END-TO-END 3D/2D REGISTRATION WITH DEEP LEARNING USING SYNTHETIC DATA

A method for the fusion of pre-op CT data on interventional XRay images

François Lecomte

Image-guided surgery

1. Pre-op CT acquisition

Segmentation -> 3D model of tumors, anatomical structures

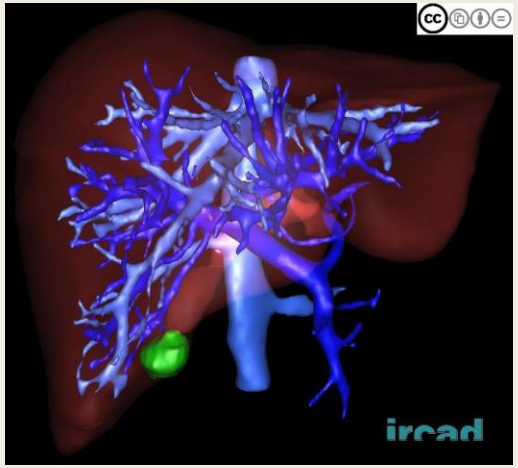
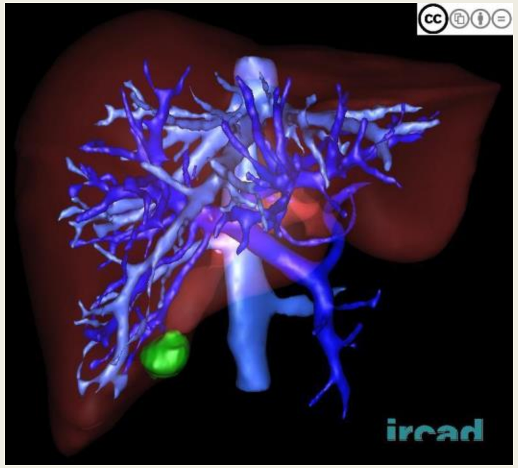


Image-guided surgery

1. Pre-op CT acquisition

Segmentation -> 3D model of tumors, anatomical structures



+

2. Interventional XRays acquisition

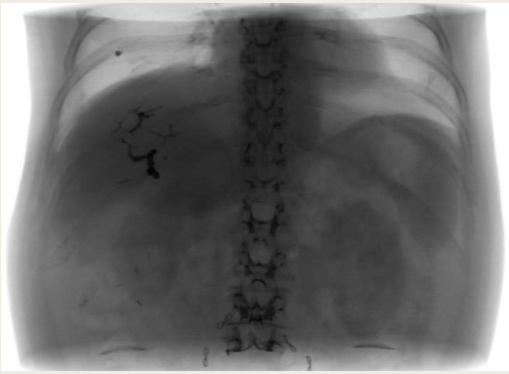
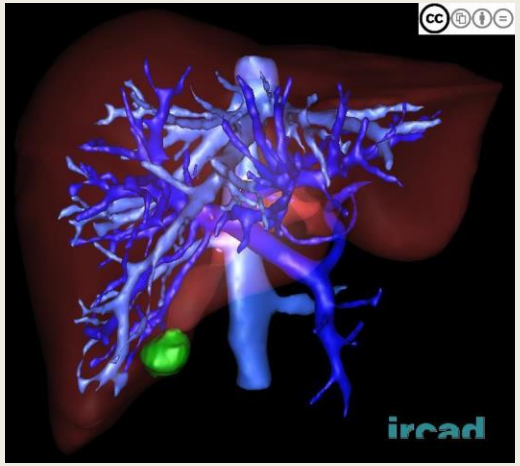


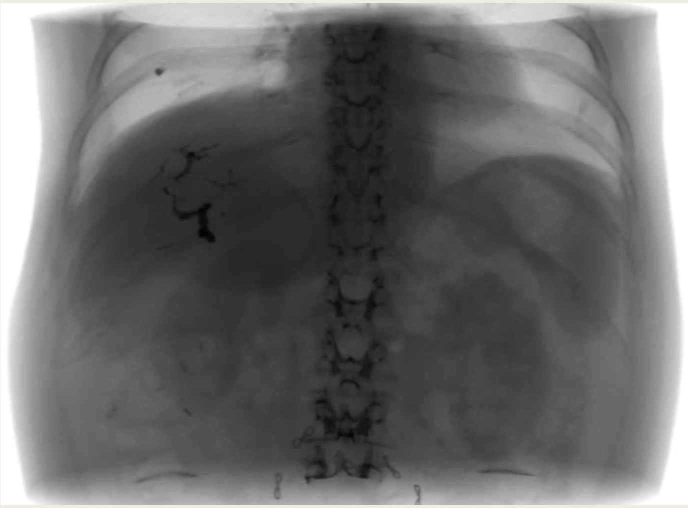
Image-guided surgery

<u>Problem :</u>

How to register in real time, taking into account deformations of the anatomy ?



+

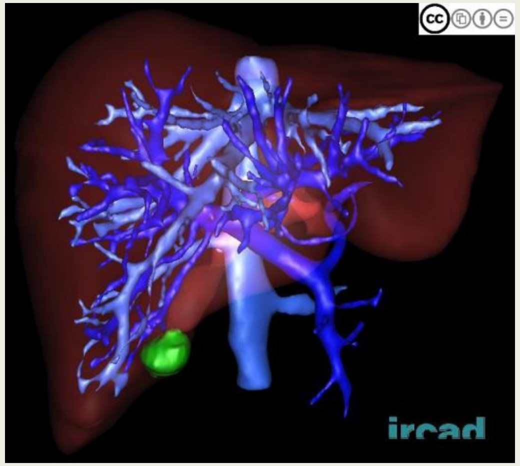


= ?

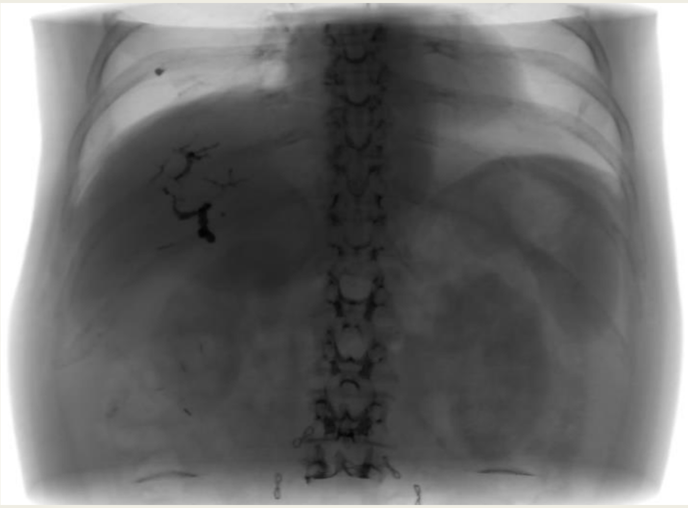
Image-guided surgery

Problem :

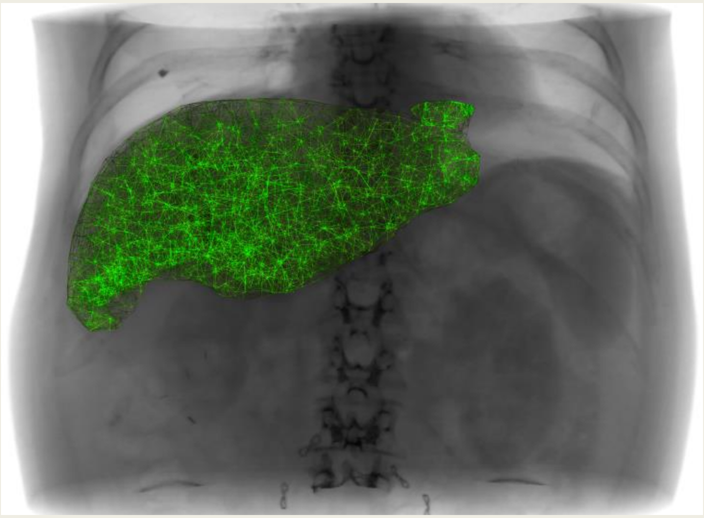
How to register in real time, taking into account deformations of the anatomy ?



