Intégration de l'humain dans les systèmes d'analyse d'images : des tentatives et des perspectives

JEAN-YVES RAMEL

LIFAT – EA 6300

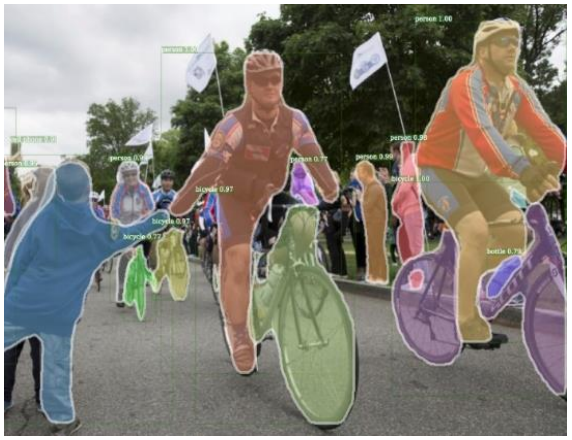
RTR DIAMS - Janvier 2023

ramel@univ-tours.fr

Starting point…

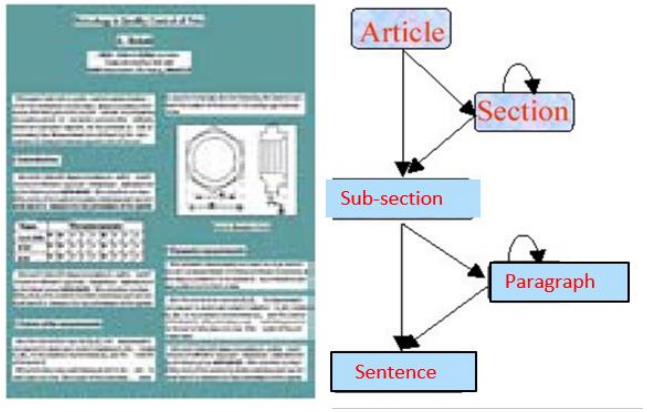
“New” goals in CV: associating semantical labels to images

- Numerous toolboxes (tensorflow, pytorch, detectron, …)
- Analysis of Spatial and Temporal relations between objects or subparts of objects, pose & emotion recognition, VQA …
- Image content interpretation, understanding



This goal is targeted since many years in DIA

- Analysis of spatial and temporal relations between elements is mandatory in OCR, layout analysis, line drawing analysis, …
- Extraction of elements of contents (EoC) at different levels: lexical, syntactical, semantical



The Challenges

- Data and Knowledge representations for a easier & better analysis of images and videos (dictionaries, models of language, …)
- Insertion of the users in the loop?

Starting point…

- CNN ➔ A **low level vision** of real world (The data are considered as a set of pixels)
- The **Machine Learning** algorithms only consider **annotated data** to set the parameters in **batch mode**

- **The human** uses an higher level of vision of the real world → merging the semantic gap?
- **Knowledge & Contextual information** should be integrated in CV systems → where is the loop?

- Future systems should be more transparent (ExAI mandatory for interaction)
- And adaptive (“on-line” plasticity)

Is it CNN and ML compatible?
What are the good directions?

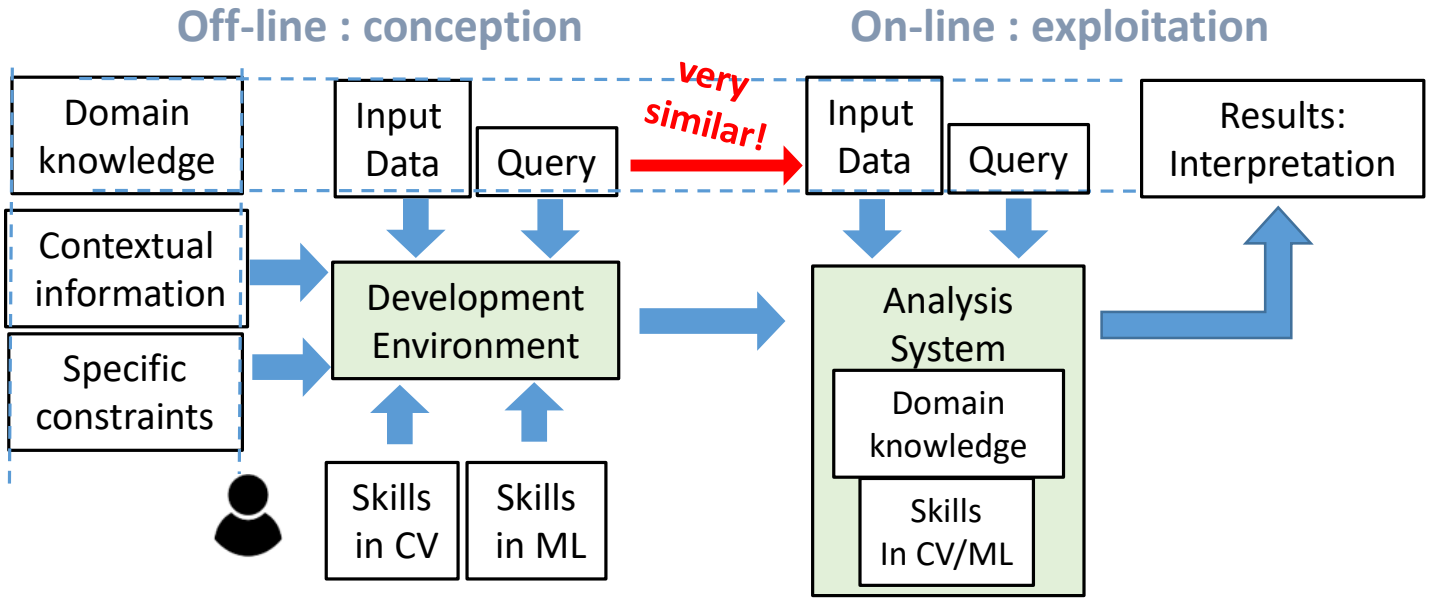
Systems and methods taxonomy

- **Categories of DIA and CV methods and systems**
  - Static systems (off-line, no learning)
  - Adaptable methods (off-line data driven and interaction)
  - Adaptive methods (on-line data driven and interaction)

Let's dive into into these categories and look for the user (and the knowledge representation)…

Static handcrafted systems

- **Inside the system, the designer encodes (off-line):**
  - All the algorithms for feature extraction and ROI (EoC) recognition
  - Using the a priori knowledge about the data
  - Regarding the known future inputs (query, images)
  - Without separations between algorithms and models, different levels …
  - New data → New development



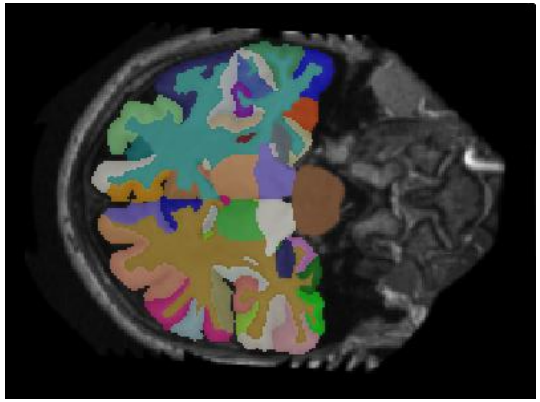
Static handcrafted systems

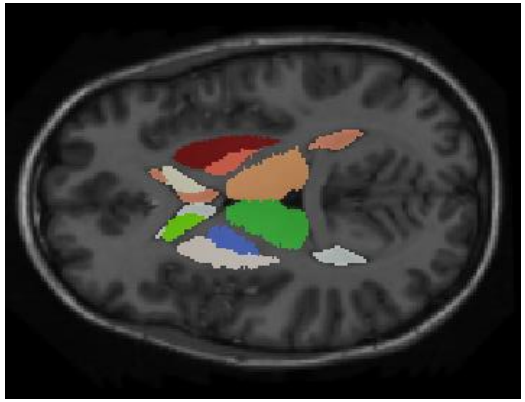
In CV, lot's of methods for segmentation and object detection

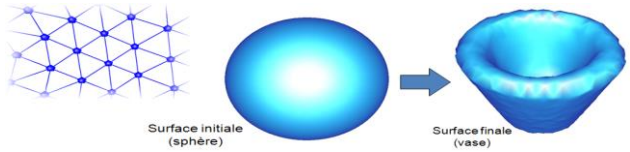
- Global approaches (atlas and scene models)
- Local approaches (active contour and shape model)
- In DIA, little more knowledge for layout analysis or OCR (document or language models → rules, grammar)

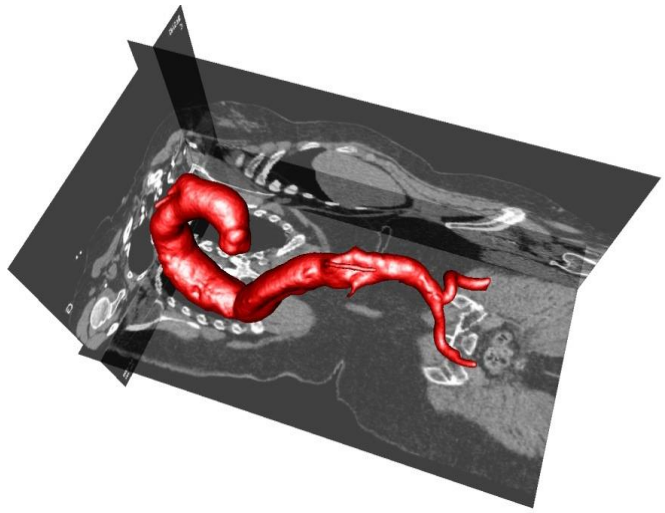
→ Users could interact only during the initialization (coarse contours, seeds, …)

