Inverse problems and machine learning in medical physics

Image registration -
Machine learning for image registration

Dr. Chiara Gianoli
14/1/2025
chiara.gianoli@physik.uni-muenchen.de

Introduction to deformable image registration

- Deformable image registration is a numerical optimization that aims at determining a spatial transformation that relates positions in one image (reference or fixed image) to the corresponding positions in another image (target or moving image)

- The aim is to compare and integrate the information given by different images

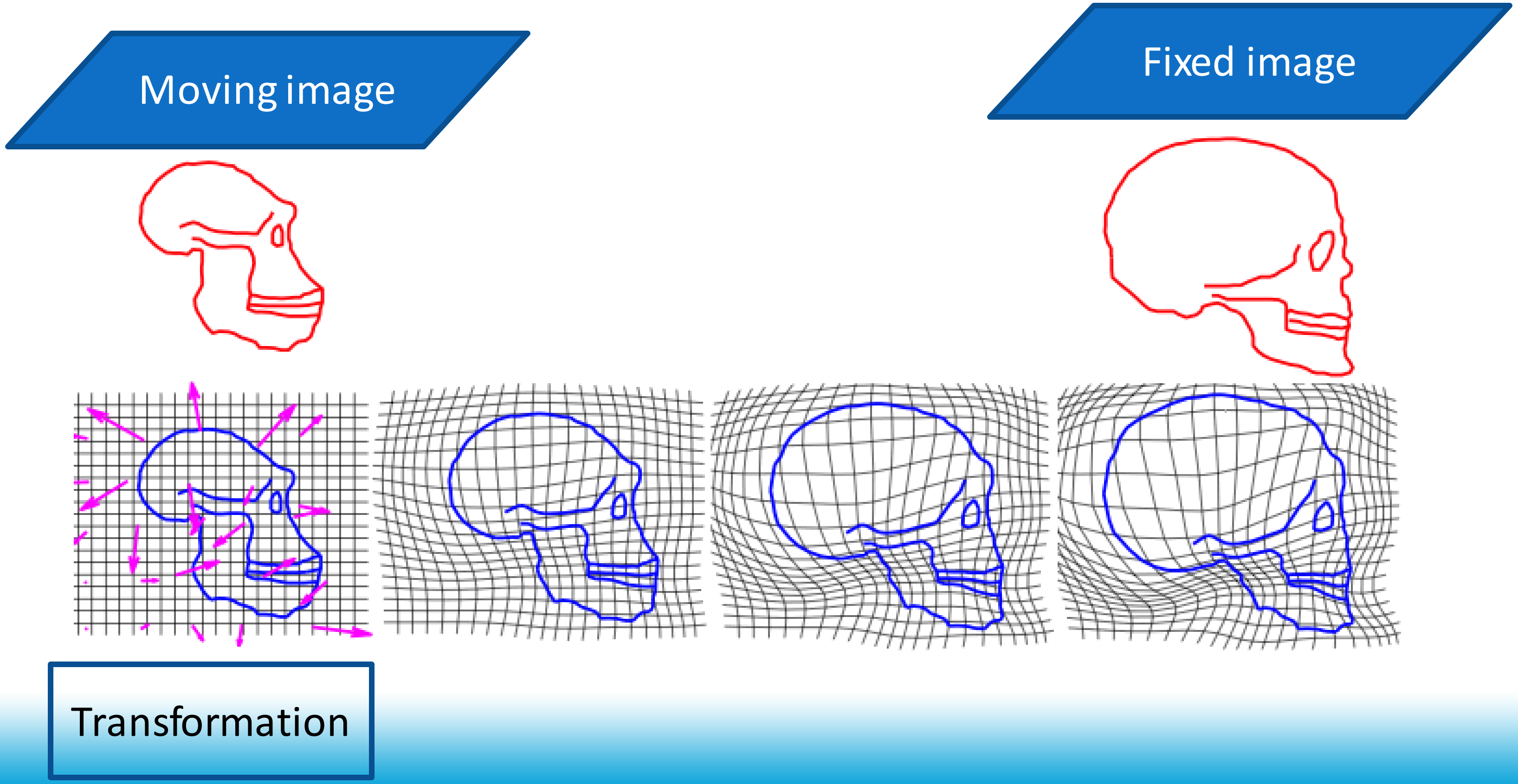
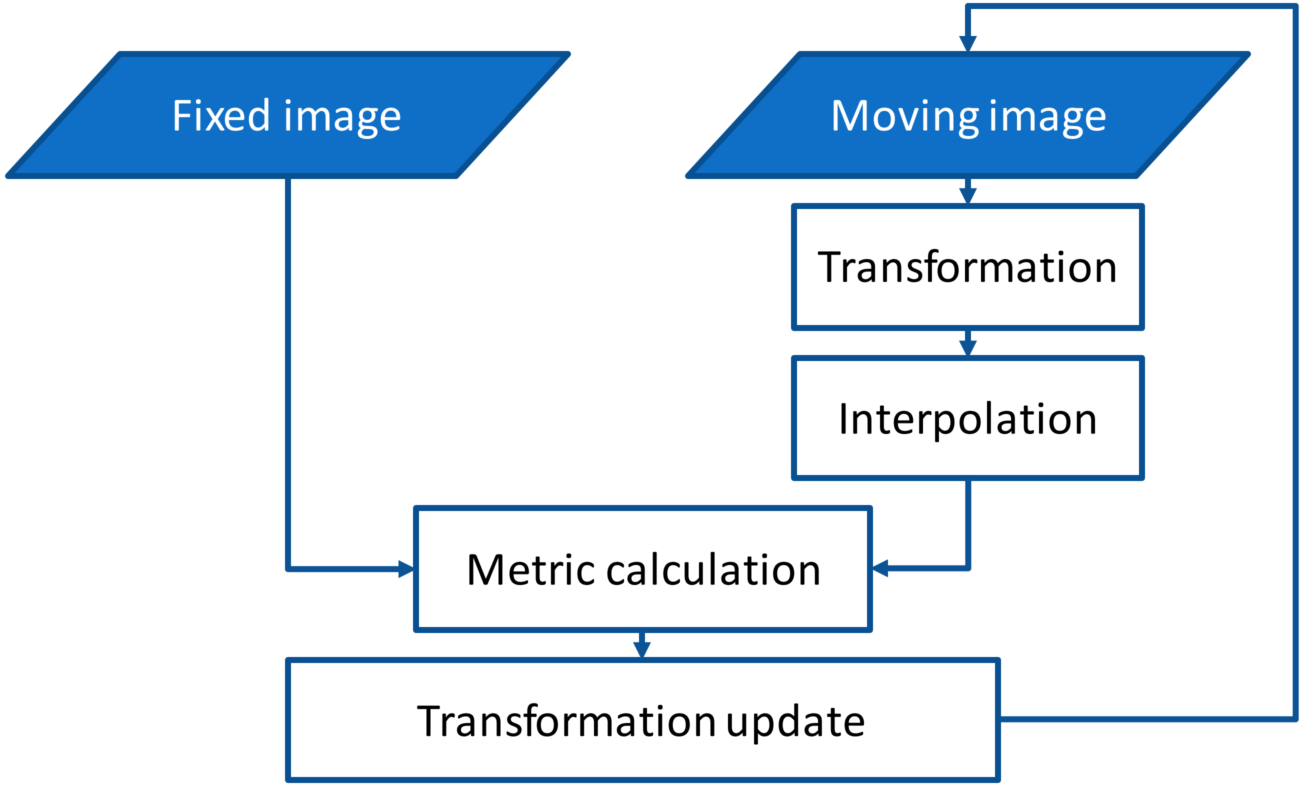


Image registration algorithm

- The numerical optimization is based on the metric, chosen according to the two image modalities, and iterative updates of the transformation parameters
  - Input: fixed image and moving image
  - Output: transformation parameters



- The moving image (undergoing transformation) requires interpolation based on the voxel grid of the fixed image

Metric calculation

- The metric is defined on the gray levels of the two images
  - The gray levels of mono-modality images expresses the same information and they are directly comparable (image differences, mean square errors, root mean square errors, correlation coefficients…)

$$MSE(x,y) = \frac{1}{IJK}\sum_{i,j,k}\left(x_{i,j,k} - y_{i,j,k}\right)^2$$

$$CC(x,y) = \frac{\sum_{i,j,k}\left(x_{i,j,k} \cdot y_{i,j,k}\right)}{\sqrt{\sum_{i,j,k} {x_{i,j,k}}^2 \cdot \sum_{i,j,k} {y_{i,j,k}}^2}}$$

  - The gray levels of multi-modality images expresses different information and “information processing” is need to compare them
    - Mutual information (MI)
    - Normalized Mutual Information (NMI)

