Distributed machine learning infrastructure for medical image segmentation

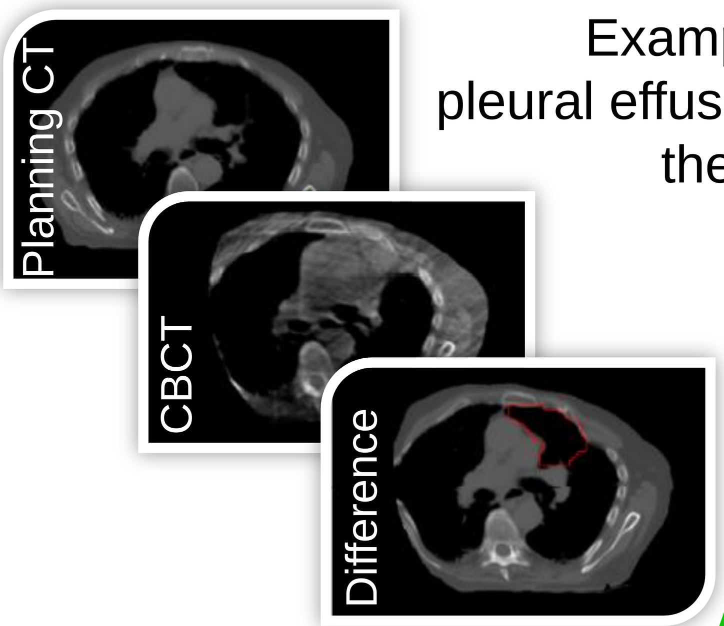
L. Hotoiu[1], E. Brion[2], J. Léger[2], R. Labarbe[1]

[1]Ion Beam Applications SA, Louvain-la-Neuve, Belgium; [2]Universite catholique de Louvain, Louvain-La-Neuve, Belgium

Poster ID: 059

Introduction

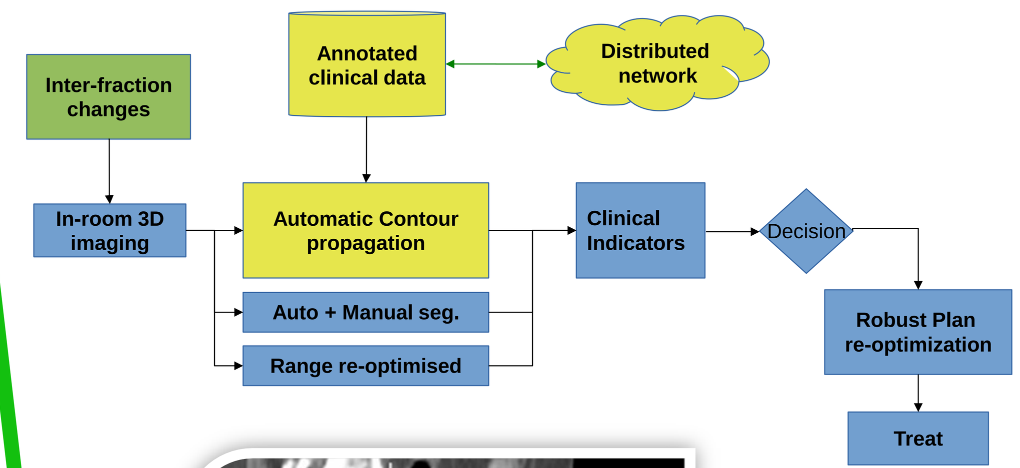
In radiation oncology, treatment planning relies on computed tomography (CT) images. During the course of treatment, a number of anatomical changes can occur in the patient's body, influencing the delivery. Inherently, the initial prescription may not be valid any longer and requires adaptation to the new conditions. Machine learning can help solving this adjustment problem.

