Deep learning on Galactic filaments

A. Zavagno

F.-X. Dupé, M. Gray, E. Schisano

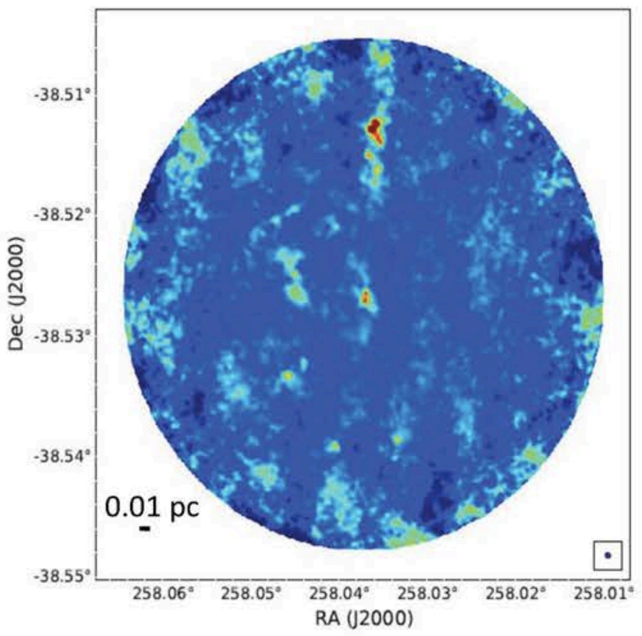
Outline

- Context
- Aims of the BigSF project. Method
- A proof-of concept study on Galactic Filaments
  - Data
  - Supervised Learning Method
  - Results
- Conclusions and perspectives

Context: Galactic & extragalactic star formation

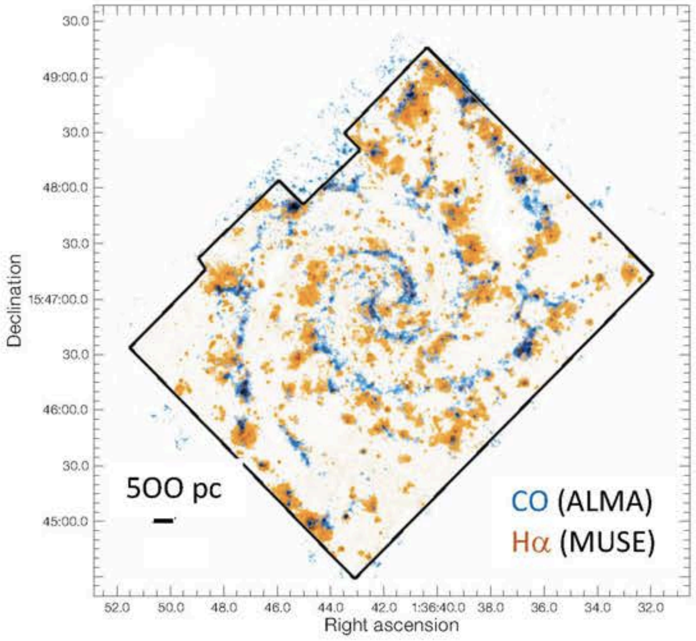
BigSF: Galactic & extragalactic SF: *Determining the role of environment on SF properties*

SF in our Galaxy



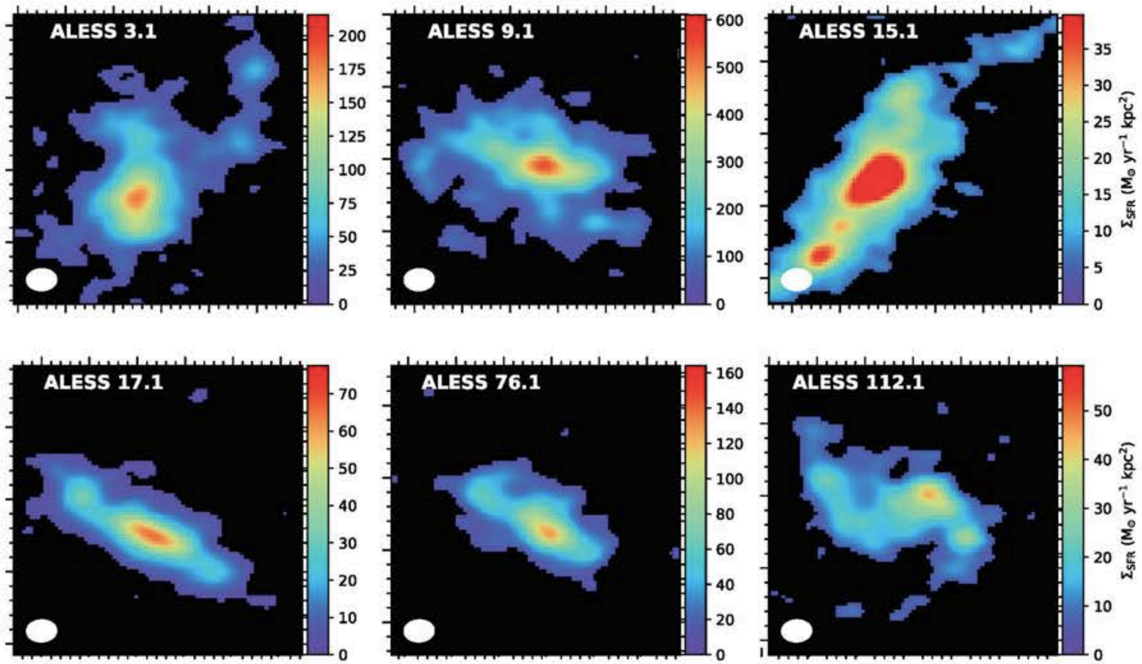
Resolved spatial scale <0.01 pc

SF in nearby galaxies (~10 Mpc)



Resolved spatial scale ~100 pc

SF in Galaxies at z ~3



Resolved spatial scale ~ 500 pc

Adapted from Figueira+2018, Schinnerer+2019, Hodge+2019

Aims of the BigSF project

- Build an empirical model of the Galactic star formation using Big Data & ML
- Star formation as a fonction of the environment (physical conditions)
- Bridging the gap between Galactic & extragalactic star formation
- Extragalactic SF: Use the built model + change the spatial resolution

→ How SF depends on the environment in galaxies ?

Method

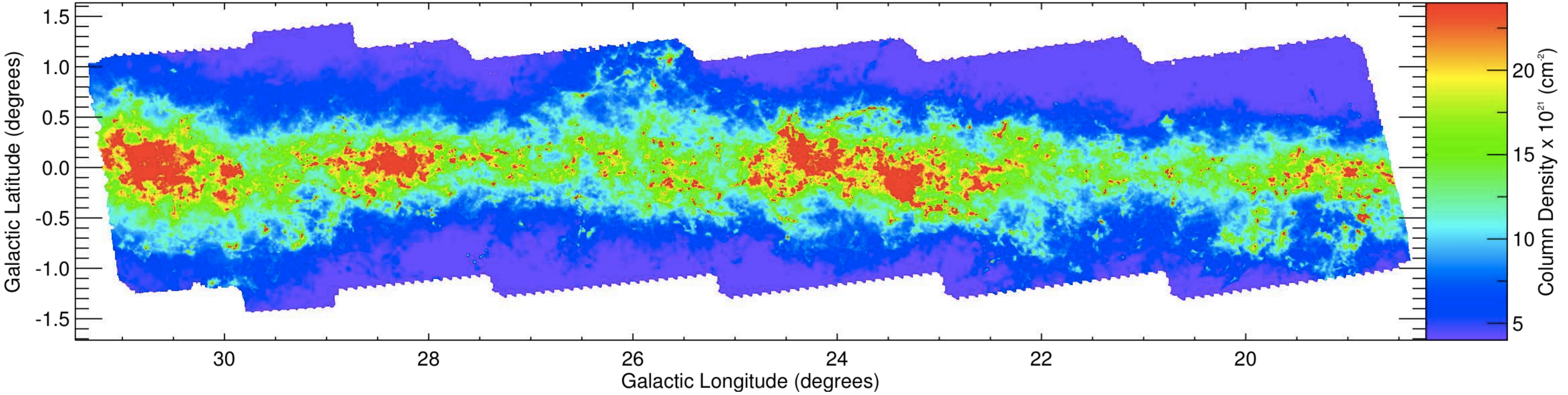
- SF components: Filaments, clumps, cores, young stars
- Galactic SF observed from 10^{-3} (ALMA) to $>10^{3}$ parsec + Information on the physical conditions
- Big data (surveys, pointed obs., images, spectra)
- Multi scale, multi wavelength → multi view
- Heterogeneous data, missing data

- Challenges
  - Build the dataset(s) for the SF components (individuals & combined)
  - Learn on the dataset(s)
  - Build the *empirical* model

Galactic filaments

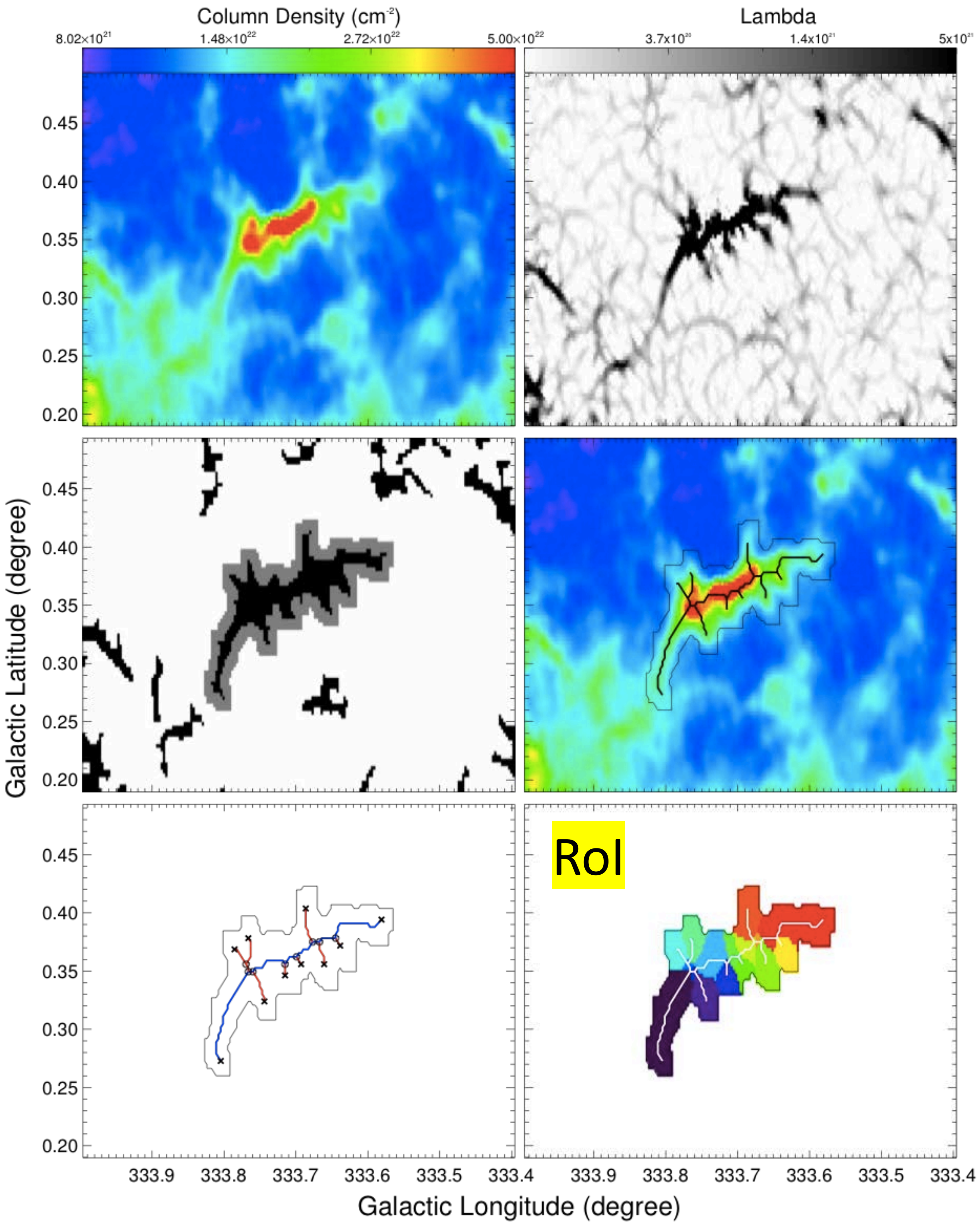
- Proof-of-concept study
- Testing the power/interest of ML on the detection of Galactic filaments

Data



Column density map of the Galactic plane with Hi-GAL (Schisano+2020, MNRAS)

Galactic Filaments (Schisano+2020, MNRAS)