Mutual information

- In information theory, the mutual information of two random variables X and Y, $I(X;Y)$, is a measure of the mutual dependence between the two variables

- Mutual information quantifies the "amount of information“ (in bit, if the logarithm base is 2) obtained about one random variable through observing the other random variable

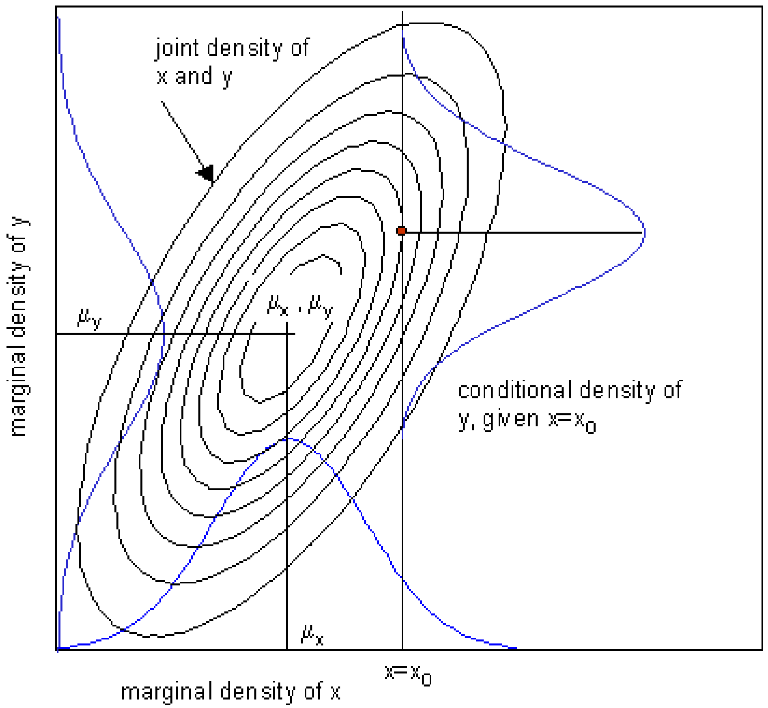
$$MI = I(X;Y) = \sum_{y \in Y} \sum_{x \in X} p(x,y) \log\left(\frac{p(x,y)}{p(x)p(y)}\right)$$

  - $p(x)$ and $p(y)$ are the marginal probability functions of X and Y, respectively

  - $p(x,y)$ is the joint probability function of X and Y

- $I(X;Y) = I(Y;X) \geq 0$ symmetric and non-negative

- $I(X;Y) = 0$ if X and Y are independent random variables, so that $p(x,y) = p(x)p(y)$



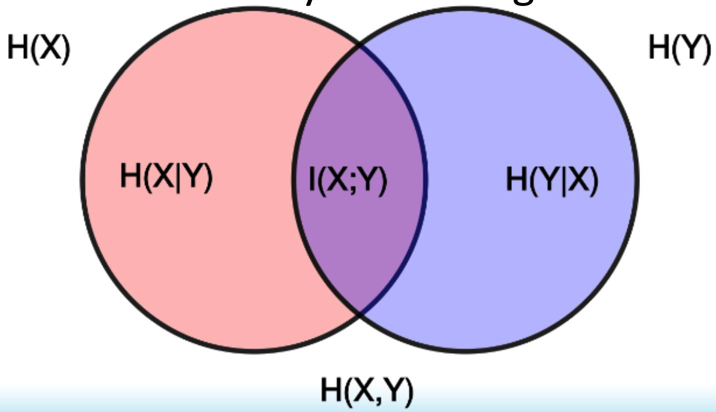
Mutual information

- The mutual information can be expressed in terms of Shannon entropy $H(X)$ as a measure of uncertainty of a random variable

$$H(X) = -\sum_{x \in X} p(x) \log\big(p(x)\big) = \sum_{x \in X} p(x) \log\left(\frac{1}{p(x)}\right)$$

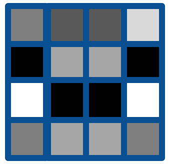
$$I(X;Y) = H(X) - H(X|Y) = H(Y) - H(Y|X) = H(X) + H(Y) - H(X,Y)$$

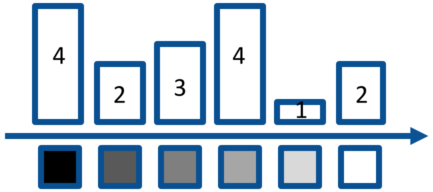
  - $H(X)$ and $H(Y)$ are the marginal entropies
  - $H(Y|X)$ is the conditional entropy of Y given X measuring the amount of uncertainty remaining about Y after X is known (and *vice versa*)
    - $p(y|x)$ and $p(x|y)$ are the conditional probability functions
    - The conditional probability functions and the joint probability function are related according to: $p(x,y) = p(y|x)p(x)$ and $p(x,y) = p(x|y)p(y)$



Mutual information

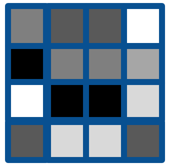
- $p(x)$ (to calculated $H(X)$) is the histogram of gray level occurrences of the fixed image

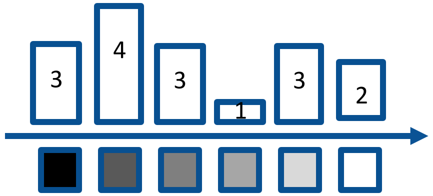




$$x = (1, \ldots N)$$

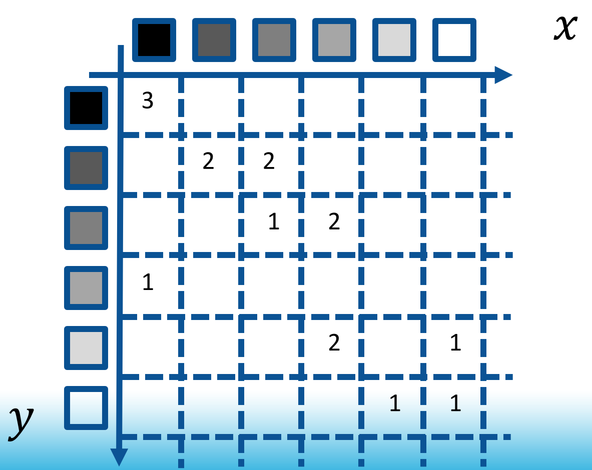
- $p(y)$ (to calculated $H(Y)$) is the histogram of gray level occurrences of the moving image

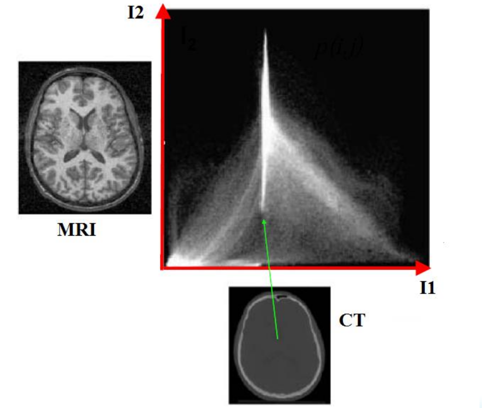




$$y = (1, \ldots N)$$

- $p(x, y)$ (to calculate $H(X, Y)$) is the joint histogram of the gray level occurrences of the two images

