Inverse problems and machine learning in medical physics

Treatment planning -
Machine learning in treatment planning

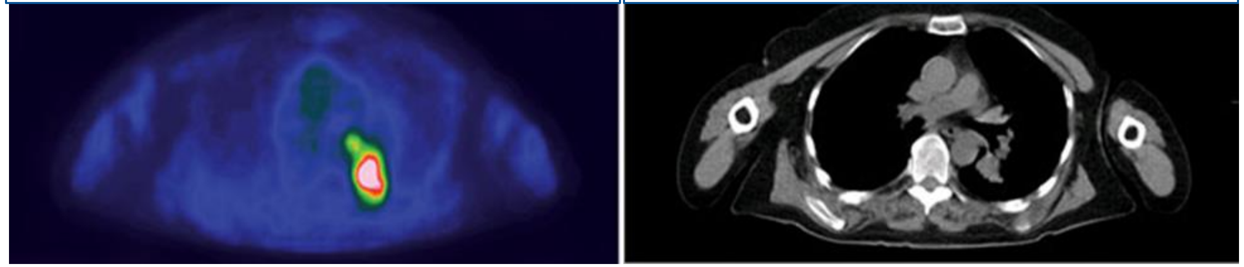
Dr. Chiara Gianoli
7/1/2025
chiara.gianoli@physik.uni-muenchen.de

Imaging in treatment planning

- Treatment planning firstly requires the identification of the radiotherapy structures using anatomical and functional information from diagnostic images
    - The target definition is based on the X-ray CT image, as primary anatomical image
    - Secondary images as positron emission tomography (PET) (i.e., functional information) and/or magnetic resonance imaging (MRI) (i.e., functional and anatomical information) can complement the target definition

see the target

PET image | CT image

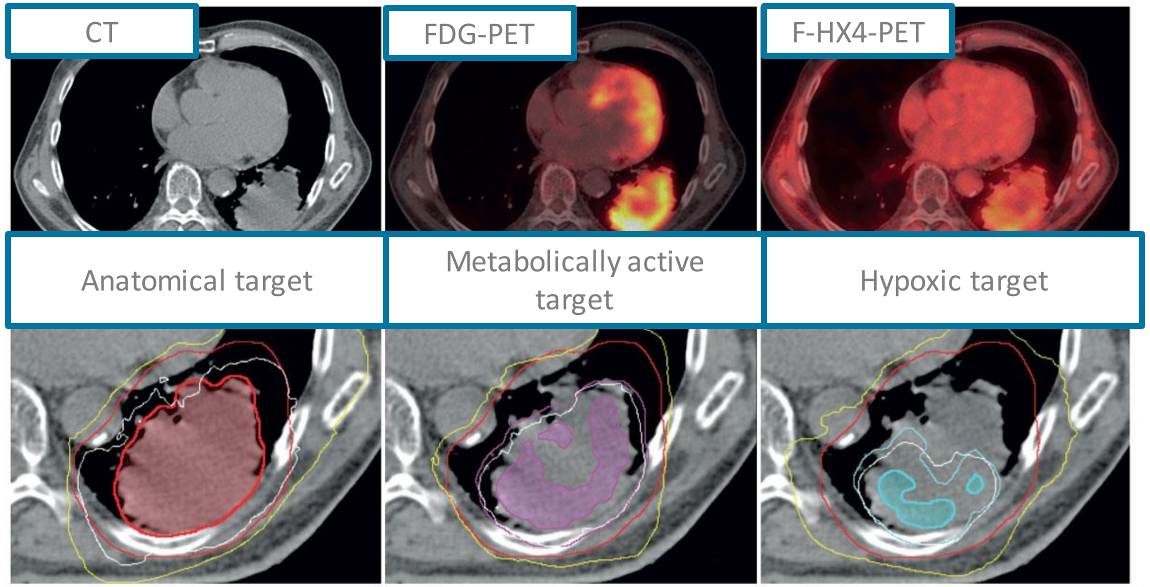


Bussink et al. *Nat. Rev. Clin. Oncol.* 2011

An archer cannot hit the bullseye if he doesn't know where the target is!

Imaging in treatment planning

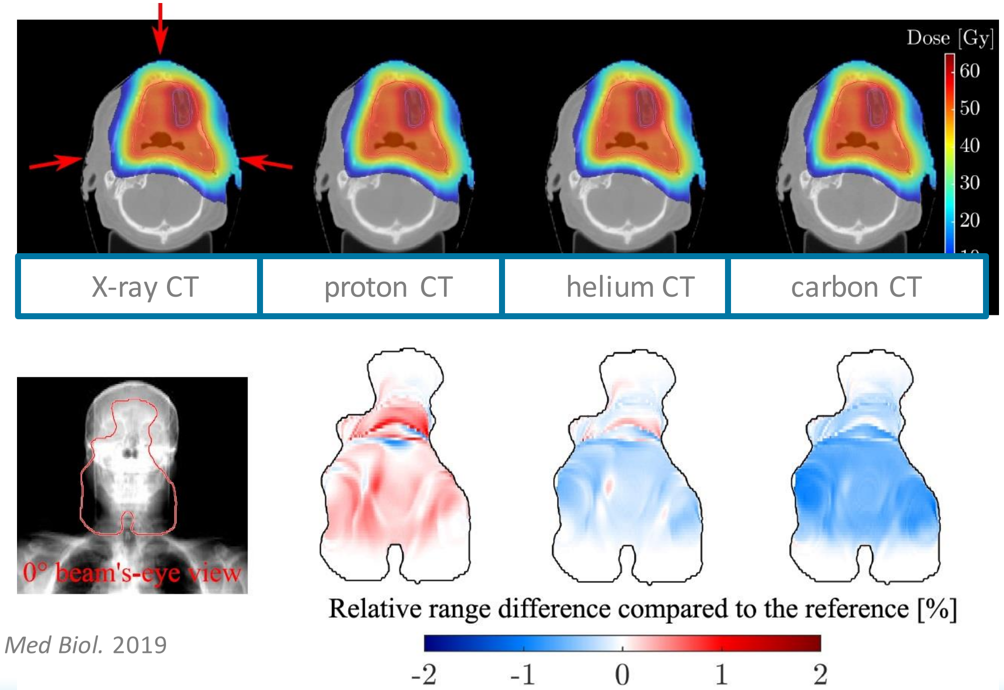
- The functional identification of the target can be used for “dose painting”
  - ^{18}FDG-PET (*Fluorodeoxyglucose*): glucose uptake and metabolism
  - ^{18}F-HX4-PET (*Fluorin-nitroimidazole*): molecular retention correlated to tumor hypoxia



Grootjans et al. *Nat. Rev. Clin. Oncol.* 2015

Fundamentals in treatment planning

- The treatment planning is an inverse problem and requires numerical optimization to define the beam parameters (i.e., inverse treatment planning), based on:
  - Definition of the treatment geometry (i.e., target and critical organs identification)
  - Patient model as physical characterization of the patient (i.e., tomographic image reconstruction of the properties of the radiation in tissue)
    - photon attenuation (X-ray CT)
    - ion stopping power relative to water (ion CT)

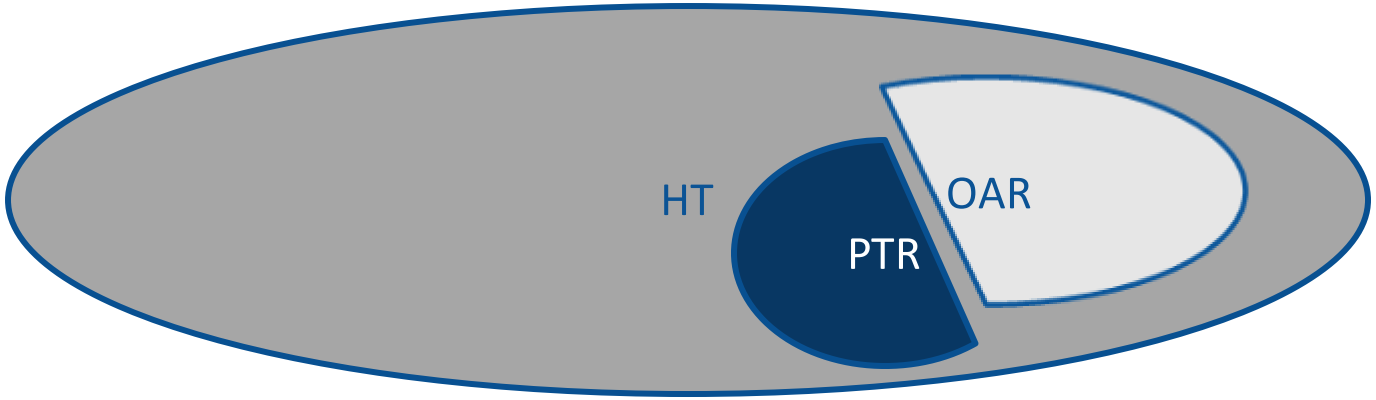


Meyer et al. *Phys Med Biol.* 2019

Fundamentals in treatment planning

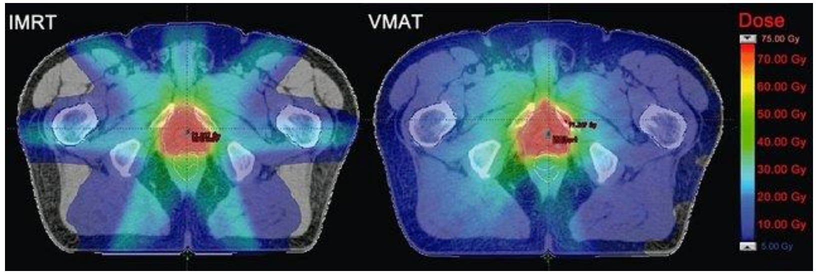
- Treatment planning aims to match the dose value of the PTR (*planning target region*) to its prescribed value while limiting dose values in the surrounding OAR (*organ at risk*) and HT (*healthy tissue*) to tolerable limits

- In particular, OARs are highly sensitive to radiation exposure and require lower dose values than HT

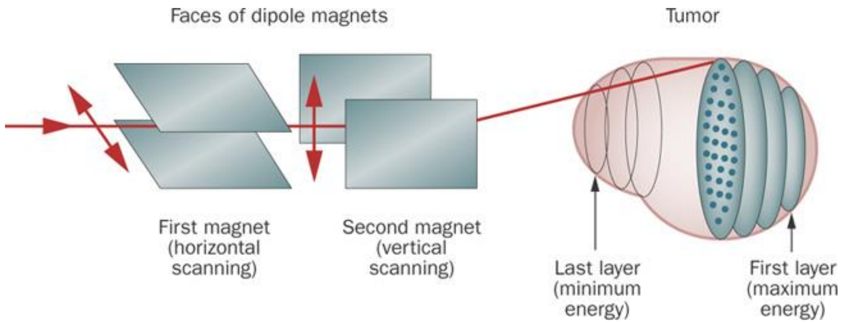


Fundamentals in treatment planning

- Intensity modulated radiation treatment (IMRT) and volumetric modulated arc therapy (VMAT)
  - High precision conformation as overlay of multiple discrete (IMRT) or continuous (VAMT) 3D dose distributions
  - The intensity of the radiation beam is subdivided in multiple beam-lets
- Ion beam therapy
  - High precision conformation as stack of multiple iso-energy 2D dose distributions
  - The intensity of the radiation beam is subdivided in multiple pencil beams



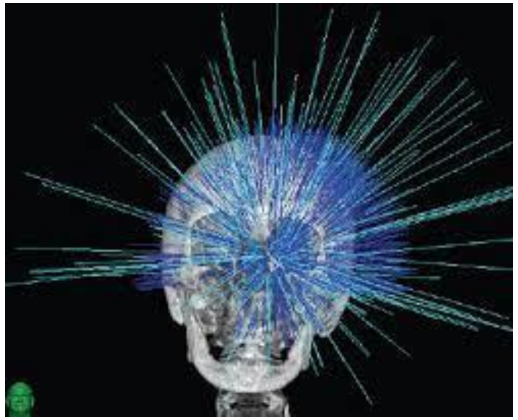
Nguyen, B. T., Hornby, C., Kron, T., Cramb, J., Rolfo, A., Pham, D., ... & Foroudi, F. (2012). Optimising the dosimetric quality and efficiency of post-prostatectomy radiotherapy: A planning study comparing the performance of volumetric-modulated arc therapy (VMAT) with an optimised seven-field intensity-modulated radiotherapy (IMRT) technique. *Journal of Medical Imaging and Radiation Oncology*, 56(2), 211-219.