→ No clear distinction between levels of knowledge



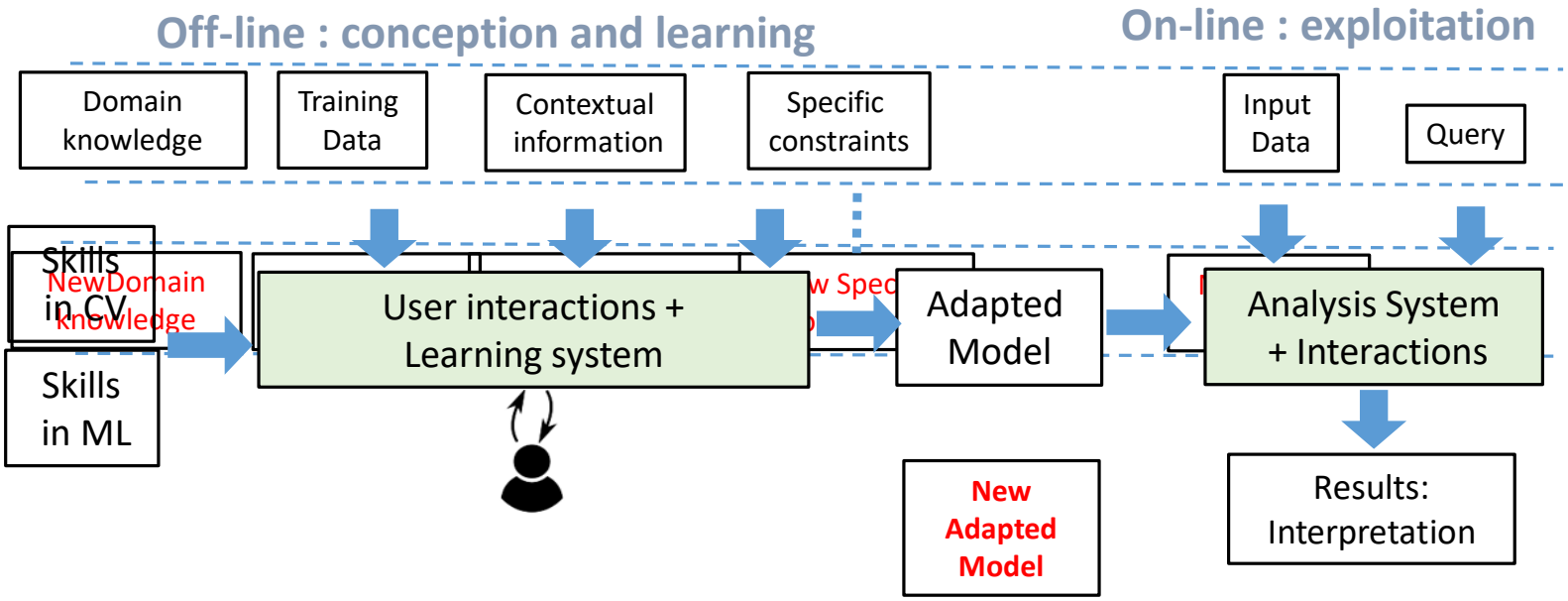






Adaptable & Interactive systems

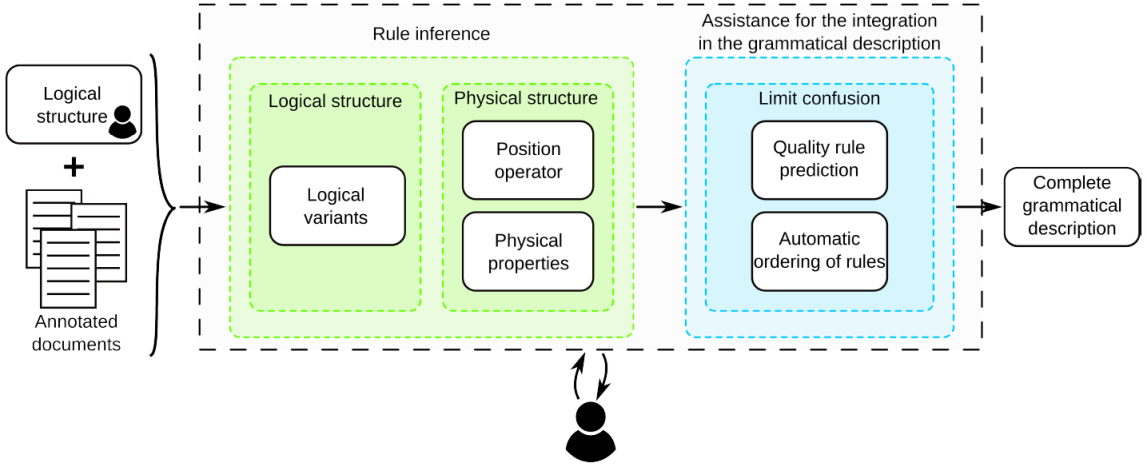
- **Inside the system,**
  - Adaptable models that can be learned or user-defined **off-line**
  - Some distinctions between levels of knowledge



Adaptable & Interactive systems

In DIA, Interactive learning for the design of rule-based systems (off-line)

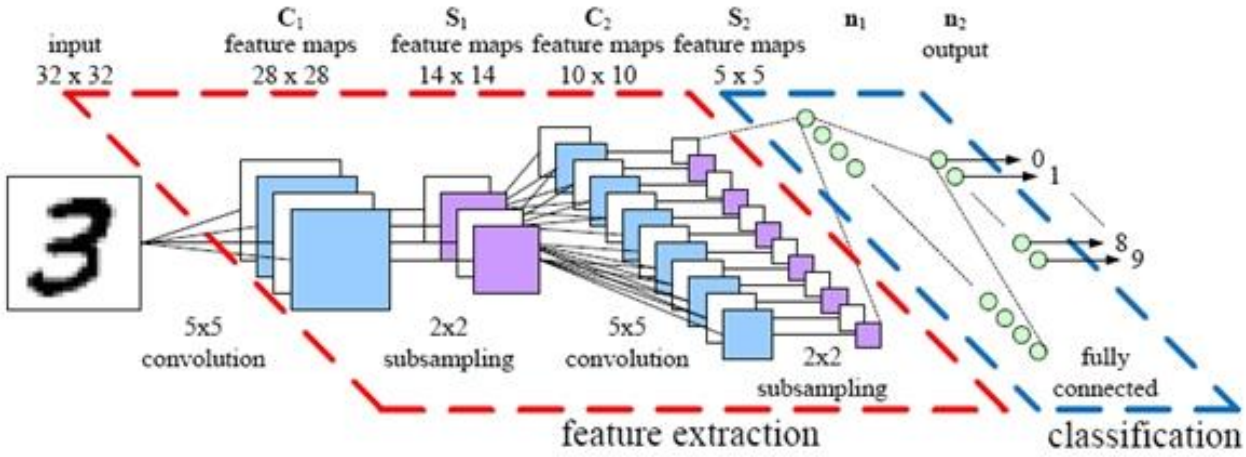
- Interactive building / learning of a full grammatical description of a set of documents (lexical and syntactical levels)
- Main steps:
  - Automatic and exhaustive analysis of an annotated data set (logical structure)
  - The rules are built progressively using a clustering algorithm
  - The interaction with the grammar writer brings semantic in the automatically inferred structures.
  - Evaluation of the pertinence of the built grammar



- Advantages of the syntactical methods → understandable, introduction of user knowledge
- Without their main drawbacks→ time needed to adapt the system to a new type of document

Adaptable & Interactive systems

- **In CV, Deep Learning systems**
  - Adaptable models that can be learned or user-defined ?

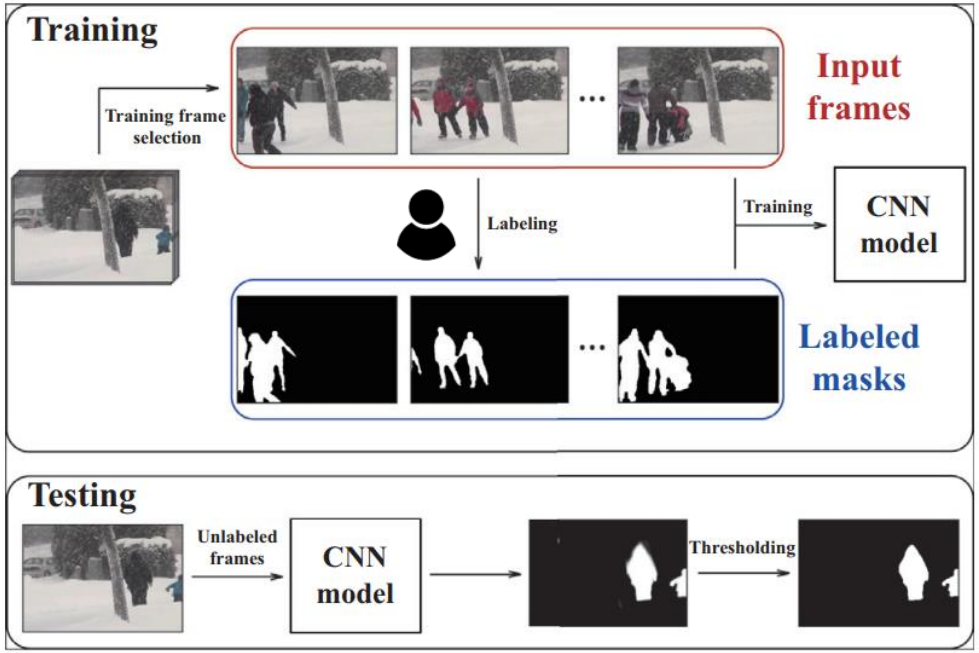


Can we do more than just automatic features selection (at the lexical level, off-line)?