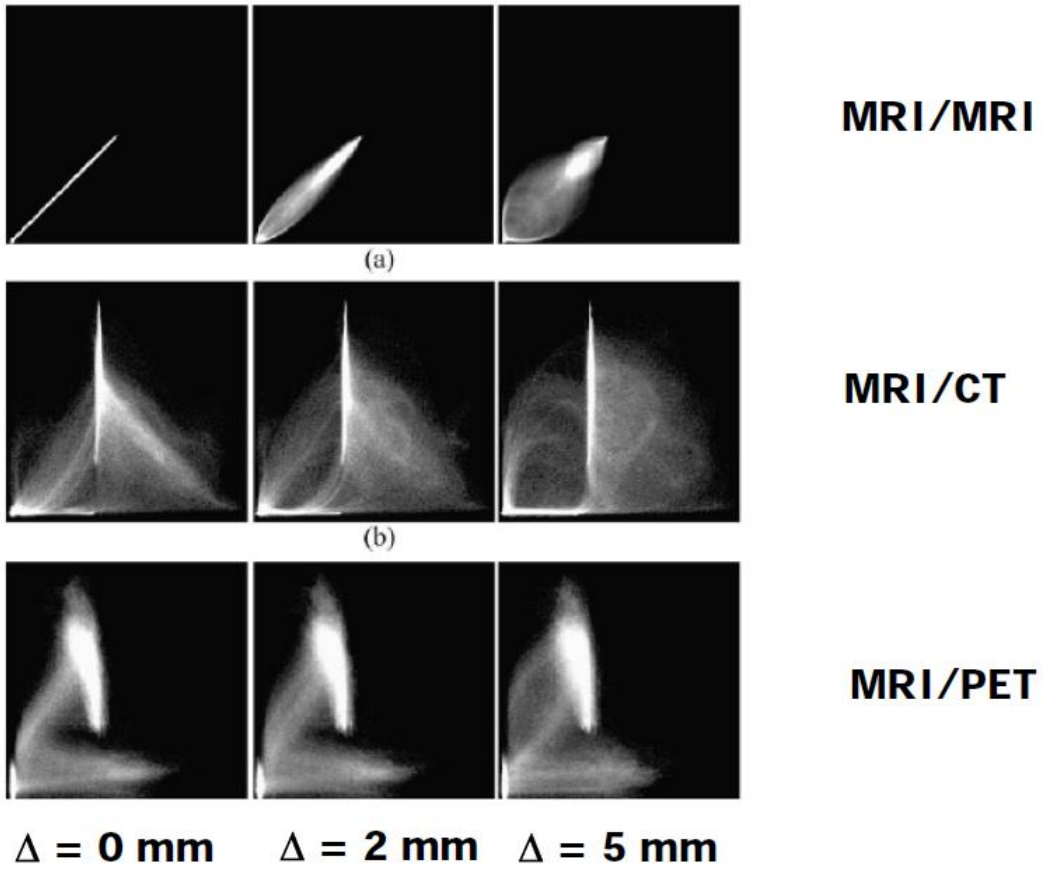




Mutual information

- Realistic joint histograms for different rigid translations (Δ)
- The mutual information is a measure of the joint histogram dispersion
- To reduce the influence of background, the normalized mutual information can be calculated

$$NMI = \frac{H(X) + H(Y)}{H(X,Y)}$$



Transformation parameters

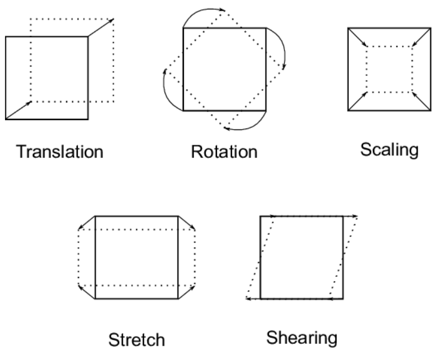
- The parametrization for rigid registration is given by 6 parameters (3 for translation and 3 for rotations)

$$T = \left(\begin{array}{ccc|c} & R(\Omega,\Phi,K) & & \begin{array}{c} X \\ Y \\ Z \end{array} \\ \hline 0 & 0 & 0 & 1 \end{array} \right)$$

$$\underbrace{\begin{pmatrix} 1 & 0 & 0 & X \\ 0 & 1 & 0 & Y \\ 0 & 0 & 1 & Z \\ 0 & 0 & 0 & 1 \end{pmatrix}}_{\text{Translation}} \underbrace{\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos(\Omega) & \sin(\Omega) & 0 \\ 0 & -\sin(\Omega) & \cos(\Omega) & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}}_{\text{Rotation across x}} \underbrace{\begin{pmatrix} \cos(\Phi) & 0 & \sin(\Phi) & 0 \\ 0 & 1 & 0 & 0 \\ -\sin(\Phi) & 0 & \cos(\Phi) & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}}_{\text{Rotation across y}} \underbrace{\begin{pmatrix} \cos(K) & \sin(K) & 0 & 0 \\ -\sin(K) & \cos(K) & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}}_{\text{Rotation across z}}$$

  - The order of operations (translation and rotations) matters…

- The parametrization for affine transformation is given by 12 parameters (3 for translation, 3 for stretching/scaling, 3 for shearing and 3 for rotations)



Translate

$$\begin{bmatrix} 1 & 0 & T_x \\ 0 & 1 & T_y \\ 0 & 0 & 1 \end{bmatrix}$$

Rotate

$$\begin{bmatrix} \cos(\Theta) & -\sin(\Theta) & 0 \\ \sin(\Theta) & \cos(\Theta) & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

Scale

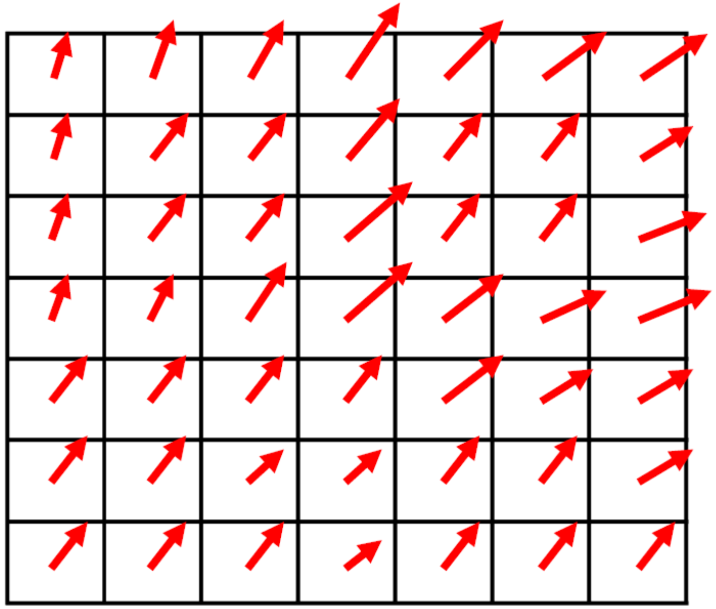
$$\begin{bmatrix} S_x & 0 & 0 \\ 0 & S_y & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

Shear

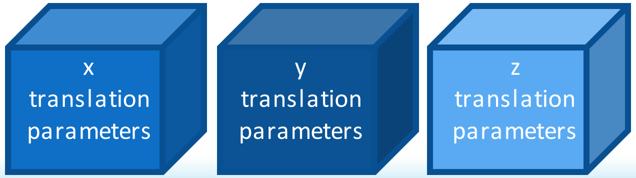
$$\begin{bmatrix} 1 & Sh_x & 0 \\ Sh_y & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

Transformation parameters

- The parametrization for deformable registration is given by 3 translational parameters for each voxel (2 translational parameters for each pixel)

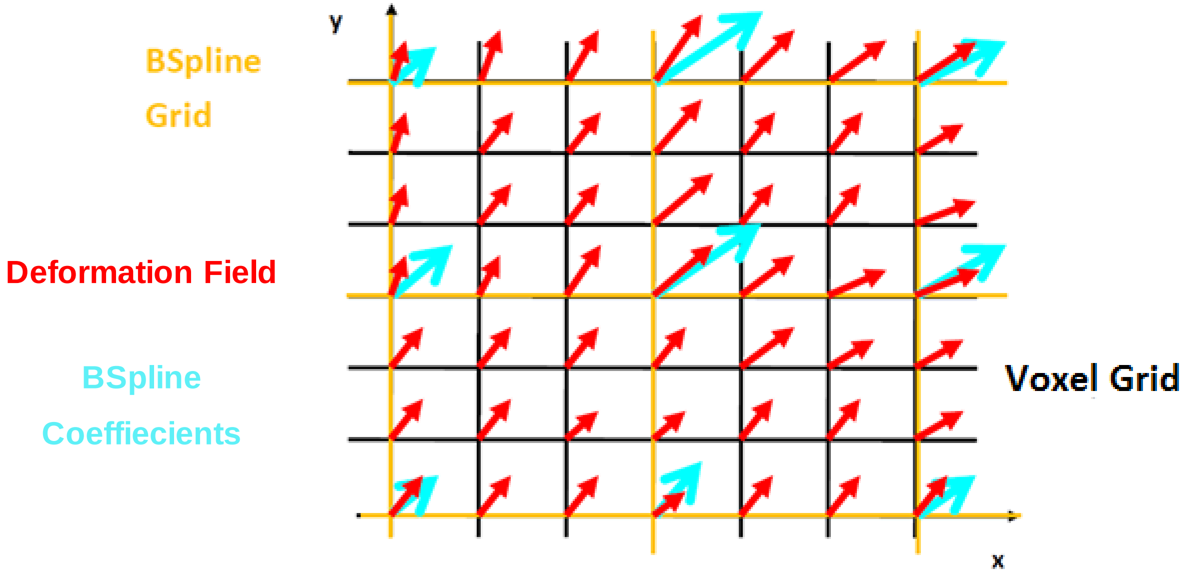


- The parameter matrix, provided with and additional dimension with respect to the image, is typically referred to as "deformation field"