+



=



Clinical context

By fusing the pre-operative data with operative fluoroscopic images, we seek to provide a better **understanding of the anatomy**, reduce **procedure time** and **eliminate the need for fiducials**.

A direct application of our method is **motion management for radiotherapy.** A variety of methods have been developed to bring solutions to this problem.

Clinical context

One of the most widely used method, CyberKnife, uses **fiducials and a dual XRay acquisition** to track the tumor with ~3mm precision[1].

Fiducials implantation is an **invasive** procedure that can lead to complications such as **pneumothorax** [2], so it is necessary to develop markerless methods to tackle this problem.

[1] Adler, J., Chang, S., et al. : The Cyberknife: a frameless robotic system for radiosurgery. Stereotact Funct Neurosurg. 69(2):124-128, 1997.

[2] Kothary, N., Heit, et al. : Safety and efficacy of percutaneous fiducial marker implantation for image-guided radiation therapy. J Vasc Interv Radiol, 20(2):235-239, 2009.

Clinical context

A more recent method uses a DeepLearning approach to bypass the need for markers in the image while still needing the dual XRay acquisition, with similar results[3].

Our goal is to eliminate this need as well, because it implies **specific equipment** and **double the radiation dose** for the patient

[3] Hirai, R., Sakata, Y., et al. : Real-time tumor tracking using fluoroscopic imaging with deep neural network analysis. Physica Medica, 59:22–29, 2019

To summarize…

- Our goal is to develop a method to **compute the anatomical deformations** corresponding to an input XRay image, in **3D and real-time**.

- We want to be the least invasive possible, so fiducials as well as dual fluoroscopic acquisitions are not possible.

- After reviewing the state of the art in the context of radiotherapy, we found **no method** able to perform 2D/3D registration **for this context**.

Assumptions

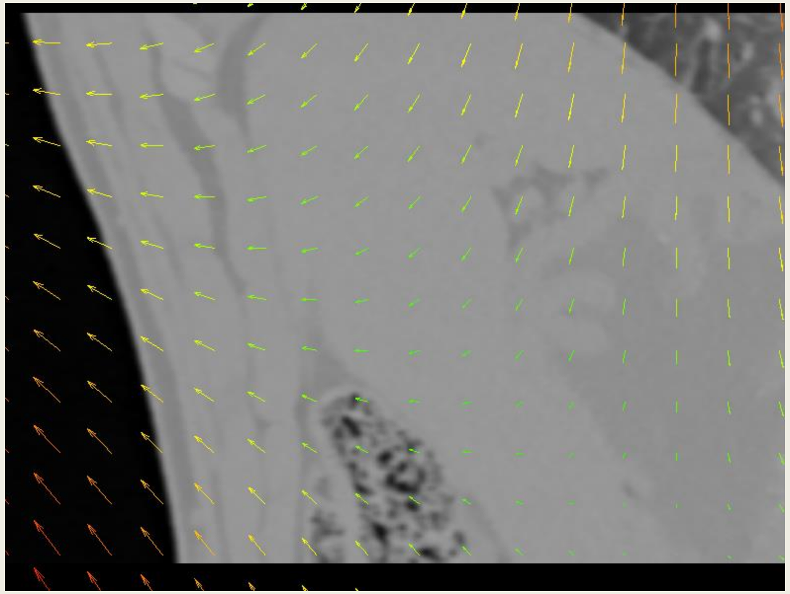
- Fluoroscopic images contain information about 3D anatomy

- This information can be translated from 2D to 3D

Data Generation: CT Deformation

- CT is a 3D-image and we want to apply geometric deformations.

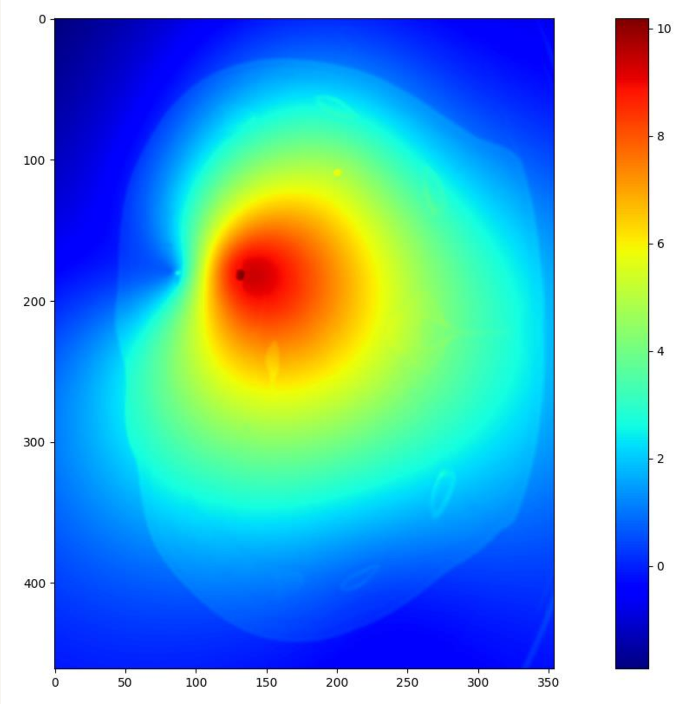
- We can start from the displacement field between two respiratory phases:



Data Generation: CT Deformation

- CT is a 3D-image and we want to apply geometric deformations.