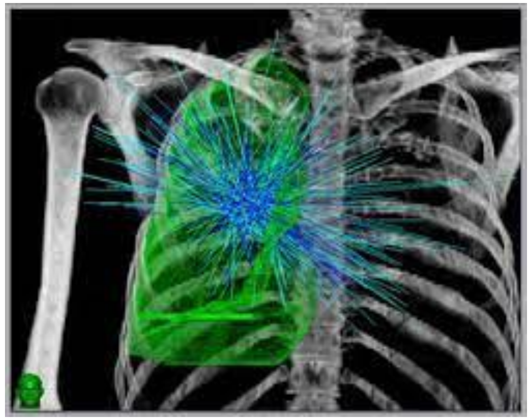


Durante, M., & Loeffler, J. S. (2010). Charged particles in radiation oncology. *Nature reviews Clinical oncology*, 7(1), 37-43.

Fundamentals in treatment planning

- Stereotactic radiation therapy (cranial SRT) and stereotactic body radiation therapy (SBRT)
  - High precision and high dose conformation as overlay of multiple 3D dose distributions, delivered from fixed points in space called nodes, arranged in spherical (intracranial applications) or ellipsoidal (extracranial applications) configurations
    - The combination of nodes and pointing vectors provides a set of "elementary beams" to plan the treatment



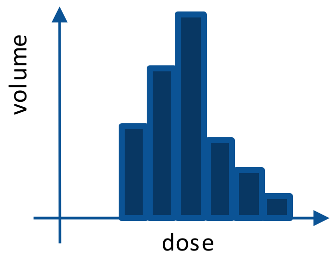


http://www.cyberknifendc.com

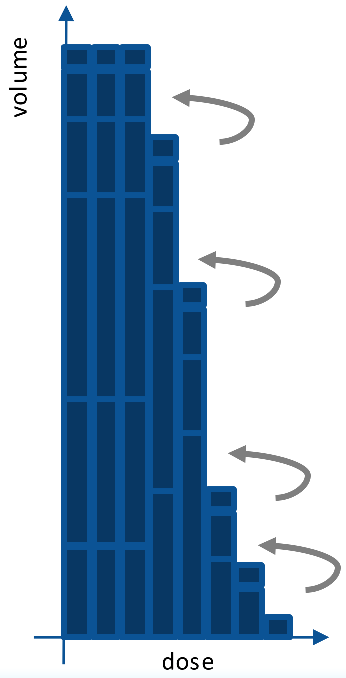
Fundamentals in treatment planning

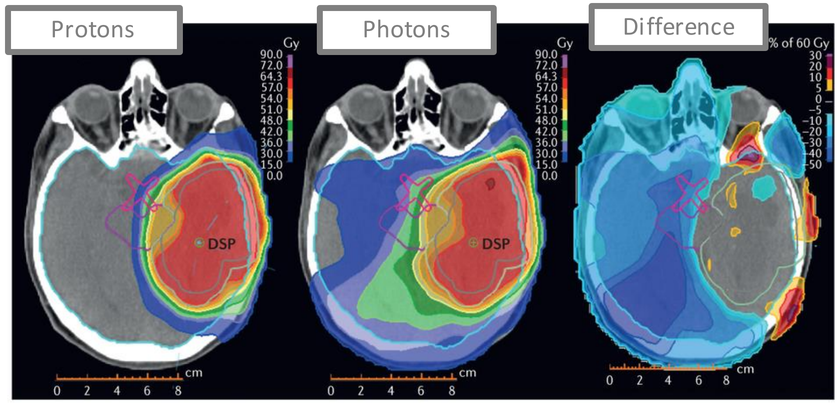
- The dose-volume histogram (DVH) is a treatment plan evaluating tool
- DVH summarizes a 3D dose distribution in a graphical 2D format
- The volumes reported in the DVH are the PTR and OAR
    - Differential DVH
        - the relative volume of PTR or OAR that receives the indicated dose
    - Cumulative DVH
        - the integral relative volume of PTR or OAR that receives at least the indicated dose

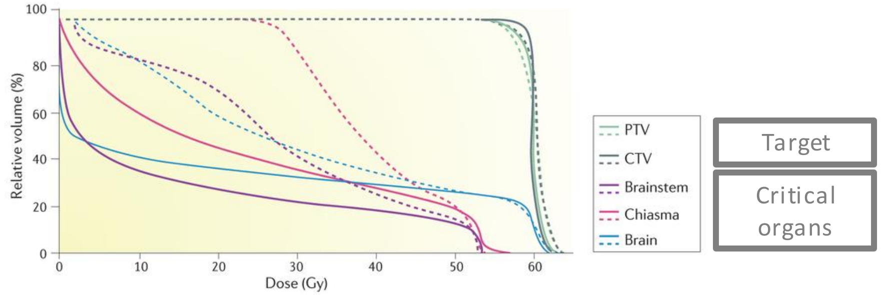
Differential DVH



Cumulative DVH





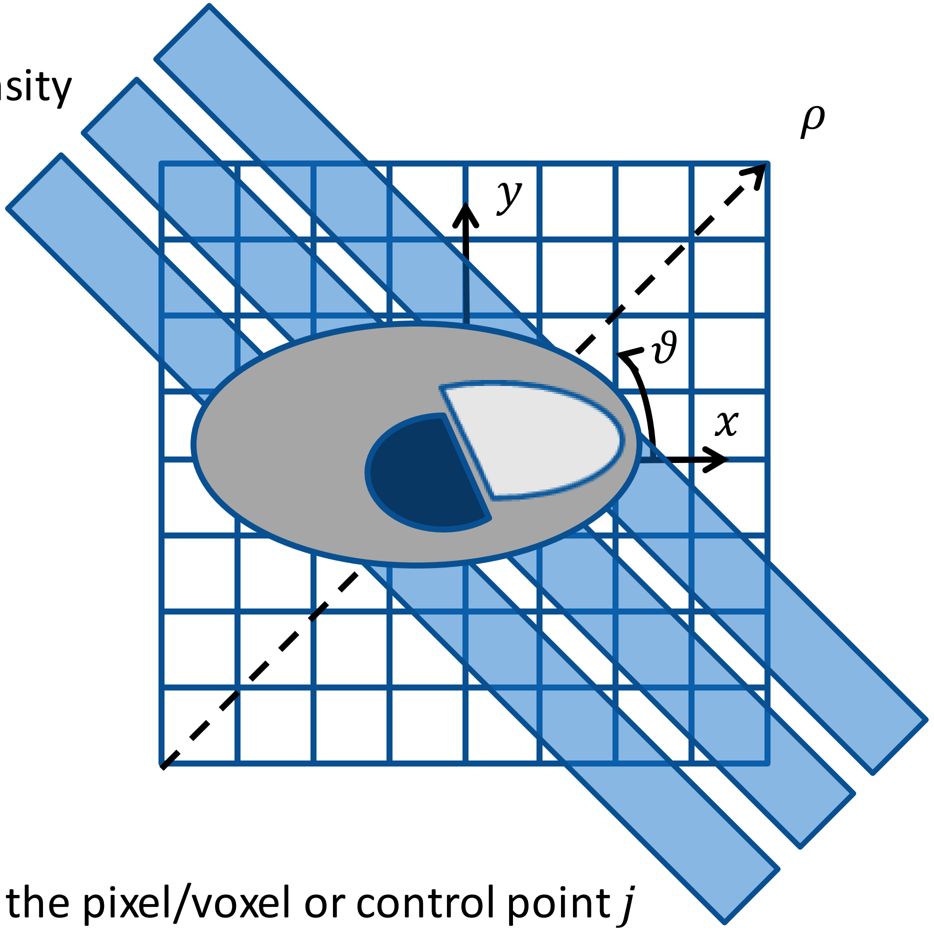


Baumann, M., Krause, M., Overgaard, J., Debus, J., Bentzen, S. M., Daartz, J., ... & Bortfeld, T. (2016). Radiation oncology in the era of precision medicine. *Nature Reviews Cancer*, 16(4), 234-249.

Fundamentals in treatment planning

- Discretization of the dose distribution into a grid of dose points
  - Beam-let or pencil beam: elementary radiation beam with fixed intensity
  - Beam: beam-lets or pencil beams with fixed angle ϑ or fixed energy
- Modeling of the inverse problem of treatment planning as matrix-vector product:

$$\bar{f}_j = \sum_i a_{ji} g_i$$



  - $\bar{f}_j$ is the dose distribution in the pixel/voxel or control point j
  - g_i is the unknown weight of the beam-let or pencil beam i
  - a_{ij} is the dosimetric contribution of the beam-let or pencil beam i to the pixel/voxel or control point j

