From Data to Segmentation Models: AI for Everybody

Aditya P. Apte, Ph.D.
Department of Medical Physics
Memorial Sloan Kettering Cancer Center, New York NY
Outline

• Motivation
• Components of segmentation pipeline
  • Hardware (Institutional HPC, Cloud services AWS, Azure, Google cloud)
  • Software (frameworks for training and deployment)
• Applications
  • Clinical (Radiation therapy treatment planning)
  • Research (Consistent data collection for research)
• Demos
Motivation

• Classification
  • predict target class at the patient level from an image or region of interest.

• Detection
  • detection tasks aim to predict the location of potential lesions, often in the form of points, regions, or bounding boxes of interest.

• Segmentation
  • identification of pixels or voxels composing an organ or structure of interest.

Motivation

➢ Screening and follow-up


➢ Treatment planning – clinical use cases

• Synthetic CT generation from MR (Jiang et al, Wolterink et al, Han et al, etc.)

• Knowledge-based prediction of 3D dose distribution (Shiraishi and Moore, etc.)

• Auto contouring for treatment planning / Adaptive RT
Building blocks of segmentation pipeline

- Domain-optimized tools
- Training framework
- Hardware
- Question and Dataset
High Performance Computing (HPC) resources

• Cloud computing with specialized hardware – AWS, Azure, Google cloud, etc.

• Institutional HPC  
  e.g., hpc.mskcc.org  
  o 91 nodes, 4640 CPU cores  
  o 372 GPUs (4 dedicated to Medical Physics)  
  o 3.6PB of fast disk storage, and 6.0PB of ‘warm’ archive storage space

• Clinical HPC  
  o 4 nodes, 144 CPU cores  
  o 20 GPUs
Software - building blocks

Domain-optimized framework – building blocks

**Pre-process**
- Parse DICOM
- Select region of interest
- Image filters
- Populate channels

**Segment**
(model and its dependencies)

**Post-process**
- Label fusion
- Morphological processing
Transformations for medical images

Resample → Crop → Resize → Transform view → Filter image → Populate channels

"channels":
[{"imageType":"original", "slice":"current-1"},
{"imageType":"original", "slice":"current"},
{"imageType":"original", "slice":"current+1"}]
Domain-optimized frameworks for medical imaging

- MONAI - PyTorch-based framework for deep learning in healthcare imaging. It provides domain-optimized foundational capabilities for developing healthcare imaging training workflows in a native PyTorch paradigm. (https://github.com/Project-MONAI/MONAI)

- NVIDIA Clara – healthcare application framework for AI-powered imaging, genomics, and the development and deployment of smart sensors (https://docs.nvidia.com/clara)

- CERR – Models for segmentation of various normal tissues used in radiation therapy treatment planning. (https://github.com/cerr/CERR)

- DeepInfer - an open-source toolkit for developing and deploying deep learning models within the 3D Slicer medical image analysis platform (http://www.deepinfer.org)
Standardizing the processing pipeline

• NVIDIA-Clara MMAR
• CERR pre/post-processing pipeline
• MONAI “transforms” module
Medical Model Archive (MMAR) in NVIDIA

• The MMAR (Medical Model ARchive) defines a standard structure for organizing all intermediate steps and results produced during the model development life cycle.
CERR Pre/post processing settings

```
{
  "strNameToLabelMap": [
    {
      "structureName": "AORTA", "value": 2,
    },
    {
      "structureName": "LA", "value": 3,
    },
    {
      "structureName": "LV", "value": 4,
    },
    {
      "structureName": "RA", "value": 5,
    },
    {
      "structureName": "RV", "value": 6,
    },
    {
      "structureName": "IVC", "value": 7,
    },
    {
      "structureName": "SVC", "value": 8,
    },
    {
      "structureName": "PA", "value": 9
    }
  ],

  "batchSize": 1,

  "inputFileType": "planC",

  "resize": {
    "method": "none",
    "size": [512, 512]
  },

  "crop": [{
    "method": "crop_to_str", "params": {
      "structureName": "lung_l", "operator": "",
    },
    {"method": "crop_to_str", "params": {
      "structureName": "lung_r", "operator": "union",
    },
    {"method": "crop_to_str", "params": {
      "structureName": "totallung", "operator": "union"
    },

  }],

  "passedScanDim": "2D",

  "reference": "https://w3.aapm.org/meetings/2019AM/programInfo/programAbs.php?sid=8013&aid=45641"
}
```
MONAI APIs