- The displacement field can also be understood as an image:

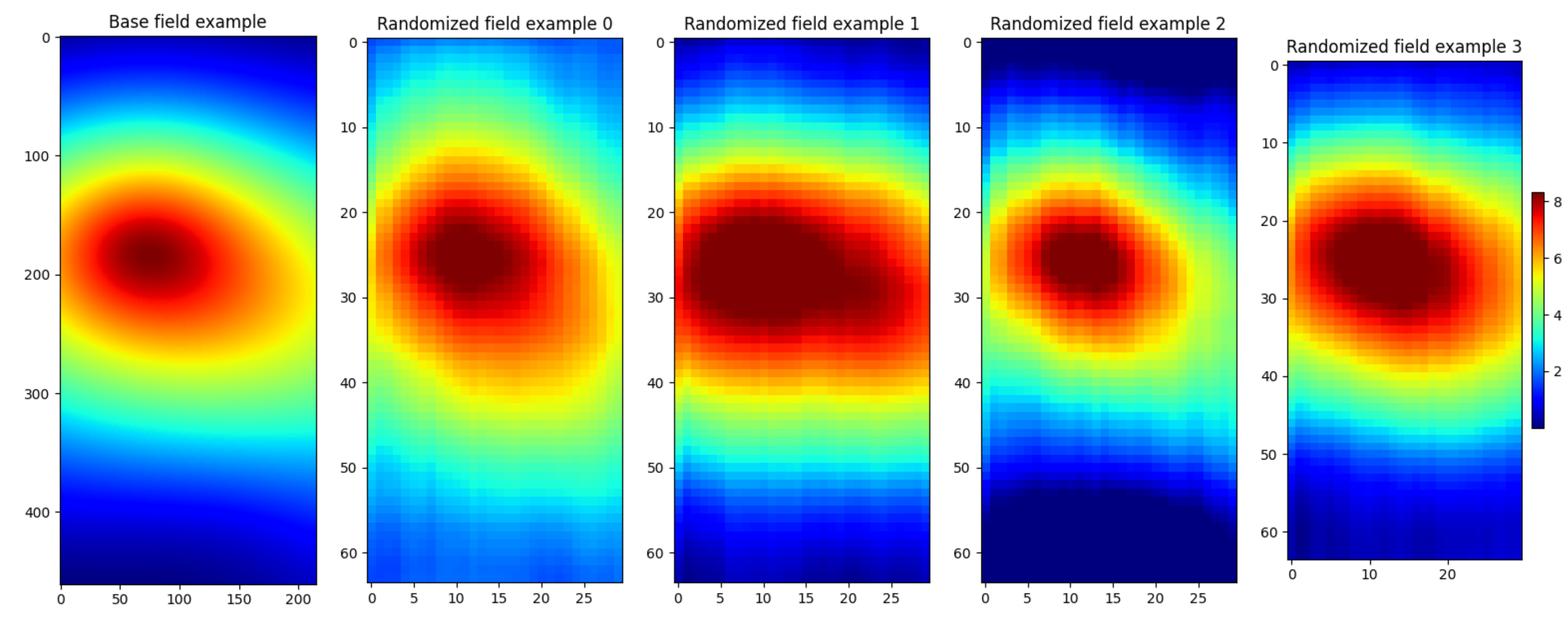


Displacement field shown on top of the CT for one slice and one direction

Data Generation: CT Deformation

- CT is a 3D-image and we want to apply geometric deformations.

- We can randomize it in the Fourier domain:



Left: Displacement field obtain by landmark-based registration
Right: Displacement fields generated by randomizing frequency components

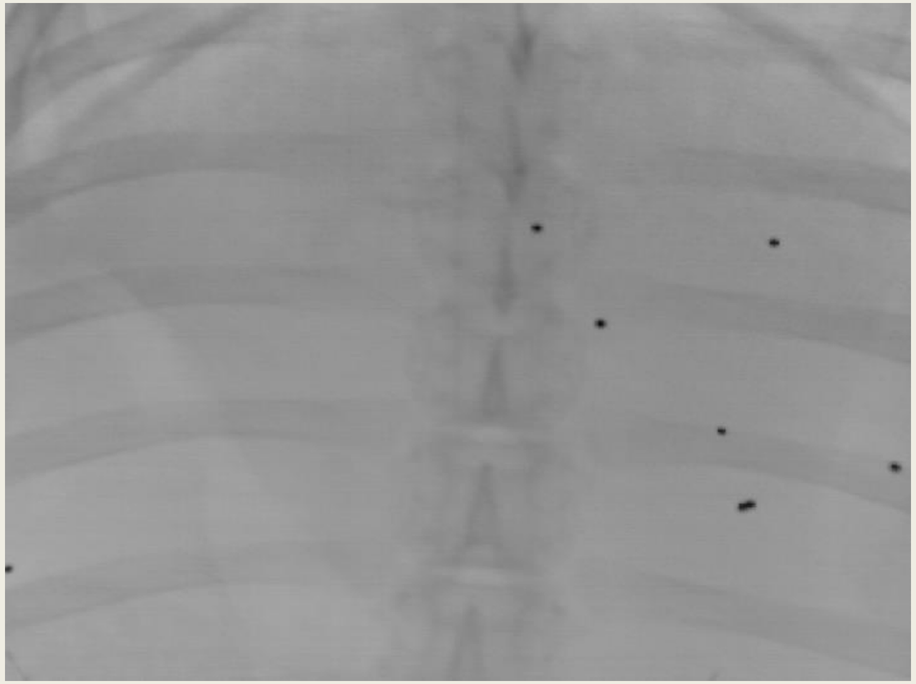
Data Generation: DRR rendering

- Uses the VolumetricXRayRendering plugin (developed by Fred Leroy)
- CT image in 3D space is projected using a ray-tracing shader
- Ray intensity diminishes as it goes through matter ($I = I_0e^{-\mu x}$)

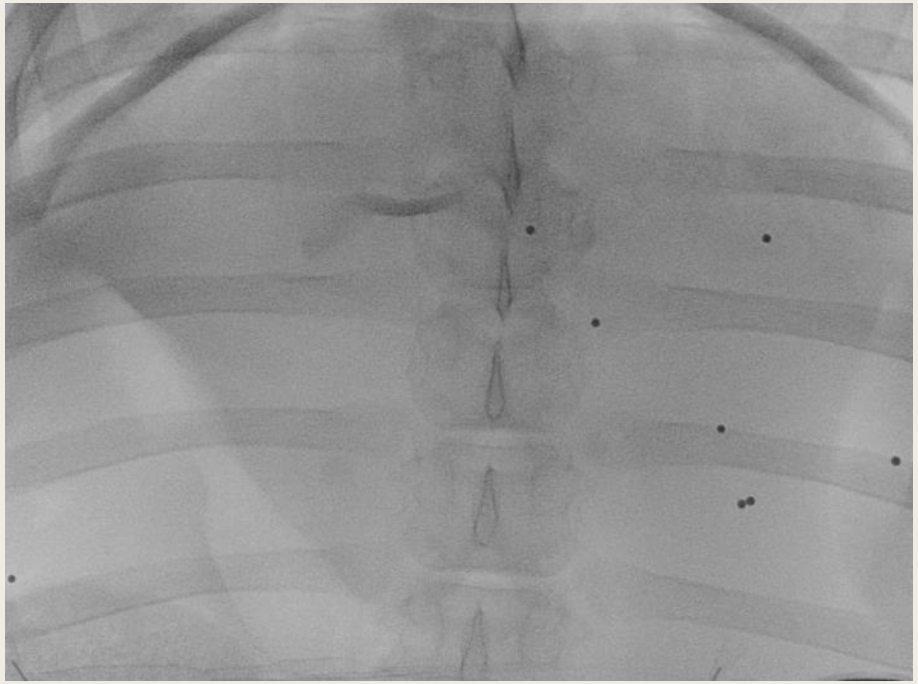
Data Generation: DRR rendering

- Uses the VolumetricXRayRendering SOFA plugin (developed by Fred Leroy)

- We obtain a Digitally Rendered Radiography (DRR):