Fundamentals in treatment planning

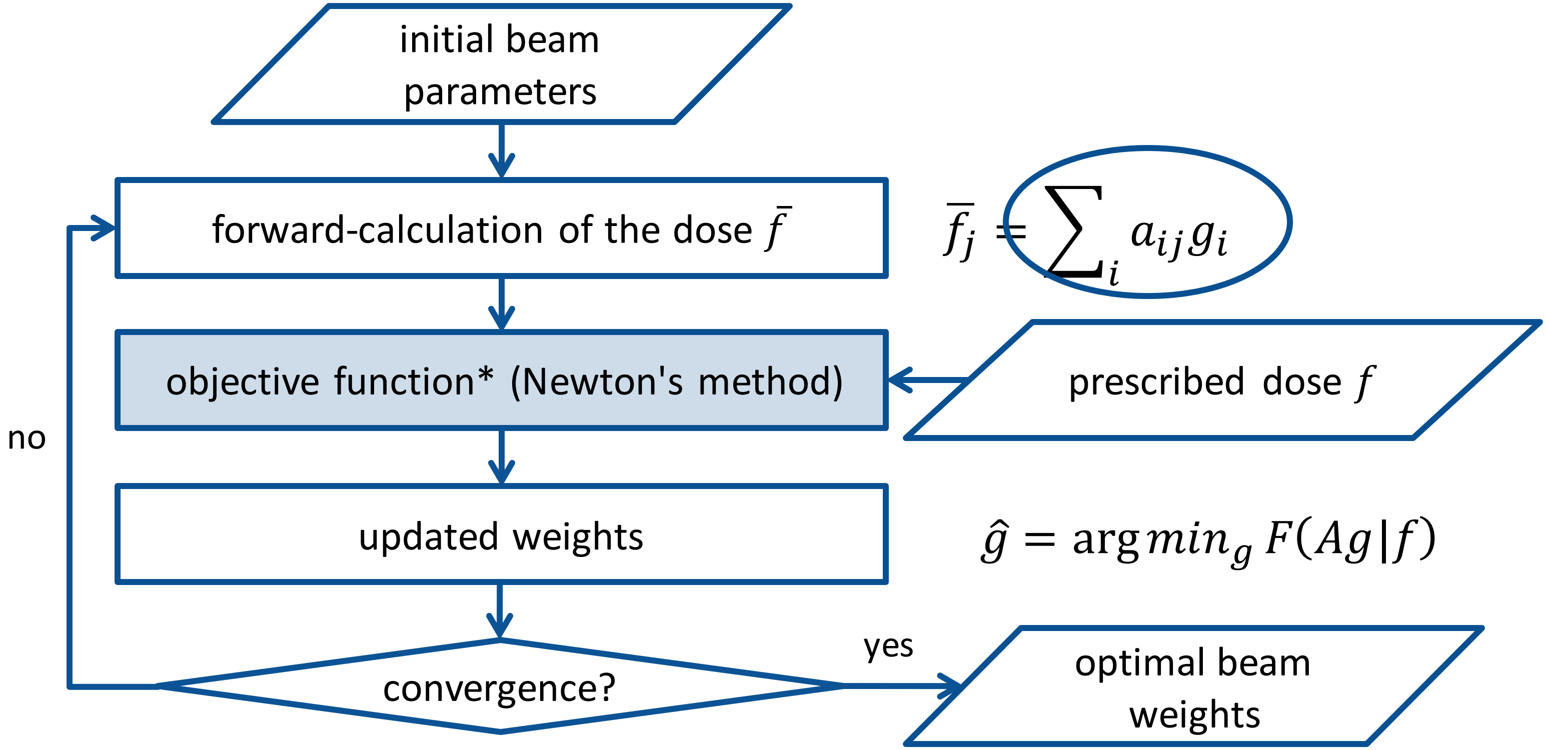
- The inverse problem of treatment planning is concerned with determining the non-negative weights g_i that results in optimal dose distribution f_j
  - a_{ij} can be interpreted as the dose per unit of time deposited at pixel or control point j by the beam-let or pencil beam i, and g_i is the time the beam-let or pencil beam i is kept on
  - a_{ij} is referred to as the dose calculation matrix
    - model-based algorithms (convolution-superposition methods based on dose kernels scaled according to the electron density or relative stopping power of the heterogeneity)
    - correction-based algorithms (semiempirical approaches to account for tissue heterogeneity)
    - Monte Carlo simulations

Oelkfe, U., & Scholz, C. (2006). Dose calculation algorithms. In New technologies in radiation oncology (pp. 187-196). Springer, Berlin, Heidelberg.

Fundamentals in treatment planning

- In photon beam therapy (IMRT), M beams (different angles) are needed to conform the dose distribution to the target
- In ion beam therapy, M beams (different energies) are needed to conform the dose distribution to the target
- The model of the inverse problem becomes a $F = AG$ matrix-matrix product with:
  - F is the matrix of optimal dose distribution for each *m=1:M* (*JxM* matrix)
  - A is the matrix of the geometrical contribution of the beam-let or pencil beam to the pixel (*JxI* matrix), referred to as dose calculation matrix
  - G is the unknown (intensity) weight matrix of each beam-let or pencil beam, for each *m=1:M* (*IxM* matrix)
- The inverse problem of treatment planning is concerned with determining the non-negative weight matrix G that results in optimal dose distribution F

Fundamentals in treatment planning



* the objective function can be either voxel-based or organ/DVH-based, the violation of the DVH constraints can be adopted as penalty function

Treatment planning

- Newton's method of objective function minimization is based on the approximation of ψ(f) as a quadratic function in the neighborhood of the minimum f_{min}
  - The objective function can be approximated by its Taylor series expansion as:

$$\psi(f_{\min} + f) \approx \psi(f_{\min}) + f^T \nabla_f \psi(f_{\min}) + \frac{1}{2} f^T \nabla^2{}_f \psi(f_{\min}) f$$

where the Gradient vector and the Hessian matrix (H) are defined as:

$$\nabla_f \psi(f_{\min}) = \left.\begin{pmatrix} \frac{\partial \psi}{\partial f_1} \\ \dots \\ \frac{\partial \psi}{\partial f_N} \end{pmatrix}\right|_{f=f_{\min}} \qquad \nabla^2{}_f \psi(f_{\min}) = \left.\begin{pmatrix} \frac{\partial^2 \psi}{\partial f_1{}^2} & \dots & \frac{\partial^2 \psi}{\partial f_N \partial f_1} \\ \dots & \dots & \dots \\ \frac{\partial^2 \psi}{\partial f_1 \partial f_N} & \dots & \frac{\partial^2 \psi}{\partial f_N{}^2} \end{pmatrix}\right|_{f=f_{\min}}$$

- Implementations of inverse treatment planning differ from objective function approximations

Treatment planning

- Newton's method finds the minimum f_{min} when the gradient of ψ(f) is equal to zero:

$$\nabla_f \psi(f_{\min}) + \nabla^2{}_f \psi(f_{\min}) f = 0$$

$$f = -\left(\nabla^2{}_f \psi(f_{\min})\right)^{-1} \nabla_f \psi(f_{\min})$$

  - The inverse Hessian matrix can be not exact ($HH^{-1} \neq$ identity matrix)
  - Iterative algorithms are adopted to compute an approximation of the inverse Hessian matrix (quasi-Newton methods)

$$f_{n+1} = f_n - \overline{H}^{-1} \nabla_f \psi(f_n)$$

- Implementations of inverse treatment planning differ from inverse Hessian matrix approximations

Treatment planning

- Being $f=Ag$, and therefore $g=A^{-1}f$, the objective function minimization is expressed as:

$$f_{n+1} = f_n - \overline{H}^{-1}\nabla_f \psi(f_n)$$

$$g_{n+1} = g_n - \overline{A}^{-1}\overline{H}^{-1}\nabla_f \psi(f_n) = g_n - \overline{A}^{-1}\overline{H}^{-1}\overline{A}^{-1}\nabla_g \psi(f_n)$$

  - A^{-1} is the inverse dose calculation matrix
  - The two gradient vectors are related according to:

$$\nabla_g \psi(f) = \frac{\partial f}{\partial g}\frac{\partial \psi(f)}{\partial f} = A\nabla_f \psi(f)$$

- Implementations of inverse treatment planning differ from inverse dose calculation matrix approximations

Treatment planning

- Closed-form least square optimization

$$g_{min} = argmin \left\| f_j - \sum_i a_{ji} g_i \right\|^2$$

$$g_{\min} = \underbrace{\boxed{(A^T A)}}_{\text{Hessian}}{}^{-1} \underbrace{\boxed{A^T f_j}}_{\text{Gradient}}$$

- Numerical (iterative) optimization or iterative inverse treatment planning

$$g_i^{n+1} = g_i^{n} + \frac{f_j - \sum_i a_{ji} g_i^{n}}{\sum_i a_{ji}^2} \cdot a_{ji}$$

Xing, L., & Chen, G. T. (1996). Iterative methods for inverse treatment planning. Physics in Medicine & Biology, 41(10), 2107.

Machine learning in treatment planning

- Conventional radiotherapy treatment planning consists of inverse optimization to determine the radiation beam weights (i.e., g_i) based on the treatment planning X-ray CT image and the dose prescription (i.e., f_j)

- The optimized radiation beam parameters need to be manually adjusted with trial and error (time-consuming and labor-intensive)

- Artificial intelligence, including machine learning and deep learning, has been recently proposed to automate radiotherapy treatment planning and improve treatment planning quality and efficiency

- Automated treatment planning includes

    - Automated beam orientation selection (i.e., pre-defined angles of the beam-lets)

    - Automated dose distribution prediction (i.e., forward-calculation of the dose)

    - Automated radiation beam parameters estimation (i.e., the weights)

Wang, M., Zhang, Q., Lam, S., Cai, J., & Yang, R. (2020). A review on application of deep learning algorithms in external beam radiotherapy automated treatment planning. *Frontiers in oncology*, 10, 2177.

Target definition

- The image segmentation of the tumor and the organs at risk (OARs) is a time-consuming process, on a slice-by-slice basis when manually performed, subject to significant inter- and intra- operator variability

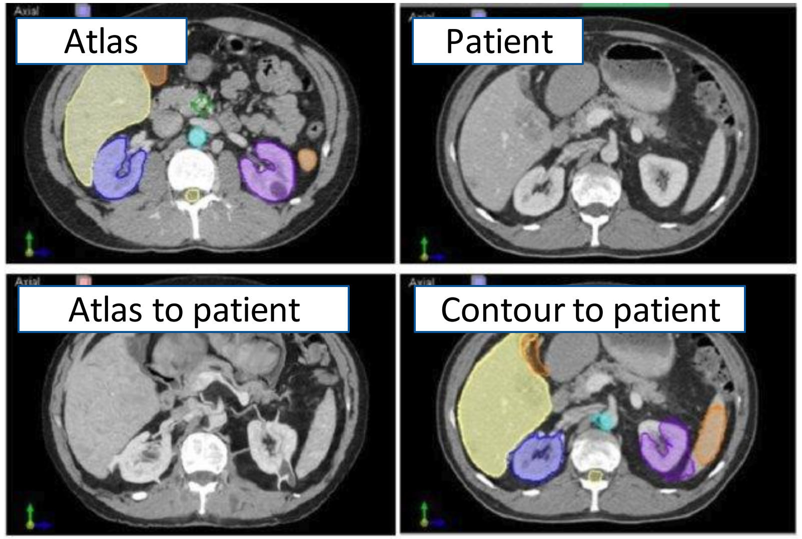
- Automatic segmentation (i.e., auto-segmentation) enables the automation and standardization of this process

  - Conventional auto-segmentation is based on the primary (and secondary) image(s) of the individual patient

  - Auto-segmentation based on atlas exploits prior knowledge from a cohort of patients as a ground truth organ segmentation

    - The segmentation is adapted to the individual patient according to deformable image registration (DIR)

  - Auto-segmentation based on DL embeds prior knowledge from the cohort of patients into a parameterized model that is optimized to match the ground truth segmentation during the training