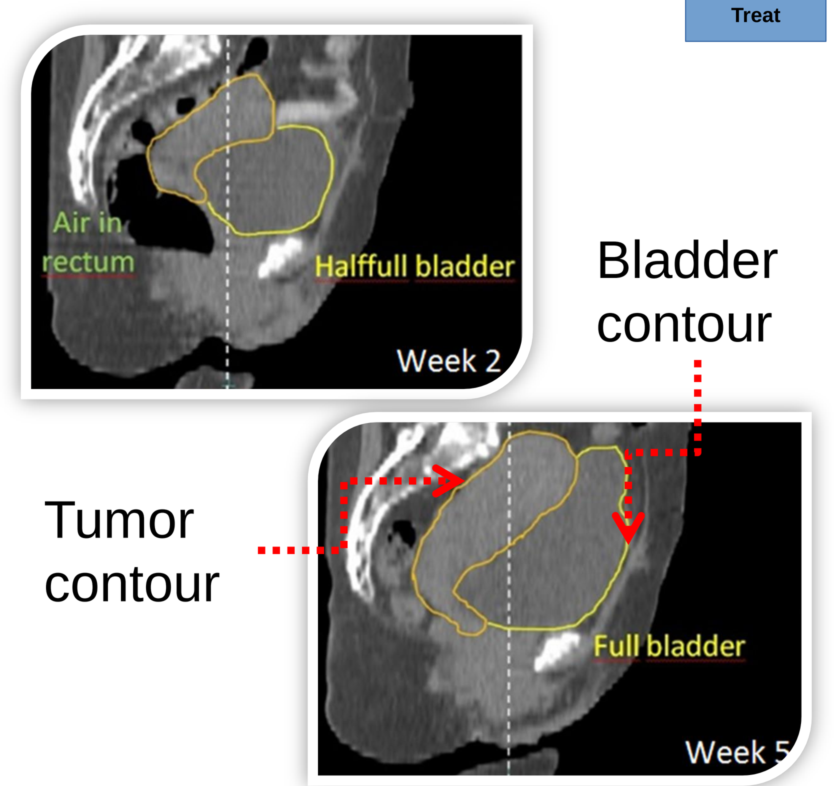


Example of pleural effusion in the lung

Problem

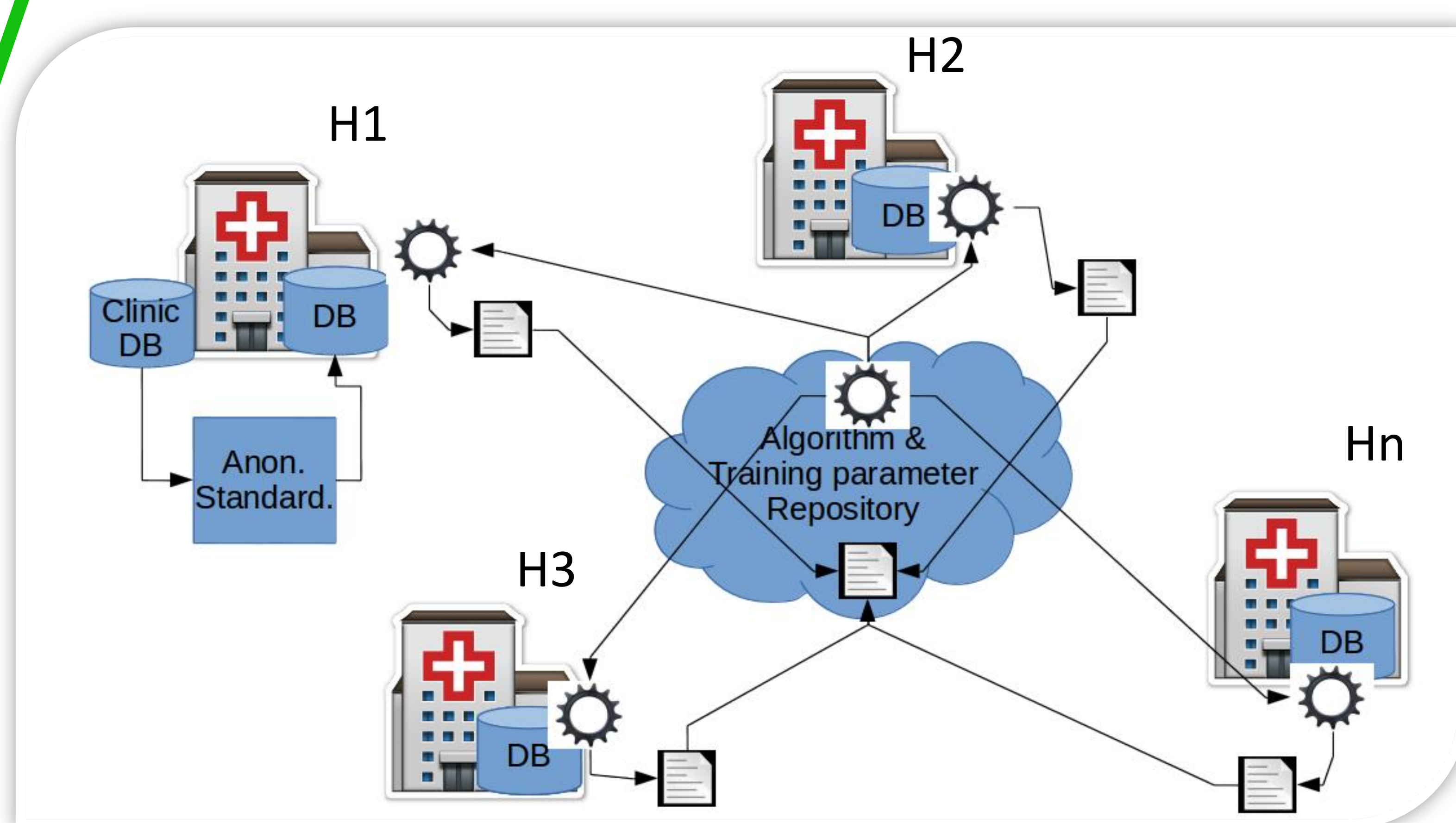
Automatic contouring. The oncologist manually delineates the tumor and the organs at risk during the planning phase, to mark the areas to be irradiated or spared. When treatment adaptation is required, the contours have to follow as well. For treatment plan adaptation to make its way in clinical practice, the process needs to be fully automated, avoiding time-consuming, manual intervention from the medical doctor.





