

Development and evaluation of automated tumour monitoring by image registration based on 3D Positron Emission Tomography/Computed Tomography (PET/CT) images

Rosana El Jurdi¹, Mohamad Ammar Said¹, Duc Thang Hoang¹, Guillaume Lamazou¹, Elsa Schalck¹, Olivier Humbert²

¹Euranova, Marseille, France; ²Université Côte d'Azur, CNRS, Inserm, iBV, Nice, France.

Introduction and contributions

Tumor tracking in PET/CT is essential for **monitoring cancer progression** and guiding treatment strategies. Traditionally, nuclear physicians **manually** track tumors, focusing on the five largest ones (PERCIST criteria), which is both time-consuming and imprecise. **Automated tumor tracking** will allow real-time and precise matching of the numerous metastatic lesions across scans, enhancing tumor change monitoring. However, research is constrained by the **limited availability of labeled medical tracking datasets**.

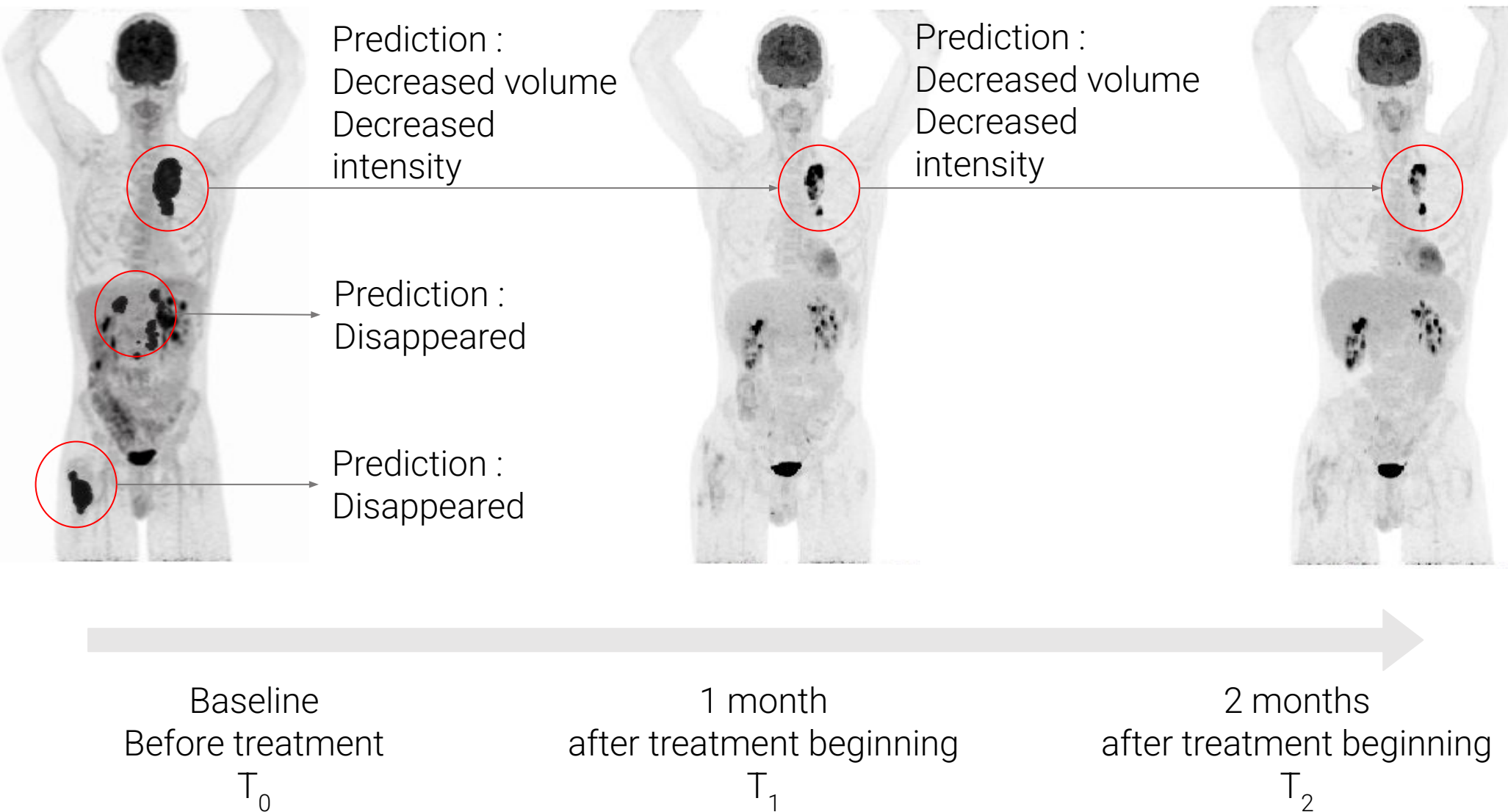
Our contributions are :