- **apps**: high level medical domain specific deep learning applications.
- **config**: for system configuration and diagnostic output.
- **data**: for the datasets, readers/writers, and synthetic data.
- **handlers**: defines handlers for implementing functionality at various stages in the training process.
- **inferers**: defines model inference methods.
- **losses**: classes defining loss functions.
- **optimizers**: classes defining optimizers.
- **transforms**: defines data transforms for preprocessing and postprocessing.
- **utils**: generic utilities intended to be implemented in pure Python or using Numpy, and not with Pytorch, such as namespace aliasing, auto module loading.
- **visualize**: utilities for data visualization.
Model deployment

• Singularity containers
  • can be run by users without root privileges.
  • checksums for making the software stacks reproducible
  • compatibility with HPC systems and enterprise architectures

• SingularityHub, DockerHub

• ModelHub (http://modelhub.ai) - repository of self-contained deep learning models pretrained for a wide variety of applications.
Model deployment

• Containers
  • Singularity unavailable for Windows.
  • Docker requires admin privilege, so not suitable for bare metal HPC.
  • Containers cannot be run within another container (e.g. Google Colab)

• Anaconda
  • Package manager for Python and R (also supports other
  • Requires dependencies (e.g., CUDA) to be installed on the host machine.
Portable framework for image segmentation


AI Segmentation example using CERR
XNAT Pipeline Engine / Event service

• XNAT is an open-source project produced by Neuroinformatics Research Group at the Washington University School of Medicine.

• Centralized data management resource to multiple investigators and research studies.
  • Import, archive, manage and securely distribute imaging datasets.

Pipeline Engine

Review

Export image

Segmentation output

OHIF Viewer

Segmentation model
XNAT segmentation pipeline demo
Fully automatic vs user-assisted contouring

• Semi-automatic:
  • Thresholding, region growing available in clinical software and research tools such as ITKSnap, 3DSlicer, CERR.

• Active-learning
  • FOVIA
  • NVIDIA Clara
  • MONAI-label

  The MONAI-label is a server-client system that facilitates interactive medical image annotation by using AI. It is an open-source and easy-to-install ecosystem that can run locally on a machine with one or two GPUs. Both server and client work on the same/different machine. (Reference: https://github.com/Project-MONAI/MONAILabel)
Example of AI assisted segmentation (DeepGrow3D) in FOVIA

Reference: https://www.youtube.com/watch?v=bllg2lwSfO4
Segmentation performance evaluation

Volumetric DICE
Hausdorff 95 distance
Surface DICE
Added Path Length

APL correlates with time required for contouring

Manual vs AI contours – time savings (Thoracic CT)

Comparison between DL and Atlas based contouring for Lung OARs. DL resulted in significant time saving over Atlas based and manual contouring.

**Manual vs AI contours – time savings (Prostate CT)**

Comparison between DL and Atlas based contouring for Prostate OARs. DL resulted in significant time saving over Atlas based and manual contouring.

Reference: W. Jeffrey Zabel, Jessica L. Conway, Adam Gladwish et al, Clinical Evaluation of Deep Learning and Atlas-Based Auto-Contouring of Bladder and Rectum for Prostate Radiation Therapy, Practical Radiation Oncology, Volume 11, Issue 1, 2021, Pages e80-e89, ISSN 1879-8500,

[https://doi.org/10.1016/j.prro.2020.05.013](https://doi.org/10.1016/j.prro.2020.05.013)

<table>
<thead>
<tr>
<th>MAN</th>
<th>DEEP</th>
<th>ATLAS</th>
</tr>
</thead>
<tbody>
<tr>
<td>1) Import CT to TPS</td>
<td>1) Import CT to DEEP software</td>
<td>1) Import CT to TPS</td>
</tr>
<tr>
<td>2) Initial contour generation (10.9 ± 4.6 min)</td>
<td>2) Initial contour generation (1.4 ± 0.1 min)</td>
<td>2) Initial contour generation (1.2 ± 0.2 min)</td>
</tr>
<tr>
<td>3) Task RO to review contours</td>
<td>3) Export contours to TPS</td>
<td>3) Task RO to review contours</td>
</tr>
<tr>
<td>4) RO contour review and editing (4.1 ± 2.5 min)</td>
<td>4) Import CT and contours to TPS</td>
<td>4) RO contour review and editing (10.2 ± 5.7 min)</td>
</tr>
<tr>
<td>5) Task RO to review contours</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6) RO Contour review and editing (4.7 ± 2.6 min)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**CT acquisition**

