Deep Learning in CT

Marc Kachelrieß
German Cancer Research Center (DKFZ)
Heidelberg, Germany
www.dkfz.de/ct
Aim

To give a coarse and critical overview of deep learning applications in CT image formation.

Conventional image post processing applications, such as image segmentation, image registration, image classification etc. as well as CAD applications are not part of this lecture.


There is a nice special issue on machine learning for image reconstruction: IEEE TMI 37(6), 2018

Conventional image post processing applications, such as image segmentation, image registration, image classification etc. as well as CAD applications are not part of this lecture.
Categories of Deep Learning Used in CT Image Formation so Far

- Replacement of missing data
  - LowRes → HighRes nice images
  - SparseView → FullView nice images
  - LowDose → HighDose nice images
  - LimitedAngle → FullAngle nice images
  - ...

- Replacement of lengthy computations
  - Reconstruction (learn denoisers, learn regularizers, learn iterations, ...)
  - Scatter estimation
  - Dose estimation
  - ...

- Other
  - Material decomposition
  - Pseudo CT from MR
  - Motion artifact recognition
  - 3D DSA from a contrast scan
  - Tomosynthesis
  - ...
**Fully Connected Neural Network**

- Each layer fully connects to previous layer
- Difficult to train (many parameters in $W$ and $b$)
- Spatial relations not necessarily preserved

$y(x) = f(W \cdot x + b)$ with $f(x) = (f(x_1), f(x_2), \ldots)$ with $f(x) = x \vee 0 = \text{ReLU}$
Convolutional Neural Network (CNN)

- Replace dense $W$ in $y(x) = f(W \cdot x + b)$ by a sparse matrix $W$ with sparsity being of convolutional type.
- CNNs consist (mainly) of convolutional layers.
- Convolutional layers are not fully connected.
- Convolutional layers are connected by small, say $3 \times 3$, convolution kernels whose entries need to be found by training.
- CNNs preserve spatial relations to some extent.

$$D_{i,j,g} = \sum_{f} S_{i,j,f} \ast K_{i,j,f}^{g} = \sum_{a,b,f} S_{i-a,j-b,f} K_{a,b,f}^{g}$$

Attention: No convolution in depth direction!
U-Net

**Input:**
- 384 x 256 x 4

**Concatenative skip connection**
- 192 x 128 x 40
- 96 x 64 x 80
- 48 x 32 x 160
- 24 x 16 x 320
- 12 x 8 x 480
- 6 x 4 x 960

**Output:**
- 12 x 8 x 480
- 24 x 16 x 320
- 48 x 32 x 160
- 96 x 64 x 80
- 192 x 128 x 40
- 384 x 256 x 4

- 6 x 4 x 960

**Operations:**
- 3 x 3 Convolution, ReLU
- 1 x 1 Convolution, ReLU
- 2 x 2 Max. Pooling
- 2 x 2 Upsampling
- Depth Concatenate
Part 1:
Replacement of Missing Data
Limited Angle Example

GT  FBP (150°)  CNN

MAR Example

• Deep CNN-driven patch-based combination of the advantages of several MAR methods trained on simulated artifacts

• followed by segmentation into tissue classes
• followed by forward projection of the CNN prior and replacement of metal areas of the original sinogram
• followed by reconstruction

(a) Reference Image

(b) Original Image

(c) BHC

(d) LI

(e) NMAR1

(f) NMAR2

(g) CNN Image

(h) CNN-MAR

= input feature 1

= input feature 2

= input feature 3

= output

= proposed method
MAR without Machine Learning: Frequency Split Normalized MAR$^{1,2}$

Patient with bilateral hip prosthesis, Somatom Definition Flash, (C=40/W=500).

Resolution Improvement Example

- 2D U-net to converts 5 mm thick images into 1 mm ones.
- E.g. to “replace a scanning protocol for a 1 mm slice with a 5 mm protocol”

“The CNN generates output images that are virtually equivalent to the ground truth.”?

No!

Sparse View Reconstruction Example

Very impressive, but...