Transformation parameters

- To reduce the amount of parameters, the deformation field can be defined in control points (grid coarser than the voxel grid) and modeled elsewhere in terms of coefficients and basis functions

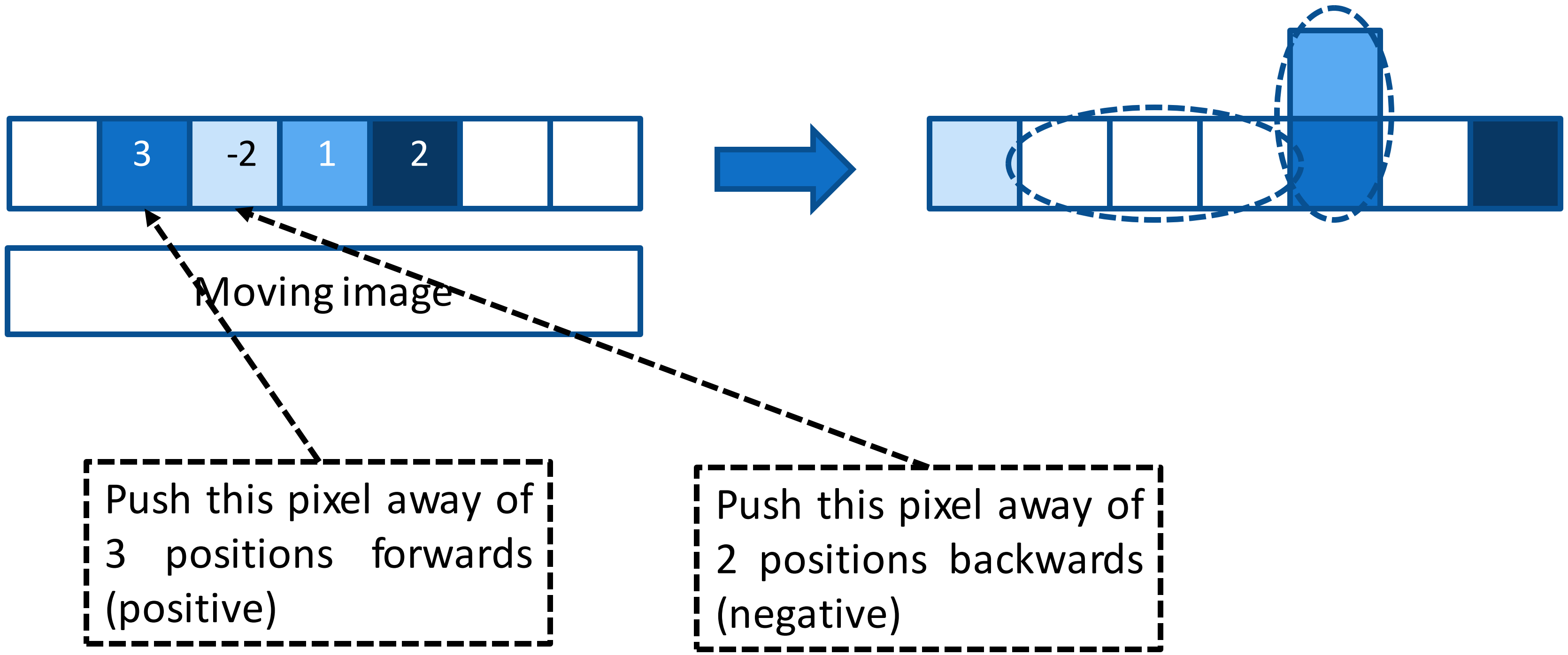


- The B-spline coefficients $P(i,j,k)$ are optimized on the B-spline grid in (i,j,k) and subsequently calculated according to the B-spline basis functions (pre-calculated) in (x,y,z)

$$v(x,y,z)=\sum_{i=1}^{4}\sum_{j=1}^{4}\sum_{k=1}^{4}P(i,j,k)B_i(x)B_j(y)B_k(z)$$

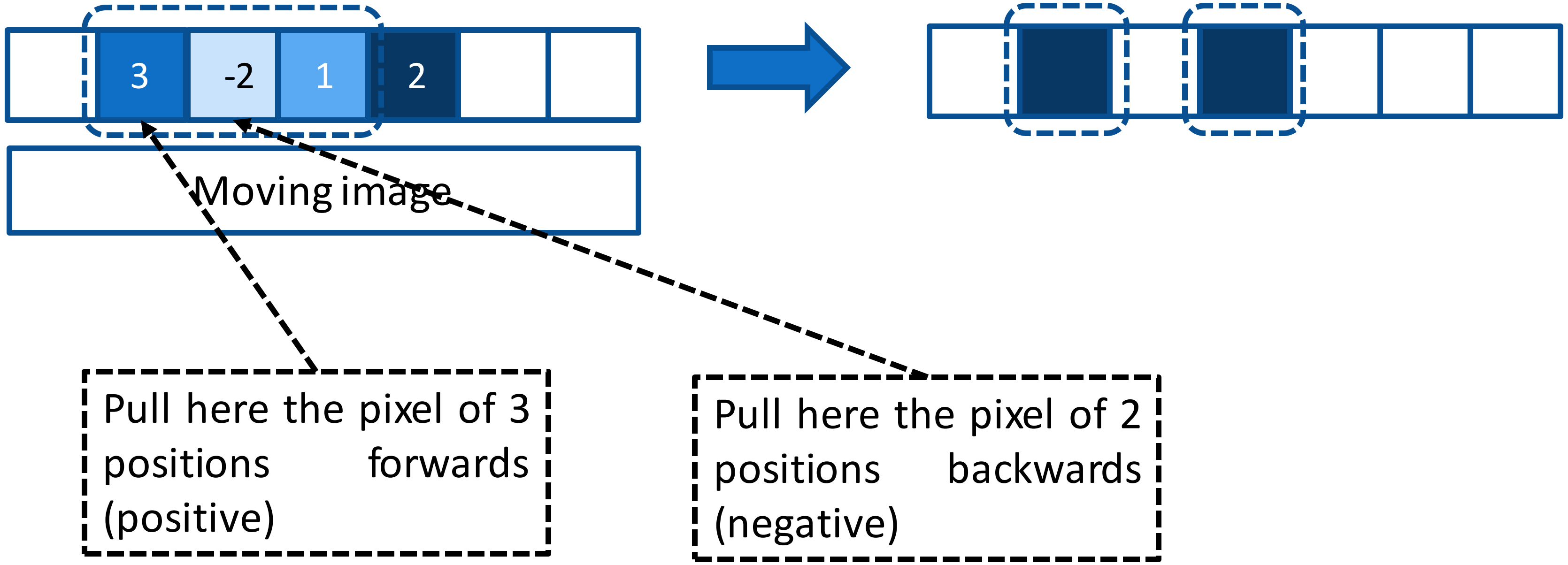
Formalism in parameterization

- Push-forward formalism: the deformation indicates the voxel of the moving image to be pushed-forward
  - The gray levels of the moving image are maintained but can create hole and overlap