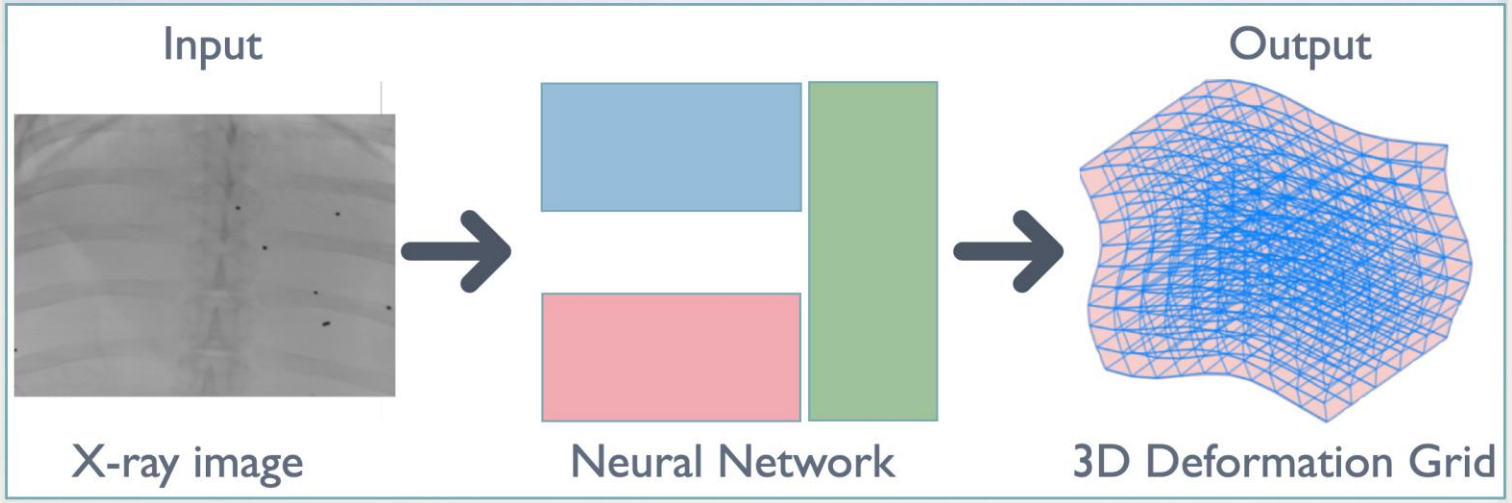
DRR



Fluoroscopic image

Deep Learning: Training procedure

1. The network is asked to find the displacement in 3D from a DRR:

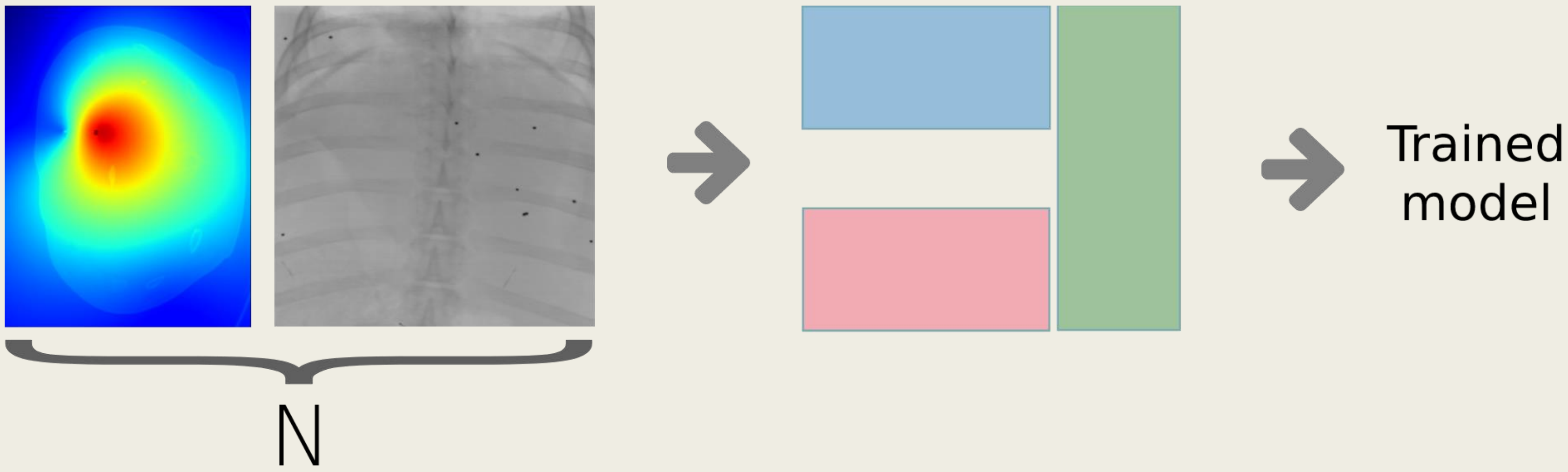


From a single projection, the encoder-decoder network generates a coarse displacement field on the 3D volume

2. The loss is computed and backpropagation is performed:

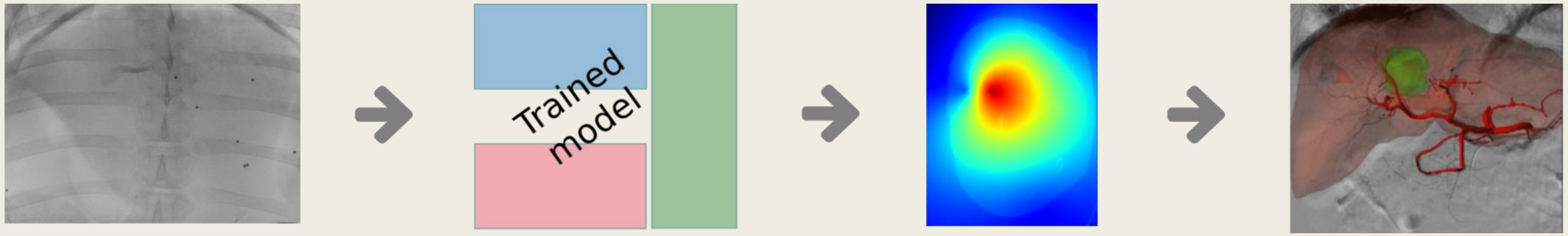
$$\mathrm{MSE} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \tilde{y}_i)^2$$

