)</td>
<td>2.00 (0.79)</td>
</tr>
<tr>
<td>(n = 50)</td>
<td>1.30 (0.37)</td>
<td>1.19 (0.25)</td>
<td>1.78 (0.45)</td>
</tr>
<tr>
<td>(n = 50)</td>
<td>1.21 (0.48)</td>
<td>1.21 (0.36)</td>
<td>1.87 (0.74)</td>
</tr>
</tbody>
</table>
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

MICCAI’16 Conference, October 2016, Athens
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.
Low and High Resolution Images

PSF kernel and patient motion

Down-sample

Sinc Filter

Sub-sampling grid

3D HR Image

Output

Input

Image Super Resolution Model

2D LAX Images

+ 3D LR Image
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
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)
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)
Anatomically Constrained Convolutional Neural Networks (ACNN): Application to Image Enhancement and Segmentation

Ozan Oktay, Enzo Ferrante, Konstantinos Kamnitsas, Wenjia Bai, Jose Caballero, Mattias Heinrich, Stuart Cook, Antonio de Marvao, Declan O'Regan, Bernhard Kainz, Ben Glocker, and Daniel Rueckert

IEEE TMI, August 2017
Analysis of Neural Networks

I. Model parameterization
II. Model capacity / receptive field
III. Loss function / objective

Standard Loss Functions

I. X-Entropy loss function

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

II. L2 or Smooth L1 loss function

\[ \sum_{i \in S} \| \Phi(x_i, \theta_r) - y_i \|^2 \]
Research Objective

1. Can we teach our models the underlying anatomical priors (eg shape)?
2. A new global training objective to teach CNN models
Standard Auto-Encoder Model
The proposed T-L Network
Proposed ACNN - Segmentation Model

Input Image ($x$) $\rightarrow$ Segmentation $\phi(\cdot)$ $\rightarrow$ Prediction $\phi(x)$ $\rightarrow$ Encoder $f(\cdot)$

ACNN-Segmentation Model

- Green arrows: Gradients for Global Loss
- Red arrows: Gradients for Pixel-Level Loss

$\frac{\partial L_{he}}{\partial \theta_s}$

$\frac{\partial L_x}{\partial \theta_s}$

X-Entropy Loss $L_x$

GT Labels ($y_s$)

Euclidean Loss $L_{he}$
Proposed ACNN – Super Res Model

ACNN-Super Resolution Model

- Green: Gradients for Global Loss
- Red: Gradients for Pixel-Level Loss

Input Image \( x \)

Super Res \( \Phi(\cdot) \)

Prediction \( \Phi(x) \)

GT Image \( y_r \)

Smooth L1 Loss \( \psi(\cdot) \)

Predictor \( p(\cdot) \)

Euclidean Loss \( L_{hp} \)
Cardiac MR Super-Resolution Experiments

Input Low Resolution Image
Cardiac MR Super-Resolution Experiments

CNN Super-Resolution Trained with Motion-Augmentation
Cardiac MR Super-Resolution Experiments

ACNN-SR (w shape model)
3D-US Segmentations Results

mean dis: 2.94
haus dis: 14.34
dice scr: 0.89
3D-US Segmentations Results

mean dis: 1.74
haus dis: 6.33
dice scr: 0.9
Learned Hidden Representations

The Proposed Regularization Model

Histogram of the Learned Codes
Learned Hidden Representations

Learned representations can be used to:

I. Predict Clinical Indices (Age, Blood Pressure, Myocardial Mass, etc.)
II. Genetic Studies / Understanding the cardiac related pathologies

PCA Codes vs T-L Codes

I. Pathology classification
   » Healthy Subjects
   » Dilated Cardiomyopathy
   » Hypertrophic Cardiomyopathy

II. Classification accuracy
   » PCA: 83.3%
   » T-L: 91.6%
   » 60 CMR Sequences
Learning Based Quality Control for Cardiac MR Images

Giacomo Tarroni, Ozan Oktay, Wenjia Bai, Andreas Schuh, Hideaki Suzuki, Jonathan Passerat-Palmbach, Antonio de Marvao, Declan P. O’Regan, Stuart Cook, Ben Glocker, Paul M. Matthews, Daniel Rueckert

IEEE TMI, November 2018
Automated MR Image Quality Assessment

Image Quality Issues

- Affecting short-axis cardiac MR image acquisition.

- Fully-automated quality control pipeline for cardiac MRI, tested on 3000 cases from the UK Biobank study.
Automated MR Image Quality Assessment

Decision tree model automatically quantifies:

- Inter-slice misalignment.
- Heart-coverage rate.
- Image contrast.