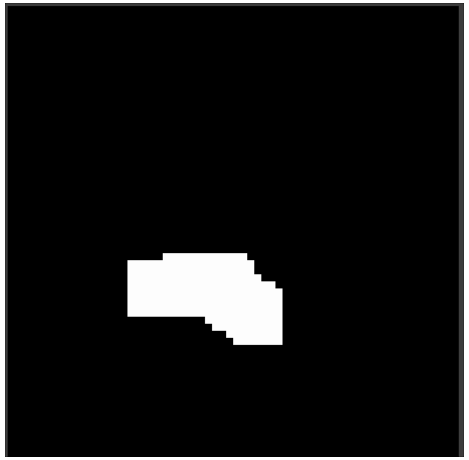
Extraction method

$N(H_2)$ images

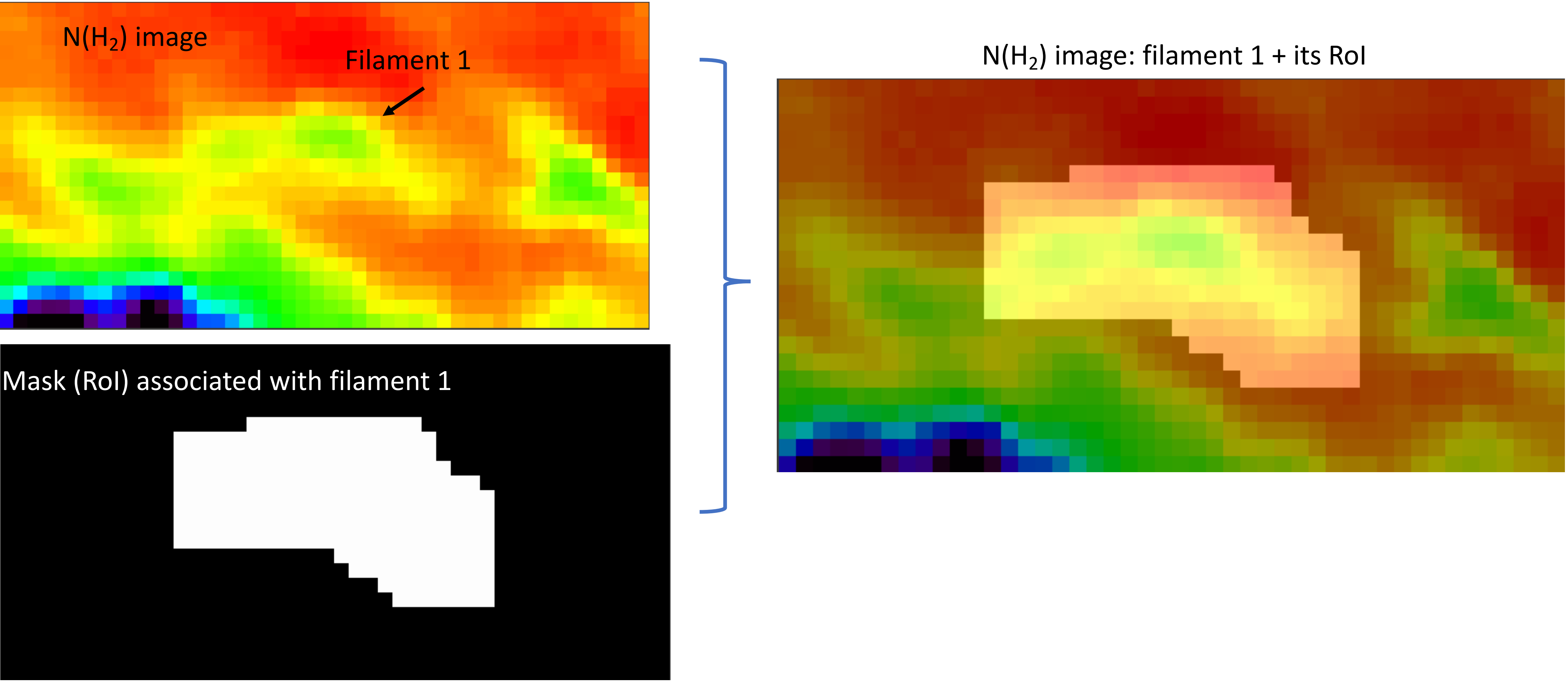
Region of Interest (RoI)

Database: 32 000 filaments (2D images)

RoI: mask (where the filament is) and 0 around



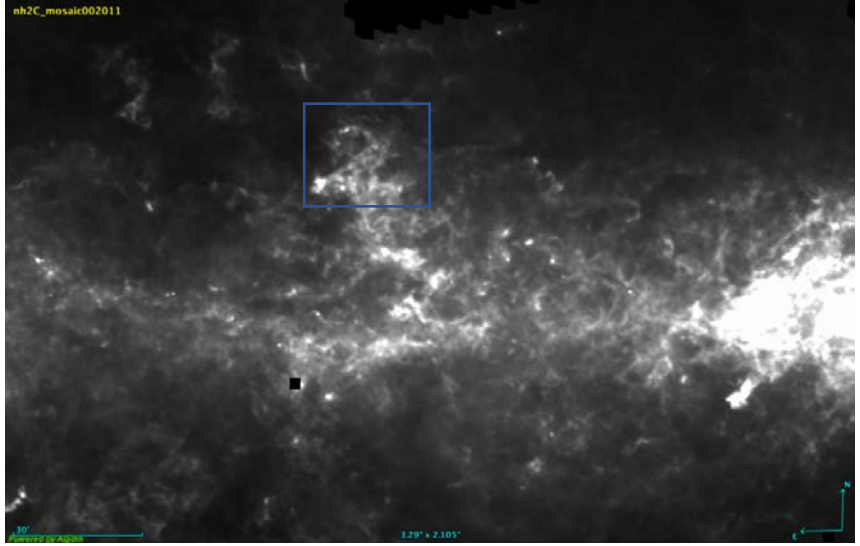
We want to learn from the data what a filament is.
→ A *representation* of a filament

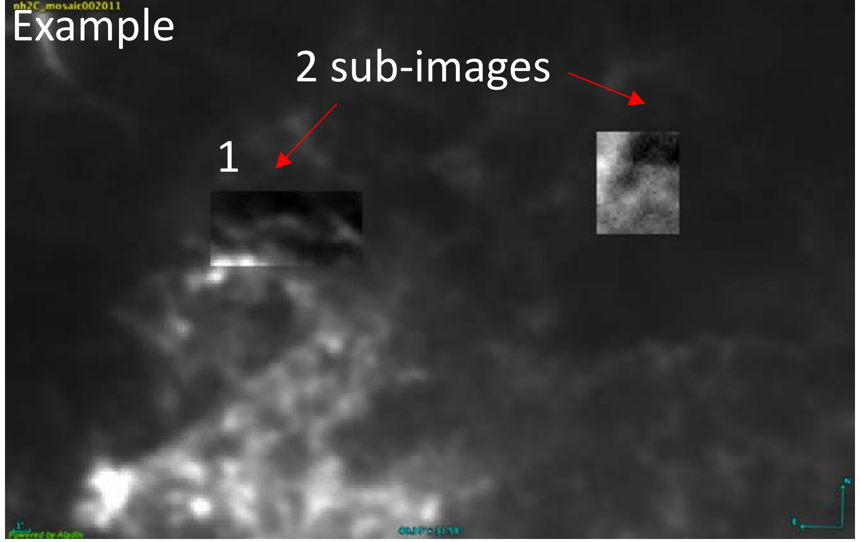


Method: supervised learning

- Database : Hi-GAL column density maps: 32 000 filaments + their associated RoI
- Images of different size
- For each filament: 3 patches (64x64) are taken from 3 random positions around the RoI
    - Must contain $\geq$ 20% of the mask
- Parts with missing data in the original image are not taken (saturation zones)
- Data augmentation: 2 flips and 3 rotations

→ ~ 900 000 patches + their associated RoI





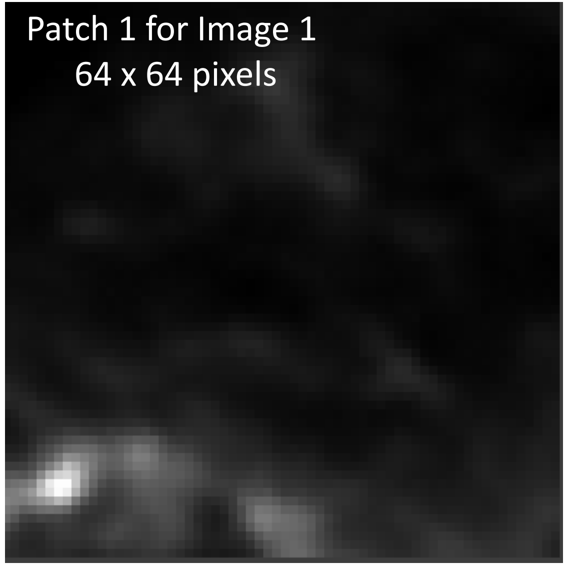


Image 1: Patch1+Mask1

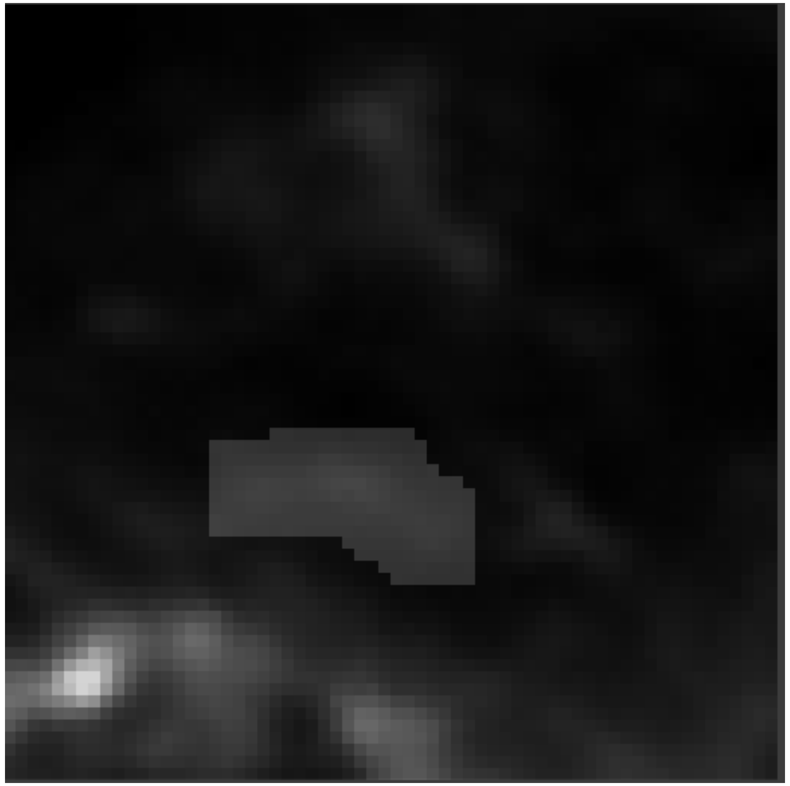


Image 1: Patch2+Mask2

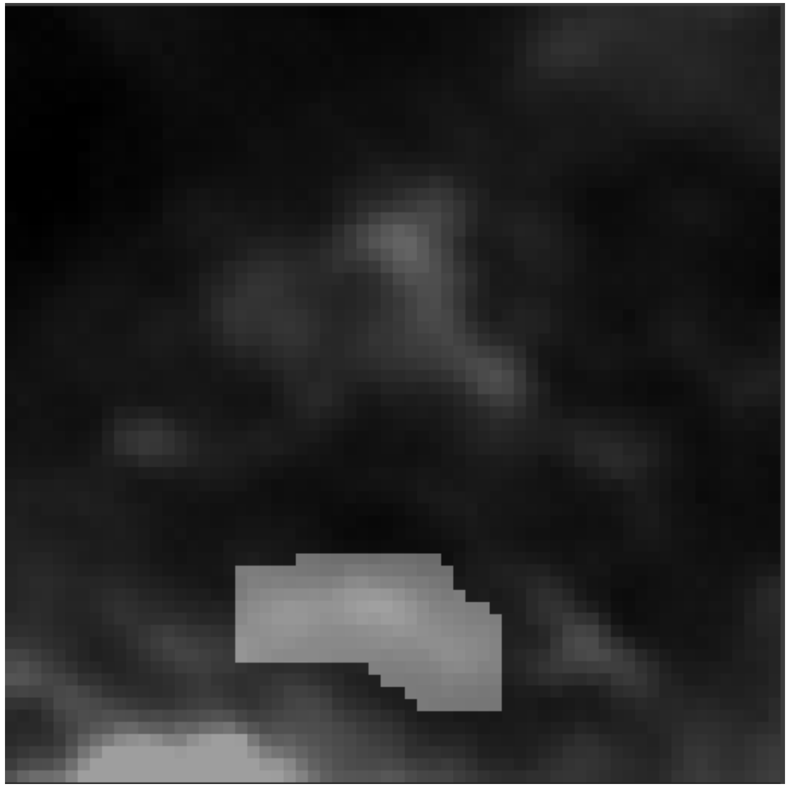
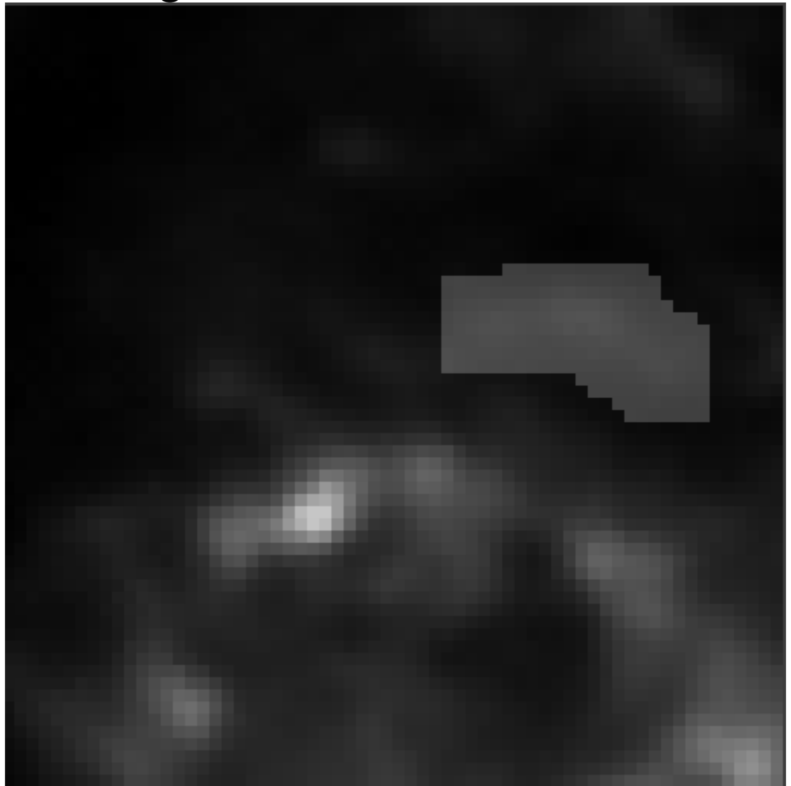