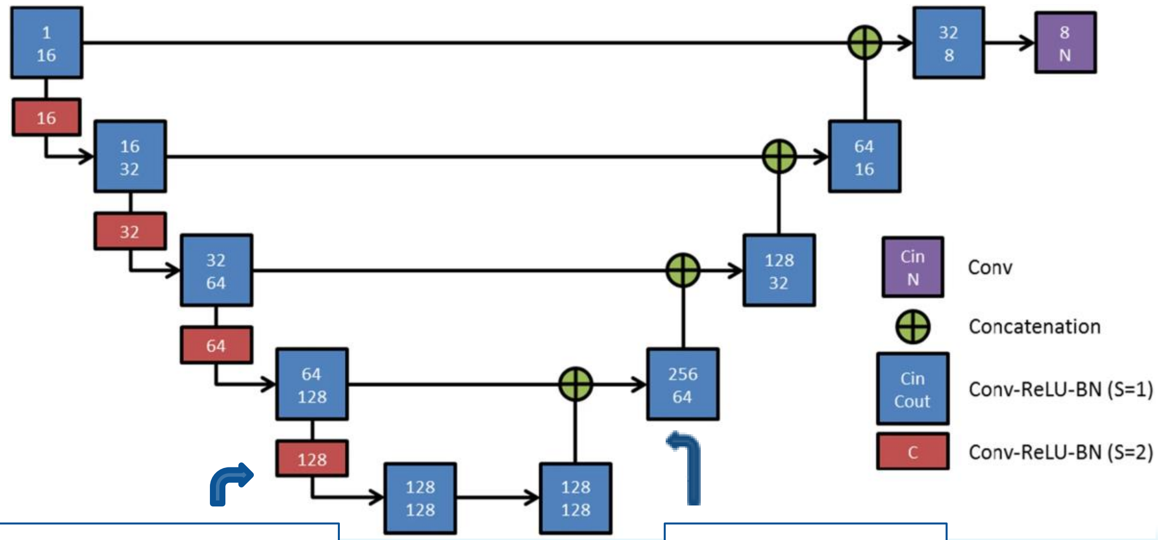


Schreibmann & Fox 2012 *J. Appl. Clin. Med. Phys.*

DL-based auto-segmentation

- Because of the local nature of the segmentation, DL-based auto-segmentation is typically based on fully convolutional neural networks

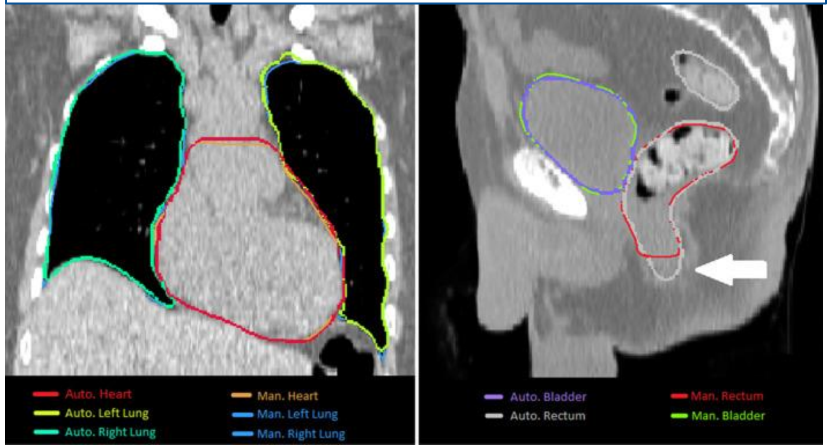
- The architecture of the commercial DL-based auto-segmentation networks is mostly undisclosed but some are reported being based on modifications of the U-net
    - The deep image-to-image network (DI2IN), commercially available in Siemens Healthineers systems, is based on a convolutional encoder-decoder architecture combined with multi-level feature concatenation



convolution with stride 2 instead of pooling

trilinear interpolation

Clinical evaluation at LMU Klinikum

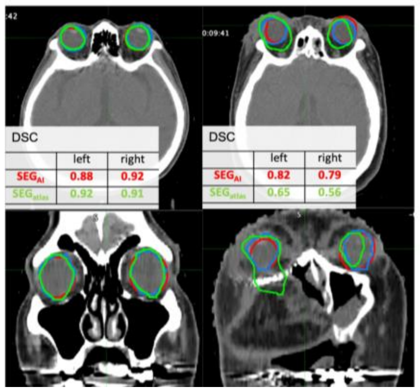


Marschner et al. 2022 *Radiat. Oncol.*

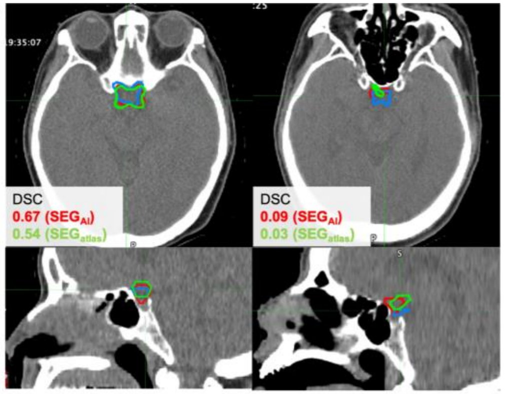
DL-based auto-segmentation

- The accuracy of the DL-based auto-segmentation is expected within the inter-operator variability, as the network cannot perform better than the manual segmentation adopted as ground truth

Eyeballs



Optic chiasm



Manual segmentation

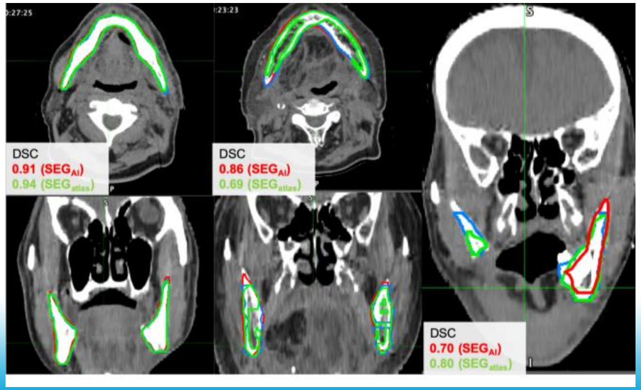
AI-based auto-segmentation

Atlas-based auto-segmentation

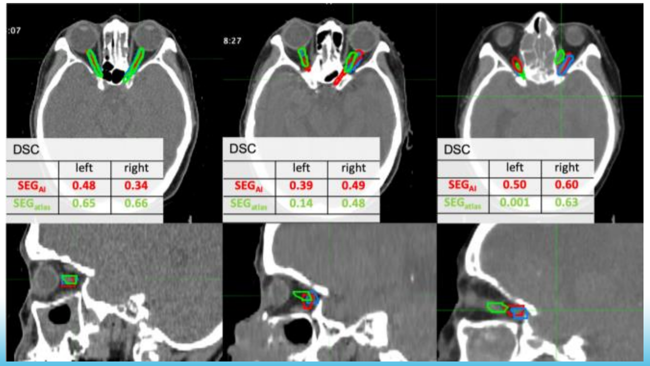
$$DSC(A,B) = \frac{2|A \cap B|}{|A| + |B|}$$

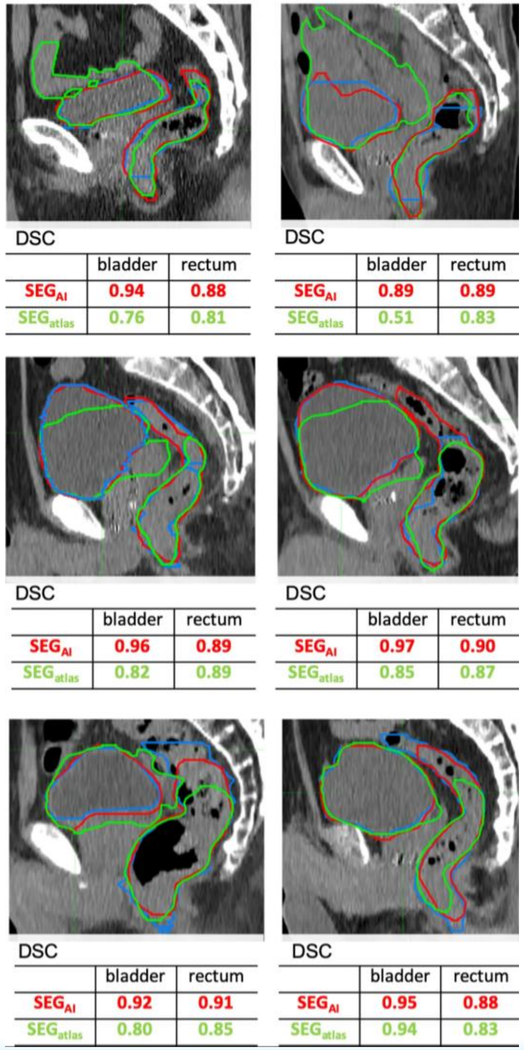
Dice-Sørensen coefficient

Mandible



Optic nerve





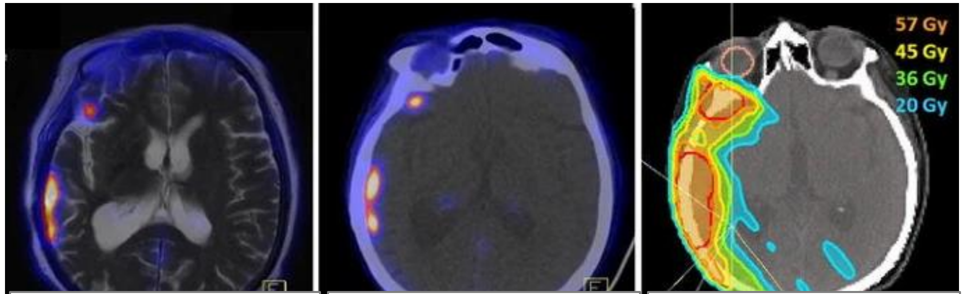
Urago et al. 2021 *Radiat. Oncol.*

Treatment planning

- Conventional treatment planning consists in the solution of an inverse problem to optimize the radiation beam parameters that match the prescribed dose on the tumor, including dosimetric constraints for OARs and normal tissue

Radiation beam → Interaction model → Dose distribution

- The direct problem is referred to as dose calculation, the inverse problem as treatment plan calculation