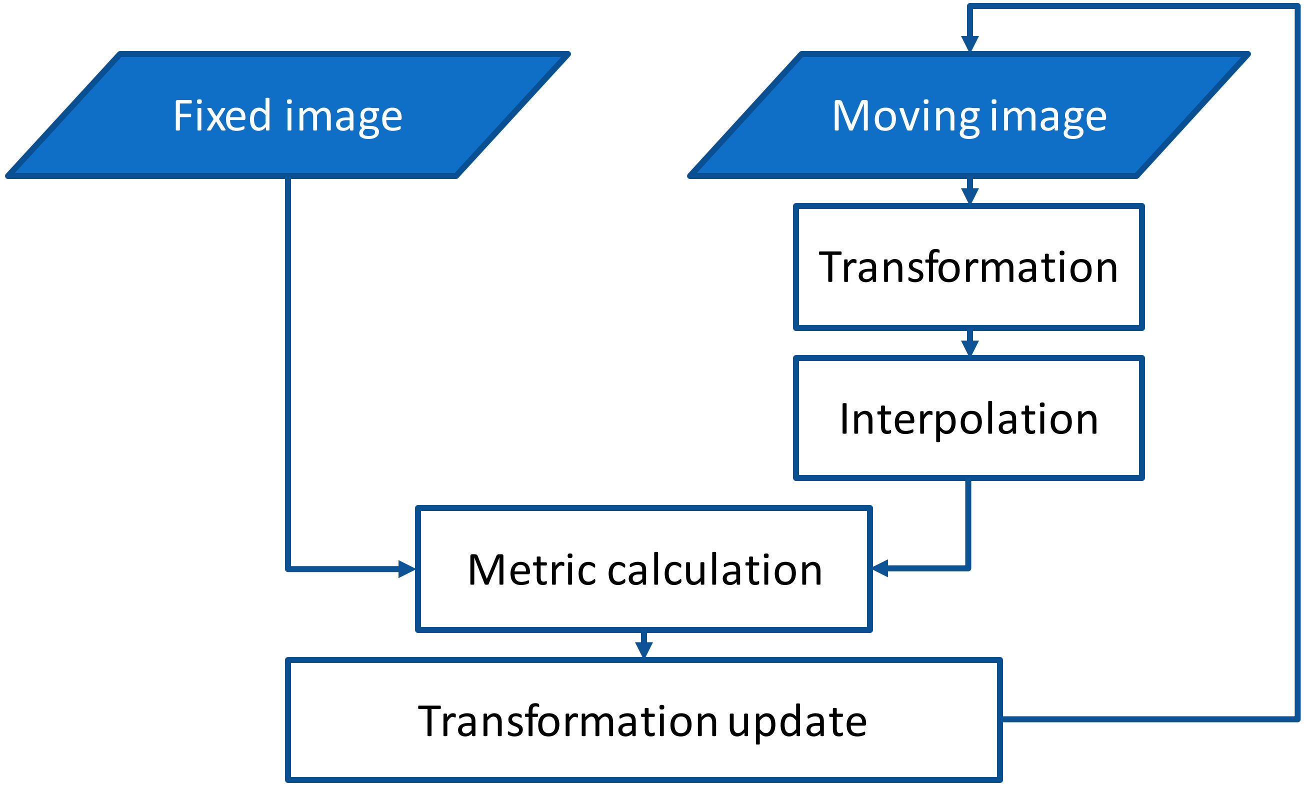
Formalism in parameterization

- Pull-back formalism: the deformation field indicates the voxel of the moving image to be pulled-back
  - The gray levels of the moving image are not maintained as they can be repeated or missed but no hole and overlap are created



Optimization

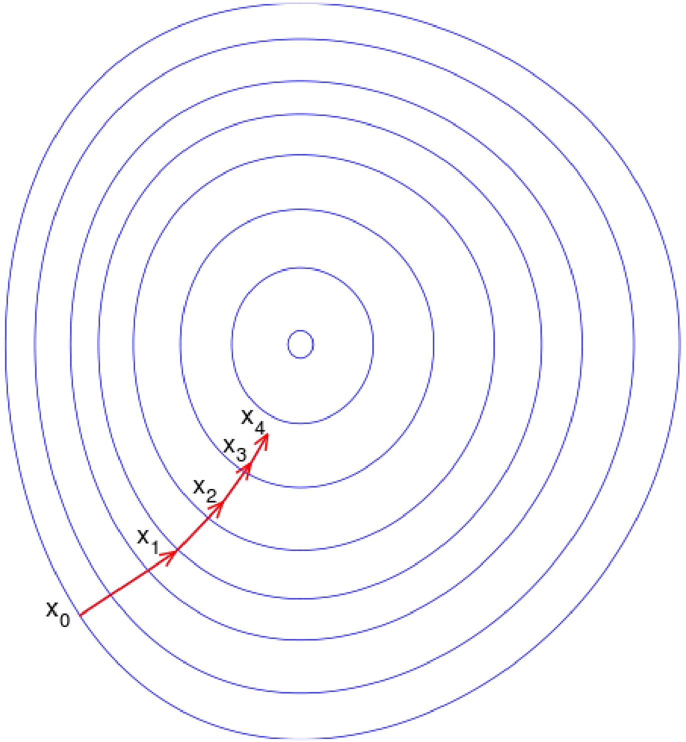
- The numerical optimization aims at finding the "best" transformation parameters according to an objective function defined by the chosen metric



- Different numerical optimization algorithms can be adopted for deformable image registration

Gradient descent

- First-order iterative optimization algorithm for finding the local minimum of the objective function using gradient descent
- Update steps proportional to the negative of the gradient (or approximate gradient) of the objective function at the current point
  - Initialize x_0
  - Compute $\nabla f(x_n)$
  - Update $x_{n+1} = x_n - \alpha \nabla f(x_n)$ (α is the step size)
  - Stop (stopping criteria on n or x_n and x_{n+1})



Newton method

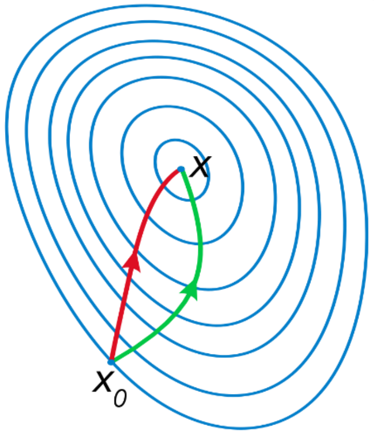
- Second-order iterative optimization algorithm for finding the local minimum of the objective function using first and second derivatives of the objective function

$$f(x_n + \Delta x) \approx f(x_n) + f'(x_n)\Delta x + \frac{1}{2} f''(x_n)\Delta x^2$$

$$\frac{d\left(f(x_n) + f'(x_n)\Delta x + \frac{1}{2} f''(x_n)\Delta x^2\right)}{d\Delta x} = f'(x_n) + f''(x_n)\Delta x = 0$$

$$\Delta x = -\frac{f'(x_n)}{f''(x_n)}$$

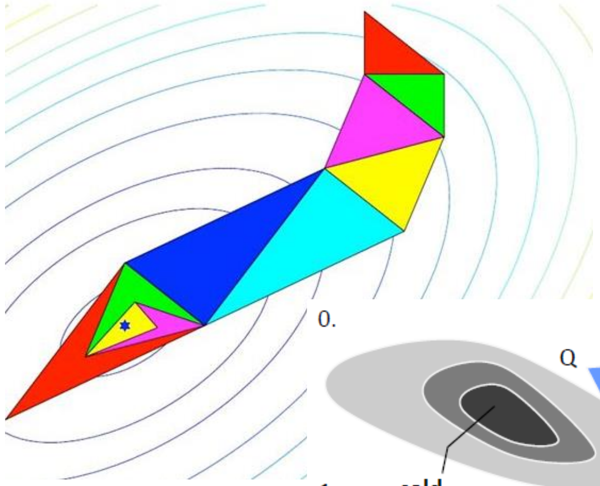
$$x_{n+1} = x_n + \Delta x$$

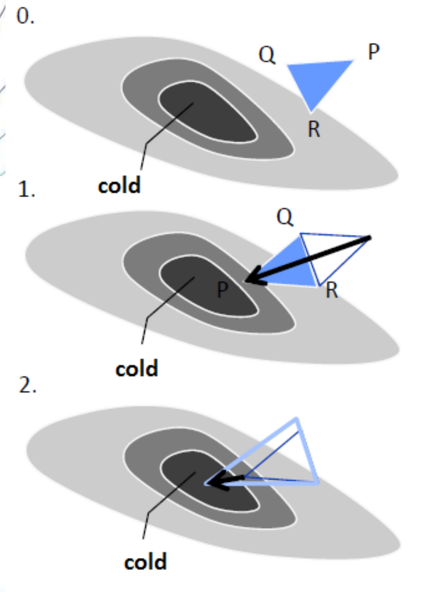


- Update steps account for also for curvature (second derivative) of the objective function at the current point

Nedler-Mead algorithm

- “Direct” (without derivative calculation) iterative optimization algorithm for finding the local minimum of the objective function based on adjustment of the simplex
  - Simplex has $n+1$ vertices, each vertex is described by $x_n \in R^n$
  - Update of the vertex position x_n based on the evaluation of the objective function $f: R^n \rightarrow R$ at each vertex position
  - Replacement of the worst vertex by the reflected centroid of the remaining n vertices across the opposite best face of the simplex
  - Expansion or contraction and shrinkage
  - The simplex moves towards the minimum of the objective function