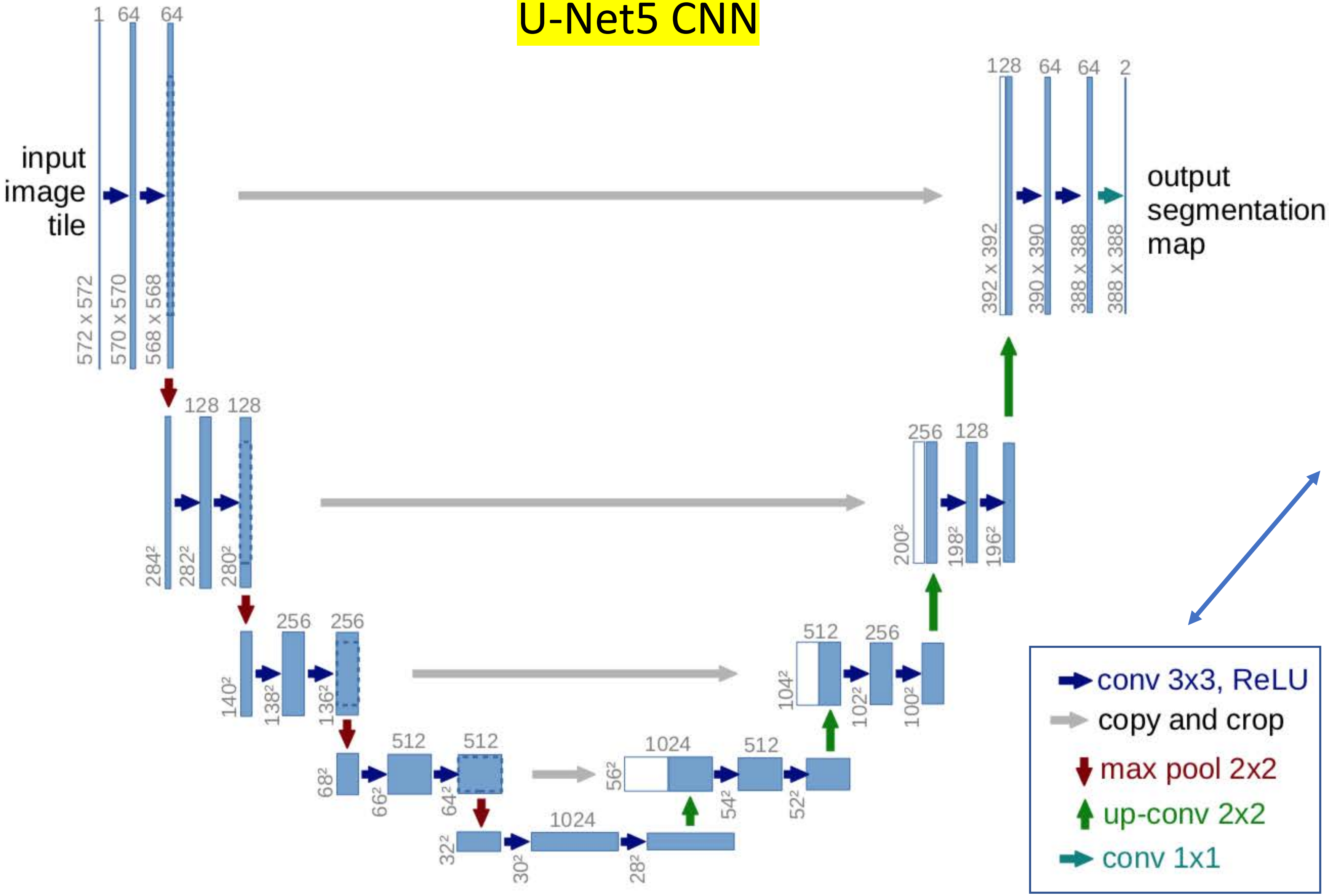
Image 1: Patch3+Mask3



Method: supervised learning

- U-Net5 Convolutional Neural Network used for image segmentation
- Base: 80% training set and 20% validation set
- Python + Keras: 20 epochs for the learning phase

U-Net5 CNN



U-net architecture (example for 32x32 pixels in the lowest resolution).

Blue box: multi-channel feature map

- *number of channels*: on top of the box
- *x-y-size*: at the lower left edge of the box.

White box: copied feature maps.

The *arrows:*the different operations.

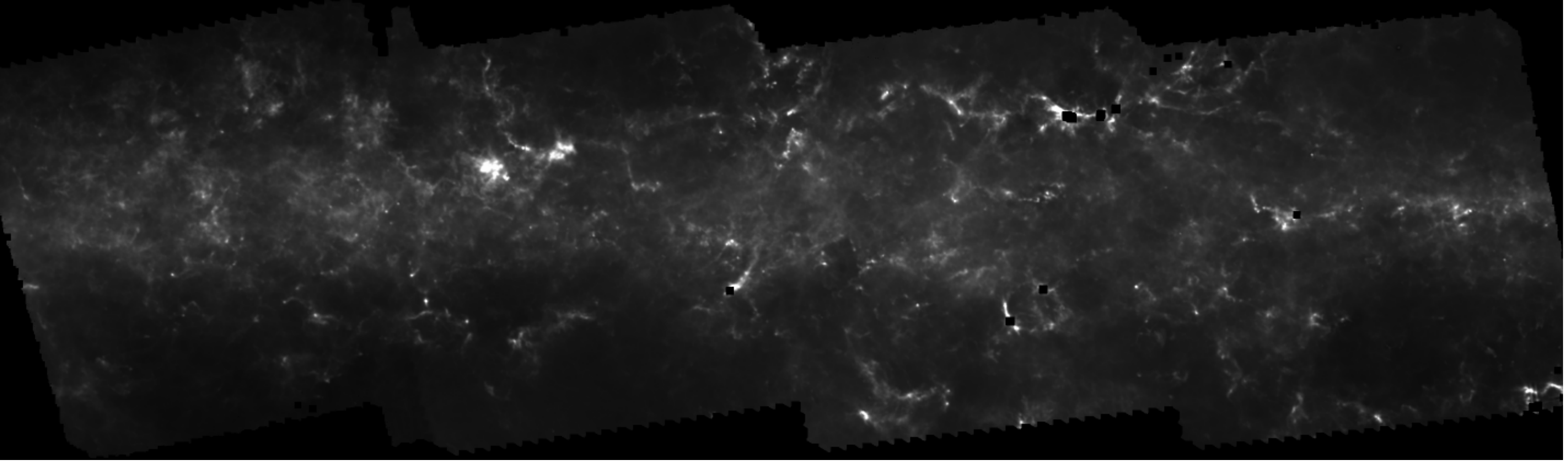
The U-Net5 we used:
5 blocks
23 layers

The machine we used:
GPU : GeForce RTX 2080Ti
CPU : Intel(R) Xeon(R) Silver 4215
CPU @ 2.50GHz

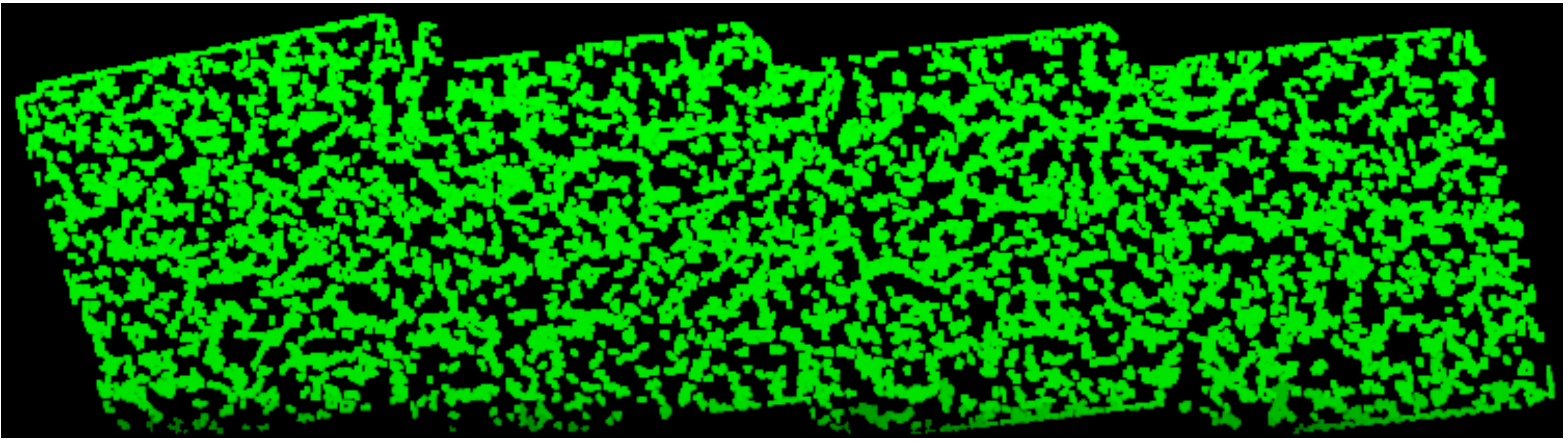
Zhang, J. https://towardsdatascience.com/unet-line-by-line-explanation-9b191c76baf5
Ronneberger+2015 arXiv:1505.04597

Results: Galactic Plane 349°-356°

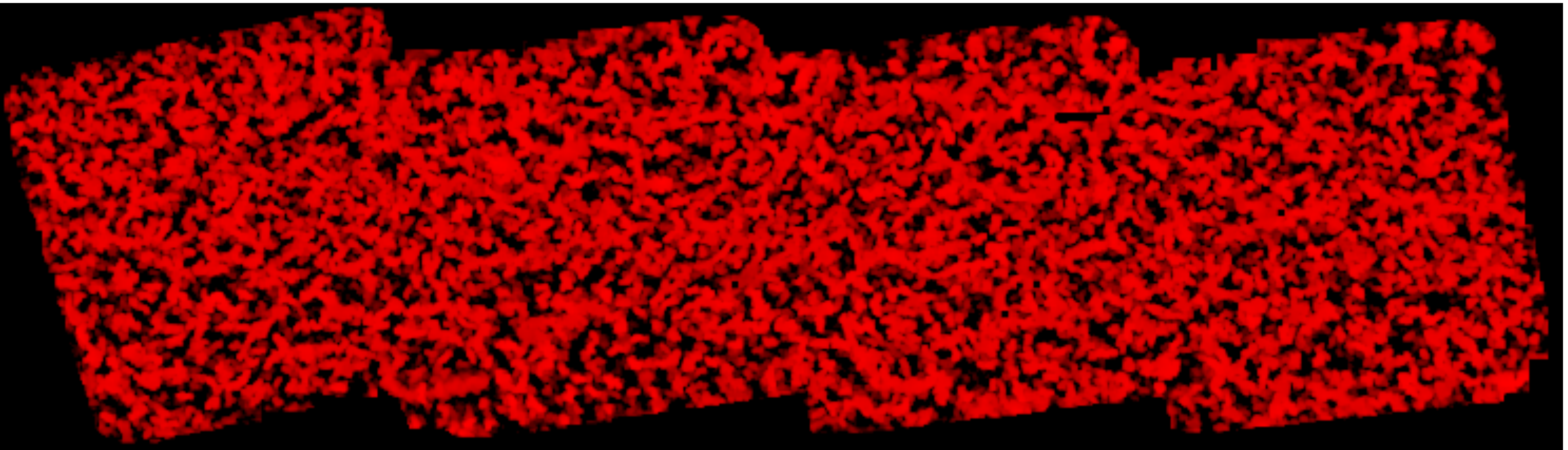
(not seen in the learning)