PET/MRI | PET/CT | treatment plan

Thorwarth et al. 2013 *Clin. Transl. Imaging*

hit the target

LMU LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN

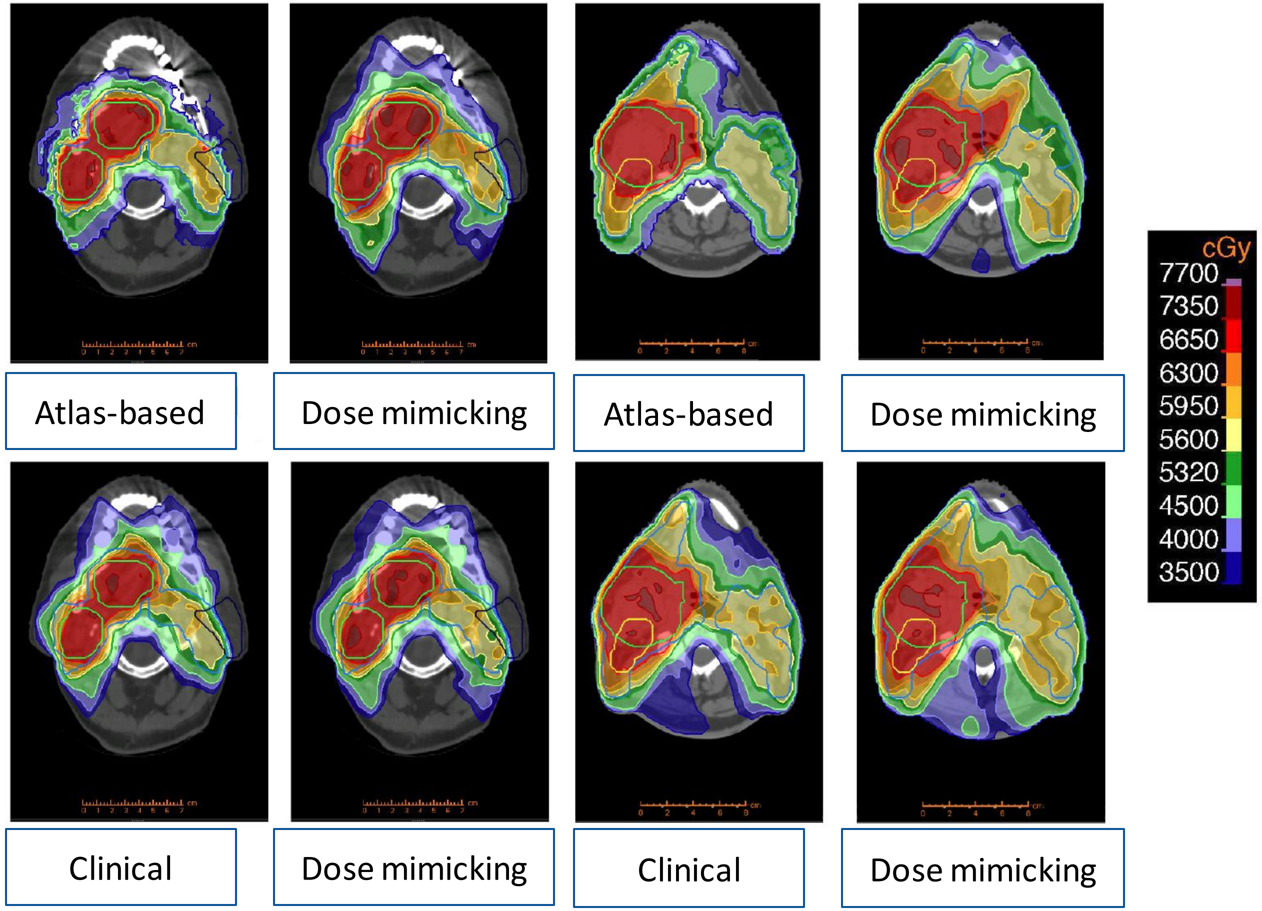
Dose estimation

- Artificial intelligence enables to automate different steps of the treatment planning and improve treatment planning quality and efficiency

- The automation is based on the anatomy-to-dose correlation inferred from a cohort of clinical treatment plans

  - The prediction of the dose distribution can be implemented as a case/atlas-based ML regression (i.e., ML-based regression from a cohort of similar cases which is usually referred to as knowledge-based radiation therapy treatment) or as DL-based inference

    - The predicted dose distribution *per se* does not account for the physics of the beams, thus, dose mimicking optimization then converts the dose distribution to a deliverable treatment plan

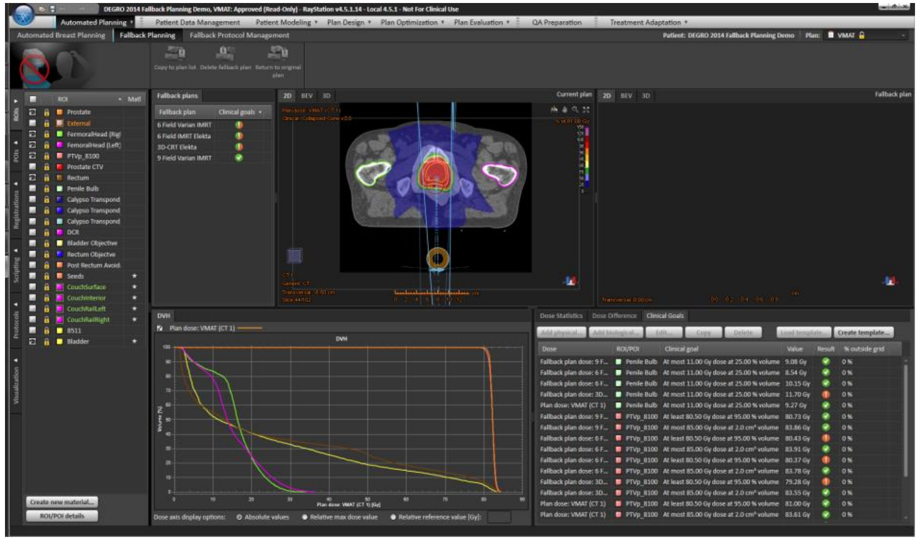


McIntosh et al. 2017 *Phys. Med. Biol.*

DL-based auto-planning

- Commercial knowledge-based radiation therapy treatment planning software are currently used in the adaptive radiation therapy workflow
  - Varian Ethos
  - Ray Station

Varian Ethos



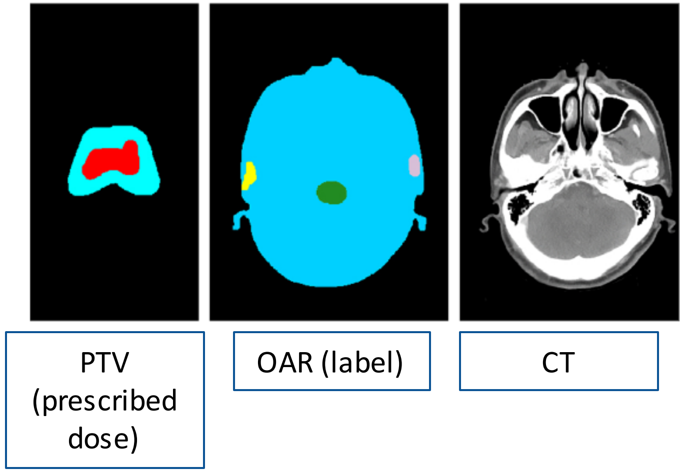
Ray Station

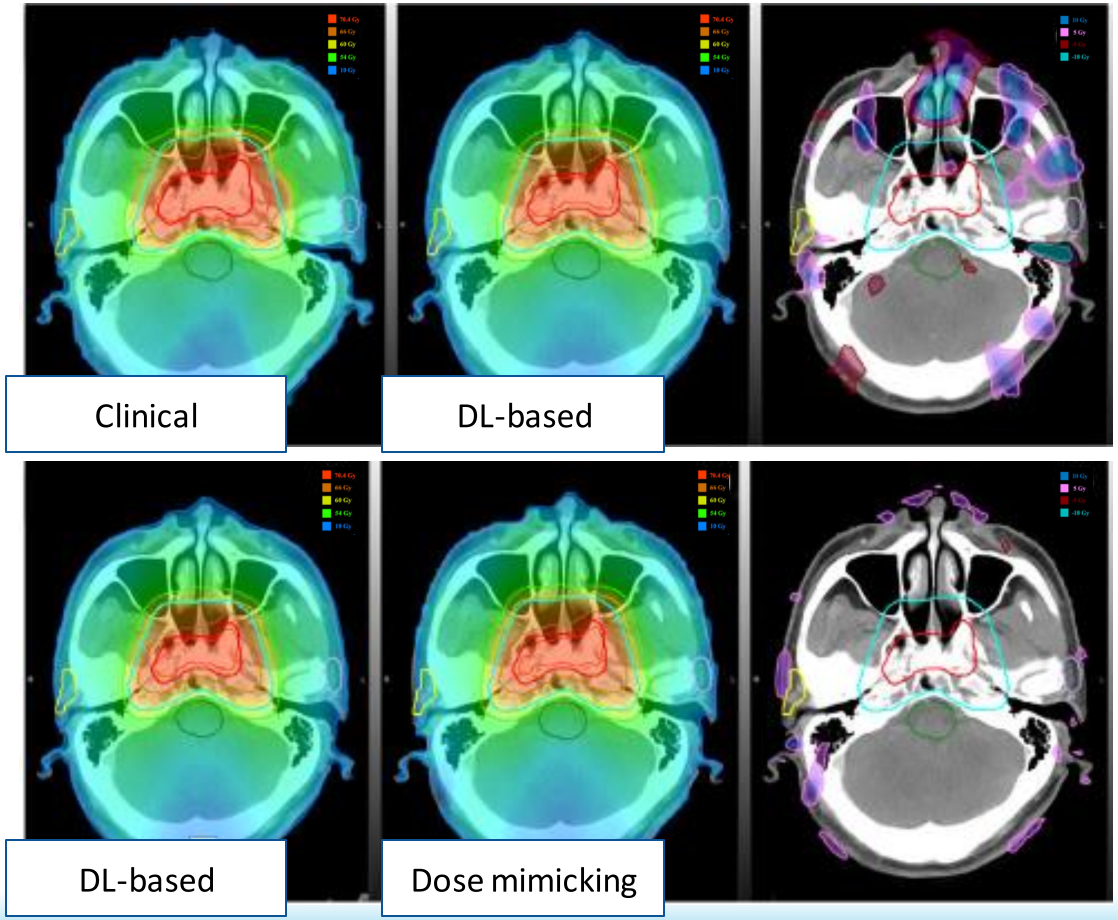
- DL-based auto-planning is typically based on deep fully convolutional neural networks combined with residual connections such as Res-Net, DoseNet and modified U-net

DL-based auto-planning

- The networks are trained on 2D or 3D images describing patient geometry in terms of CT image, segmented target and OARs (input) and the manually optimized ground truth dose distribution (target)

- The Res-Net - deep residual neural network - is trained on 3D images for intensity-modulated radiation therapy (IMRT) in head-and-neck cancer cases