Adaptable & interactive systems

Interactive (deep) learning ➔ **Only off-line and at the lexical / feature level**

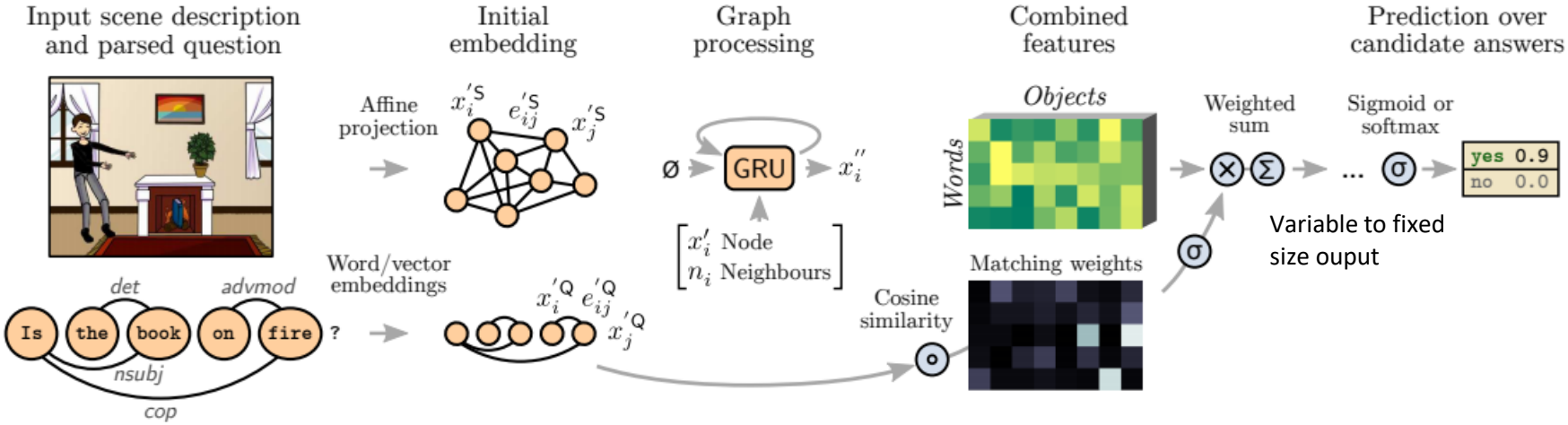
- The users can interact with the training data (off-line)
- Transfer learning (off-line): multi-task learning, featuriser, …
- Curriculum learning (off-line ordering of the training exemples)
- …



Can DL deal with syntactical level?

VRD (learning relations between objects) using graph representations

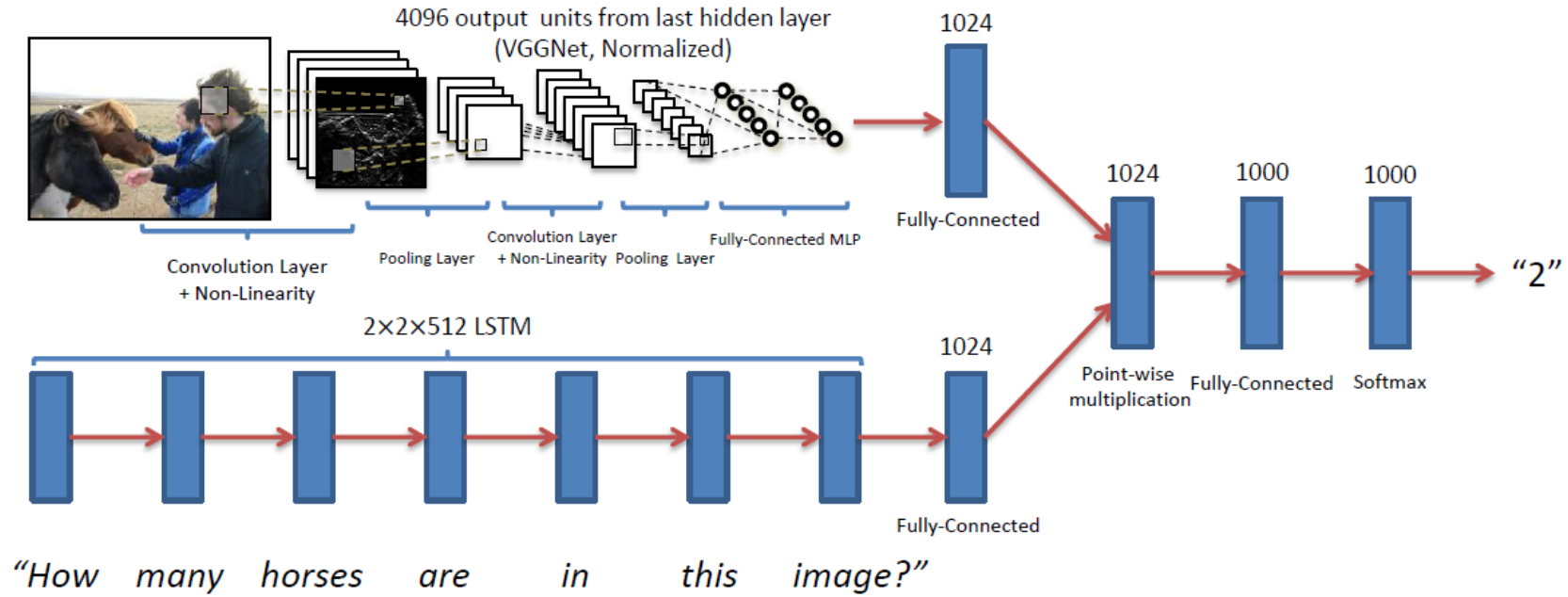
- A scene graph with attributed nodes (objects) and edges (spatial relationships)
- A question graph with node (words) and edges (type of syntactic)
- A recurrent unit (GRU) transform the 2 graphs into word and object features
- Both features are concatenated pairwise (inside a matrix)
- Objects and words are “matched” using learned weights
- A final classifier predicts scores over a fixed set of candidate answers



Can DL deal with semantical level?

Is there semantic in Visual Question Answering ?

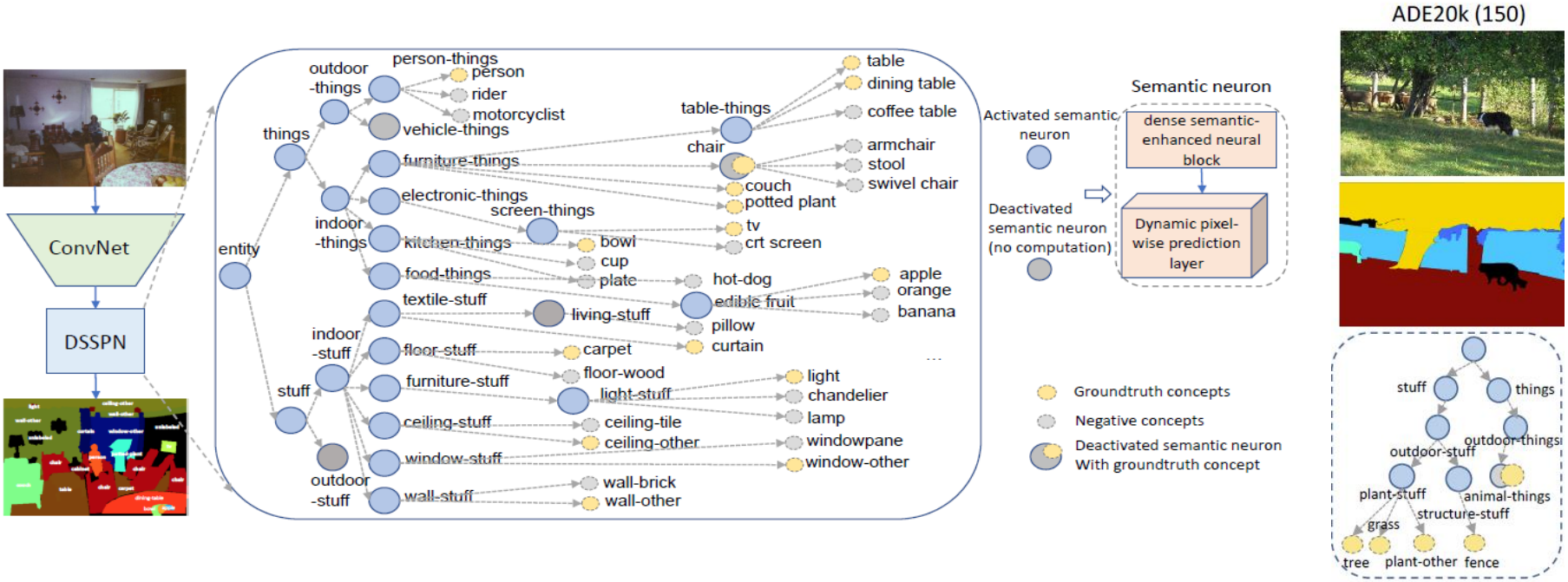
- VGGNet to encode the image content
- LSTM to encode the question
- Question and images features are transformed, put into a common space and pass through a FCL to select the best answers
- Off-line, no loop, no user, no semantic here…



Can DL deal with semantical level?