- The creation of a **unique dataset fully annotated for tumour tracking**.
- The introduction of a **robust evaluation framework** for tumor tracking assessment.
- The validation of an automated tumor tracking method based on **image registration** followed by distance-based matching for simple cases.

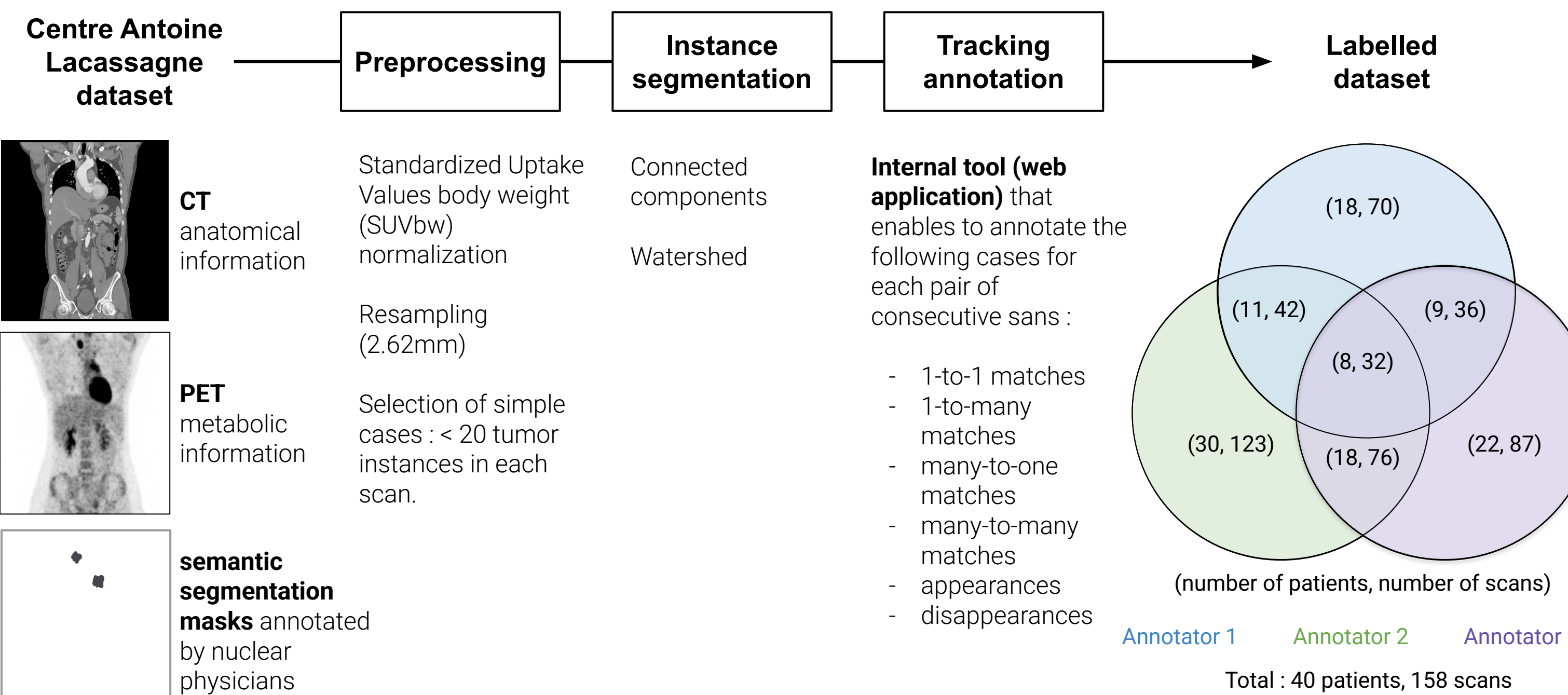
Objective

Complete tracking of the evolution of each individual tumor between baseline and follow-up scans :

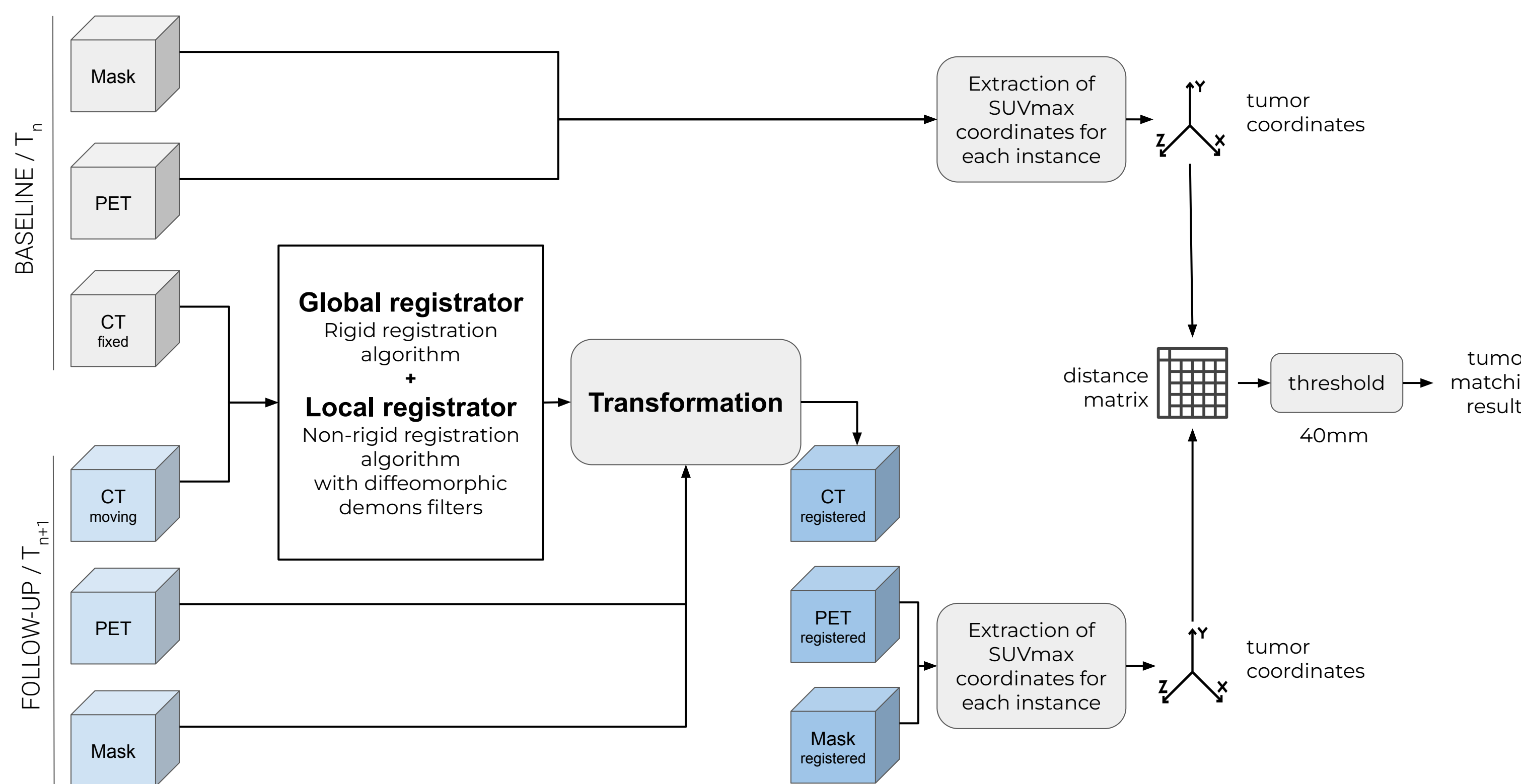
- one or multiple tumors are matched to one or multiple tumors
- tumor disappearance
- tumor appearance