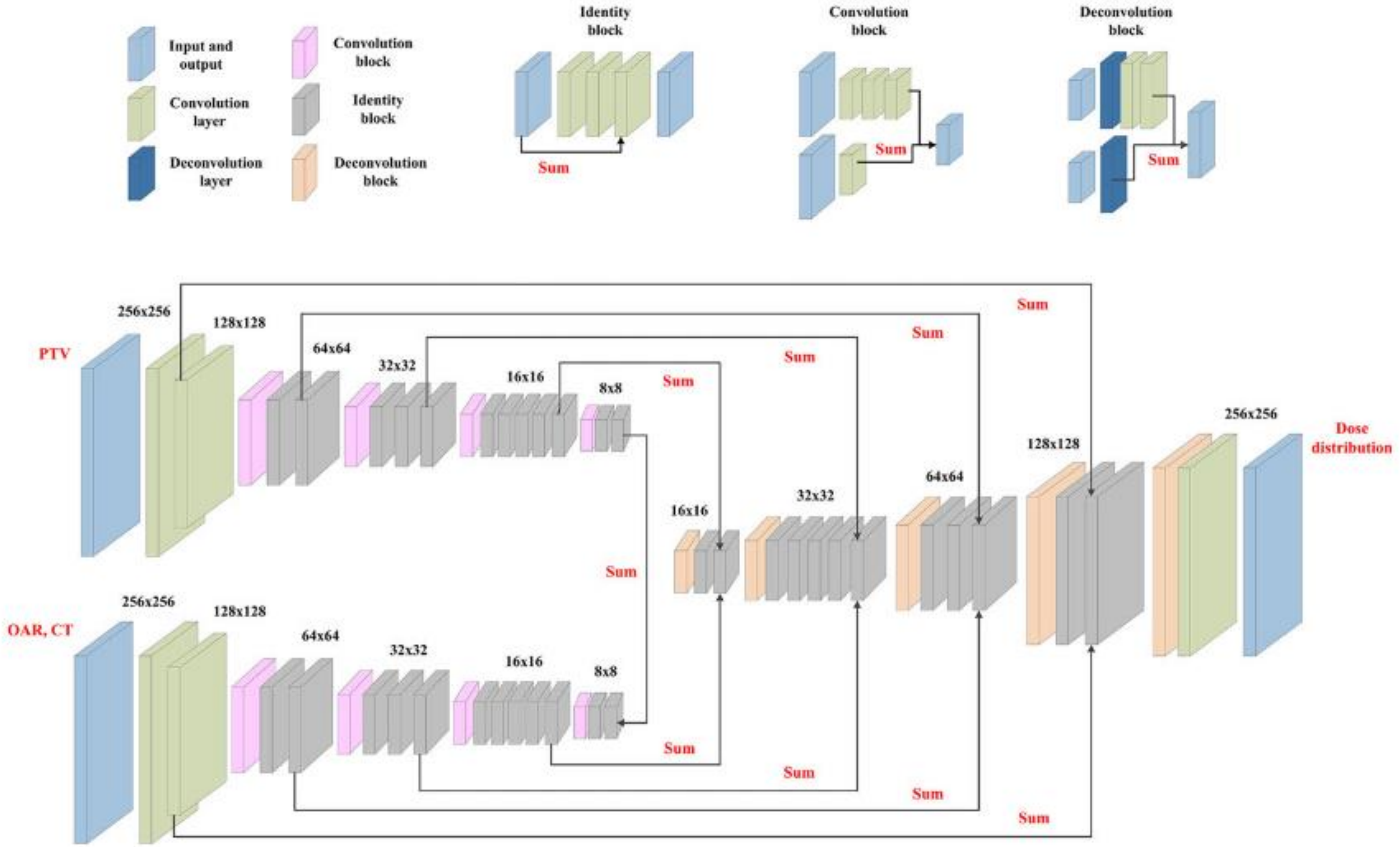




Fan et al. 2019 *Med. Phys.*

DL-based auto-planning

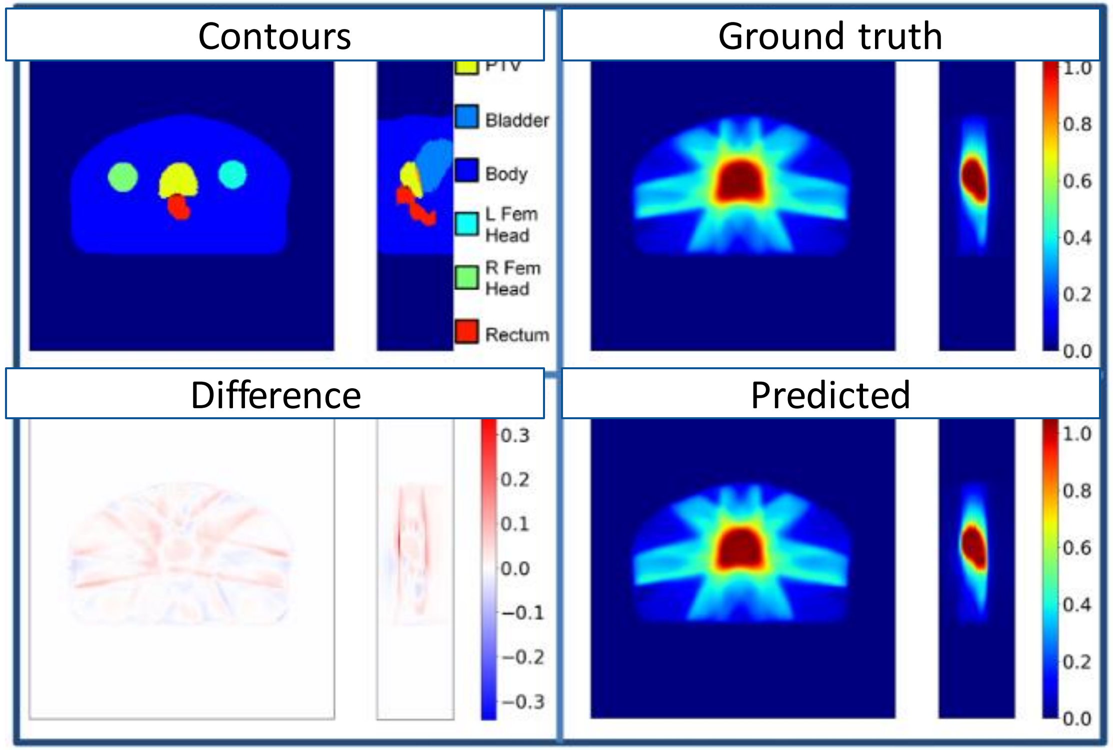
- Convolutional layers to down-sample the feature maps
- Deconvolutional layers to up-sample the feature maps and recover the image details
- Links between convolutional and deconvolutional layers with multiple skip-layer connections (tackling the problem of gradient vanishing and passing of image details)



Fan et al. 2019 *Med. Phys.*

DL-based auto-planning

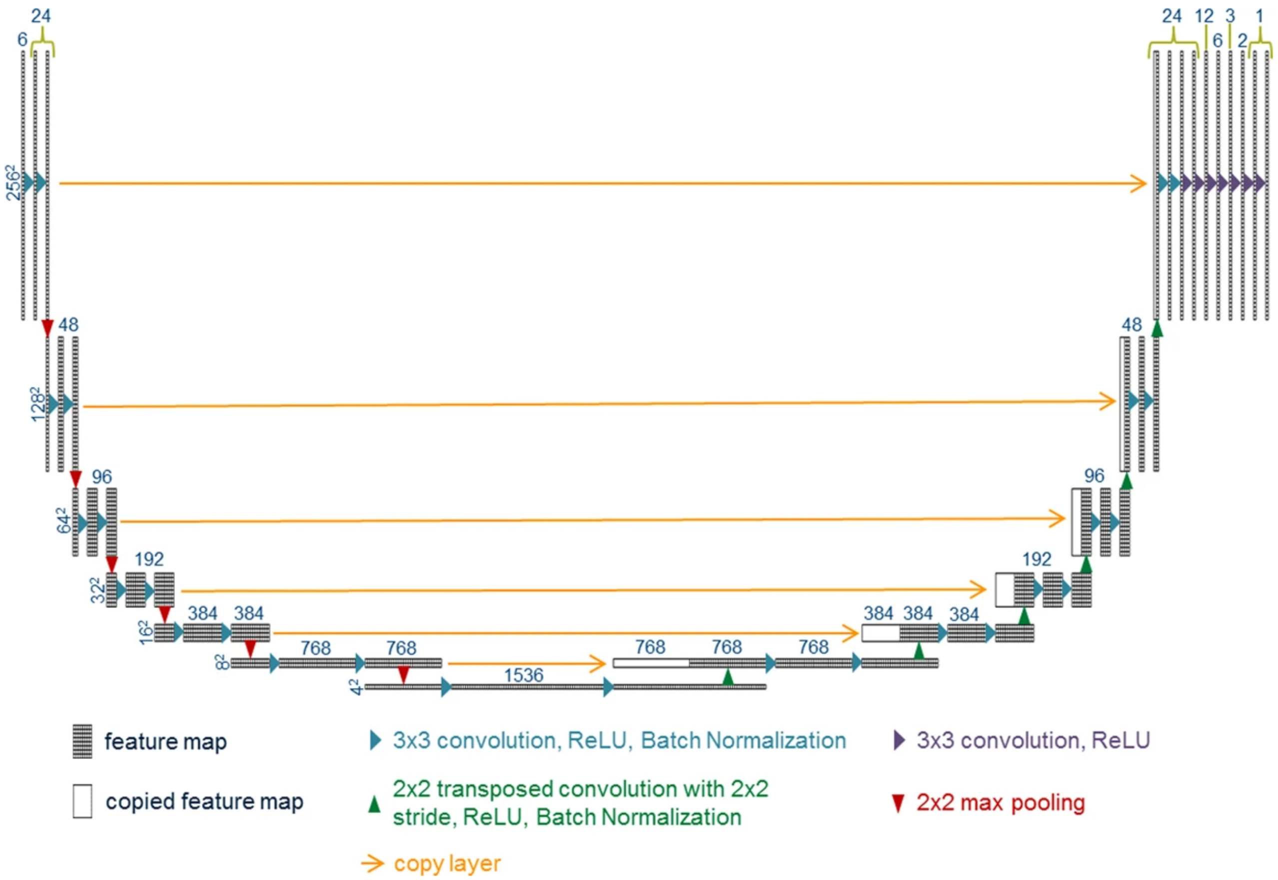
- A modified U-Net – well known deep neural network made of several hierarchical levels - is trained on 2D images treated as channels for a slice-by-slice prediction of the 3D dose distribution of intensity-modulated radiation therapy (IMRT) for prostate cancer patients
  - The ground true dose is the dose distribution explicitly informed about the physics of the beamlets
  - The predicted dose is the dose distribution informed about the physics of the beamlets through the treatment planning data