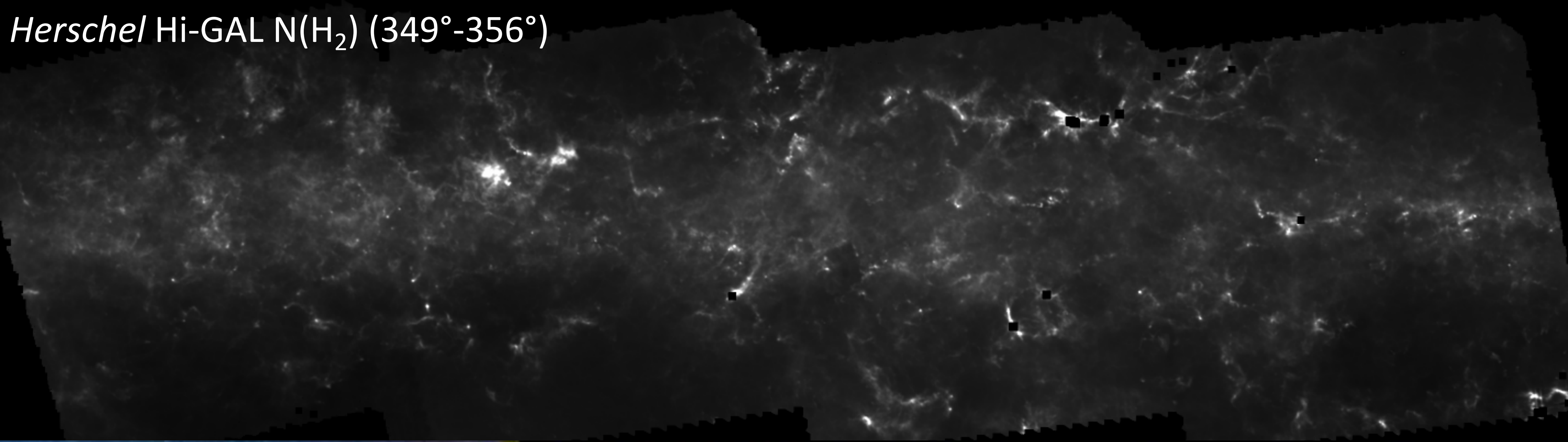
Original $N(H_2)$ image
Column density map
Herschel Hi-GAL
349°-356°
(Schisano+2020)



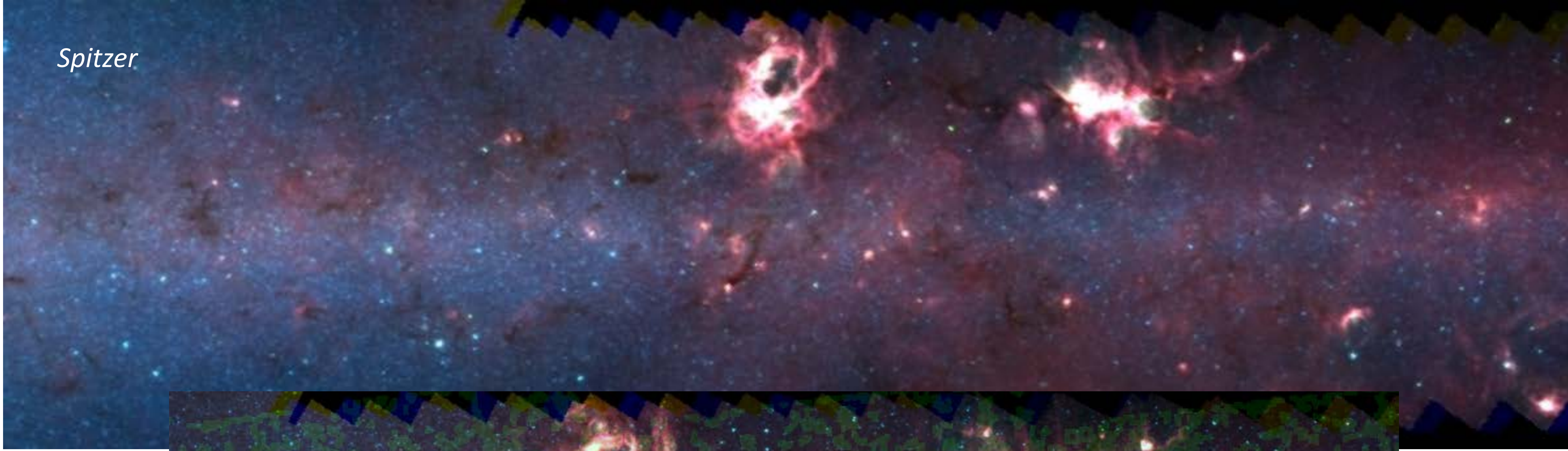
Original RoI image



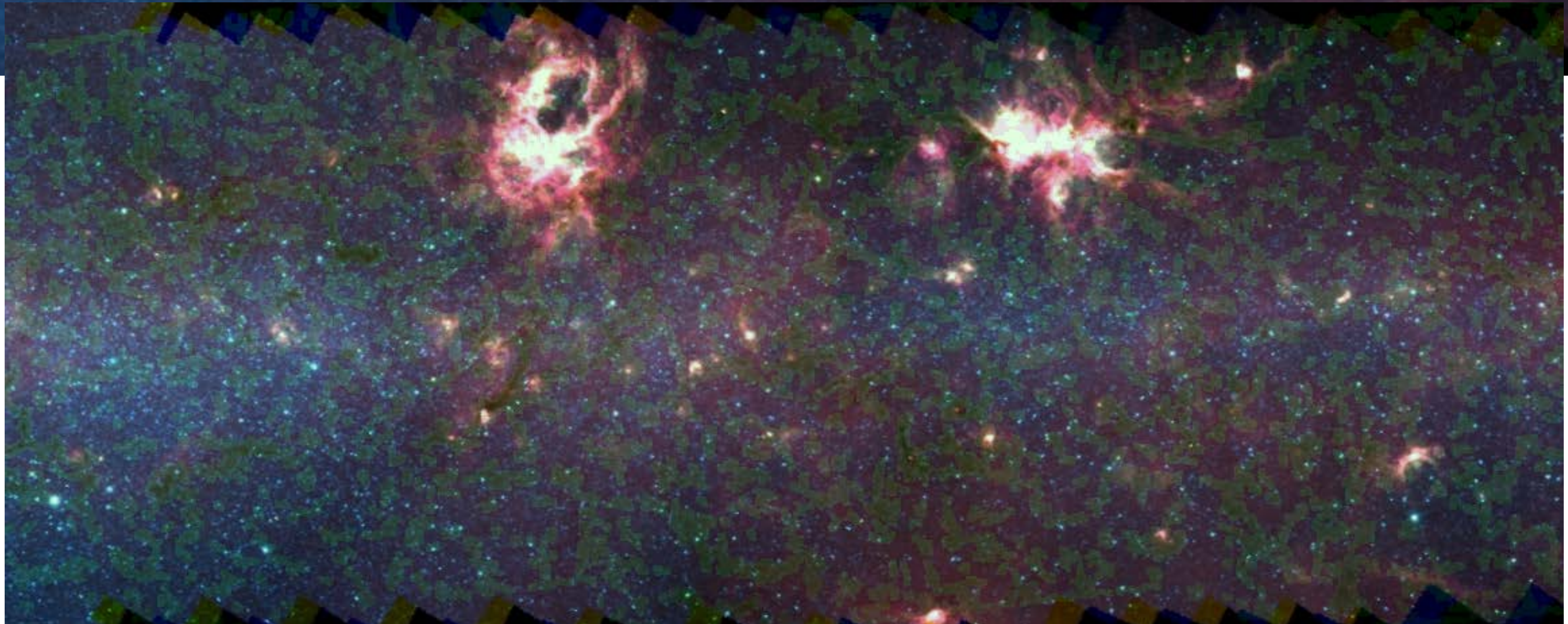
Segmented RoI image

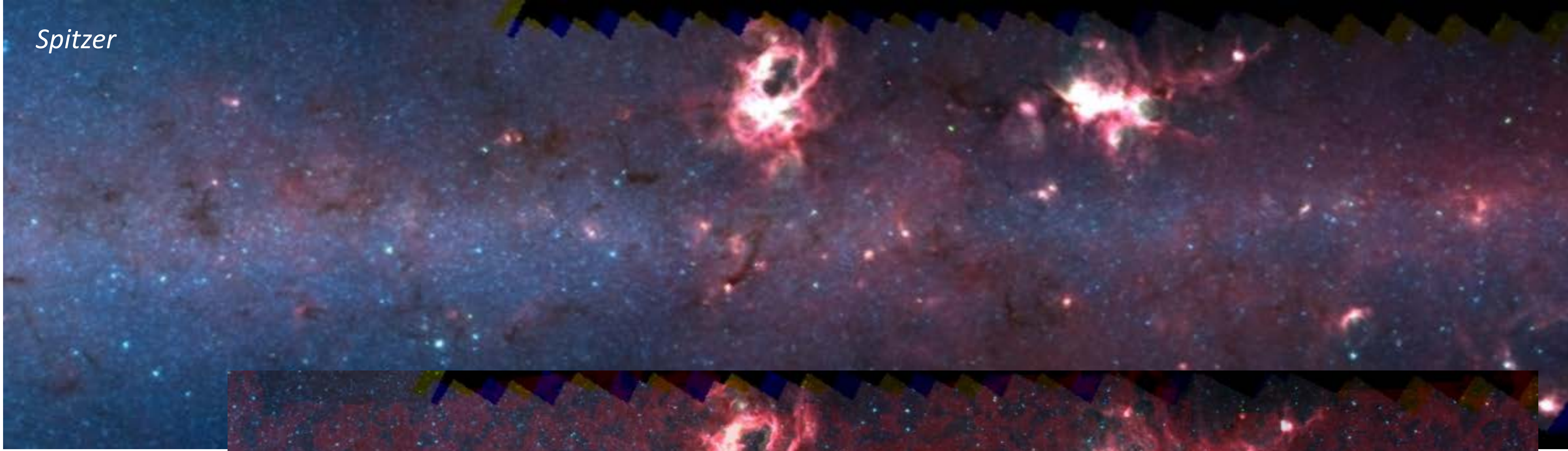

Herschel Hi-GAL $N(H_2)$ (349°-356°)