Automatic contouring via image segmentation is key.

Distribution principle

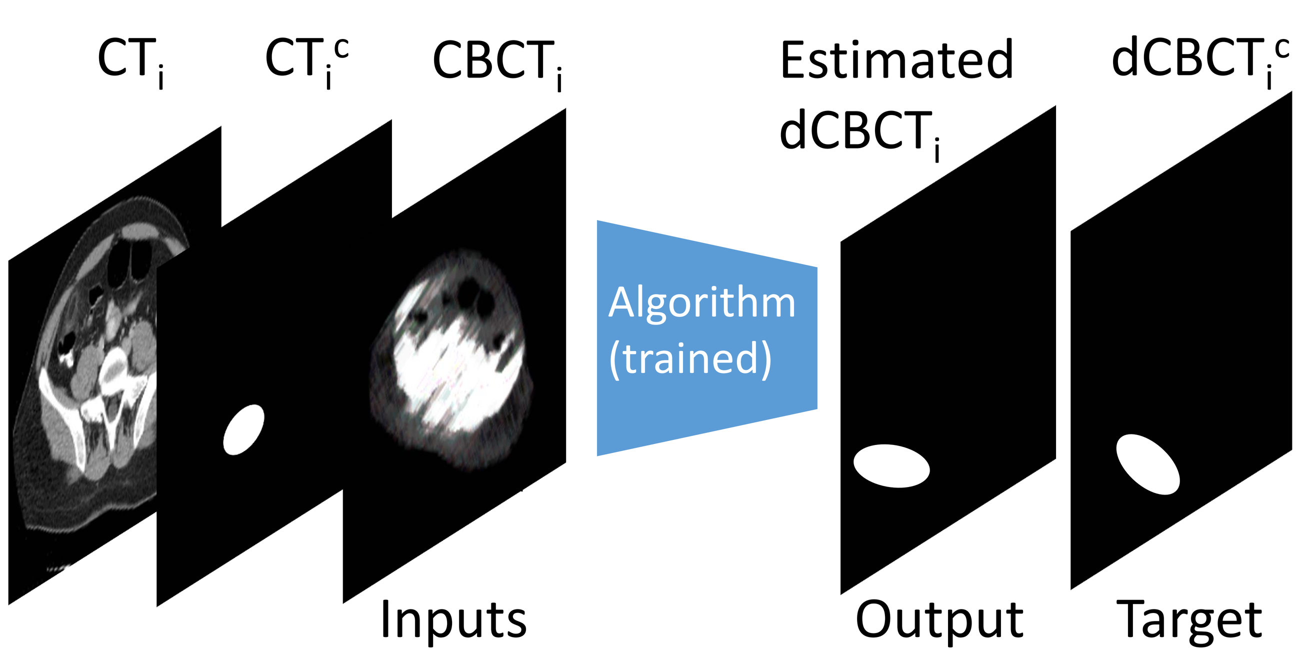


- Patient data is protected and non-standard.
- A distributed infrastructure deploying a CNN makes that medical records never quit hospital's infrastructure. Only learning data retrieved & shared between hospitals [2].
- A standard data format for machine learning is needed. Extensive communal effort is paramount for such a system to work.