Dataset



Tumor tracking method



Evaluation method

The model performance is evaluated at the level of a match in 2 different ways.
Definition of a match : correspondence of tumors/group of tumors between a baseline and a follow up scan.
Correct match : the predicted tumor/group of tumors is included in the ground truth tumor/group of tumors or the inverse. If not, it is as a mismatch error (mme).

Confusion matrix for tumor matching

Positive : tumor with match
Negative : tumor without match (disappeared or appeared)

		Ground truth	
		Positive	Negative
Prediction	Positive	TP correct match (if incorrect: mme)	FP disappeared or appeared tumor matched by the model
	Negative	FN tumor with match predicted as disappeared or appeared tumor	TN disappeared or appeared tumor predicted as disappeared or appeared tumor

Confusion matrix for tumor appearances/disappearances

Positive : tumor without match (disappeared or appeared)
Negative : tumor with match

		Ground truth	
		Positive	Negative
Prediction	Positive	TP disappeared or appeared tumor predicted as disappeared or appeared tumor	FP tumor with match predicted as disappeared or appeared tumor
	Negative	FN disappeared or appeared tumor matched by the model	TN correct match (if incorrect: mme)

mme : mismatch error : the predicted match is not correct (the predicted tumor/group of tumors is not included in the ground truth tumor/group of tumors or the inverse).

Results

	Annotator 1	Annotator 2	Annotator 3
Annotator 1	-	0.93	0.92
Annotator 2	0.93	-	0.94
Annotator 3	0.92	0.94	-

Table 1. Inter-rater Variability (accuracy)

→ Annotations provided by different experts are consistent and reliable

Metric		Annotator 1	Annotator 2	Annotator 3	Average ± std
Overall	accuracy	0.88	0.83	0.77	0.83 ± 0.05
	Error _{mme}	0	0.01	0.02	0.01 ± 0.00
Tumor matching metrics	f1-score	0.93	0.89	0.83	0.88 ± 0.04
	precision	0.89	0.84	0.77	0.83 ± 0.05
	recall	0.98	0.97	0.95	0.97 ± 0.01
Tumor appearances/disappearances metrics	f1-score	0.52	0.44	0.42	0.46 ± 0.04
	precision	0.74	0.74	0.69	0.72 ± 0.02
	recall	0.45	0.36	0.34	0.38 ± 0.05

Table 2. Performance of the automatic tracking method at patient level.

- Small differences of performance between annotators
→ **consistent and reliable results across annotators**
- High overall accuracy, tumor matching metrics and low Error_{mme}
→ **the model performs well in identifying matches**
- Low f1-score and recall for tumor appearances/disappearances
→ **the model fails to properly detect tumor appearances/disappearances**

Conclusion and future work

The proposed tumor tracking method by image registration, though just as a baseline, performs well on the annotated dataset, composed of simple cases.

Future work will focus on :

- **In-depth study of the results** : visualization, failure cases.
- Annotation and evaluation of the method on a **larger dataset** : complex cases, multiple cohorts.
- Development and evaluation of **deep learning methods** : siamese networks.

Acknowledgement

ANR-19-P3IA-0002, Canceropôle PACA and GIRCI Méditerranée.