Spitzer IRAC mid-IR

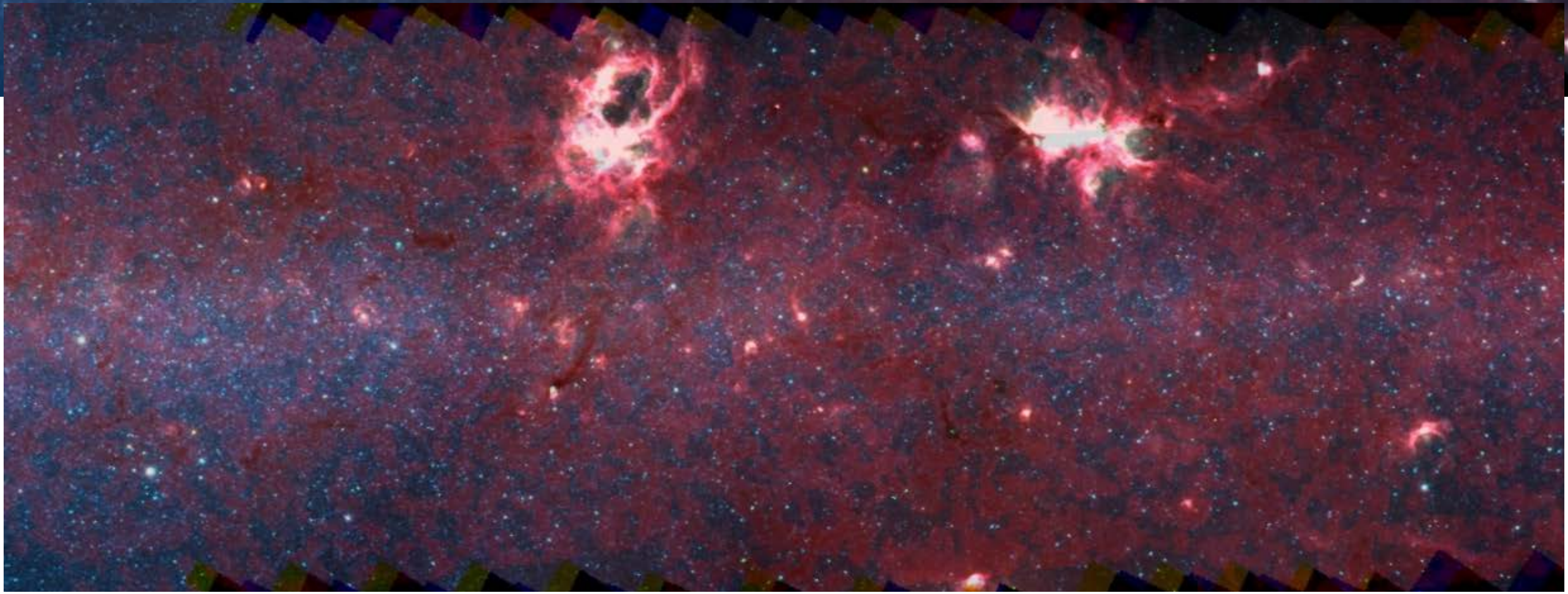


Spitzer
+
Original RoI

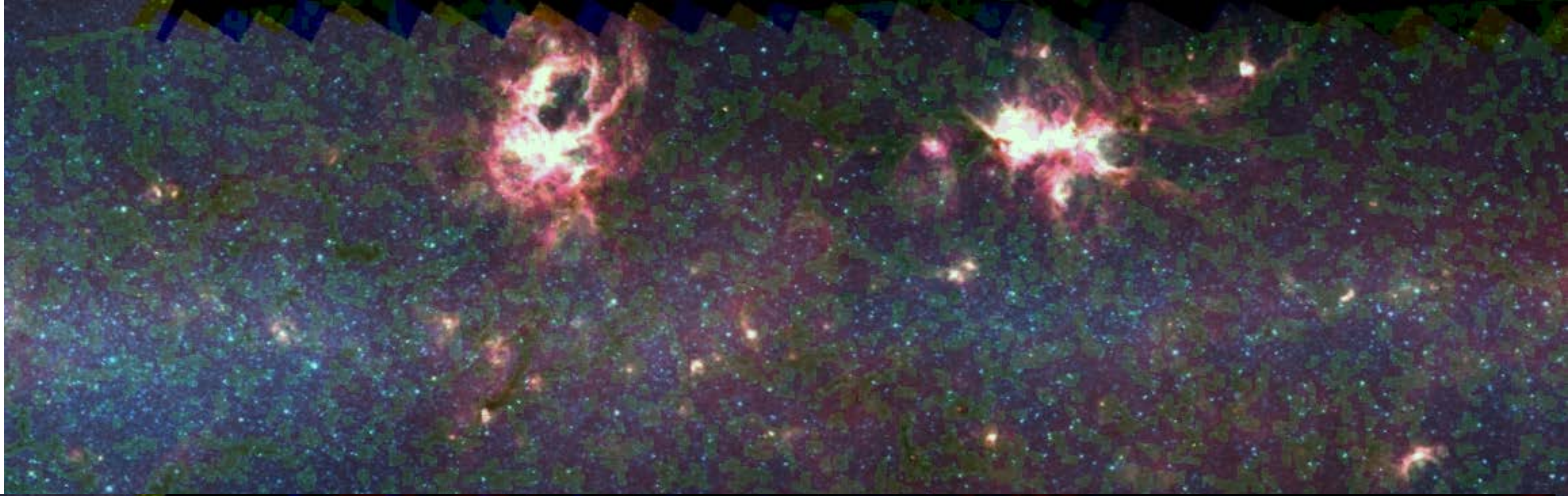




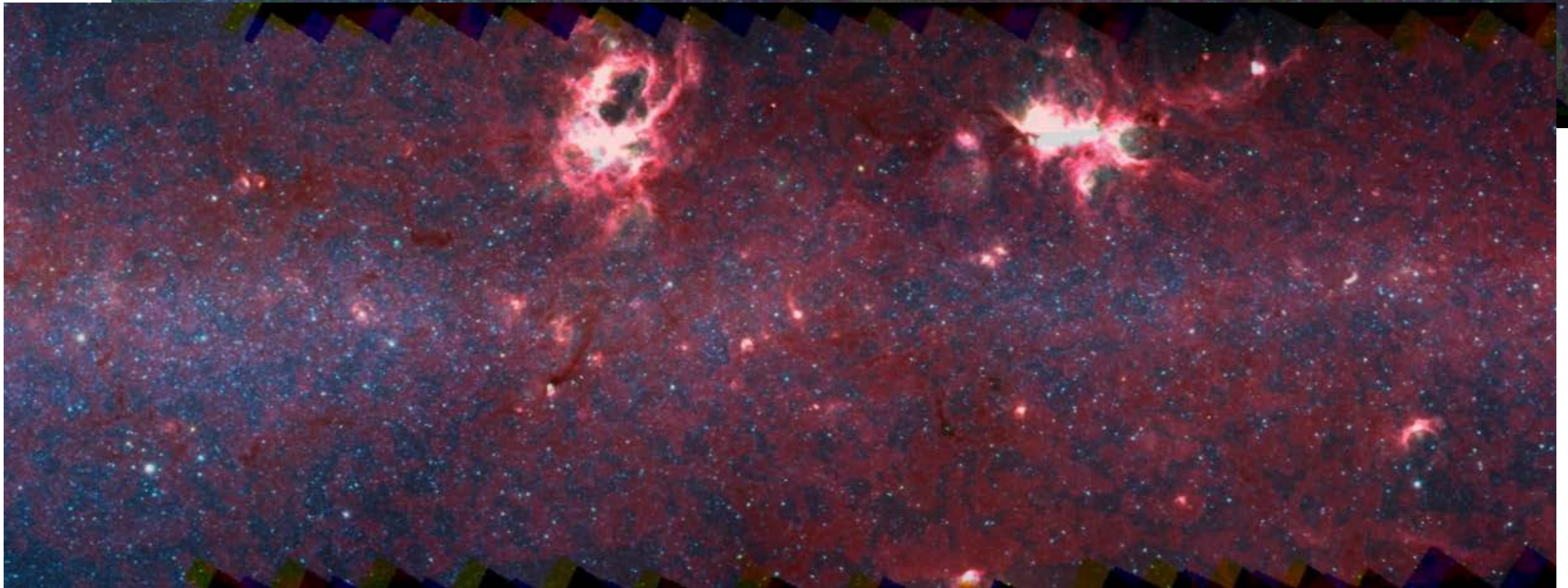
Spitzer
+
Segmented RoI



Spitzer
+
Original RoI



Spitzer
+
Segmented RoI

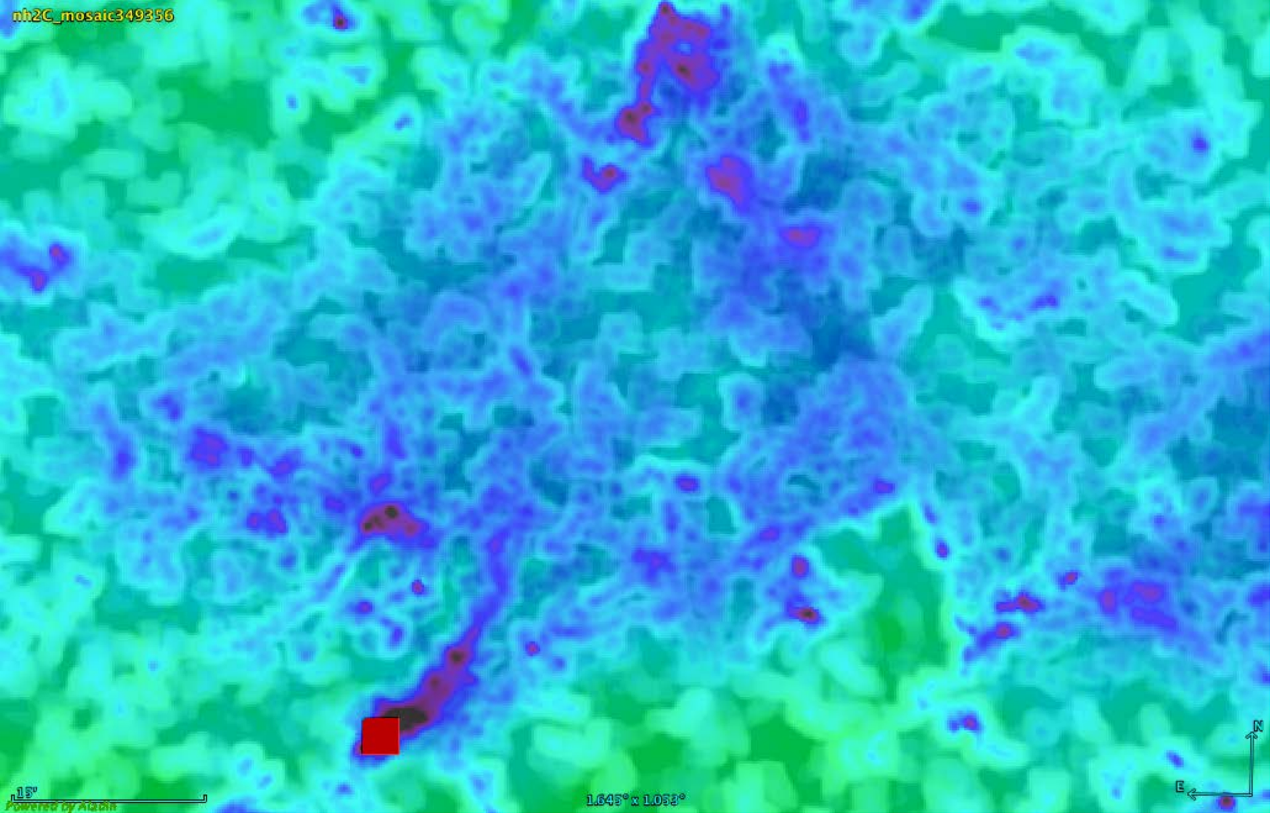


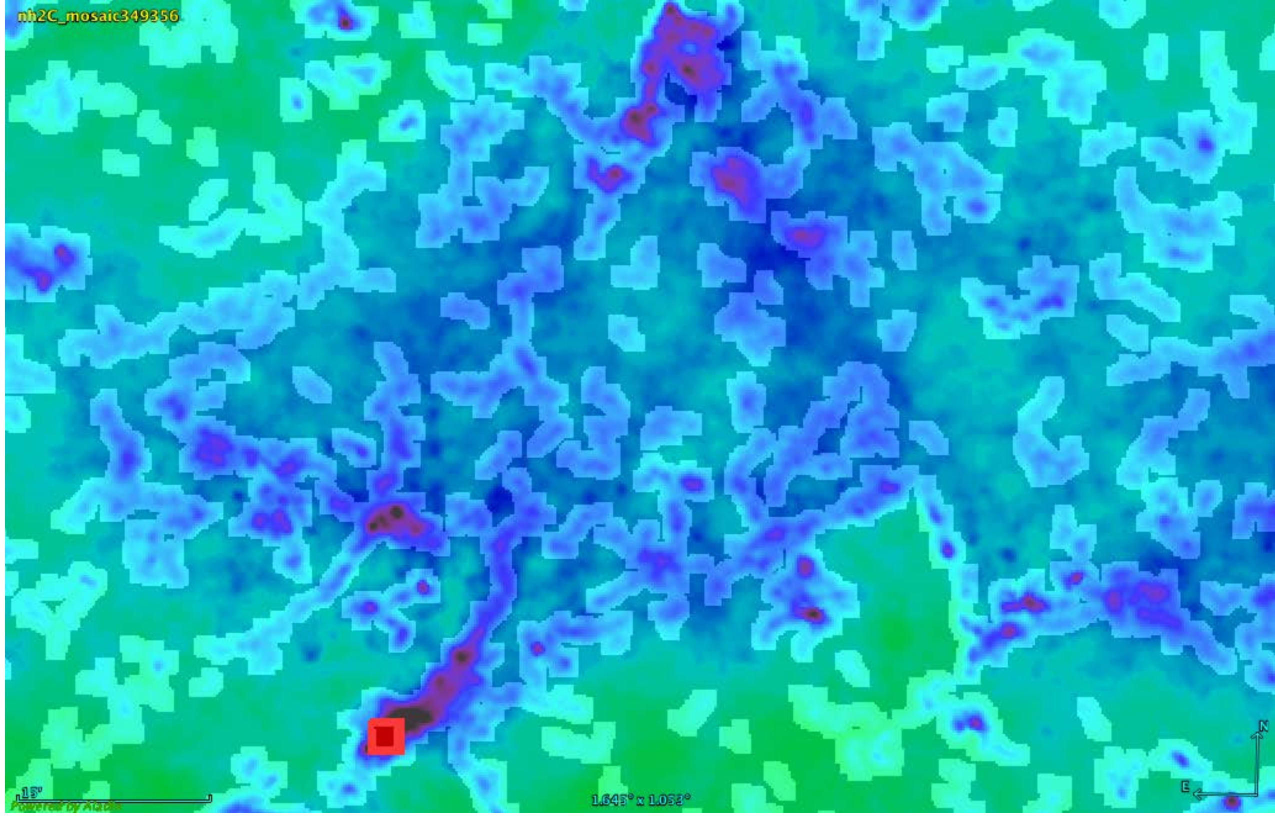
Result

Higher coverage for the segmented RoI image

but

are the low level identified structures « real » filaments ?