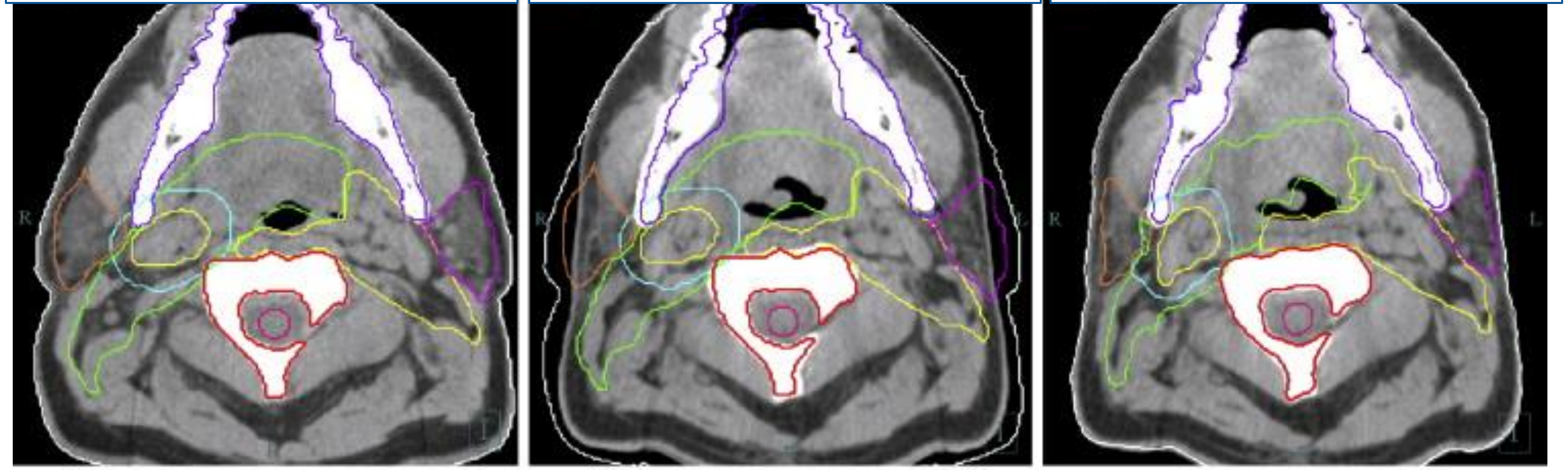




Deformable image registration for treatment adaptation

- Deformable image registration is applied between the treatment planning CT (moving image) and the “CT of the day” (fixed image) to obtain a deformation field to be applied to the treatment planning CT and the contours (*contour propagation*) for adaptive radiation therapy

Treatment planning | CT of the day | CT of the day



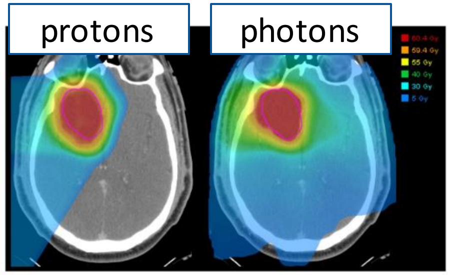
Contours | Rigid alignment of contour | Contour propagation

Schwartz et al. 2012 *Int. J. Radiat. Oncol. Biol. Phys.*

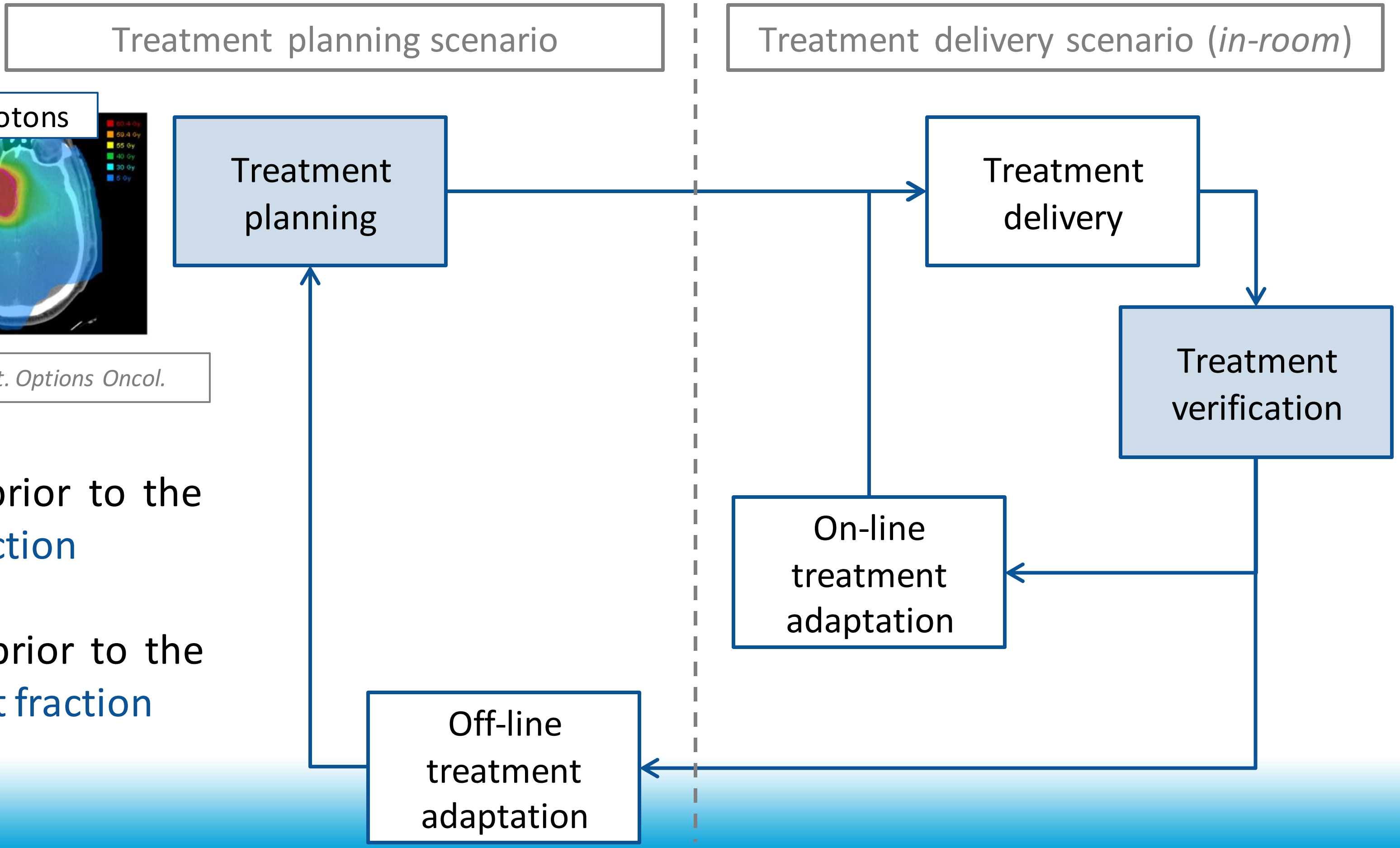
- Radiotherapy is typically administered in a fractionated treatment course entailing a few days (for hypofractionated treatment regiments) up to several weeks (for standard fractionation) of almost daily dose applications

Adaptive radiation therapy

- Adaptive radiation therapy is a closed-loop radiotherapy workflow, applicable both to conventional photon therapy and to ion beam therapy, where the treatment is adapted to the patient based on anatomical information provided by in-room imaging within the fractionated treatment course



Winter et al. 2023 *Curr. Treat. Options Oncol.*



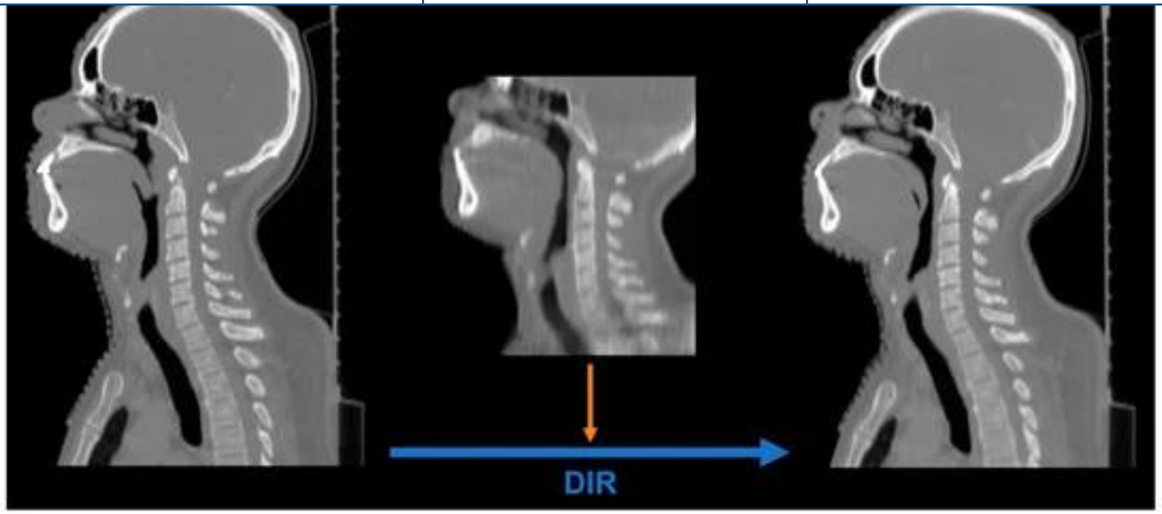
- On-line – prior to the current fraction
- Off-line – prior to the subsequent fraction