Using graph representations (to learn relations between objects)

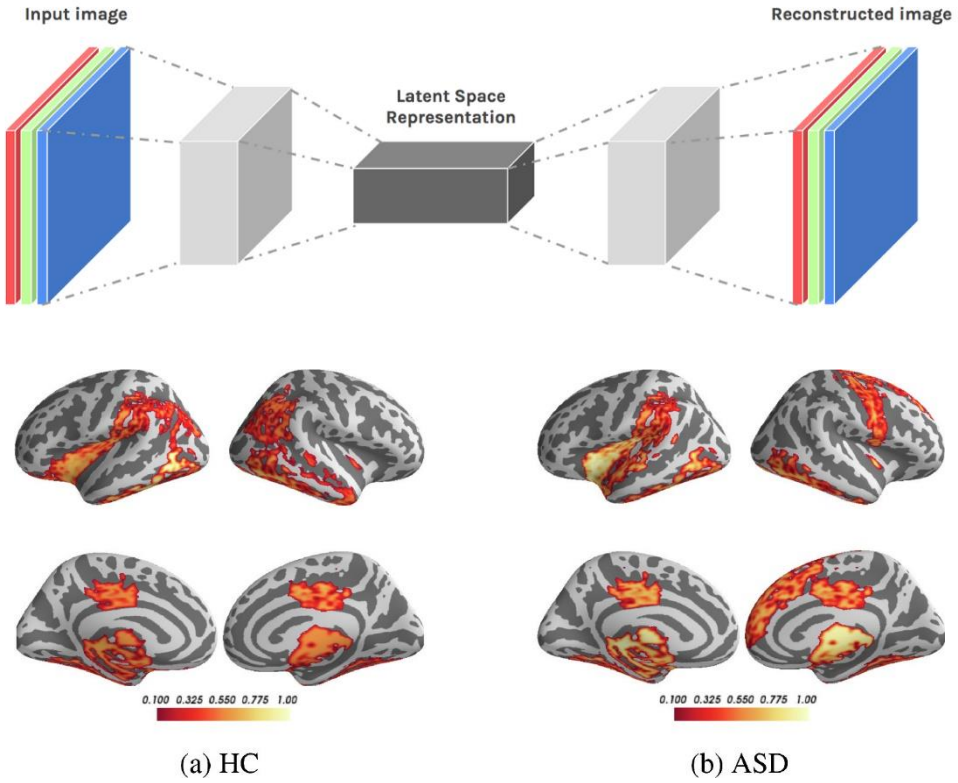
- Proposition of a **Dynamic-structured Semantic Propagation Network**
- A semantic hierarchy (neuron graph network) ➔ Model of the world (manually built?)
- CNN features are propagated into a graph for hierarchical pixel-wise recognition
- Sub-graphs activation during training/testing (feed-forward and back propagation)
- Learn and use of a Hierarchical description of the world/scene

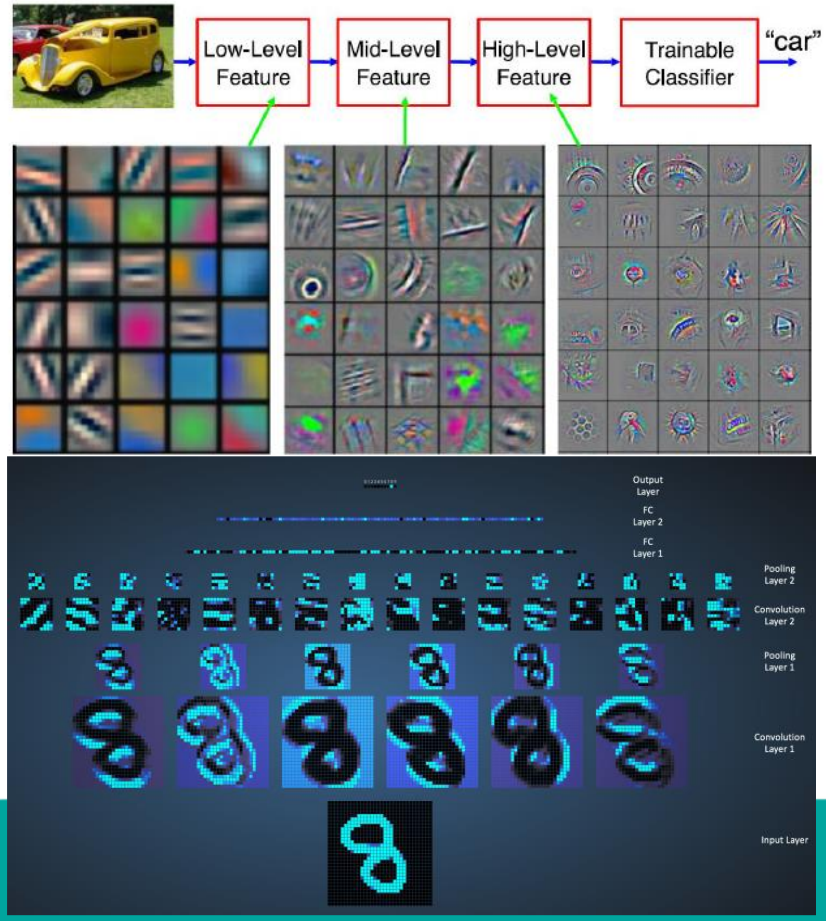


DL and Explainable AI

Feature Maps and Latent spaces visualization and interpretation (off-line)

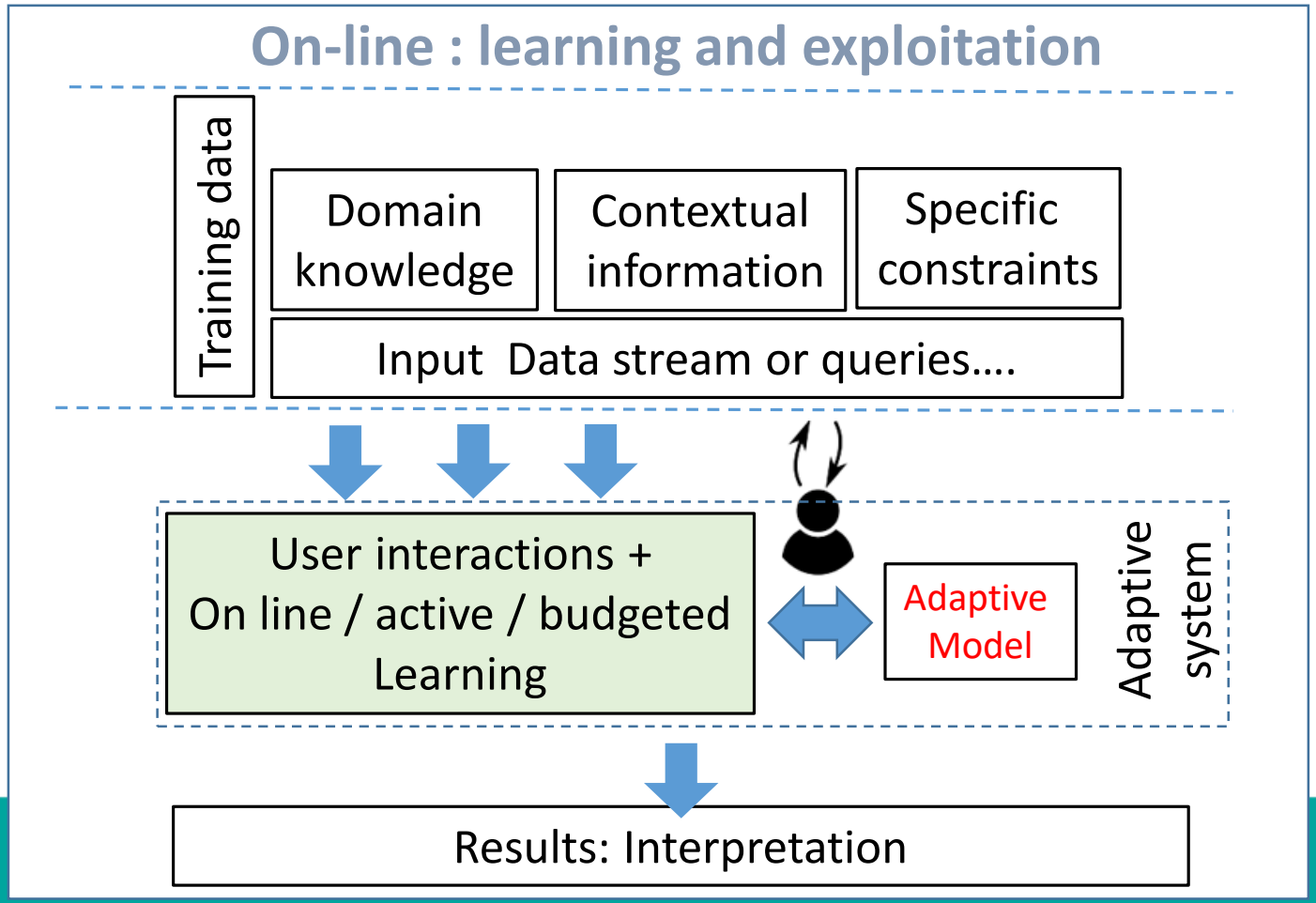
- Very interesting properties as well for CNN as for GNN → Salient ROI detection
- Users can make **a posteriori analysis** done on classical deep architectures
- **The users are not really in the loop**





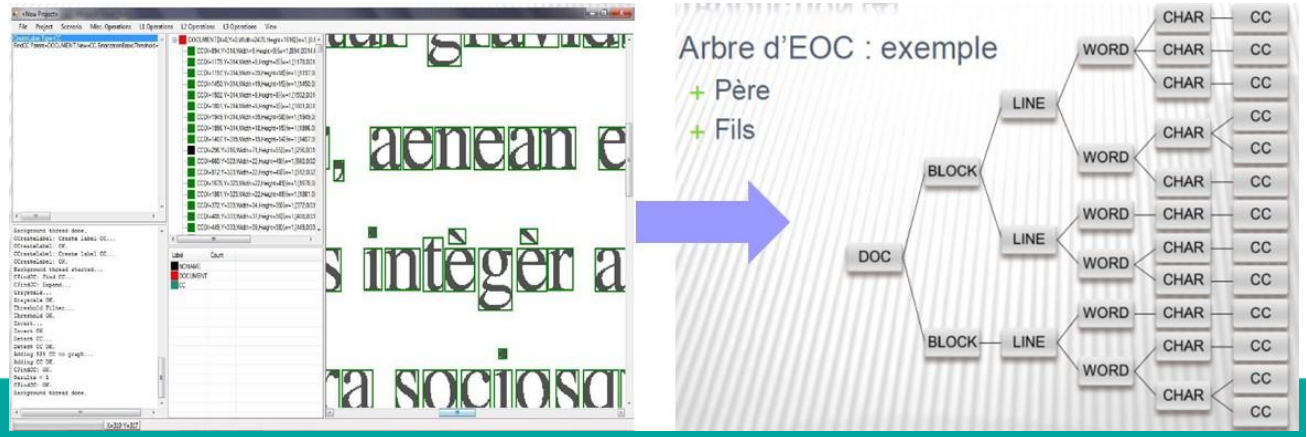
Adaptive & interactive systems

- **Inside the system,**
  - Adaptive models are updated **ON-LINE → the LOOP is here !!!**
  - **Adaptations are supervised by the system or by the USER (in the loop)**