Training summary



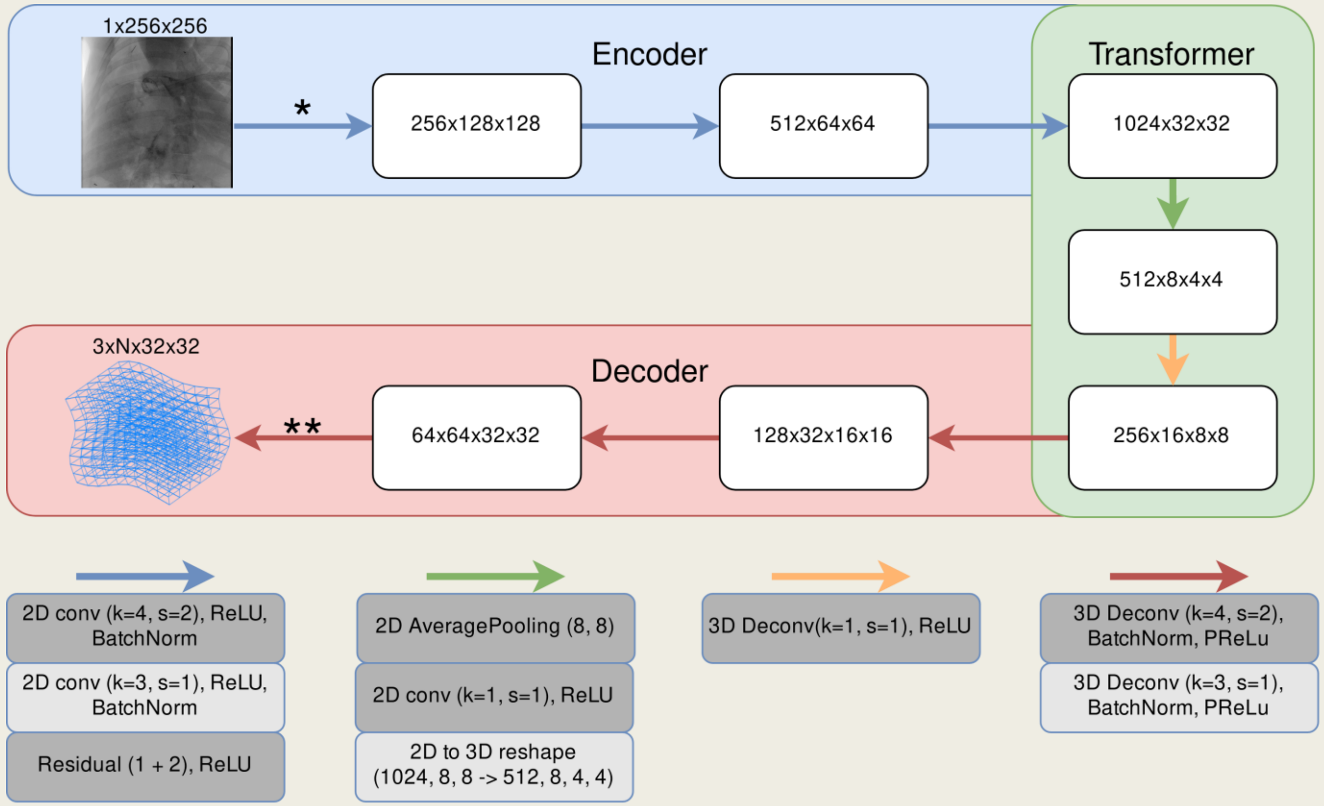
The encoder-decoder is trained on *N* data samples for >10 hrs

Usage



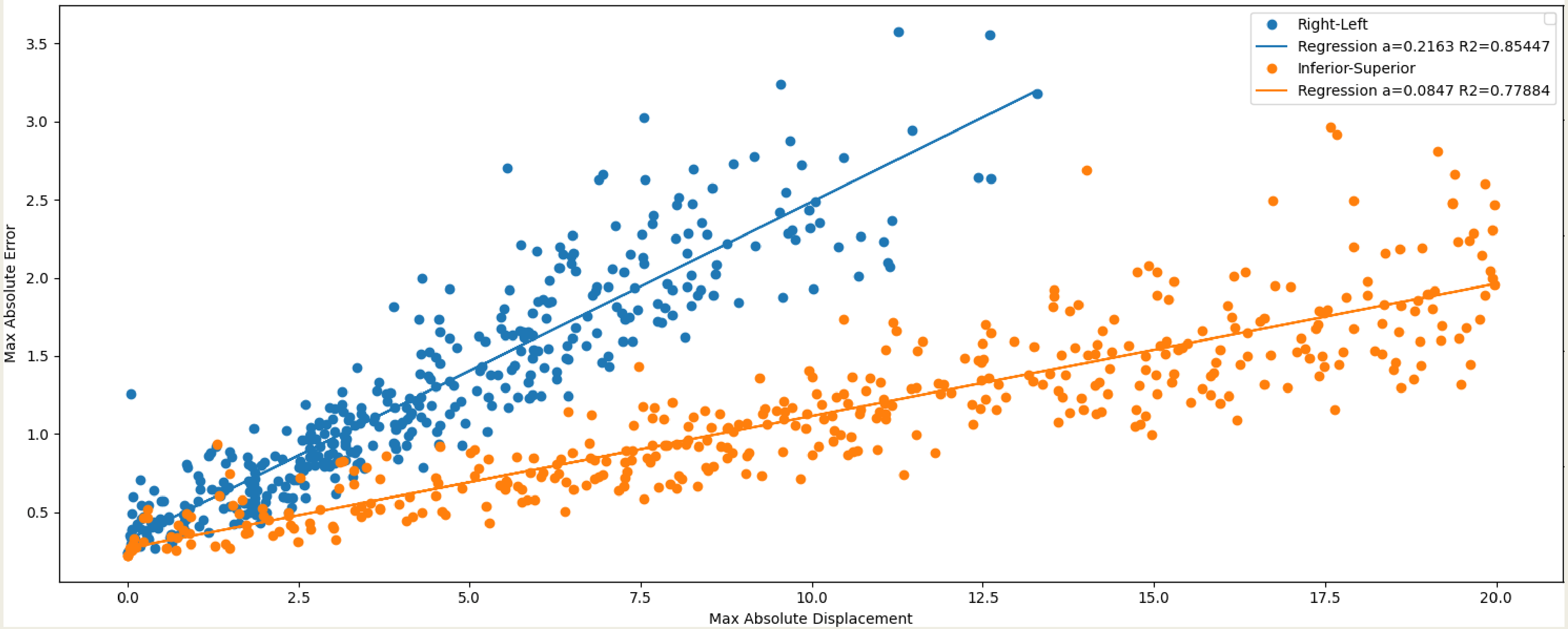
1. The encoder-decoder predicts a displacement field in ~30ms from a fluoroscopic image.
2. The displacement is applied to 3D data for augmented reality in real-time

Deep Learning: Network Architecture



1. The **Encoder** features from the input image to build a feature map.
2. The **Transformer** module reshapes the feature map to obtain a 3D representation of the image.
3. The **Decoder** converts the feature representation back to intensity space.