Adaptive radiation therapy

- When in-room volumetric imaging such as cone beam computed tomography (CBCT) is available, the treatment plan can be updated on-line based on this imaging

- CBCT is clinically used for patient position verification and treatment adaptation based on the in-room CBCT requires correction for scattering and noise, either based on deformable image registration (DIR) as virtual CT image, along with contour propagation, or scatter correction techniques

Treatment planning CT	CBCT	Virtual CT



Nesteruk et al. 2021 *Cancers*

- The role of AI is relevant to the definition of models for converting the in-room imaging into a suitable image for treatment planning while accounting for the occurred anatomical changes, as synthetic CT image

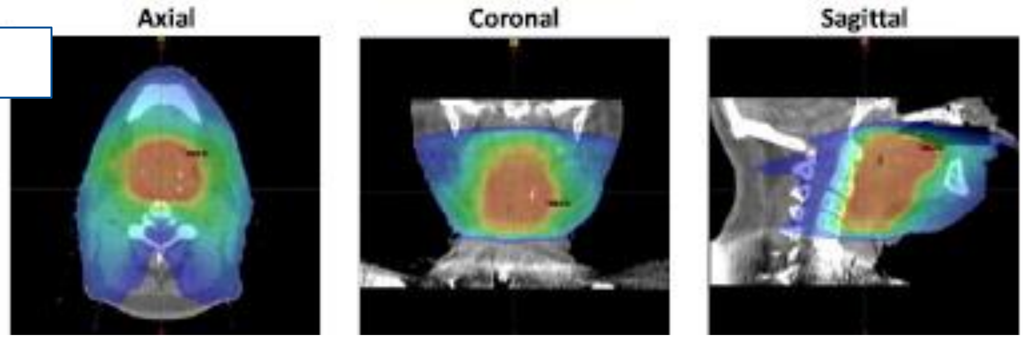
Synthetic CT image

- A cycle generative adversarial network (CycleGAN) is proposed as unsupervised learning without fully relying on paired CT-CBCT images (supervised training is very difficult in these scenarios)

- Synthesized CT images are obtained from CBCT images for adaptive radiation therapy with artifacts removed or greatly reduced and intensities corrected while keeping the anatomical accuracy

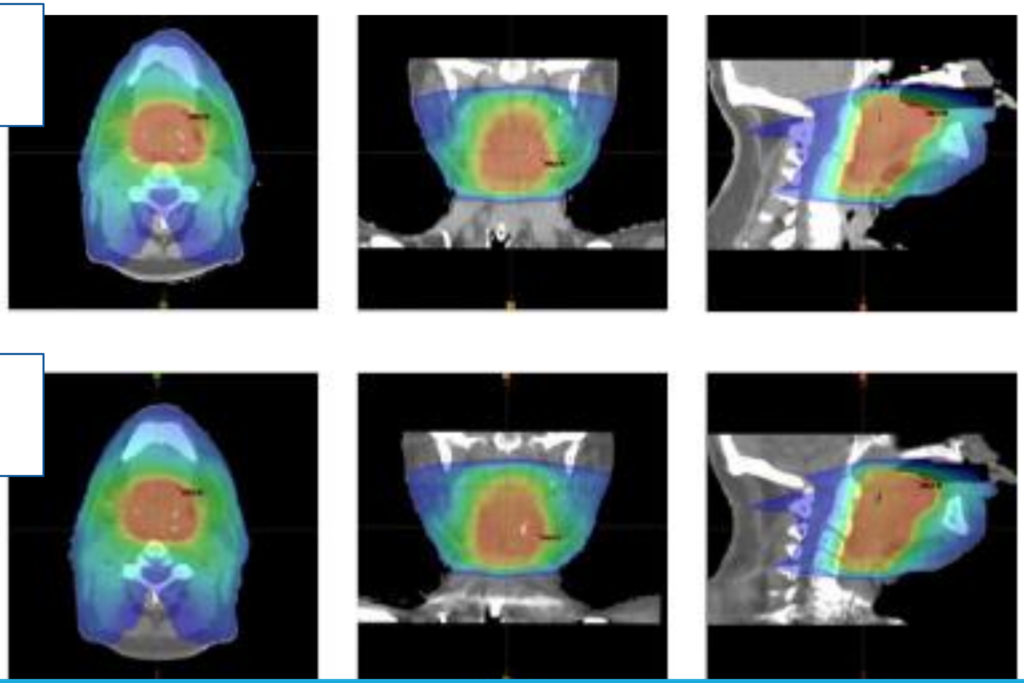
  - The synthesized CT images are used for dose calculation in adaptive radiation therapy

CBCT



deformed CT

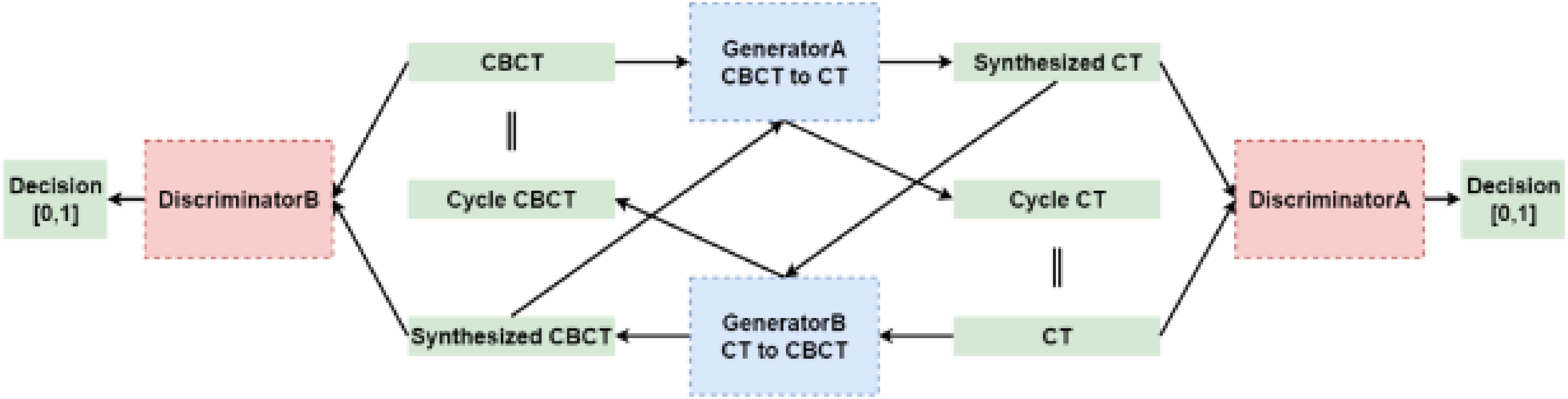
CycleGAN CT



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

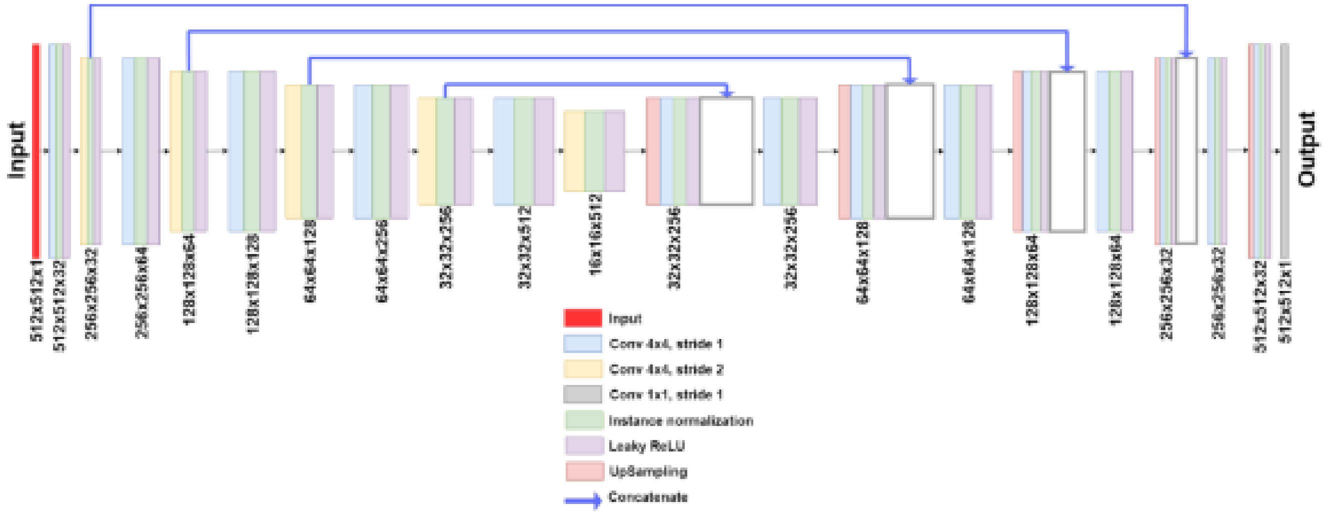
Synthetic CT image

- Cycle-GAN includes two generators
  - The first aims to map from CBCT to CT
  - The second aims to map from CT to CBCT
- Cycle-GAN includes two discriminators
  - The first aims to distinguish real CT from fake CT
  - The second aims to distinguish real CBCT from fake CBCT