[G. Tarroni, O. Oktay et al. IEEE TMI 2018]
Structured Decision Forests For Multi-modal Ultrasound Image Registration

Ozan Oktay, Andreas Schuh, Martin Rajchl, Kevin Keraudren, Alberto Gomez, Mattias Heinrich, Graeme Penney, and Daniel Rueckert

MICCAI’15 Conference, October 2015, Munich
Image Guided Cardiac Interventions

Pre-Operative Stage CT and MR Image Acquisitions

Spatially Aligned Pre and Intra Operative Images
Advantages of Probabilistic Edge Maps

I. Modality independent (e.g. CT, MRI, US)

II. Computationally efficient (20s per image)

III. Target organ specific image registration

IV. Accurate and smooth anatomical representation

V. Same training and testing configuration is applied to all three modalities.
Structured Decision Forest

Input Image

Structured Decision Tree

Input Space: $x_i \in \mathcal{X}$

Image features

- Each voxel is voted for $N_t \times (M_e)^3$
- $N_t$ is the number of trees.
- All the votes are aggregated by averaging.

$\psi_1, \theta_1$

$\psi_2, \theta_2$

$\mathcal{L}_1$

$\mathcal{L}_2$

$\mathcal{L}_3$

Output Space: $y_i \in \mathcal{Y}$

$M_e$

Output edge patch labels

Dollar et al.: “Structured forests for fast edge detection.” ICCV 2013
## Proposed Multi-Modal Registration Framework

<table>
<thead>
<tr>
<th>Input cardiac images</th>
<th>PEM representation</th>
<th>Initial Alignment of the images</th>
<th>Global alignment with robust block matching [2]</th>
<th>B-spline FFD based non-rigid registration [1]</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image1.png" alt="PEM-CT Image" /></td>
<td><img src="image2.png" alt="PEM US Image" /></td>
<td><img src="image3.png" alt="Initial Alignment Image" /></td>
<td><img src="image4.png" alt="Global Alignment Image" /></td>
<td><img src="image5.png" alt="B-spline FFD Image" /></td>
</tr>
</tbody>
</table>

### Computation Time

- **Input cardiac images**: ~20s per image
- **PEM-CT**: ~21s per image
- **PEM-US**: ~73s per image

---


Ourselin et al.: “Reconstructing a 3D structure from serial histological sections.” Image and Vision Computing ’01
US/CT & US/MR Image Alignment
3DUS Image Registration with PEMs

- Initial alignment
  - atlas image (in green)
  - # Atlases = M₁

- Affine Registration with Block Matching
  - # Atlases = M₂ < M₁

- Deformable registration and globally weighted label fusion
Endocardial Surface Distance Errors

Surface to Surface Distance Errors (30 Subjects both ED & ES Frames)

- Mean Surface Distance Error
- Hausdorff Surface Distance Error

- Model based Level-Set: 2.61 mm
- Graph-Cut: 2.5 mm
- B-Spline Active Contour: 2.35 mm
- Standard Multi-Atlas: 2.33 mm
- PEM Multi-Atlas: 2.08 mm

Inter-Observer Manual Segmentation Error is 1.01 mm (mean) and 3.34 mm (Hausdorff)
Stratified Decision Forests for Accurate Anatomical Landmark Localization

Ozan Oktay, Wenjia Bai, Ricardo Guerrero, Martin Rajchl, Antonio de Marvao, Declan Regan, Stuart Cook, Mattias Heinrich, Ben Glocker, and Daniel Rueckert

IEEE TMI, September 2016
I. Anatomical landmark localization

II. Extracted boundaries regress the location of each landmark point

III. Similar approaches can be easily formulated with CNN models (e.g. YOLO v2)
Structured Regression Forest

Attention Gated Networks: Learning to Leverage Salient Regions in Medical Images

Ozan Oktay, Jo Schlemper, Michiel Schaap, Mattias Heinrich, Bernhard Kainz, Ben Glocker, Daniel Rueckert

Medical Image Analysis Journal, Jan 2019
Cascaded Models in Image Analysis

Cascaded models:
- Strategy: First localise then classify.
- GPU memory constraints.
- Solving simpler problems.
- Additional context information from preceding models.

Potential Drawbacks:
- Parameter & computation redundancy
- Multiple training schemes might be required
Attention Gates in CNN Models
Proposed Soft-Attention Gates

\[
\alpha = \sigma_2(\psi^T x) = \begin{cases} 
    \sigma_2(x^T g + W_g^T x + W_g^T g + b_g) + b_\psi, & \text{concat} \\
    \psi^T \tanh(W_c [x; g]), & \text{general}
\end{cases}
\]

Concatenation (additive) based attention function
Final activation function has an influence on the training behaviour
Attention Coefficients
Across Different Training Epochs

Input Image
Epoch-3
Epoch-6
Epoch-10
Epoch-60
Epoch-150

Attention coefficients across different training epochs (kidneys, spleen, pancreas)
Attention Gates in Image Classification

Input

Conv3x3 + ReLU  Max-Pool /2
Attention Unit  Global Avg. Pooling

Aggregation

prediction
Attention Maps at Different Scales

Adaptive pooling of feature maps with attention gates instead of using global aggregation
Acknowledgments

Thank you, Questions?