Very impressive, but...

Very impressive, but...
Sparse CT Recon with Data Consistency Layers (DCLs)

Noise Removal Example 1

- 3-Layer CNN uses low dose and corresponding normal dose image patches for training

![Normal dose](image1)
![Low dose](image2)
![ASD-POCS](image3)

![KSVD](image4)
![BM3D](image5)
![3-Layer CNN](image6)

• Architecture based on state-of-the-art networks for image classification (ResNet).
• 32 conv layers with skip connections
• About 2 million tunable parameters in total
• Input is arbitrarily-size stack of images, with a fixed number of adjacent slices in the channel/feature dimension.
Noise Removal Example 2

Low dose images (1/4 of full dose)

Noise Removal Example 2

Denoised low dose

Noise Removal Example 2

Full dose

Noise Removal Example 2

Denoised full dose

Part 2:
Replacement of Lengthy Computations
X-ray scatter is a major cause of image quality degradation in CT and CBCT. Appropriate scatter correction is crucial to maintain the diagnostic value of the CT examination.

CT image scatter
Primary intensity
CT reconstruction
CT reconstruction
C = 0 HU, W = 800 HU
Scatter Correction

Scatter suppression
• Anti-scatter grids
• Collimators
• ...

Scatter estimation
• Monte Carlo simulation
• Kernel-based approaches
• Boltzmann transport
• Primary modulation
• Beam blockers
• ...

Measured intensity - Scatter estimate
Monte Carlo Scatter Estimation

- Simulation of photon trajectories according to physical interaction probabilities.
- Simulating a large number of photon trajectories well approximates the actual scatter distribution.

1 to 10 hours per tomographic data set
Deep Scatter Estimation (DSE)

Train a deep convolutional neural network (CNN) to estimate scatter using a function of the acquired projection data as input.

Input: $T(p)$

Scatter estimate

Convolutional neural network

0.1 to 1 minute per tomographic data set

Deep Scatter Estimation
Network architecture & scatter estimation framework

Input:
384 × 256 × 4

Projection data

Downsampling and application of operator $T(p)$

192 × 128 × 40

3 x 3 Convolution, ReLU

96 × 64 × 80

1 x 1 Convolution, ReLU

48 × 32 × 160

2 x 2 Max. Pooling

24 × 16 × 320

2 x 2 Upsampling

12 × 8 × 480

Depth Concatenate

6 × 4 × 960

3 x 3 Convolution, ReLU

Output: scatter estimate

12 × 8 × 480

2 x 2 Max. Pooling

6 × 4 × 960

Upsampling to original size

Training the DSE Network

CBCT Setup

Primary intensity

+ MC scatter simulation

+ Poisson noise

→ Input

Desired output

- Simulation of 12000 flat detector projection using data of different heads.
- Simulate different tube voltages.
- Splitting into 80% training and 20% validation data.
- Optimize weights of the CNN to reproduce the Monte Carlo scatter estimates:

\[(w, b) = \arg \min_{w, b} \|DSE_{w, b}(T(p)) - I_{MC}\|_2^2\]

- Training on a GeForce GTX 1080 for 80 epochs.

### Results on Simulated Projection Data

<table>
<thead>
<tr>
<th>View #1</th>
<th>View #2</th>
<th>View #3</th>
<th>View #4</th>
<th>View #5</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image1.png" alt="Image" /></td>
<td><img src="image2.png" alt="Image" /></td>
<td><img src="image3.png" alt="Image" /></td>
<td><img src="image4.png" alt="Image" /></td>
<td><img src="image5.png" alt="Image" /></td>
</tr>
<tr>
<td>Primary intensity</td>
<td>Scatter ground truth (GT)</td>
<td>(Kernel – GT) / GT</td>
<td>(Hybrid - GT) / GT</td>
<td>(DSE – GT) / GT</td>
</tr>
<tr>
<td>C = 0 %, W = 50 %</td>
<td></td>
<td>14.1% mean absolute percentage error over all projections</td>
<td>7.2% mean absolute percentage error over all projections</td>
<td>1.2% mean absolute percentage error over all projections</td>
</tr>
<tr>
<td>C = 0.5, W = 1.0</td>
<td>C = 0.04, W = 0.04</td>
<td>C = 0 %, W = 50 %</td>
<td>C = 0 %, W = 50 %</td>
<td>C = 0 %, W = 50 %</td>
</tr>
</tbody>
</table>