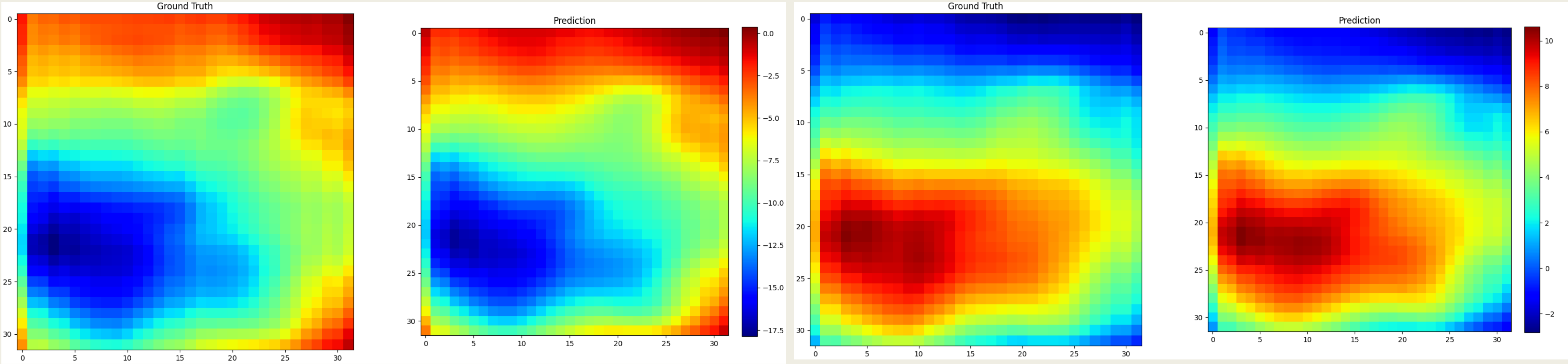
Synthetic results :



Plot of the maximal error of the network against the maximal displacement value per validation sample (N = 400).

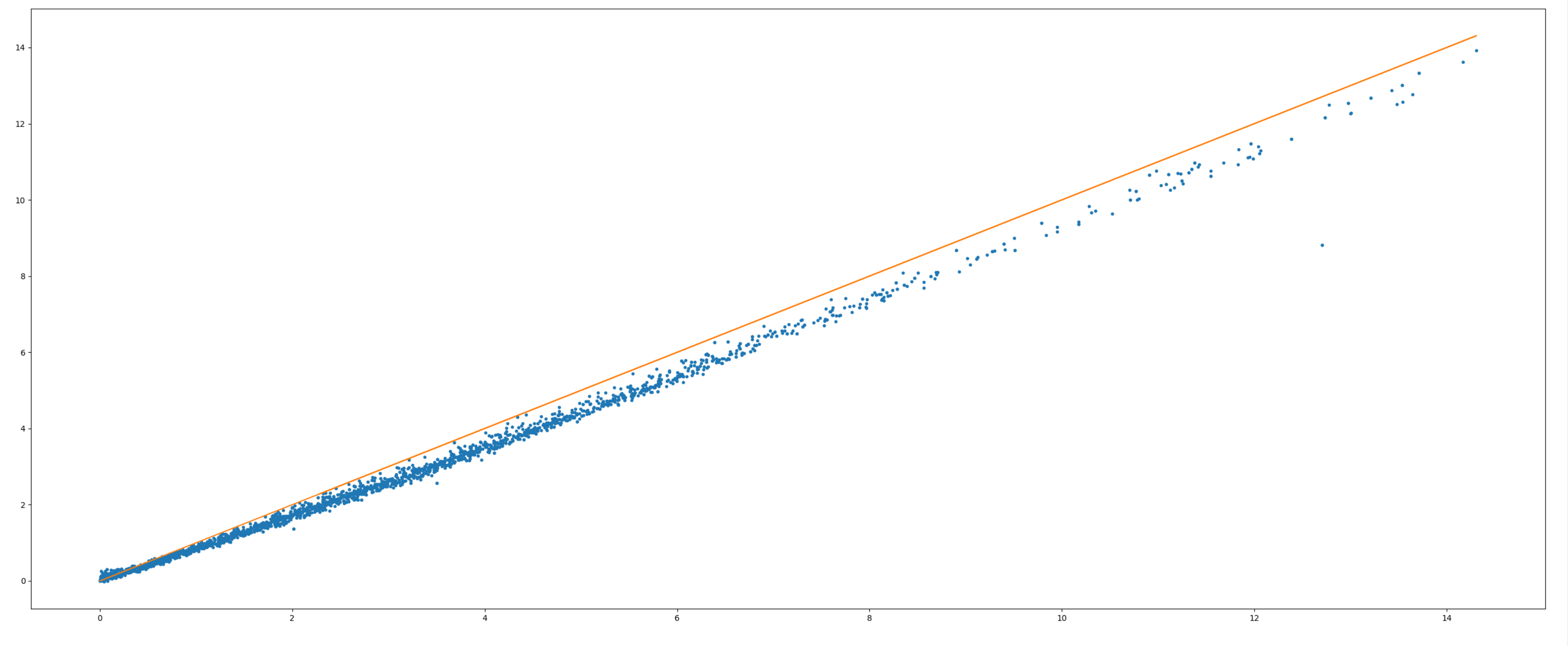
The blue and orange dots represent the two displacement directions the network predicted.

Synthetic results :



Examples of GT vs Prediction for the same slice for two validation samples (Inferior-Superior displacement)

Synthetic results :



Regularization accuracy gain vs. Displacement amplitude
As the predicted displacement error increases (orange line), the elastic model (blue dots) compensates for most of it.

Discussion

- Our method is capable of achieving state-of-the art results (error < 4mm), for a computation time of 30ms.

- It also demonstrates that it is possible to predict 3D data from a 2D projection.

- However, it only works on synthetic images and has trouble generalizing to unseen deformations.

- The physical regularization is necessary for the prediction of displacements perpendicular to the projection plane but hasn't been tested in conjunction with the network

Next steps

- The performance on realistic data depends on the quality of the training data :
  - The DRR quality can be improved by using updated DRR tools.
  - The displacement fields generation needs to be perfected.

- FEM regularization needs to be implemented (preliminary results demonstrate proof of concept)

- Moving forward, the framework can be generalized to be robust for any projection direction by including camera pose information in the training.

Conclusion

- The method we are developing aims to compute the 3D deformation field that would allow the registration of the CT on the input XRay image, in real-time.

- The goal is to be the least invasive, so fiducials as well as dual fluoroscopic acquisitions are not possible.

- After reviewing the state of the art in the context of radiotherapy, we found no method able to perform 2D/3D registration for this context.