Adaptive & interactive systems → Agora

- **Interactive Indexation and transcription of old documents**
  - **User-driven methods** for layout analysis of old books
  - **User defined Content** extraction in historical documents
- **Our proposal : Interactive definition of analysis procedures**
  - Allow users to use their own knowledge about the documents to process
  - Allow users to define the adequate order and criteria for the extraction and recognition of the elements considered as relevant at one time
  - Allow an interactive construction of scenario with direct feedback on a typical image
  - Avoid the encoding of a static and specific model of documents
  - User-guidance during the processing procedure → easier first, avoid mistakes

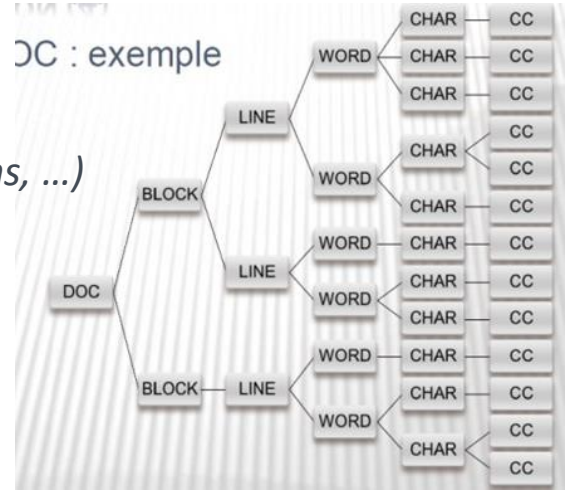


- Google Award Project: https://sites.google.com/site/paradiitproject/

Adaptive & interactive systems → Agora

Understandable Data Representation and Processing Operators

- **Data representation**
  - A dynamic **tree of EoC**
  - EoC = user-defined element *(char, word, graphic, noise, captions, …)*
  - Tree of EoC → Document contents + Layout
- **Processing Operators**
  - Low level primitives extraction *(connected components)*
  - EoC Relabeling *(small CC → Noise)*
  - EoC Merging *(Char → word → sentences…)*
- **Three types of criteria**
  - Geographic position of the EoC
  - Neighborhood relationship
  - Features (size) of the EoC
- **Customized analysis sequences (scenario)**
  - Build by the user according to the book specificities & user needs



1. Extraction des composantes connexes CC par Expand
2. Détection des IMAGES (CC de taille supérieure à 100 pixels)
3. Détection et suppression des petites CC = NOISE afin de conserver uniquement de quoi construire grossièrement des lignes de texte et les images
4. CHAR = CC restantes
5. Construction grossière des LINE (= ensemble de CHAR alignés horizontalement)
6. Affinage de ces lignes : découpage d'une LINE en plusieurs LINE2 si CHAR très espacés
7. Elimination des LINE2 trop petites
8. Construction des BLOCK à partir des LINE2, et simplification des BLOCK qui s'intersectent
9. Construction d'un bloc englobant LAYOUT et différenciation PROSE/VERS selon l'alignement avec le bloc englobant
10. Exportation Alto

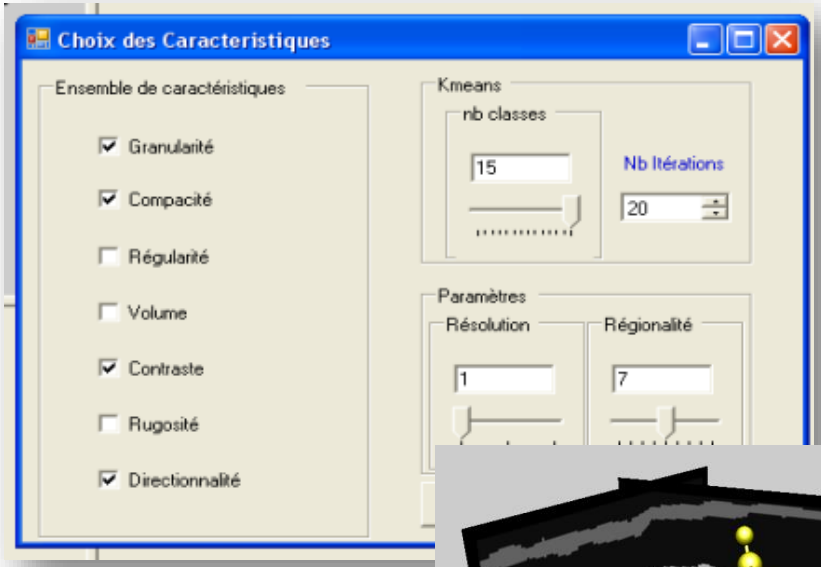
Adaptive & interactive systems → 3DimgSeg

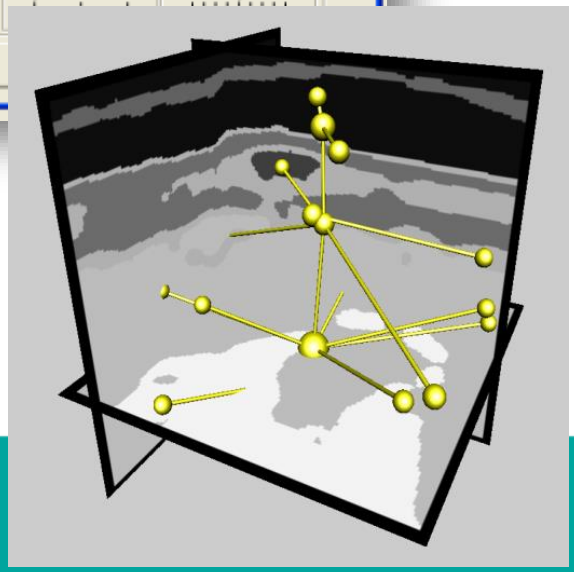
Interactive Segmentation of 3D ultrasound images (on-line)

- As in Agora → Incremental segmentation using understandable operators

- **Data representation**
  - A graph of segmented Regions

- **Processing operators**
  - Split Region & Merge Regions
    - **Merge = 1 clic on 1 edge**
    - **Split = 1 clic on 1 node**
  - Split = Kmeans defined by the user using **understandable features**
  - Direct visualization of the results
  - Undo or not

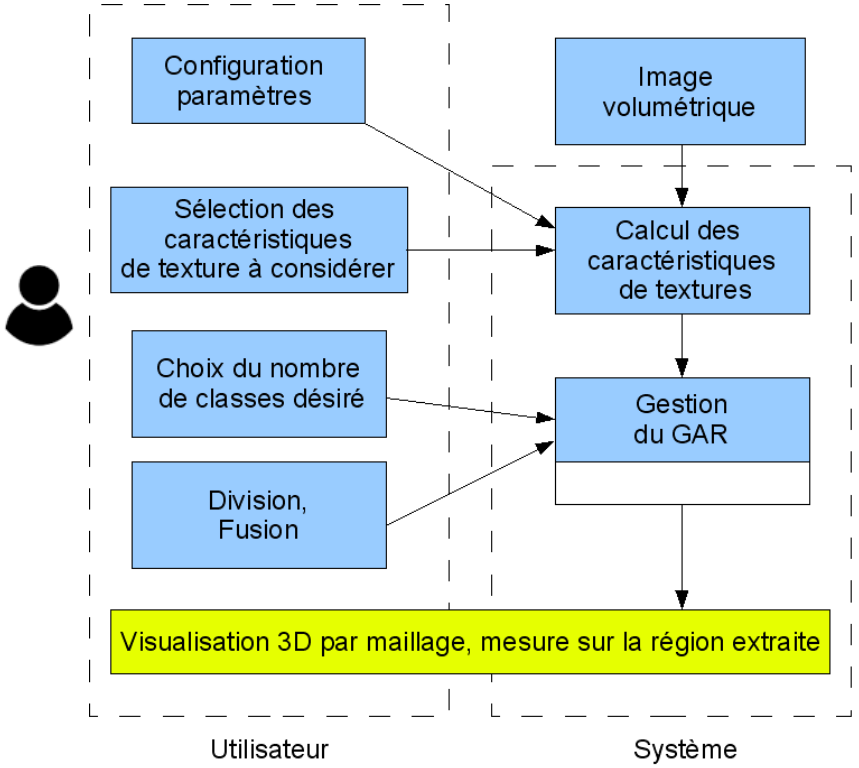


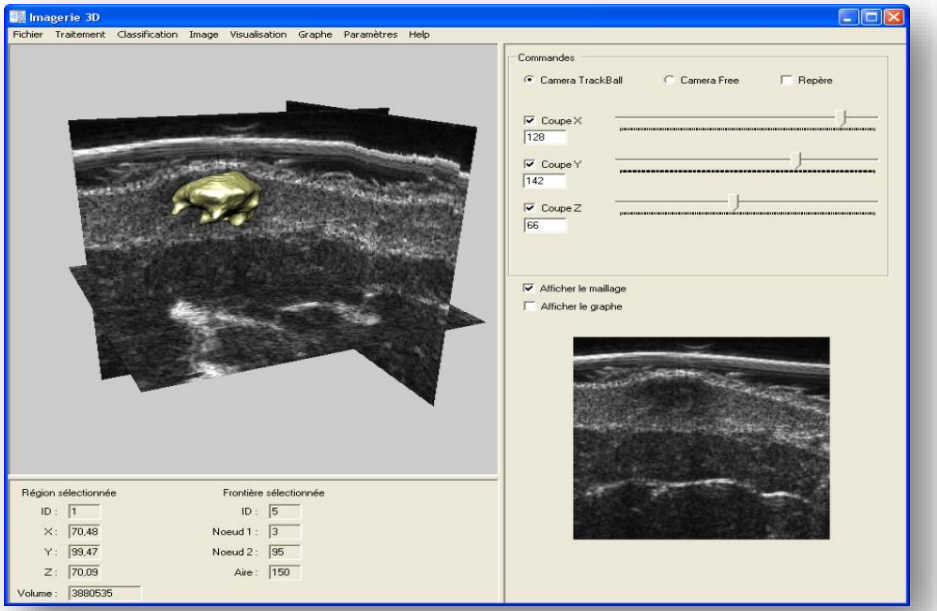


L. Paulhac, J.-Y. Ramel, et P. Makris, « A combined topological and statistical approach for interactive segmentation of 3D images », *Machine Vision and Applications*, Vol 24(6). Sept. 2013.