DSE trained to estimate scatter from **primary plus scatter**: High accuracy
Reconstructions of Simulated Data

Ground Truth  No Correction  Kernel-Based Scatter Estimation  Hybrid Scatter Estimation  Deep Scatter Estimation

C = 0 HU, W = 1000 HU

Testing of the DSE Network for Measured Data (120 kV)

- Measurement of a head phantom at our in-house table-top CT.
- Slit scan measurement serves as ground truth.

Reconstructions of Measured Data

- Slit Scan
- No Correction
- Kernel-Based Scatter Estimation
- Hybrid Scatter Estimation
- Deep Scatter Estimation

C = 0 HU, W = 1000 HU

A simple detruncation was applied to the raw data before reconstruction. Images were clipped to the FOM before display. $C = -200$ HU, $W = 1000$ HU.

Truncated DSE

To learn why MC fails at truncated data and what significant efforts are necessary to cope with that situation see [Kachelrieß et al. Effect of detruncation on the accuracy of MC-based scatter estimation in truncated CBCT. Med. Phys. 45(8):3574-3590, August 2018].

Conclusions on DSE

• DSE needs about 20 ms per projection. It is a fast and accurate alternative to Monte Carlo (MC) simulations.
• DSE outperforms kernel-based approaches in terms of accuracy and speed.
• Interesting observations
  – DSE can estimate scatter from a single (!) x-ray image.
  – DSE can accurately estimate scatter from a primary+scatter image.
  – DSE cannot accurately estimate scatter from a primary only image.
  – DSE may thus outperform MC even though DSE is trained with MC.
• DSE is not restricted to reproducing MC scatter estimates.
• DSE can rather be trained with any other scatter estimate, including those based on measurements.
DSE for PET

Input feature 1: \( p = \ln \text{ACF} \)

Input feature 2: \( 1/p \)

Input feature 3: Prompts

Bed position f55d49, NMAE: 1.09 %, NMSE: 0.00 %
252 projection angles, 25 fps. DSE filtered in angular direction (Gaussian, FWHM 3.5 projections) for display

DSE for PET

Bed position f55d49, NMAE: 1.09 %, NMSE: 0.00 %
Reconstruction, transaxial (a.u.)

Deep Dose Estimation (DDE)

<table>
<thead>
<tr>
<th></th>
<th>MC</th>
<th>DDE</th>
</tr>
</thead>
<tbody>
<tr>
<td>48 slices</td>
<td>1 h</td>
<td>0.25 s</td>
</tr>
<tr>
<td>whole body</td>
<td>20 h</td>
<td>5 s</td>
</tr>
</tbody>
</table>

MC uses 16 CPU kernels.
DDE uses one Nvidia Quadro P600 GPU.
DDE training took 30 h for 200 epochs, 720 samples, 48 slices per sample.

Conclusions on Deep Learning for CT Image Formation

• Machine learning will play a significant role in CT image formation.

• High potential for
  – Artifact correction
  – Noise and dose reduction
  – Real-time dose assessment (also for RT)
  – ...

• Care has to be taken
  – Underdetermined acquisition, e.g. sparse view or limited angle CT, require the net to make up information!
  – Nice looking images do not necessarily represent the ground truth.
  – Data consistency layers may ensure that the information that is made up is consistent with the measured data.
  – ...
Thank You!

This presentation will soon be available at www.dkfz.de/ct.
Job opportunities through DKFZ’s international PhD or Postdoctoral Fellowship programs (marc.kachelriess@dkfz.de).
Parts of the reconstruction software were provided by RayConStruct® GmbH, Nürnberg, Germany.