Nguyen et al. 2019 *Sci. Rep.*

DL-based auto-planning

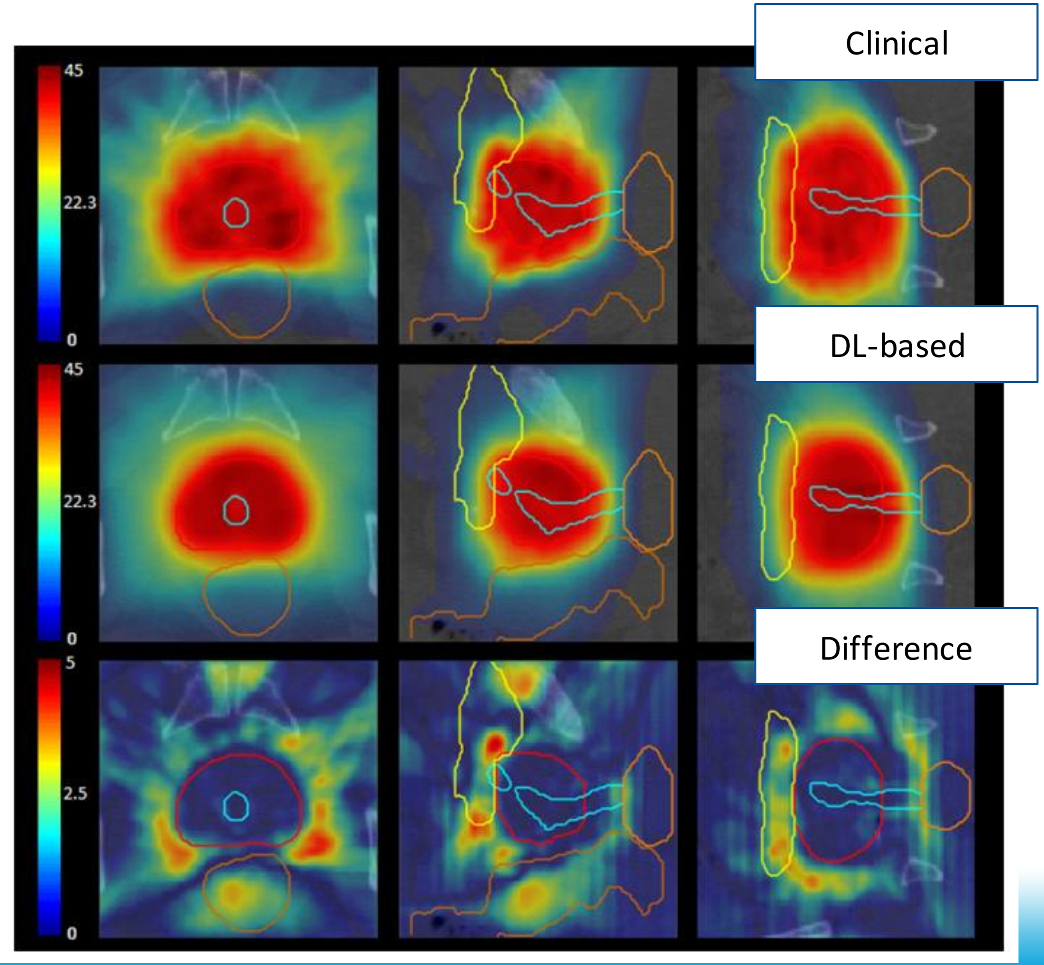
- Contours of PTV and OARs treated as channels
- Fully convolutional network, originally designed for segmentation purposes
- Large number of max pooling operations to allow for the convolution filters to find higher level global features
- Transposed convolution operations (i.e., deconvolution or up-convolution) to return the image to its original size
- Copying the maps from the first half of the U-net in order to preserve the lower-level local features

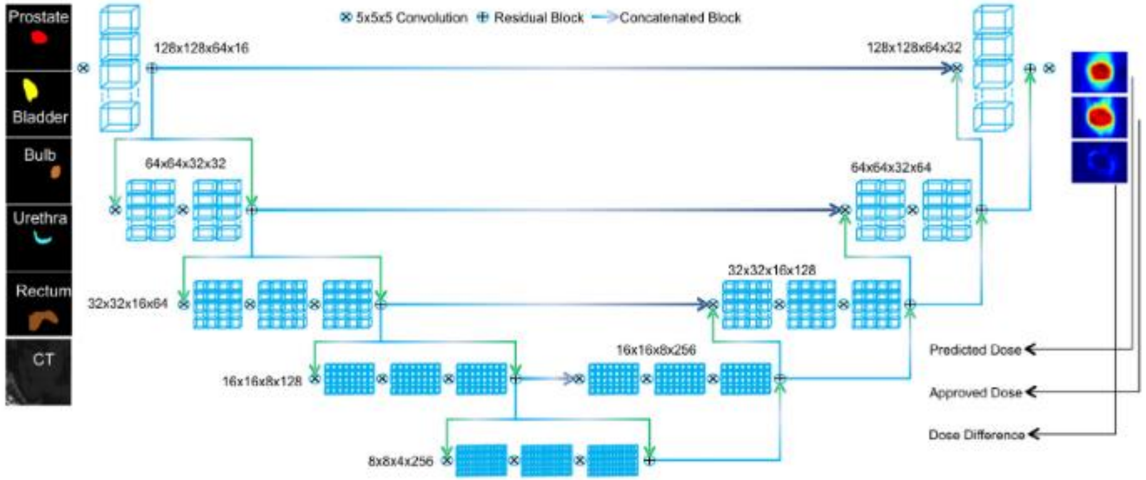


Nguyen et al. 2019 *Sci. Rep.*

DL-based auto-planning

- The DoseNet - deep residual neural network based on convolutional down- and up- sampling - is trained on 3D images for non-coplanar prostate stereotactic body radiotherapy (SBRT) patients

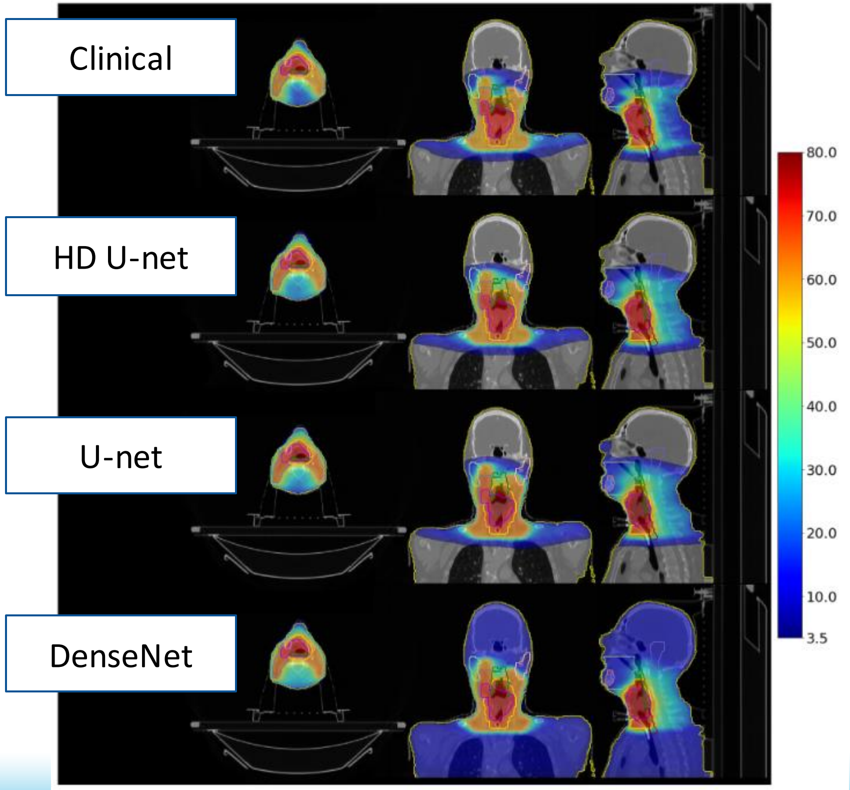


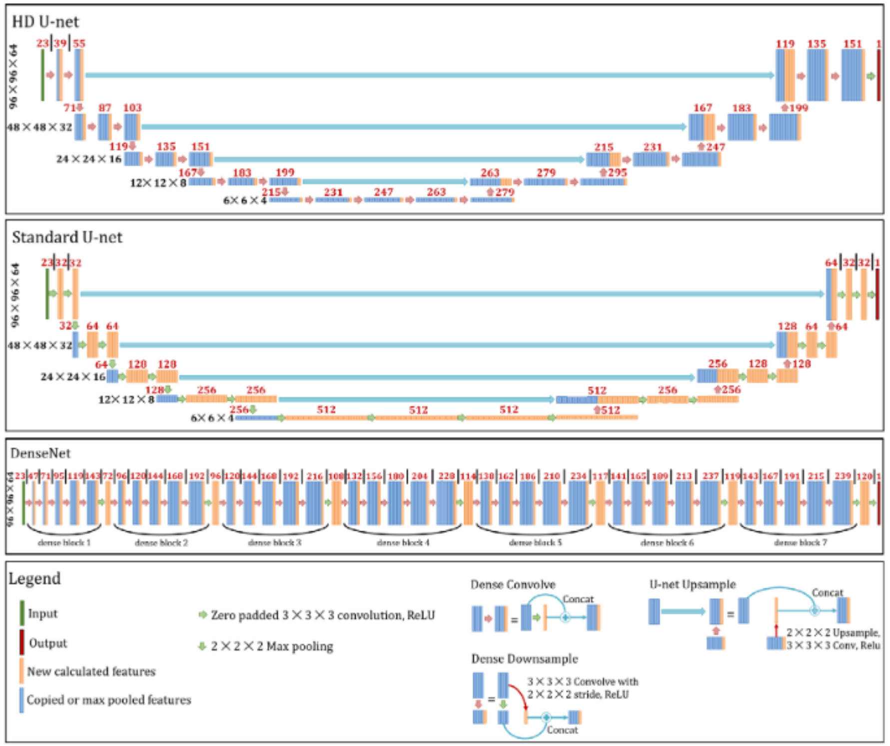


Kearney et al. 2018 *Phys. Med. Biol.*

DL-based auto-planning

- HD U-net - Hierarchically Densely connected U-net based on U-net and DenseNet architectures - is trained o 3D images for head and neck cancer patients treated with volumetric-modulated arc therapy (VMAT)
  - DenseNet is similar to ResNet, but the convolution output is concatenated, rather than added

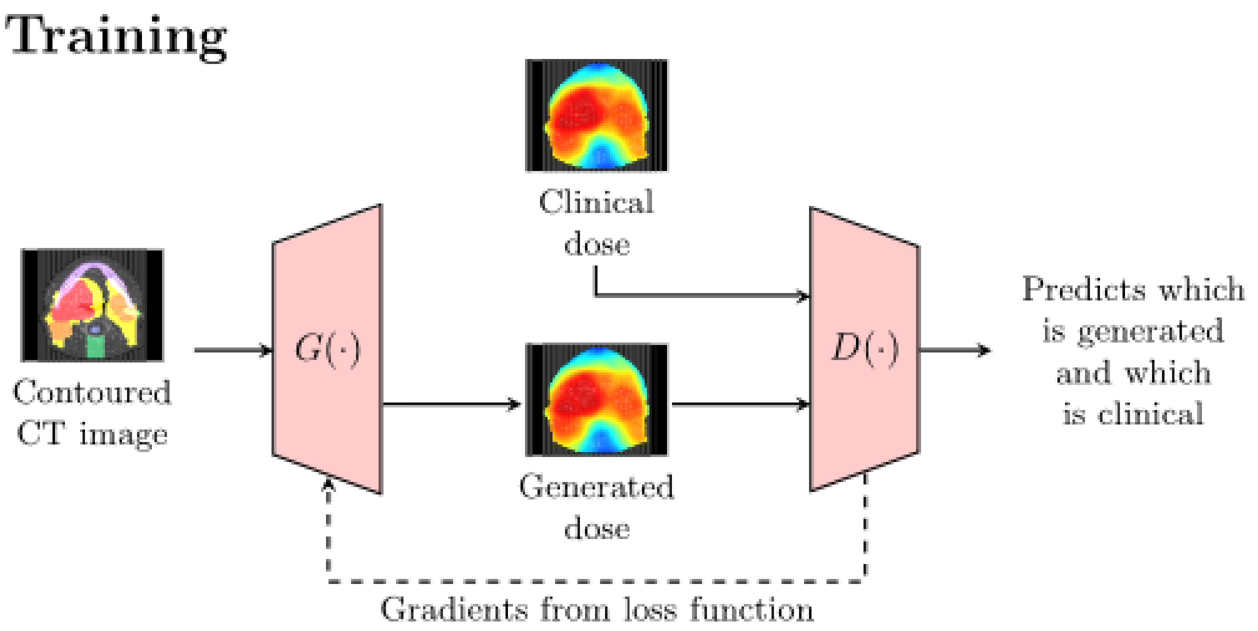




Nguyen et al. 2019 *Phys. Med. Biol.*

DL-based auto-planning

- The GAN architecture is proposed to replicate the role of the treatment planner (the generator that performs the task) and the role of the radiation oncologist that evaluate the treatment planner (i.e., the discriminator that evaluates the performance of the generator)