Adaptive & interactive systems → 3DimgSeg

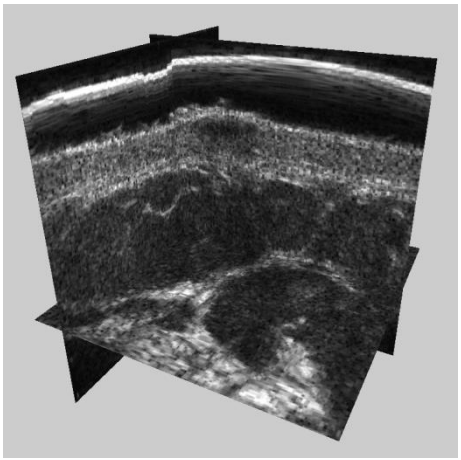
- **Interactive definition of analysis procedures (on-line)**
  - Allow users to define the adequate order and criteria for the extraction and recognition of the elements considered as relevant at one time

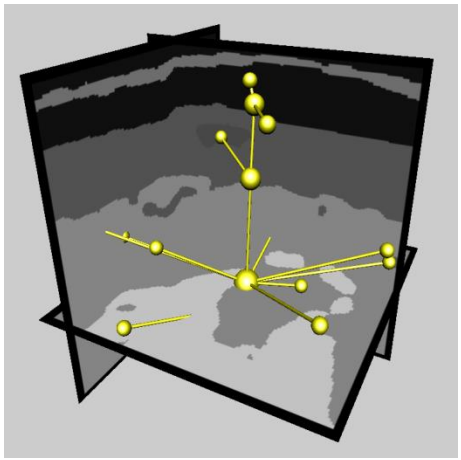




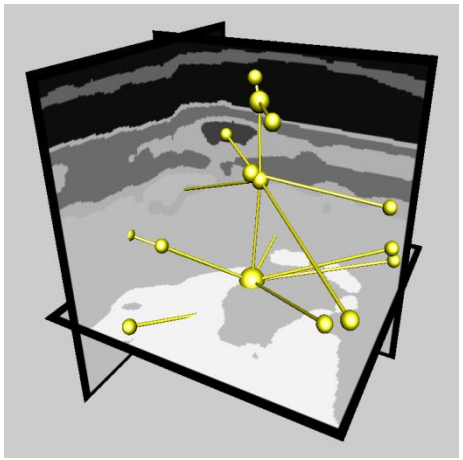
L. Paulhac, J.-Y. Ramel, et P. Makris, « A combined topological and statistical approach for interactive segmentation of 3D images », *Machine Vision and Applications*, Vol 24(6). Sept. 2013.