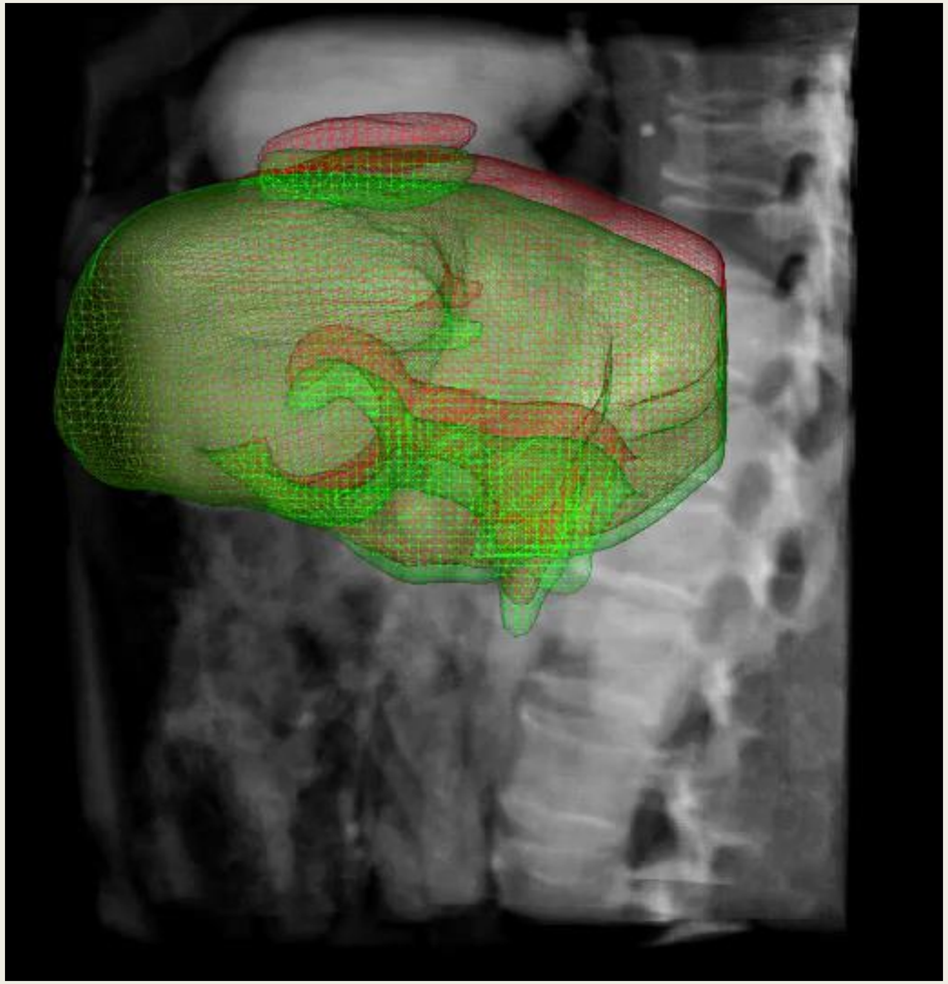
- Moving forward, the framework can be generalized to be robust for any projection direction by including camera pose information in the training.

Thank you for your attention !

Please don't hesitate to ask questions.

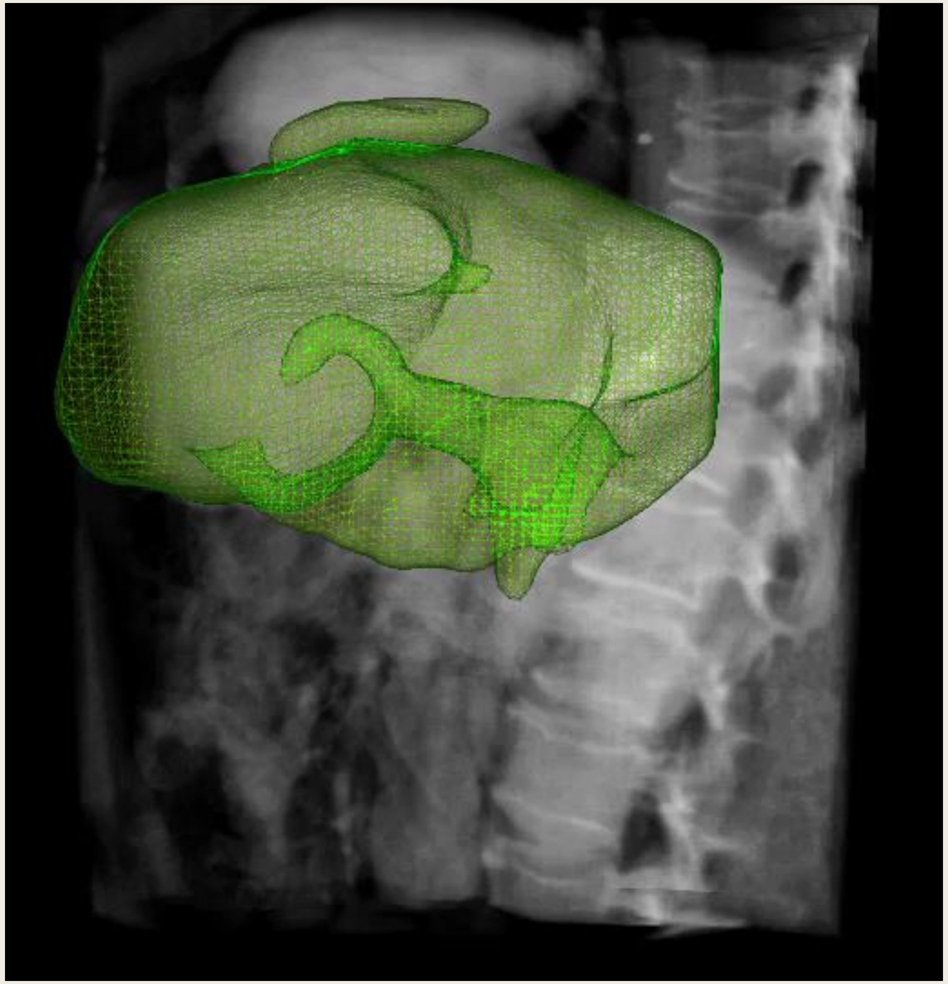
The End

Synthetic results :



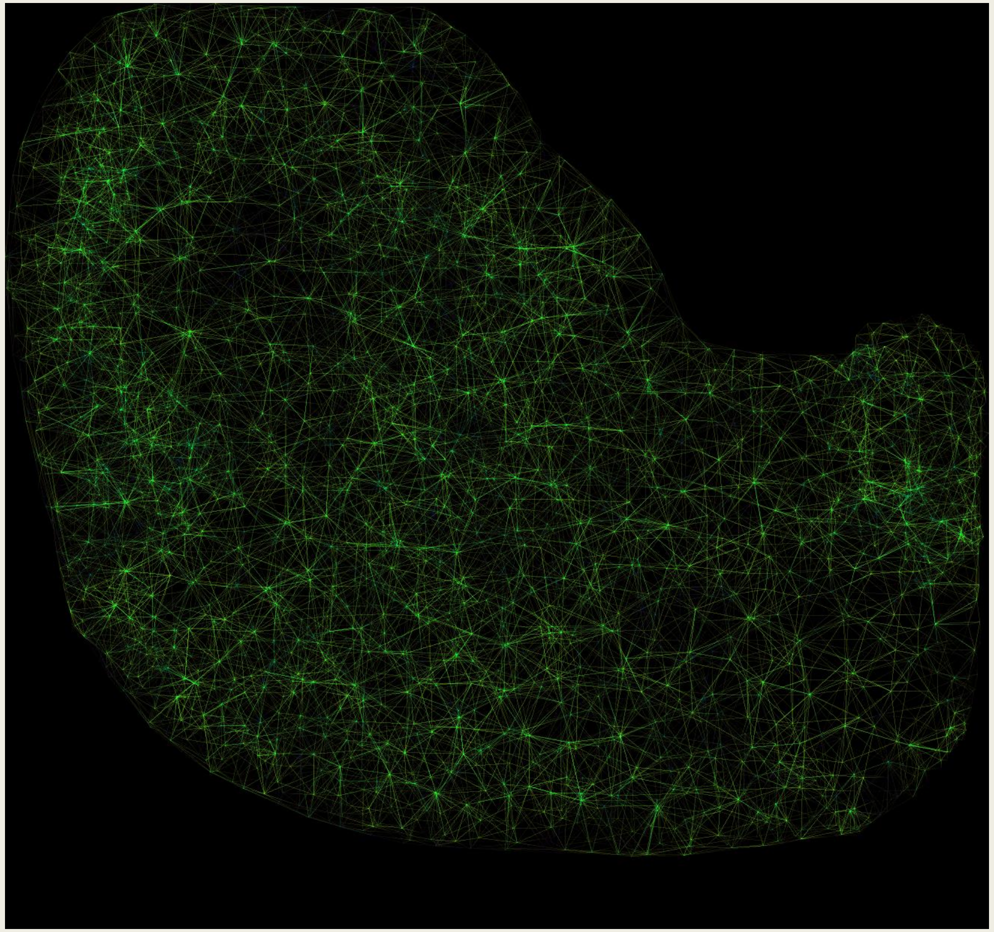
Simulated respiratory motion : target (red), fixed (green)