Deep learning

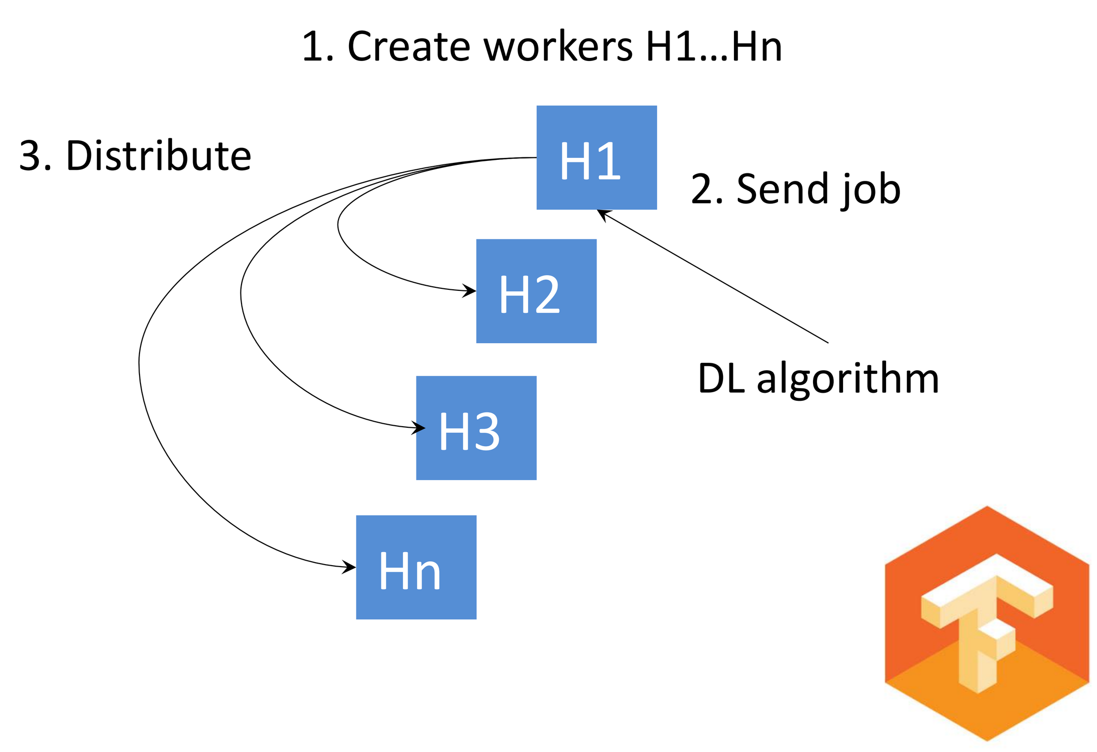
- 1000 patients with pelvic region cancer.
- 1000 planning CT_i and contours CT_i^c.
- Literature estimates ~5000 samples/pathology for excellent results [1].
- 3D U-Net CNN efficient in segmenting volumetric medical images [3].

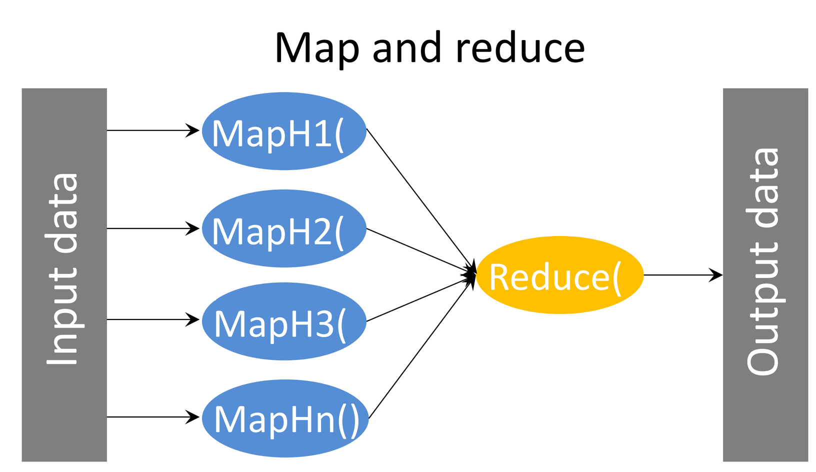
See poster ID 097



Infrastructure

- Use Tensorflow distributed learning capabilities:
  - Distributed computing using workers.
  - Map and reduce to combine results [4].





Conclusion

- Keeps the data locally, retrieves only intelligence, alleviating the legal problem of patient data protection.
- Beneficial for all participants which will gain access to a larger data-pool in exchange of reciprocity.
- Distributes the computation load in between the participating institutions. Avoids the need of a "massive" central cluster.

Outlook

- Develop the distributed capacity based on Tensorflow.
- Attract more institutions into actively participating.
- Persuade participating institutions to standardize their medical records (actively & retroactively) to help machine learning and general automation.

[1] Goodfellow, et and al., Deep Learning, MIT Press, 2016.
[2] Skripcak, et al., Radiother & Onc, v. 113, p. 303-309, 2014.
[3] Özgün Çiçek, et al., In MICCAI, pp. 424-432 Springer, 2016.
[4] https://learningtensorflow.com/lesson11/.

Contact: lucian.hotoiu@iba-group.com
IBA research

DS3 Summer School, 28th August - 1st of September 2017, Paris