Scénarios de segmentation

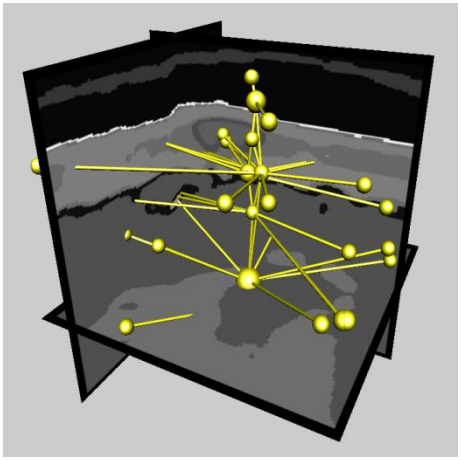




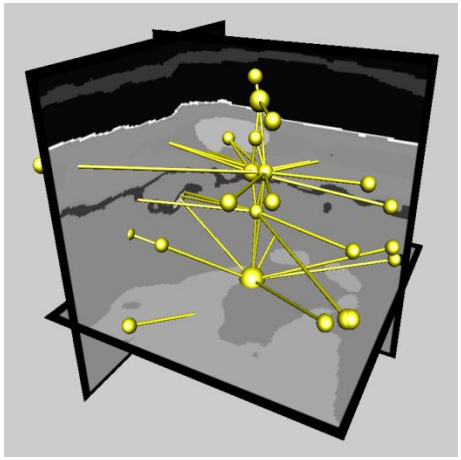
Division, K=2



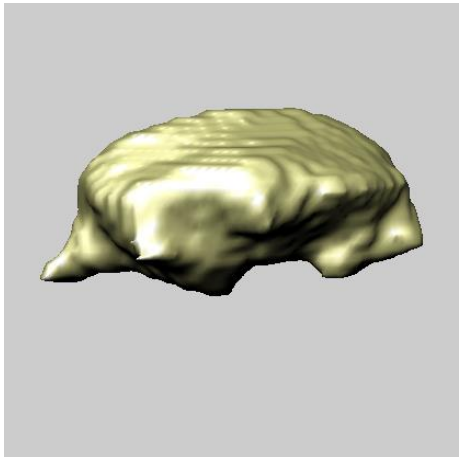
Division, K=2



Division, K=2

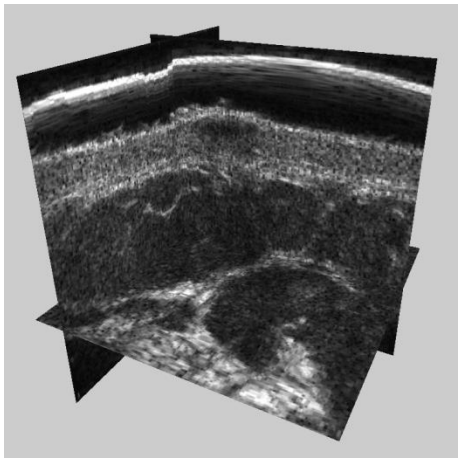


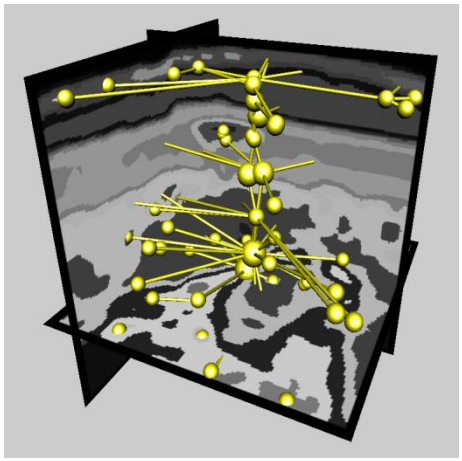
Fusion



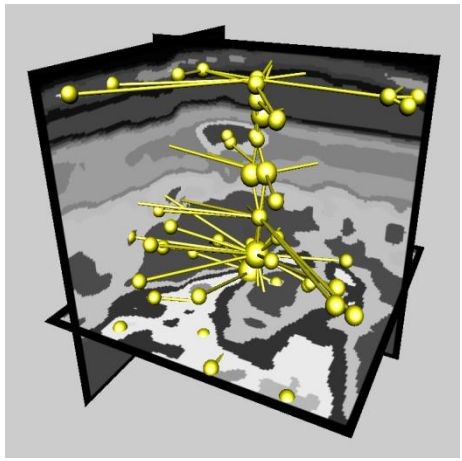
Résultat scénario 1

Scénarios de segmentation

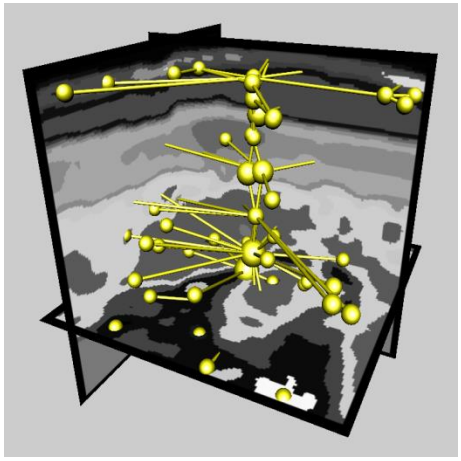




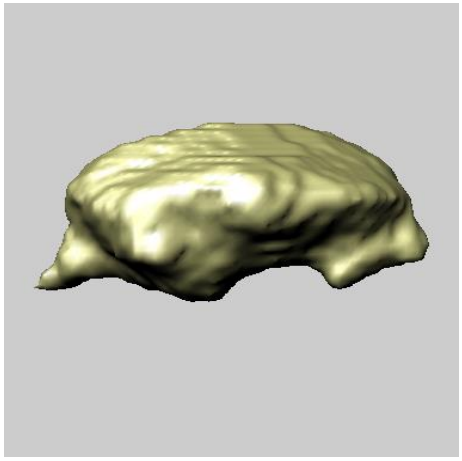
Division, K=6



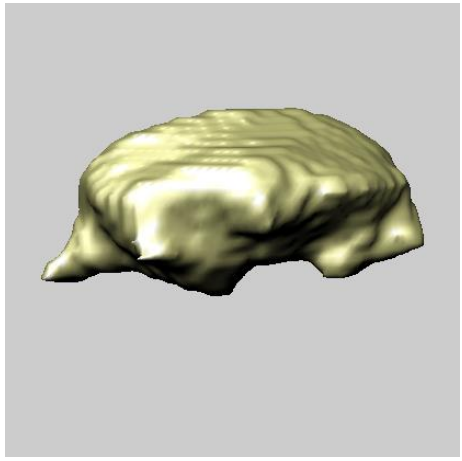
Fusion



Fusion



Résultat scénario 2



Résultat scénario 1

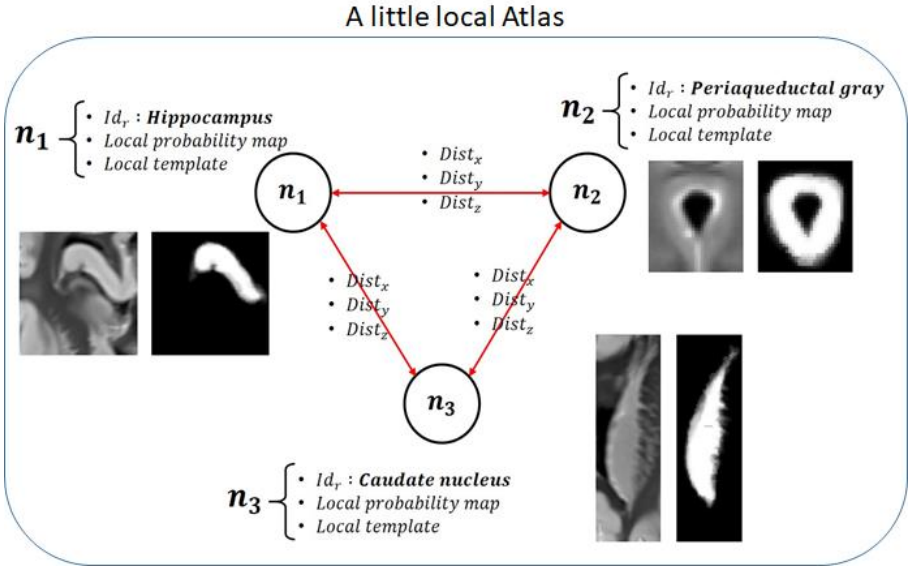
Adaptive & interactive systems → SILA3D

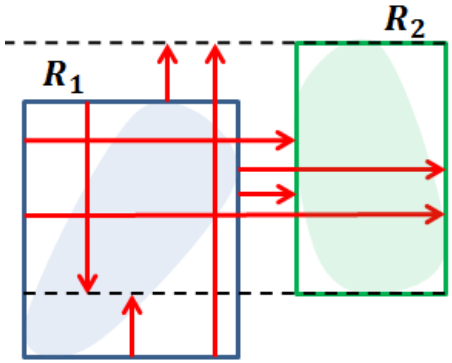
SILA-3D : Interactive and incremental Segmentation (semantic segmentation)

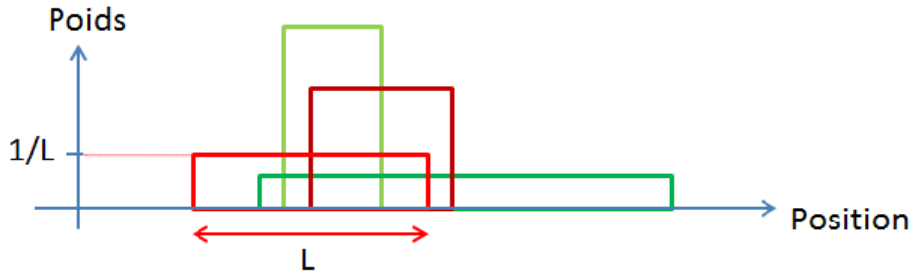
Learnable Knowledge Representation

- A Graph Representation (GAL)
- 1 node = 1 ROI = 1 dedicated classifier
- 1 edge = spatial relationships between 2 ROI

- Learning local classifiers for each ROI
- Learning spatial relationships between ROI





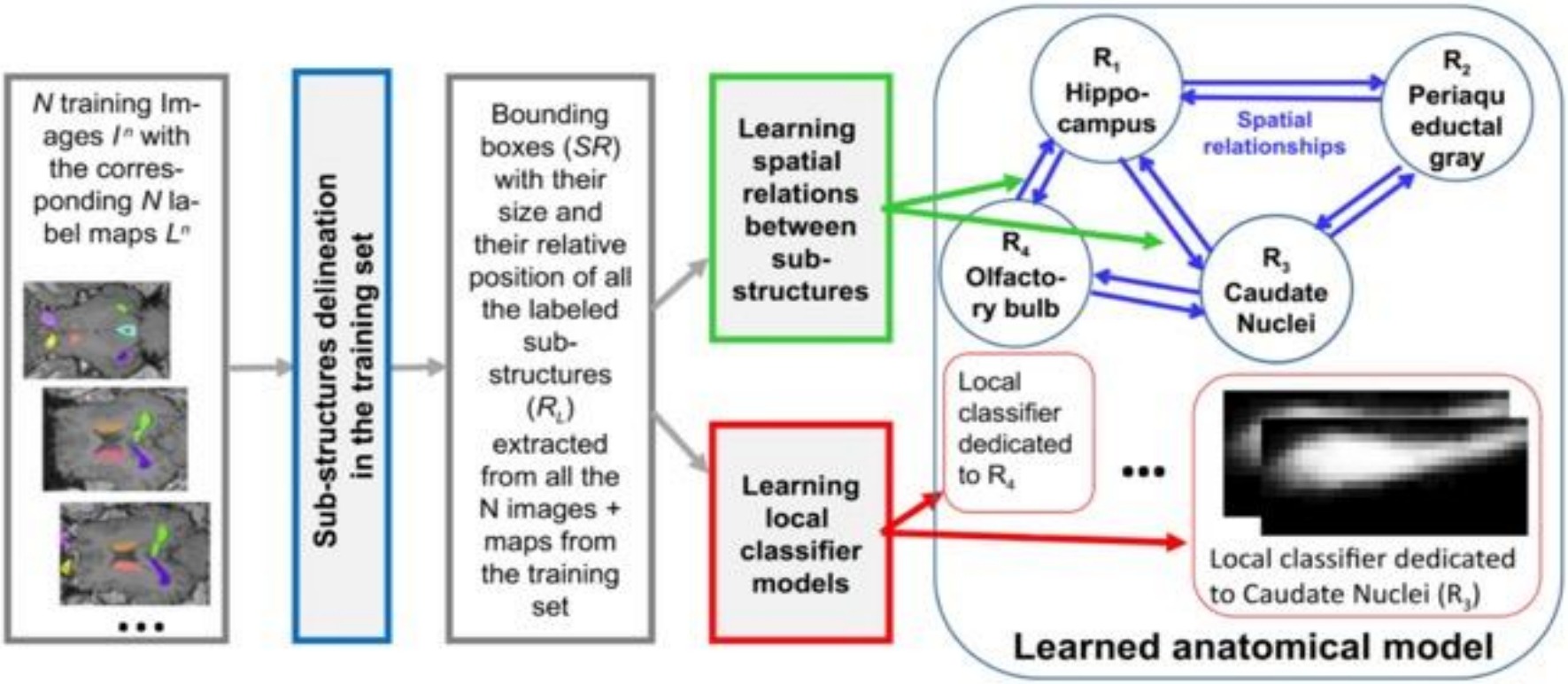


Gaëtan Galisot, aJean-Yves Ramel, Thierry Brouard, Elodie Chaillou, Barthélémy Serres Visual and structural feature combination in an interactive machine learning system for medical image segmentation. Machine Learning with Applications. Volume 8, 15 June 2022, URL

Adaptive & interactive systems → SILA3D

SILA-3D : Interactive and incremental Segmentation (semantic segmentation)