Synthetic results :



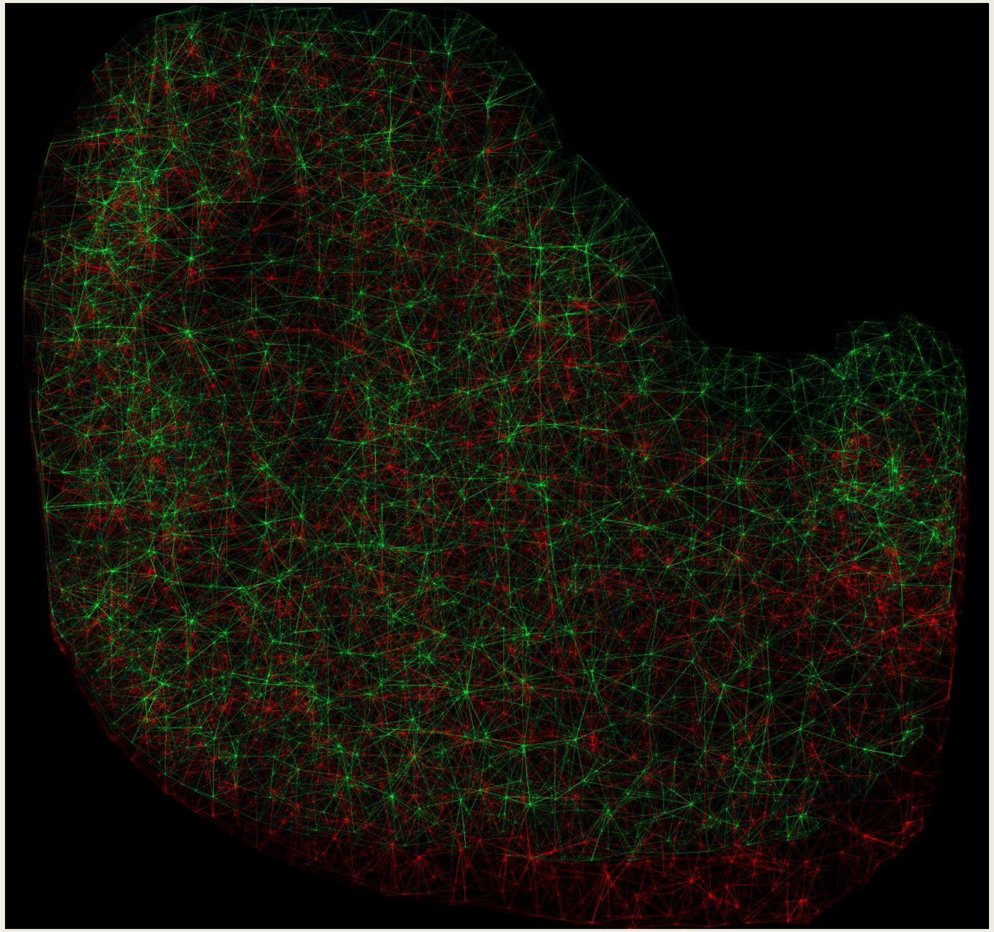
Simulated respiratory motion : target (red), predicted (green)

Synthetic results :



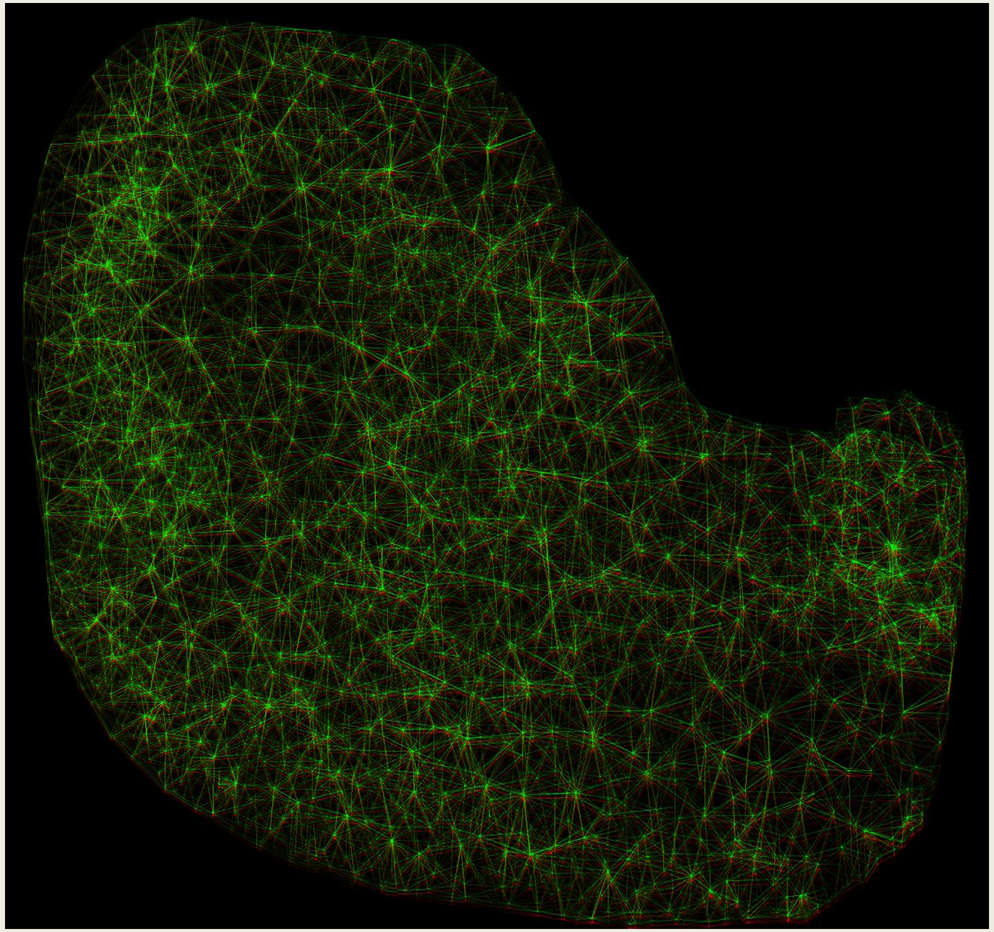
Green: Rest FEM mesh extracted from pre-op CT

Synthetic results :



Green: Rest FEM mesh extracted from pre-op CT
Red: FEM mesh after registration for only two directions

Synthetic results :



Green: Rest FEM mesh extracted from pre-op CT
Red: FEM mesh after physical regularization