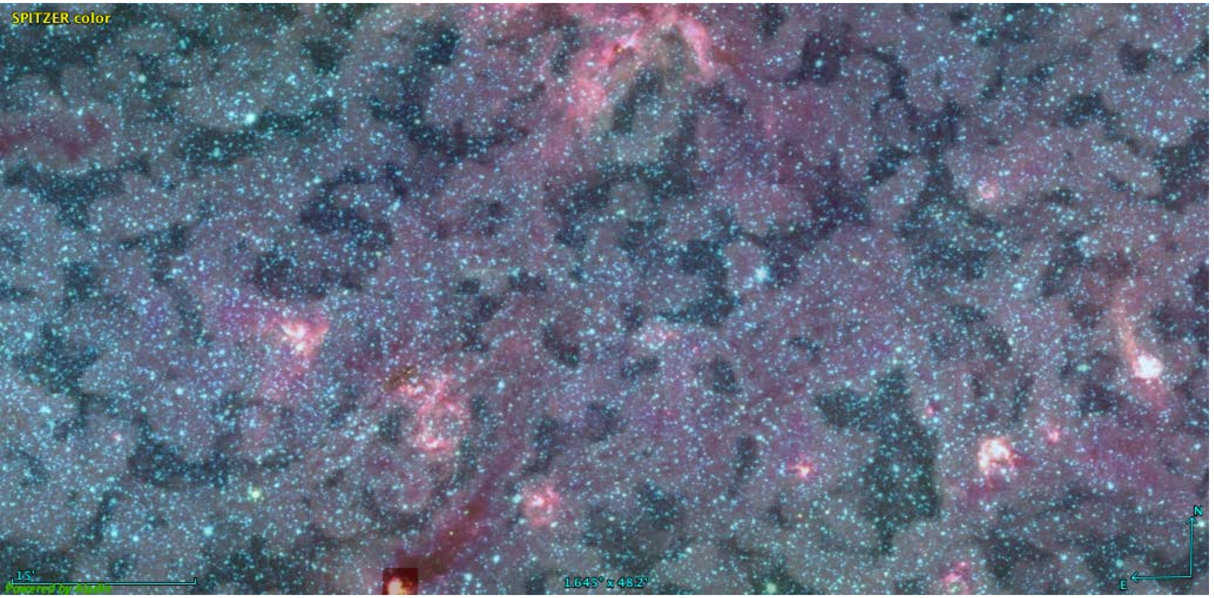


$N(H_2)$ + segmented RoI
Smoother coverage of the structures

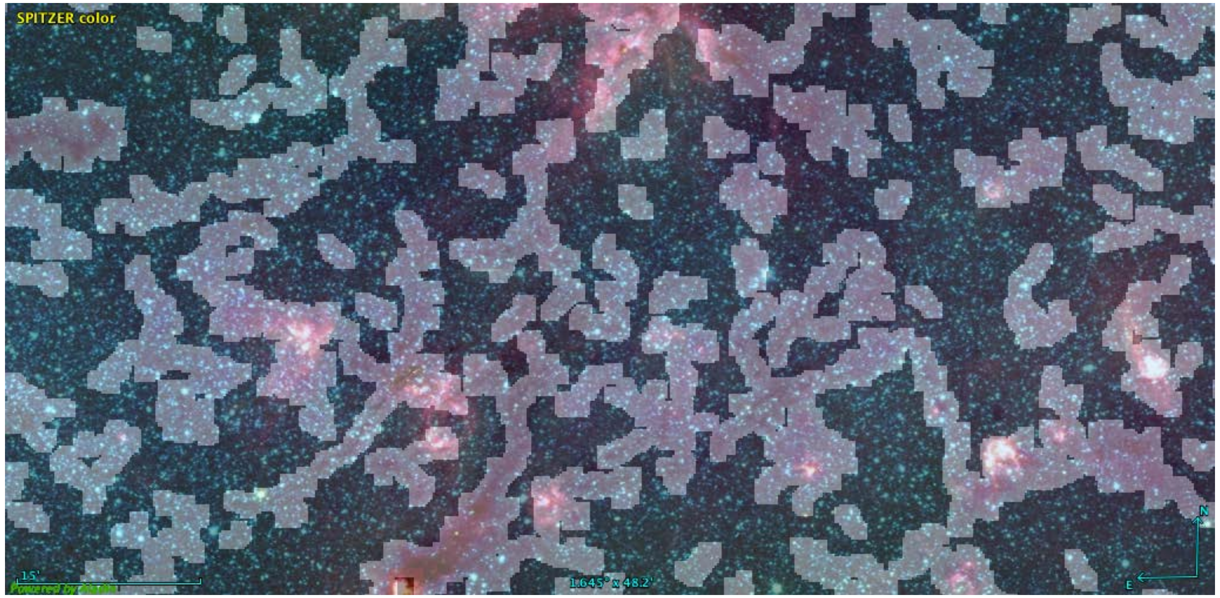
$N(H_2)$ + original RoI

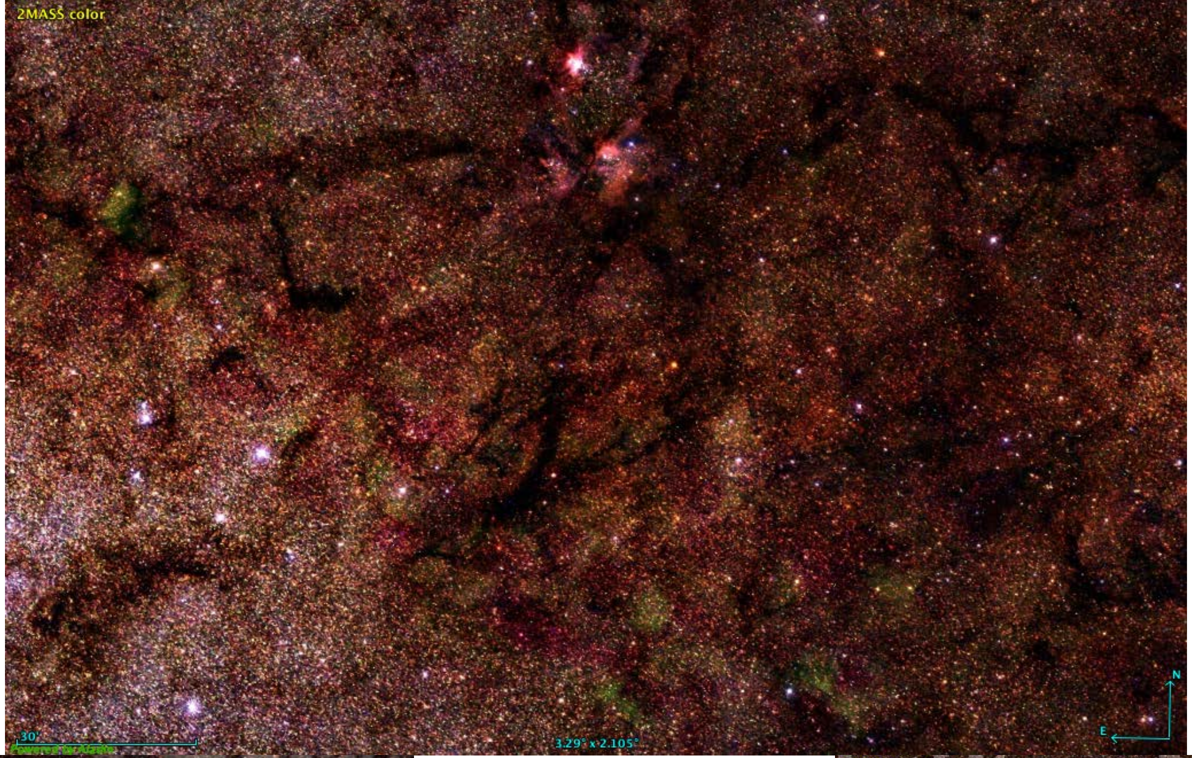
Spitzer image
Filaments in dark

RoI segmented



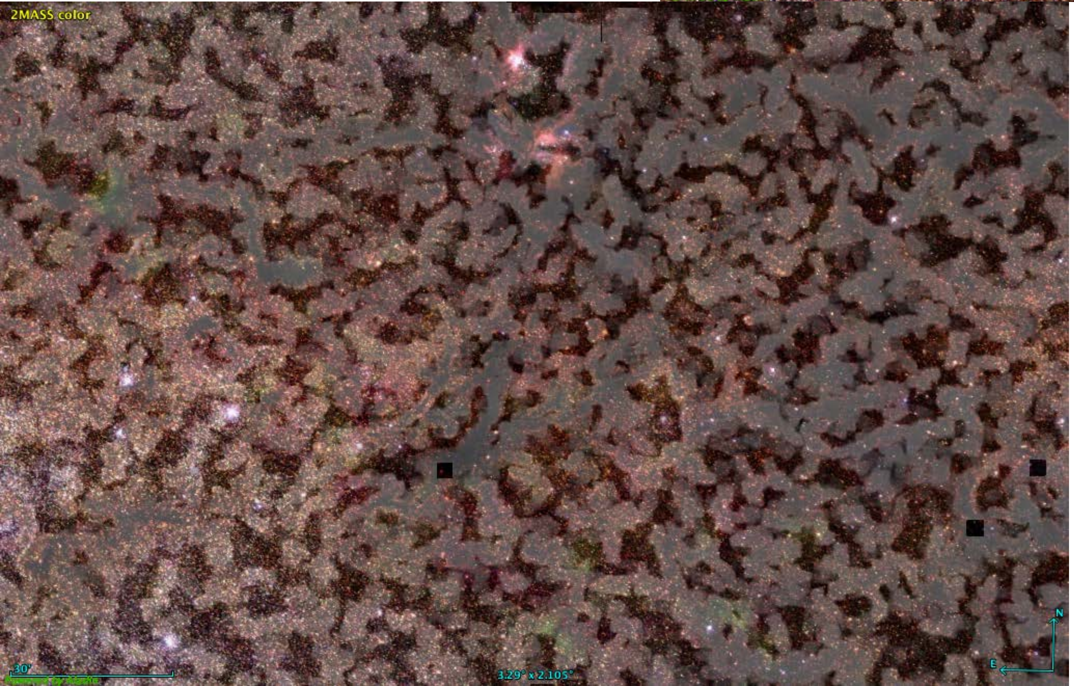
RoI original



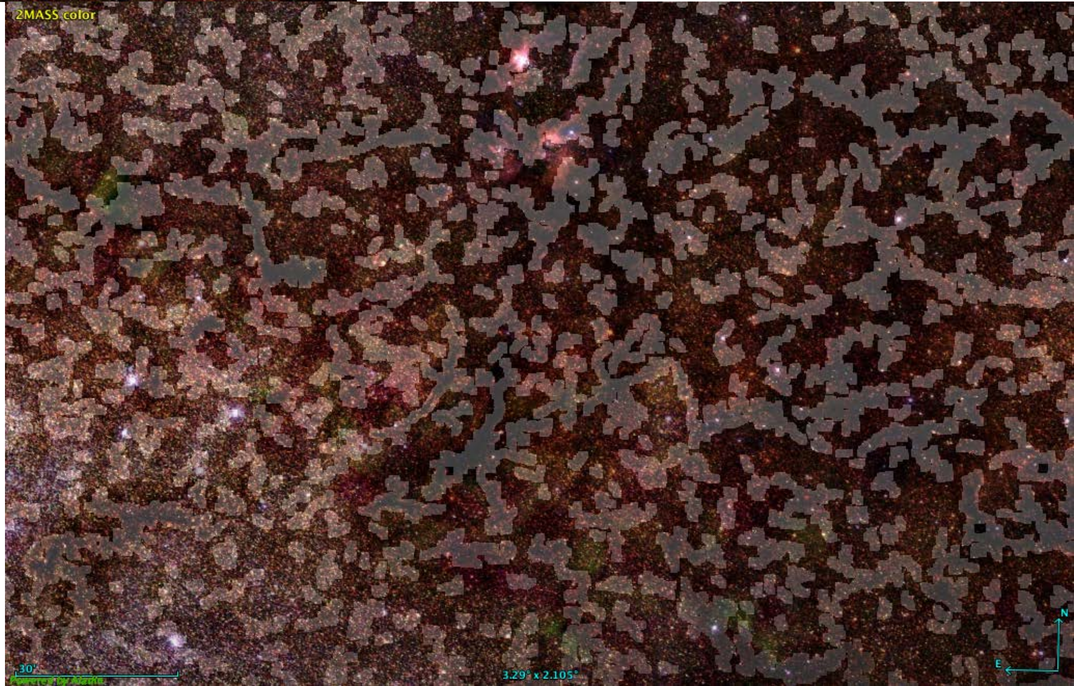


2MASS image
Filaments in dark

RoI segmented



RoI original



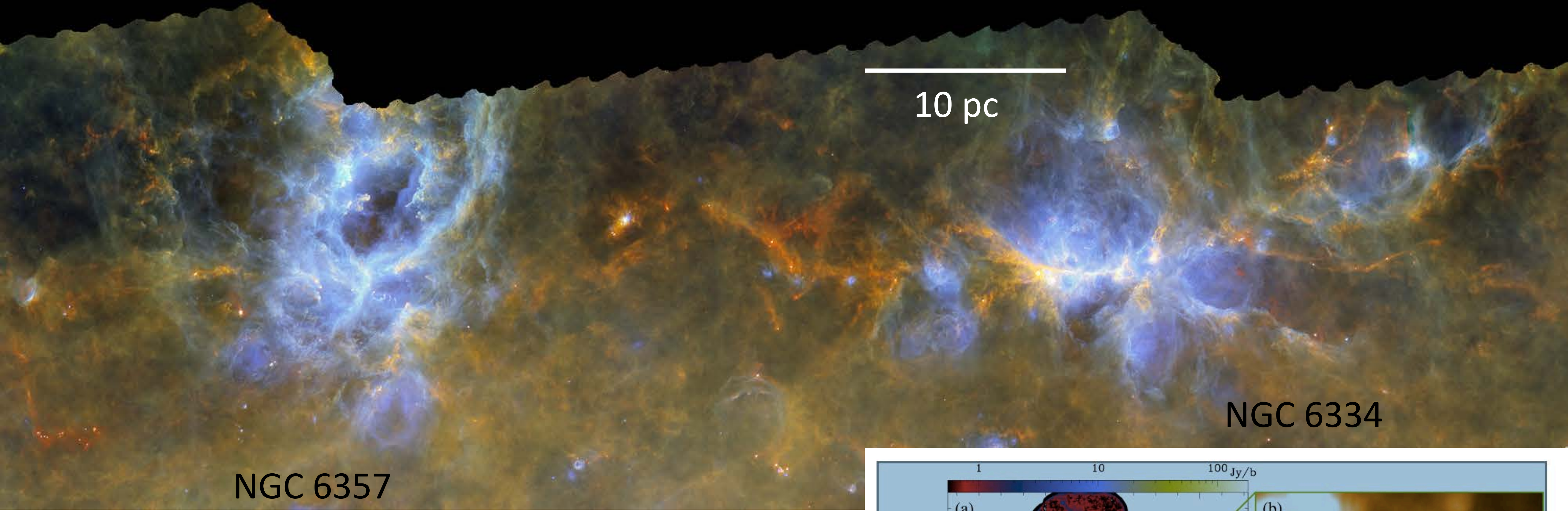
Conclusions

2D images and supervised learning

- Comparaison between 2 methods (classical extraction and ML)