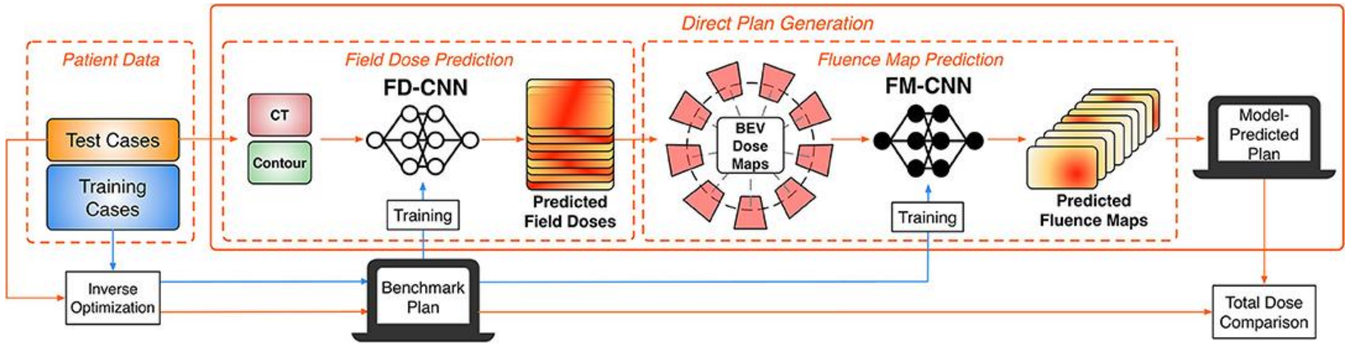


Babier et al. 2020 *Med. Phys.*

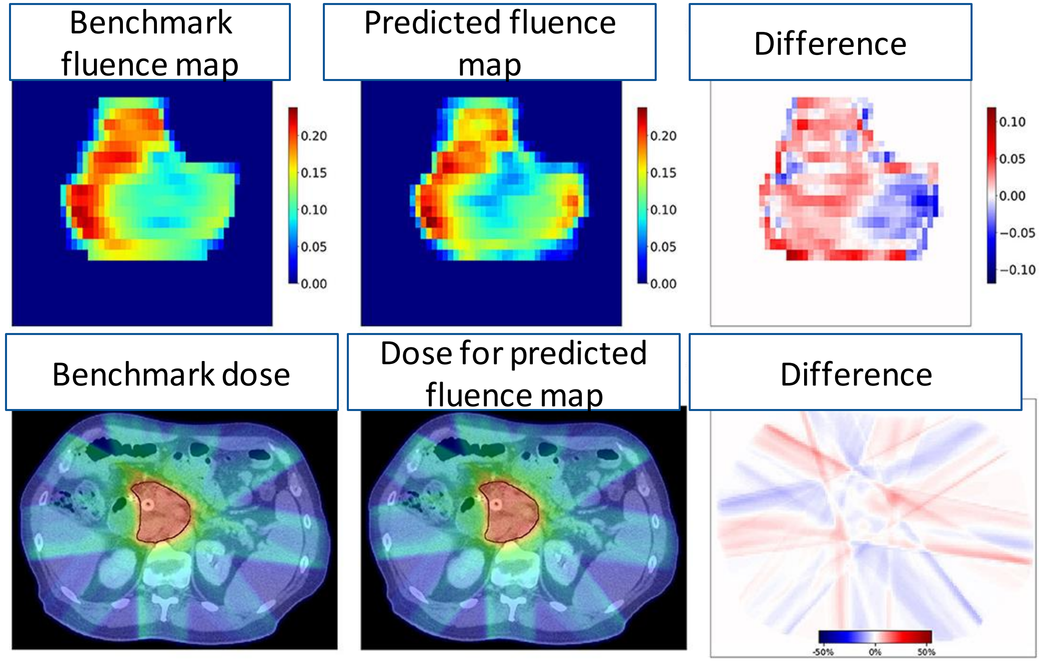
DL-based auto-planning

- AI-based auto-planning is also reported to estimate the radiation beam parameters without inverse optimization (i.e., the inverse problem)

  - The prediction of fluence map per beam requires the predicted field dose projected onto the beam's eye view

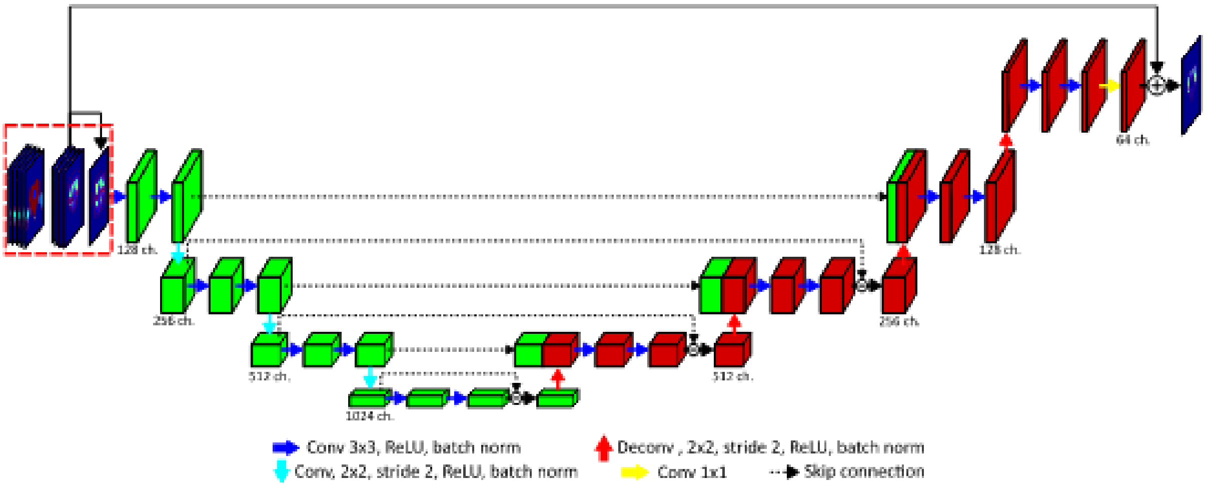


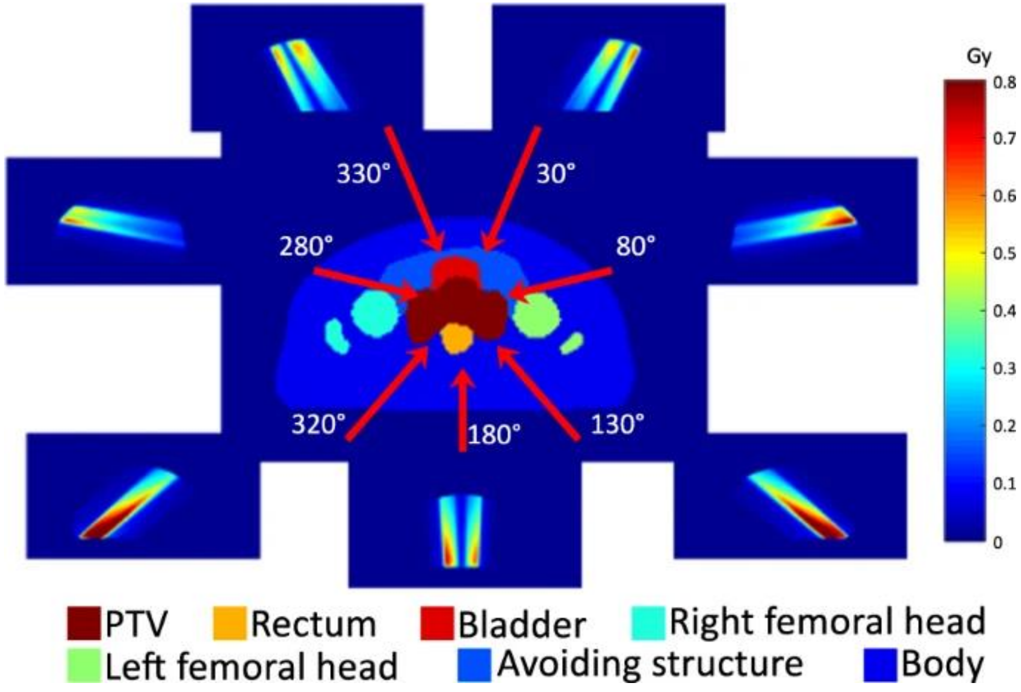
Wang et al. 2020 *Front. Artif. Intell.*



DL-based auto-planning

- The prediction of fluence map per beam requires the contours (i.e., target and OARs) and the volumetric dose distributions viewed from the beam's eye view (BEV) of a single beam

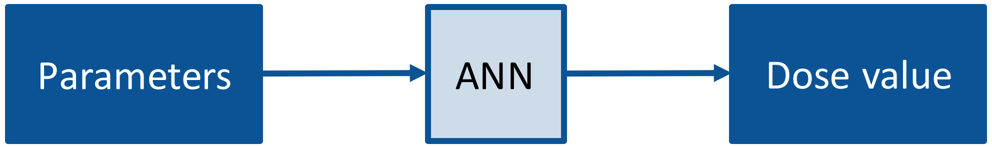




Lee et al. *Sci rep* 2019

ML-based auto-planning

- A 3D dose distribution can be predicted by training artificial neural networks based on patient-specific geometric (i.e., based on CT image and structures) and planning (i.e., the closest distance to planning target volume (PTV) and organ-at-risks (OARs), number of beams irradiating the voxel …) parameters
    - Feed-forward networks with a limited number of layers and nodes
        - Two-layer feed-forward network, ten nodes in the first layer, one single node in the second layer[1]
        - Multiple feed-forward networks with 1-3 hidden layers, each layer with 10-50 nodes[2]
    - Two separated artificial neural networks are trained for voxels within and outside the PTV due to very different dose distribution patterns in the two regions
        - Weak generalizability



[1]Shiraishi, S., & Moore, K. L. (2016). Knowledge-based prediction of three-dimensional dose distributions for external beam radiotherapy. *Medical physics*, 43(1), 378-387.
[2]Campbell, W. G., Miften, M., Olsen, L., Stumpf, P., Schefter, T., Goodman, K. A., & Jones, B. L. (2017). Neural network dose models for knowledge-based planning in pancreatic SBRT. *Medical physics*, 44(12), 6148-6158.

Outlook

- Treatment planning consists in the solution of an inverse problem
- Treatment planning in high precision 3D conformal radiotherapy relies on optimization algorithms (analytical methods can only be applied to geometrically simple cases)
  - Numerical methods are required for geometrically complex cases
    - Many degrees of freedom
    - Many beam-lets or pencil beams
    - High degree of flexibility in dose distribution
- The role of machine learning in treatment planning is relevant to the automation of tasks to support (or accomplish) the planning of the treatment
  - Auto-segmentation
  - Auto-planning as dose estimation (i.e., solution of the forward-problem) or actual inverse problem solution

Exam

- Day: February 25th
- Time:10.00-13.00 tbc
- Room: tbd