Learnable Knowledge Representation

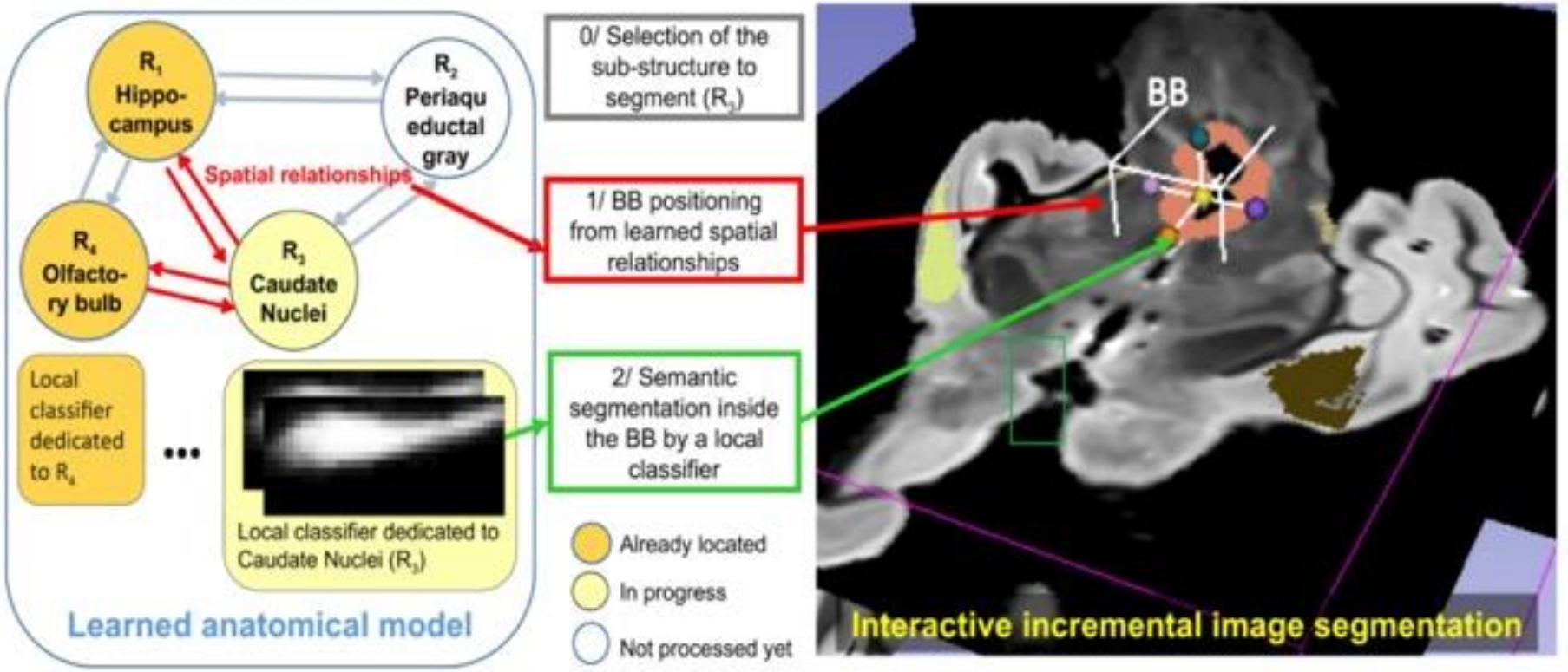


Gaëtan Galisot, aJean-Yves Ramel, Thierry Brouard, Elodie Chaillou, Barthélémy Serres Visual and structural feature combination in an interactive machine learning system for medical image segmentation. Machine Learning with Applications. Volume 8, 15 June 2022, URL

Adaptive & interactive systems → SILA3D

SILA-3D : Interactive and incremental Segmentation (semantic segmentation)

Incremental interactive segmentation



Gaëtan Galisot, aJean-Yves Ramel, Thierry Brouard, Elodie Chaillou, Barthélémy Serres Visual and structural feature combination in an interactive machine learning system for medical image segmentation. Machine Learning with Applications. Volume 8, 15 June 2022, URL

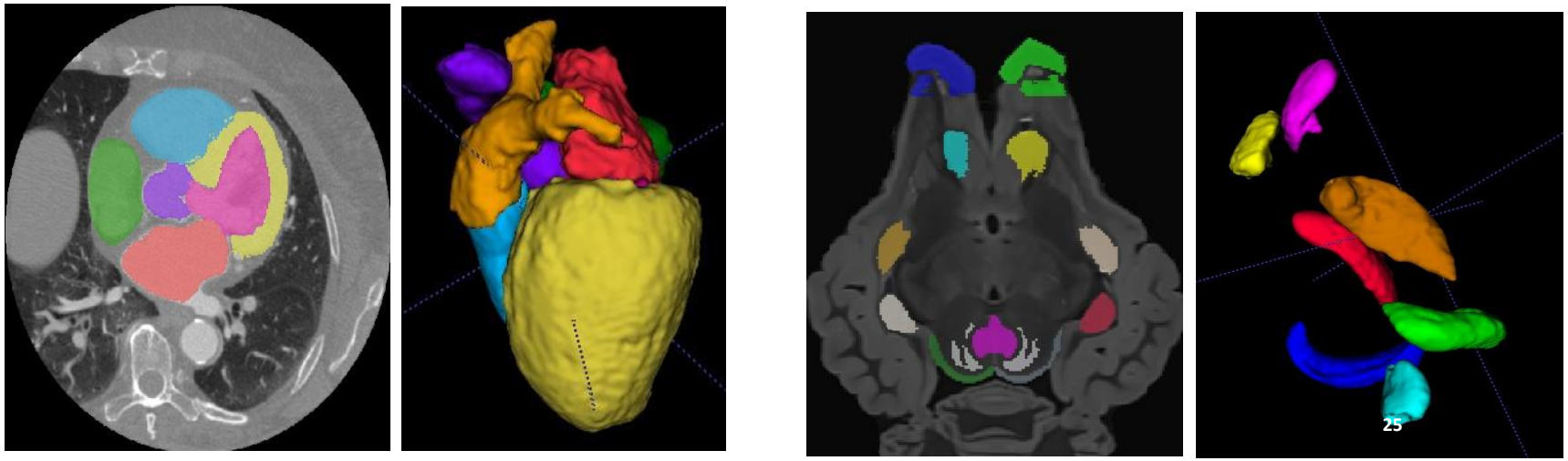
Adaptive & interactive systems → SILA3D

SILA-3D : Interactive and incremental Segmentation (semantic segmentation)

Incremental interactive segmentation

- Utilization of one of the learned GAL
- Selection of the order of ROI segmentation
- Visualization, correction and validation of each intermediate results (for each ROI)

- **On-line leaning supervised by the user**
  → GAL updating or not



Gaëtan Galisot, aJean-Yves Ramel, Thierry Brouard, Elodie Chaillou, Barthélémy Serres Visual and structural feature combination in an interactive machine learning system for medical image segmentation. Machine Learning with Applications. Volume 8, 15 June 2022, URL

Adaptive & interactive systems → ML / DL ?

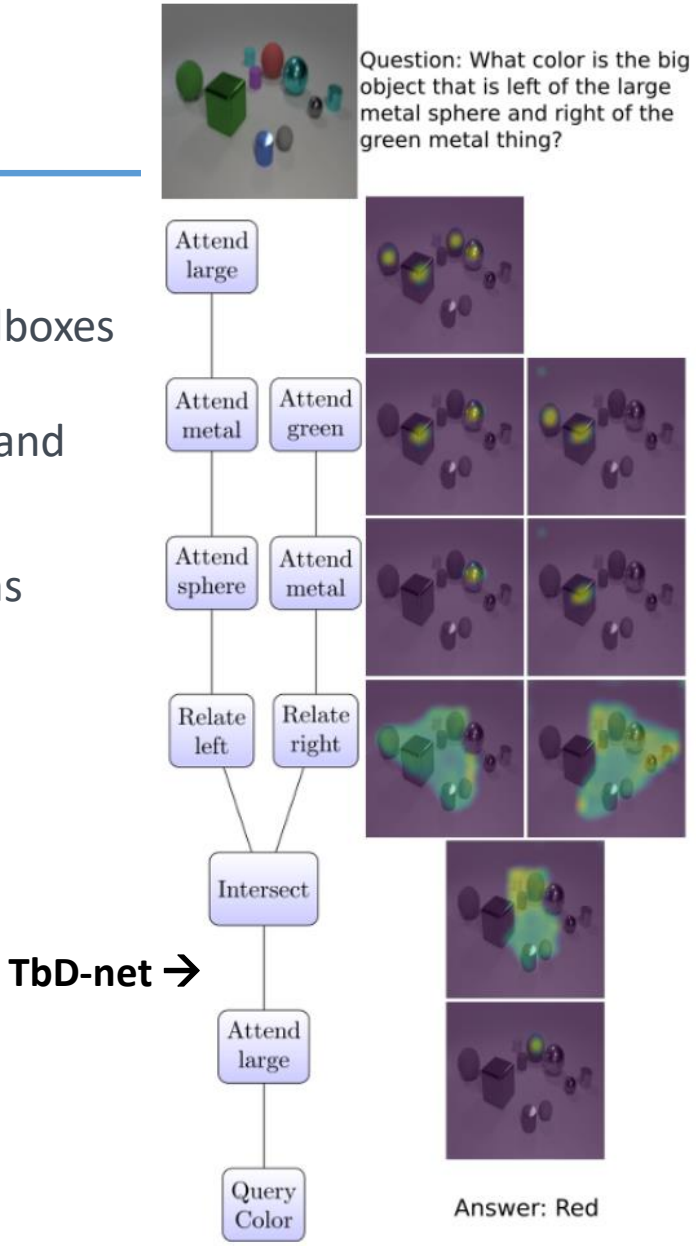
Requirements for online evolving ML / DL systems

- **Online learning**
  - Incremental learning from few initial learning data
  - Adapt models according to new data without requiring all the original data
  - Preserve previously acquired knowledge (no catastrophic forgetting)
  - Memory and computing time must be limited
  - System learning can be interrupted and its quality shouldn't be altered
- **On-line active learning**
  - A classifier can achieve equivalent performance with only part of the learning data, if those data have been correctly chosen
  - The learning system itself will choose which data samples will be used
  - Need method to evaluate the classifier confidence during recognition
  - Ask the users to decide when to query the label of the sample
  - Decide the label of the new samples (Semi supervised learning)
  - Ask the users to label data samples for which the system is likely to make a recognition error and which will be very interesting for the evolving classifier learning
- **Budgeted Learning & incremental classification**
  - New systems need problem resolution under time and memory constraints
  - Main ideas At test time, compute & use costly features only if necessary (utility scores)
  - New strategy → cost vector associated to the features → weak classifiers like in Adaboost

Conclusion

- **What should we remember?**
  - Adapted & static methods → a lot of operational toolboxes in CV, PR, ML, …
  - Adaptable methods → Off-line learning (from datasets) and from human interaction
  - Adaptive, incremental, interactive explainable systems
    → Human supervision, active learning
  - Time and memory constraints
    → Anytime, budgeted & distributed systems
  - Transparent by design DL Systems, active DL
    → the LOOP is not present so often

- My keywords for the future
  - Graph (Neural Networks)
  - Active, Budgeted, Interactive, Incremental **but less sequential more dynamic** (perceptive cycles)



TbD-net →

Online Deep Learning: Learning Deep Neural Networks on th… y. Doyen Sahoo, Quang Phan, Jing Lu, Steven C.H. Hoi. arXiv:1711.0370 Sequential Labeling with online Deep Learning. Gang Chen, Ran Xu, Sargur Srihari. arXiv:1412.339 Transparency by Design: Closing the Gap Between Performance and Interpretability in Visual Reasoning David Mascharka∗1 Philip Tran2 Ryan Soklaski1 Arjun Majumdar. arXiv:1803.0526

MERCI