  - Hi-GAL images database
  - Comparison with « never seen » images at other wavelengths: impressive recovery of the filamentary structures by the learning. To be confirmed.

Development of supervised and unsupervised learning (ongoing)

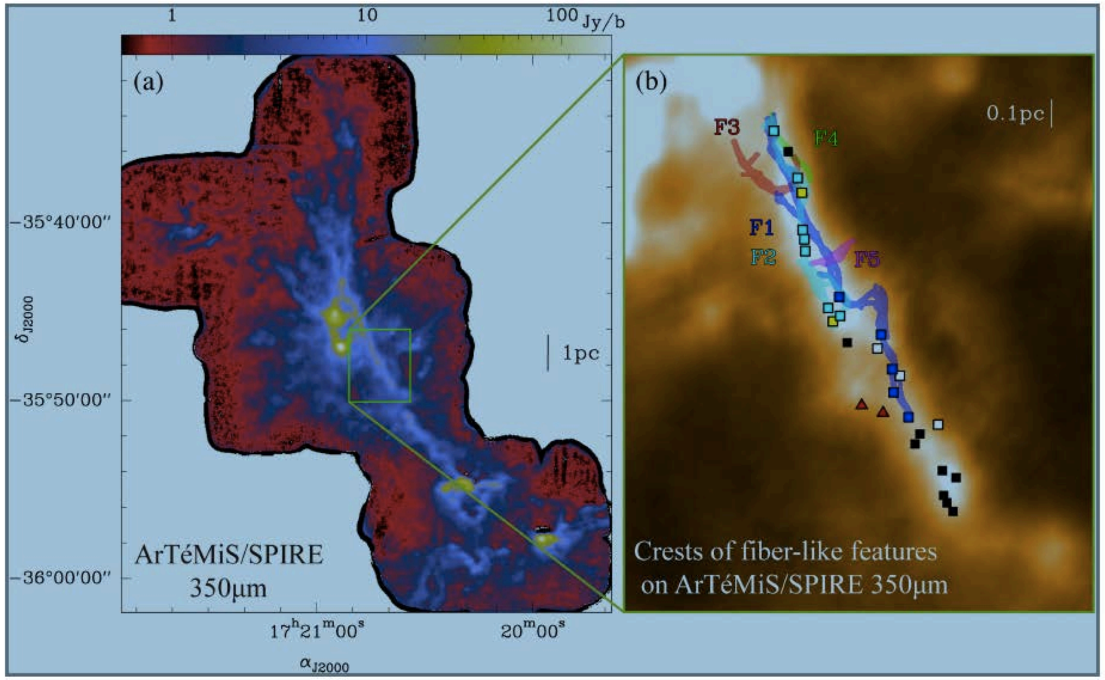
  - Refine the database: filaments properties (distance, physical properties, etc…)
  - → test the learning with the new base

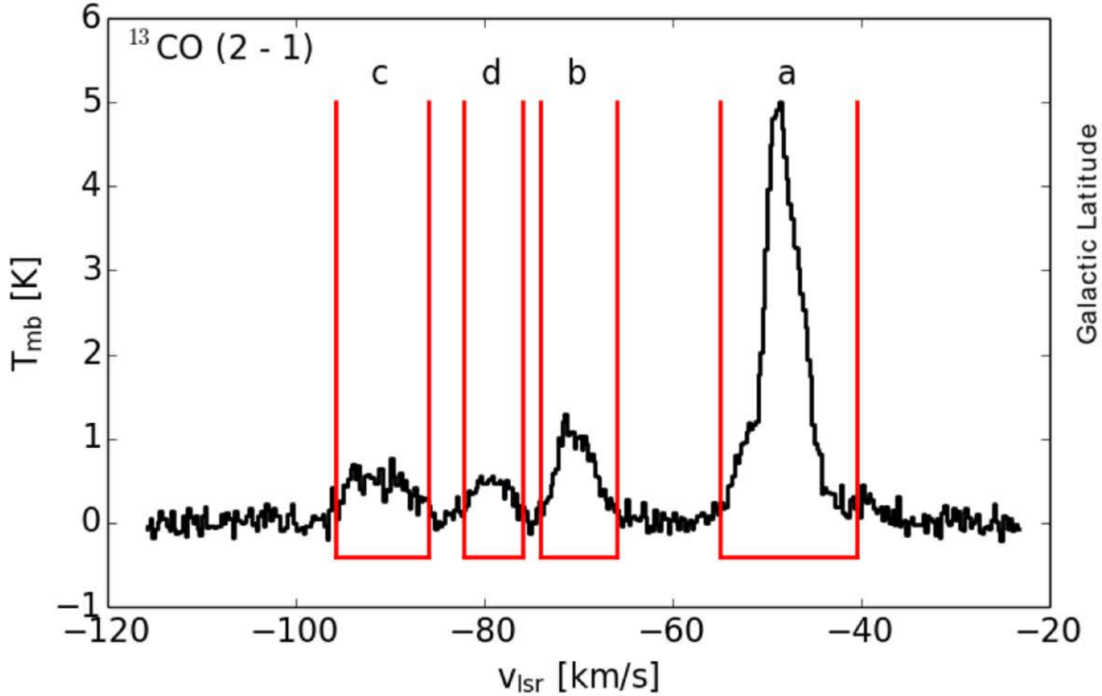


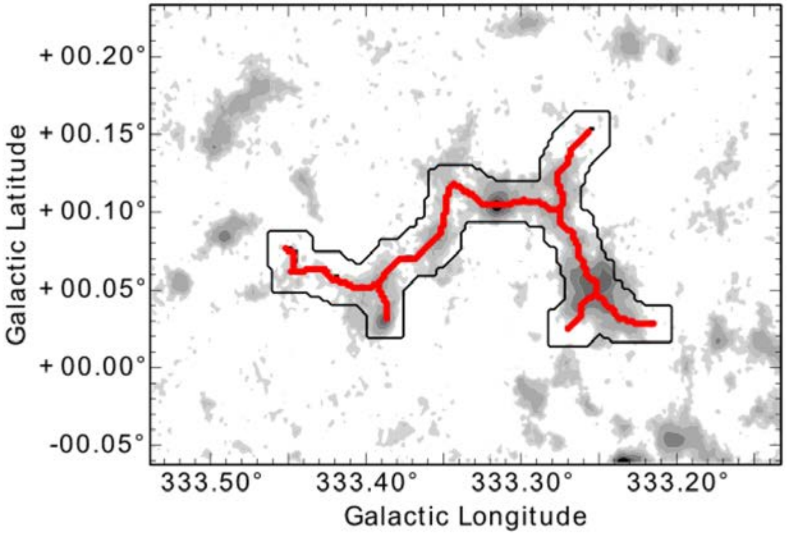
Herschel view of NGC 6357 and NGC 6334.