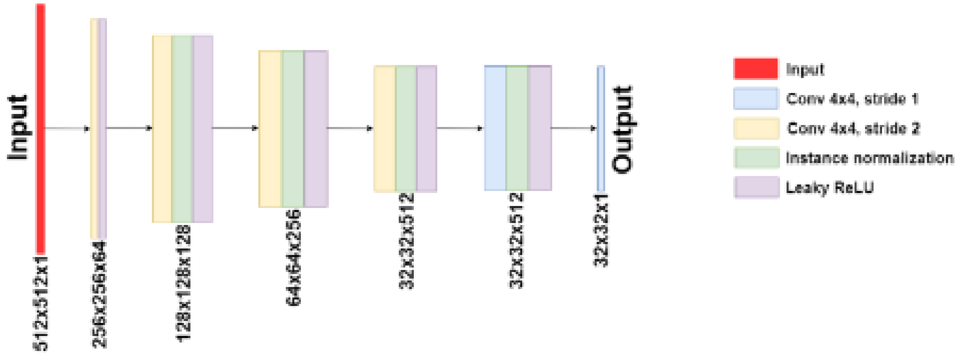


Synthetic CT image

- Generator (U-Net)
  - U-Net-type architectures with encoder/decoder arms
  - Fully convolutional framework provided with skip connections



- Discriminator (patchGAN)
  - Encoder classifiers



Synthetic CT image

- Two cycles are included in Cycle-GAN
  - In the first cycle, the CBCT is used as input to the first generator, which generates the synthetized CT. Then, the second generator takes the synthetized CT as input and generates the cycle CBCT, which is supposed to be equal to the CBCT
    - Meanwhile, the first discriminator identifies real and synthetized CT images
      - The CT label is 1 and the synthetized CT label is 0
  - In the second cycle, the CT is used as input to the second generator which generates the synthetized CBCT. Then, the first generator takes the synthetized CBCT as input and generates the cycle CT, which is supposed to be equal to the CT
    - Meanwhile, the second discriminator identifies real and synthetized CBCT images
      - The CBCT label is 1 and the synthetized CBCT label is 0

Synthetic CT image

- The Cycle-GAN is a variant of the GAN that introduces a cycle-consistency loss using two generators and two discriminators

- GAN training proceeds in alternating:

  - (1) the discriminator is trained for one or more epochs while keeping the generator constant (i.e., minimization of the discriminator loss) to optimize the faking capability of the generator

  - (2) the generator is trained for one or more epochs while keeping the discriminator constant (i.e., maximization of the adversarial loss) to optimize the fooling capability of the discriminator

  - Repeat (1) and (2)

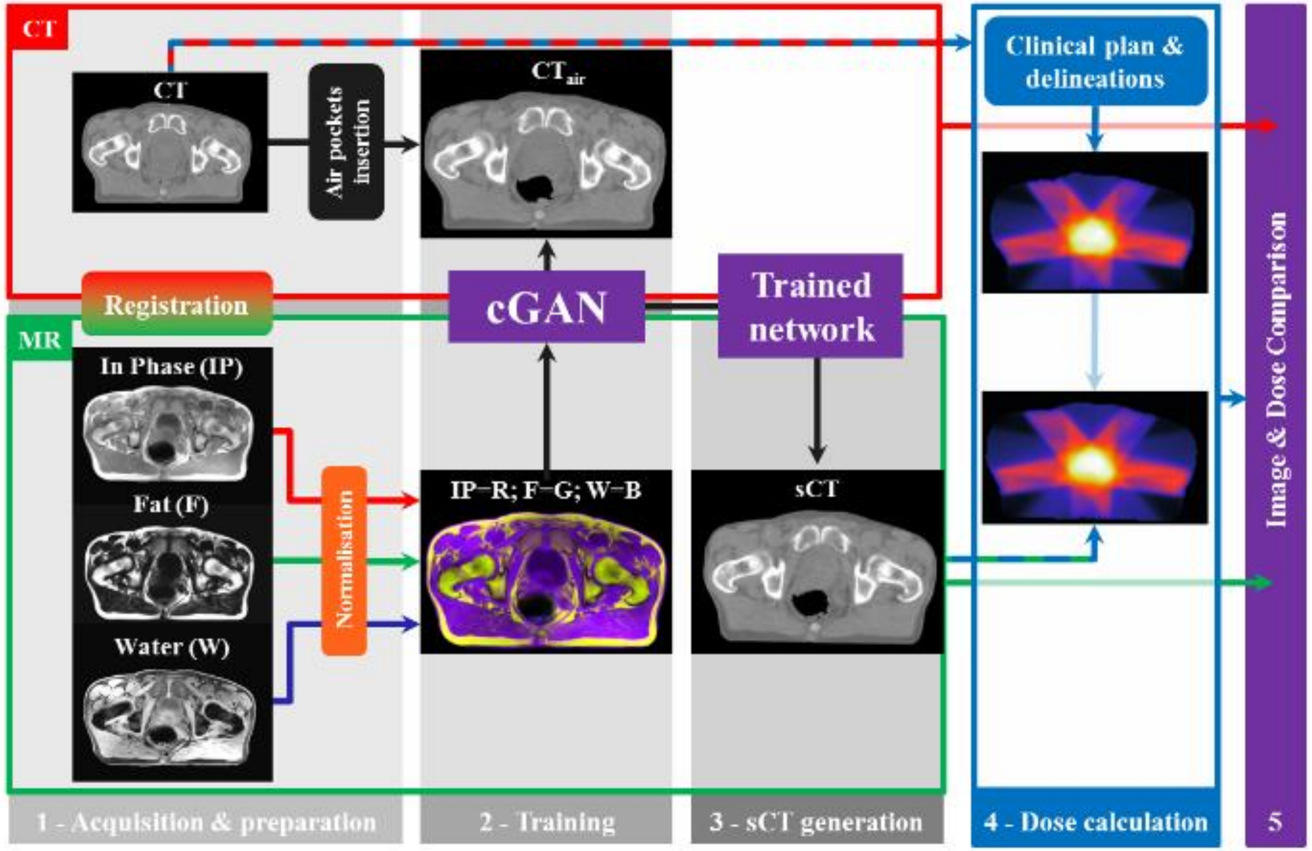
MRI-only radiotherapy

- The soft tissue contrast offered by MRI is exploited without recurring to inter-modality image registration

- The patient exposure to ionizing radiation is reduced, along with treatment cost and workload, thus enabling on-line daily re-planning in MR-guided radiotherapy systems

Tenhunen, M., Korhonen, J., Kapanen, M., Seppälä, T., Koivula, L., Collan, J., ... & Visapää, H. (2018). MRI-only based radiation therapy of prostate cancer: workflow and early clinical experience. Acta Oncologica, 57(7), 902-907.

  - Conditional generative adversarial network (cGAN) as supervised version of GAN based on paired images (i.e., DIR)
    - Generator based on U-Net architecture, producing data consistent with the "condition"
    - Discriminator based on convolutional "PatchGAN" classifier, receiving information about the "condition"



Maspero, M., Savenije, M. H., Dinkla, A. M., Seevinck, P. R., Intven, M. P., Jurgenliemk-Schulz, I. M., ... & van den Berg, C. A. (2018). Dose evaluation of fast synthetic-CT generation using a generative adversarial network for general pelvis MR-only radiotherapy. Physics in Medicine & Biology, 63(18), 185001.

Outlook

- Image registration is used at different stages in radiation oncology
  - Patient positioning is based on rigid registration
  - Multi-modality treatment planning, atlas-based segmentation in treatment planning and treatment adaptation are based on deformable image registration
- Deep learning is adopted to replace the role of deformable image registration with advantages in term of quality and efficiency
  - Automatic contouring (i.e., auto-segmentation) of targets and organs at risks for treatment planning
  - Treatment adaptation as “synthetic CT image” generation based on CBCT imaging
    - Image quality of the CBCT is inappropriate for treatment planning (compromised by artifacts and scattering effects)
  - Treatment planning and treatment adaptation as “synthetic CT image” generation based on MRI (i.e., “MRI-only radiotherapy”)