Total: 15.3 min  Total: 6.8 min  Total: 11.7 min
Clinical experience at MSKCC (Elguindi et al)

• Rollout for clinical use:
  • Prostate T1 axial MR – 2019 (Elguindi et al, Physics and Imaging in Radiation Oncology, Vol.12, pp. 80-86, 2019, )
    • Released May 2020
    • Updated Model: Mid Feb 2021
    • Dataset: 174 patients collected (106 v1, 64 v2), out of approximately 350 patients in total
    • Structures contoured:
      • Left/Right parotids, Left/Right submandibular glands, brainstem, mandible, Cord, Oral Cavity, Left/Right medial pterygoid and Left/Right masseter muscles
    • Average APL: 194.6 cm, average time savings of 8.6 minutes with Left/Right parotids, Left/Right submandibular glands, brainstem.
Clinical Workflow: HPC + Containers
Deep learning-based segmentation of OARs in H&N CT scan
Consistent segmentation in clinical trials

# Models available in CERR

<table>
<thead>
<tr>
<th>SITE</th>
<th>MODALITY</th>
<th>ORGAN/S</th>
<th>MODEL / FRAMEWORK</th>
<th>REFERENCE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lung</td>
<td>CT</td>
<td>Nodules</td>
<td>Incremental MRRN / Keras, Tensorflow</td>
<td>Jiang et al, IEEE Transactions on Medical Imaging, 38(1): 134 – 144 <a href="https://doi.org/10.1109/TMI.2018.2857800">https://doi.org/10.1109/TMI.2018.2857800</a></td>
</tr>
<tr>
<td>Head &amp; Neck</td>
<td>CT</td>
<td>Masseters (left, right), Medial Pterygoids (left, right), Constrictor muscles (superior, middle, inferior) and Larynx</td>
<td>DeepLabV3+ / Pytorch</td>
<td>Iyer et al, <a href="https://www.biorxiv.org/content/10.1101/772178v2.full">https://www.biorxiv.org/content/10.1101/772178v2.full</a></td>
</tr>
<tr>
<td>Head &amp; Neck</td>
<td>CT</td>
<td>Parotids (left, right), Mandible, Sub-mandibular glands (left, right), Brain stem</td>
<td>Self attention / Pytorch</td>
<td>Jiang et al, <a href="https://arxiv.org/abs/1909.05054">https://arxiv.org/abs/1909.05054</a></td>
</tr>
</tbody>
</table>
Models available in open-source frameworks

• **NVIDIA Clara** ([https://docs.nvidia.com/clara/deploy/BundledPipelines/index.html](https://docs.nvidia.com/clara/deploy/BundledPipelines/index.html))
  • Pipelines for Liver, Spleen, Lung, Lung tumor, Pancreas tumor, Colon tumor, Brain tumor.

• **DeepInfer** ([http://www.deepinfer.org/pages/models/](http://www.deepinfer.org/pages/models/))
  • Prostate segmentation for T2 MRI, Brain tumor segmentation for Axial T1 MRI and FLAIR

• **ModelHub** ([http://modelhub.ai/](http://modelhub.ai/))
  • Liver and liver tumor segmentation for CT, Brain tumor segmentation, prediction of dental artifacts, Right ventricle segmentation from MRI scan,
Deep Learning Criticisms

• Long on data but short on knowledge, medical reasoning, and new insights. (Partho P. Sengupta and Y.S. Chandrashekhar, “Building Trust in AI: Opportunities and Challenges for Cardiac Imaging”, J Am Coll Cardiol Img. 2021 Feb, 14 (2) 520–522)

• External validation to explore the model generalizability in external patient cohorts obtained from different institutions, geographic boundaries, and different time periods.

• Approaches to probe and interpret AI techniques have been proposed recently with the emergence of techniques like symbolic AI, knowledge graphs, and their underlying semantic technologies that can provide reasoning mechanisms.
Conclusion

• Numerous open-source frameworks to train and deploy models for clinical and research use.

• Standardizing data processing pipelines enable reproducible application of models

• Encapsulating models in containers makes them readily deployable, hence facilitating their validation on external/unseen datasets.

• AI is already helping to boost productivity in clinic. This is just the beginning!
Acknowledgement

• Eve LoCastro, M.S.
• Aditi Iyer, M.S.
• Sharif Elguindi, M.S.
• Maria Thor, M.S.
• Jung Hun Oh, Ph.D.
• Harini Veeraraghavan, Ph.D.
• Joseph O. Deasy, Ph.D.