70 microns, 160 microns and 350 microns

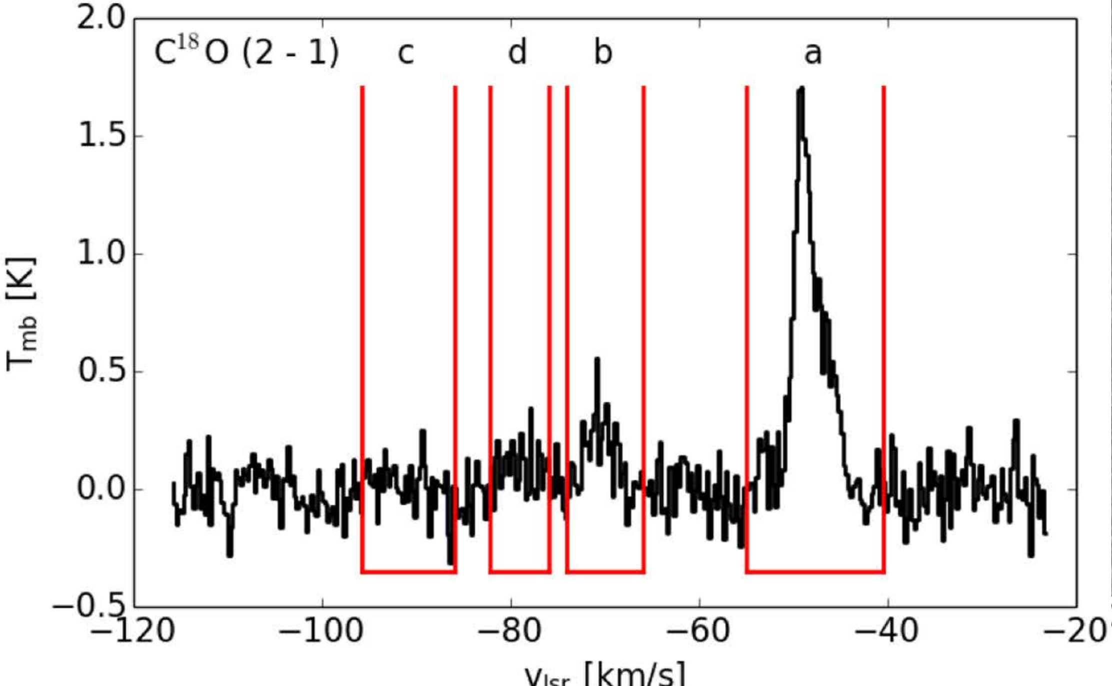
ArTéMiS-APEX+SPIRE et ALMA

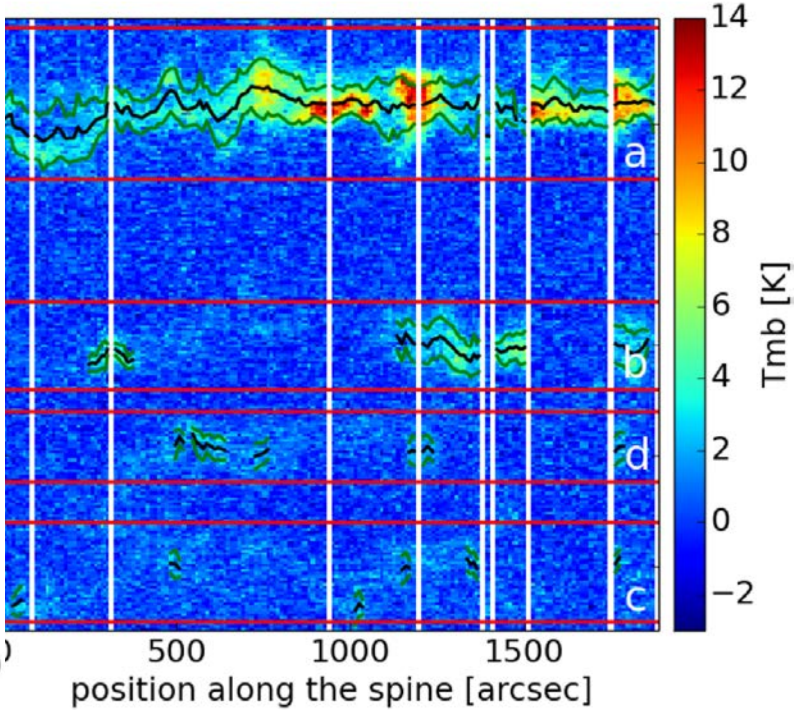




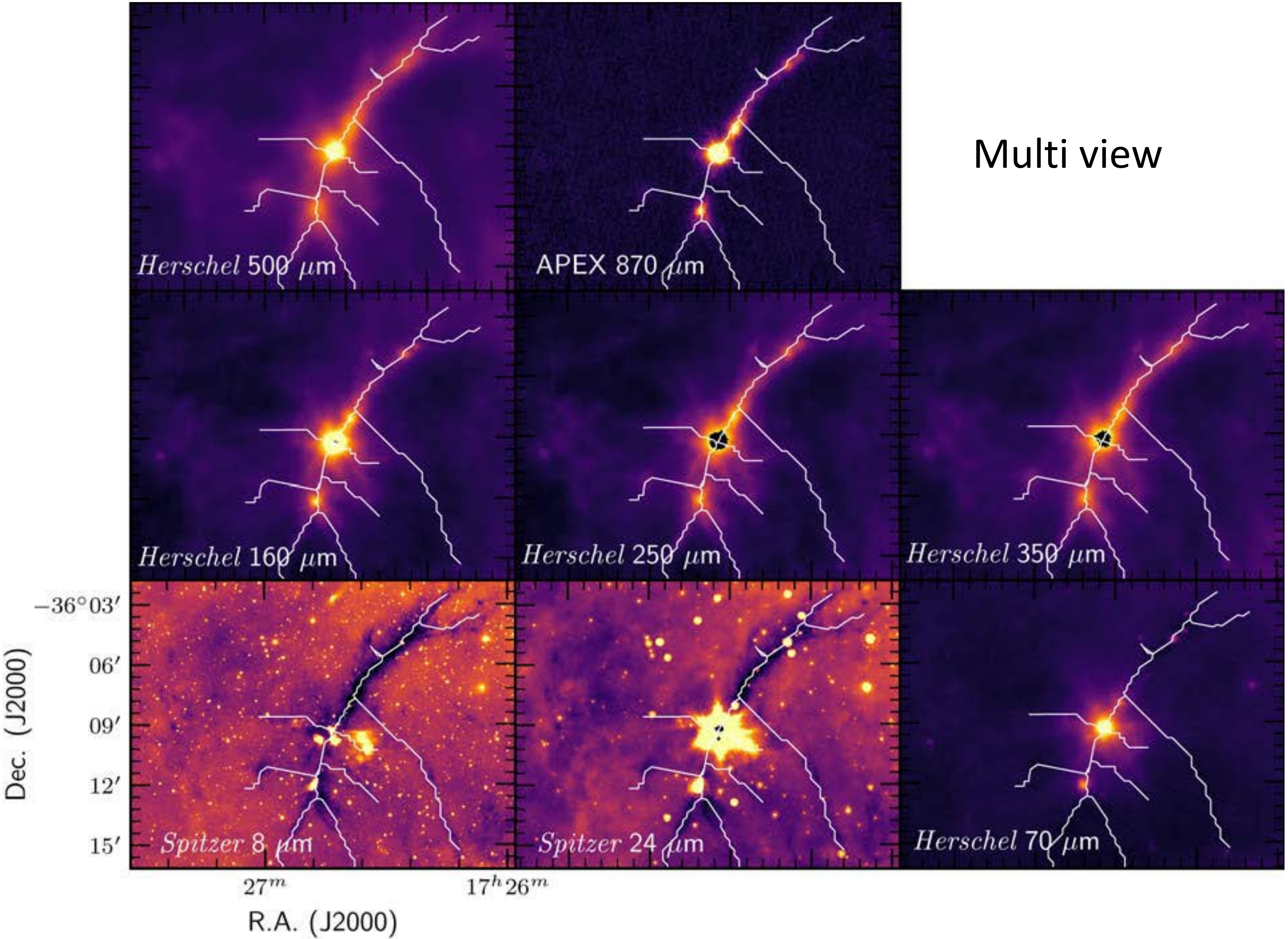


Mattern+2018

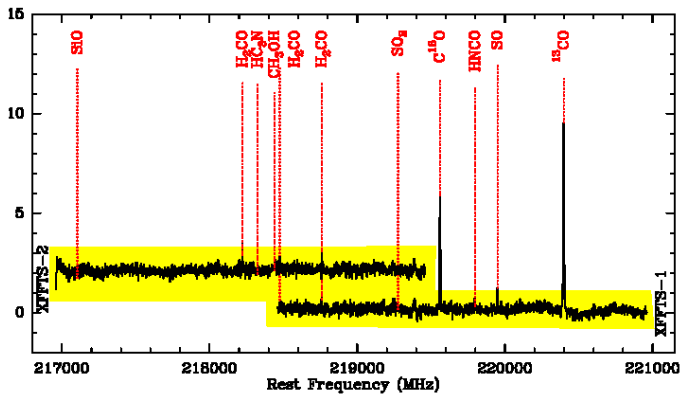




- Galactic filaments (Leurini et al. 2019, A&A)



Multi view

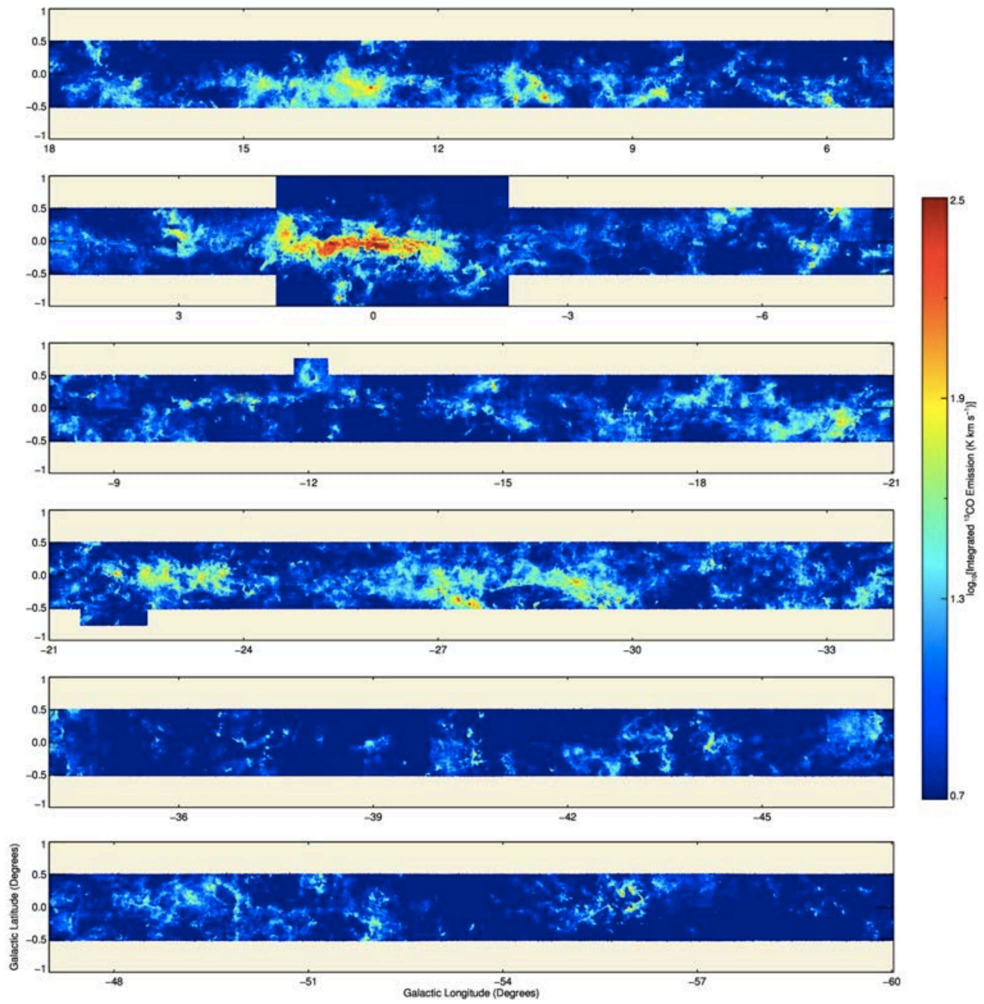


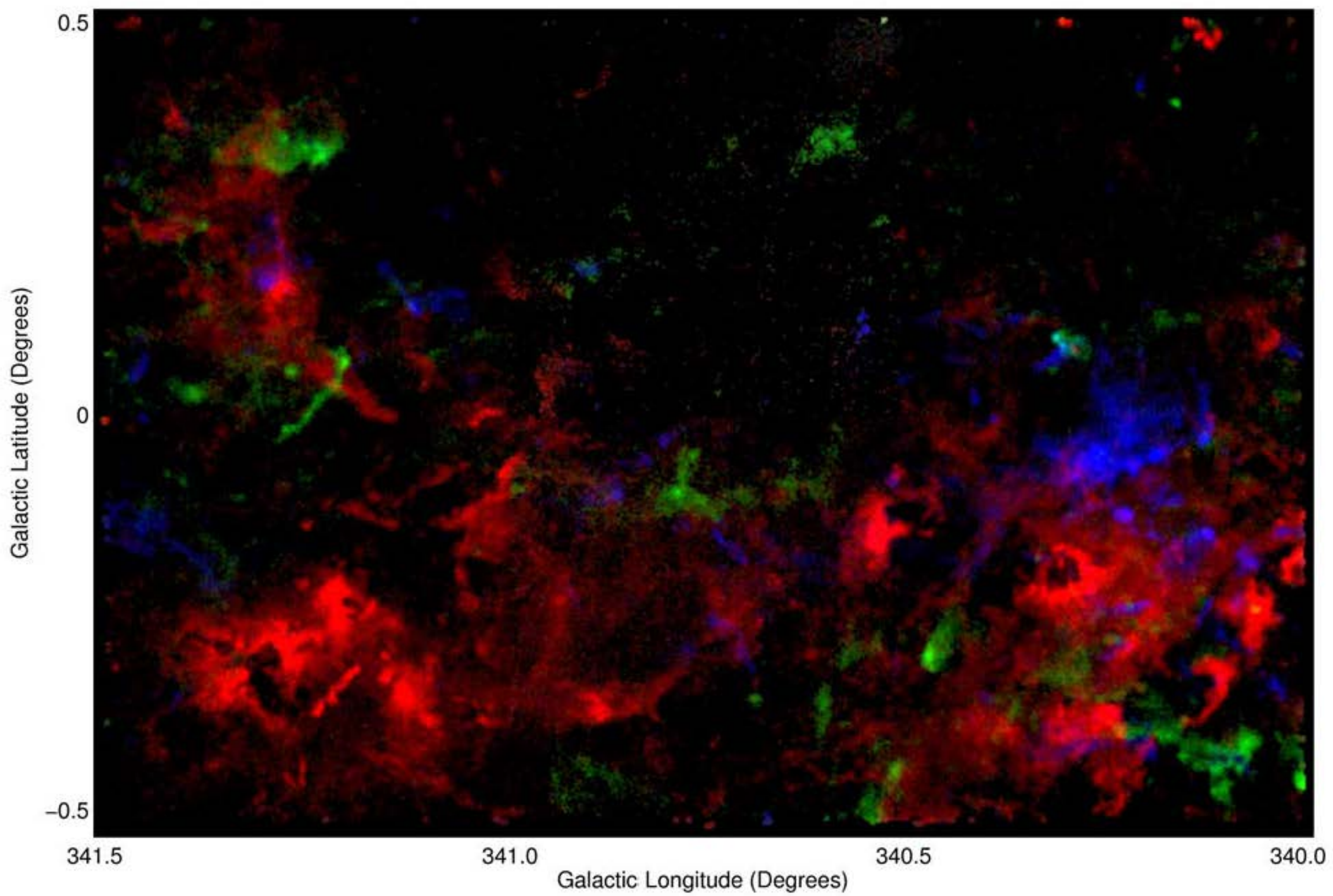
Spectroscopy → Velocity → 3D vision

SEDIGISM data

Schuller+2020

Duarte-Cabral+2020





Perspectives

- Multi view → to be done
  - Challenges: Big Data, heterogeneous data (2D, 3D) multi resolution
- Hyperspectral deep learning (Signoroni+2019, Sun & Bourennane 2020)
  - Take into account the wealth of existing data (multi scale, multi wavelength)
  - To be included in the learning
    - Numerical simulations
    - New extraction method *getsf*