Utilization of Convolutional Neural Networks for HI Source Finding

Team FORSKA-Sweden approach to SKA Data Challenge 2

Henrik Håkansson, Anders Sjöberg, M. Carmen Toribio, Magnus Önnheim, Michael Olberg, Emil Gustavsson,
Michael Lindqvist, Mats Jirstrand, and John Conway

FRAUNHOFER CHALMERS
RESEARCH CENTRE FOR INDUSTRIAL MATHEMATICS

SKA Data Challenge 2

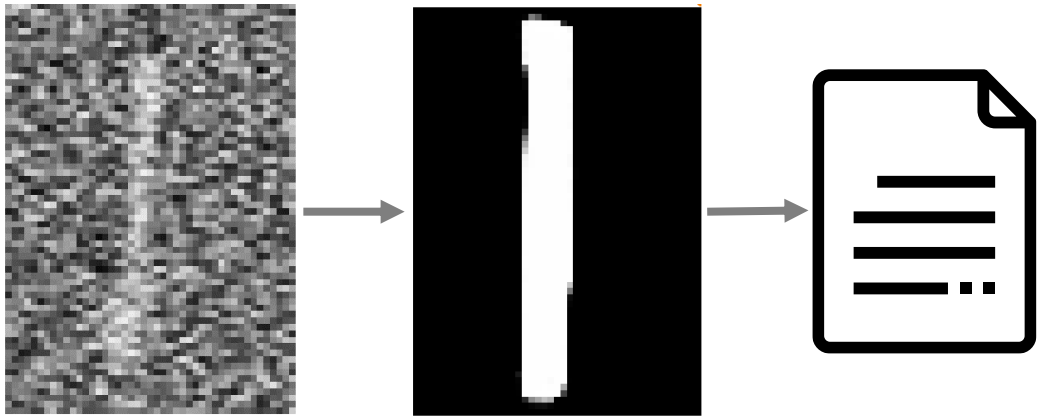
- Generate catalog of simulated HI observation
- Finished July 2021
- Approaches
  - SoFiA: highly specialized source finder
  - Deep learning: generic machine learning models

Group	Points	Deep learning	SoFiA
MINERVA	23254	Yes	No
FORSKA-Sweden	**22489**	**Yes**	**Yes**
SoFiA	16822	No	Yes
Naoc-Tianlai	14416	No	Yes
HI-FRIENDS	13902	No	Yes
EPFL	8515	No	No
Spardha	5614	No	Yes
Starmech	2095	No	Yes
JLRAT	1079	Yes	No
Coin	-2	Yes	No
Hiraxers	-2	Yes	No
SHAO	-471	No	No

Håkansson et al., 2023

Machine learning setup

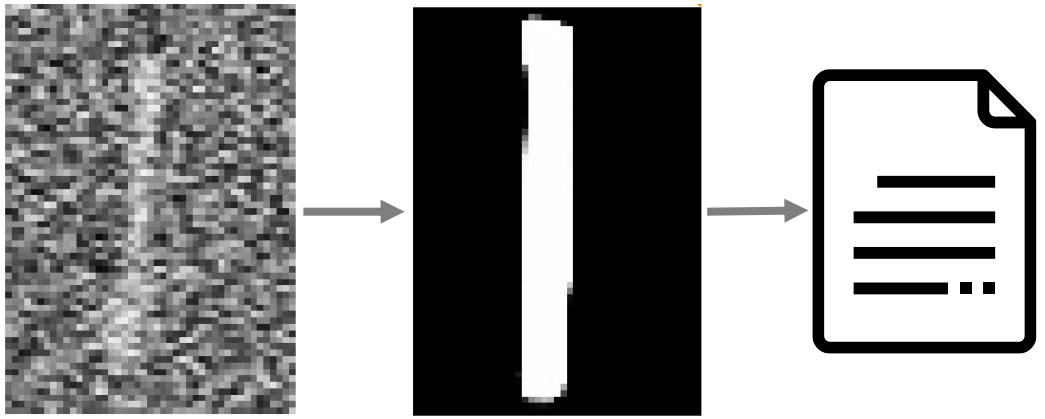
- Established procedure:
  1. Create a binary mask
  2. Estimate source attributes from mask

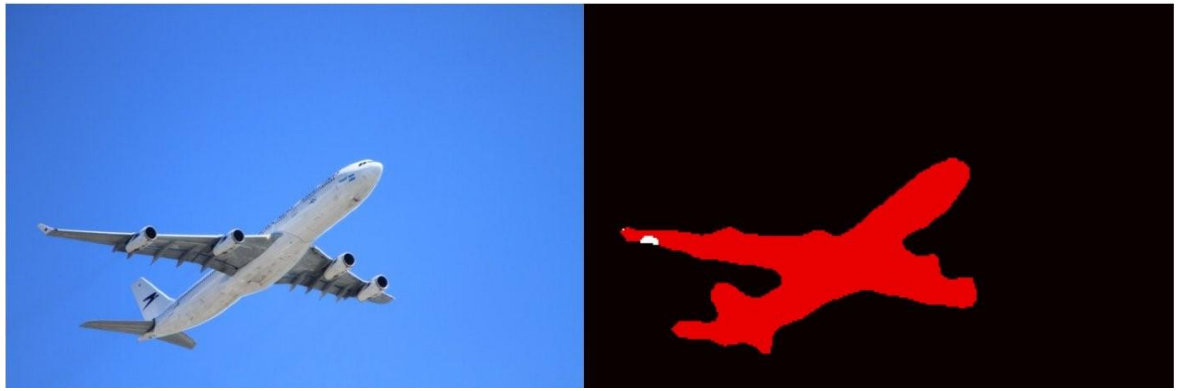


Håkansson et al., 2023

Machine learning setup

- Established procedure:
  1. **Create a binary mask**
  2. Estimate source attributes from mask
- Similar to semantic segmentation
- Why not learn the mask instead?
  - Access to true source catalog

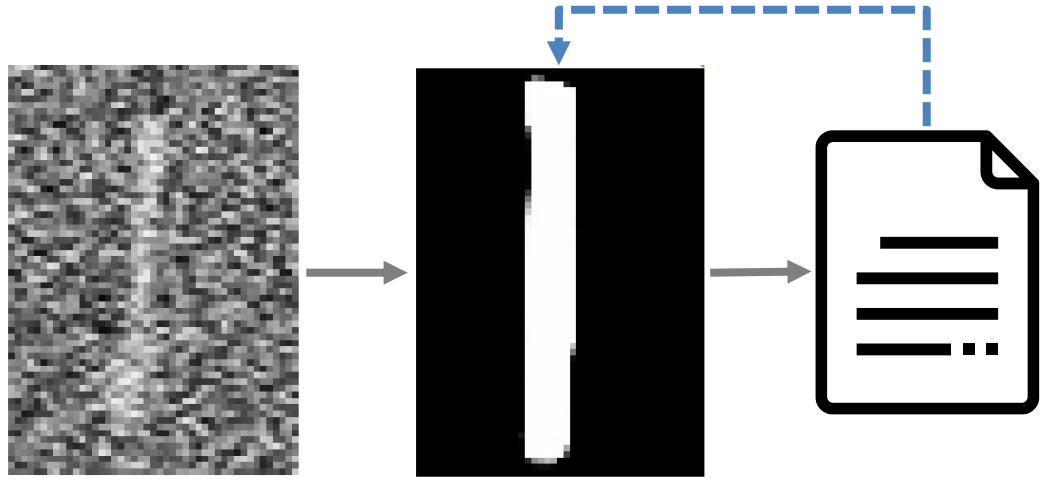


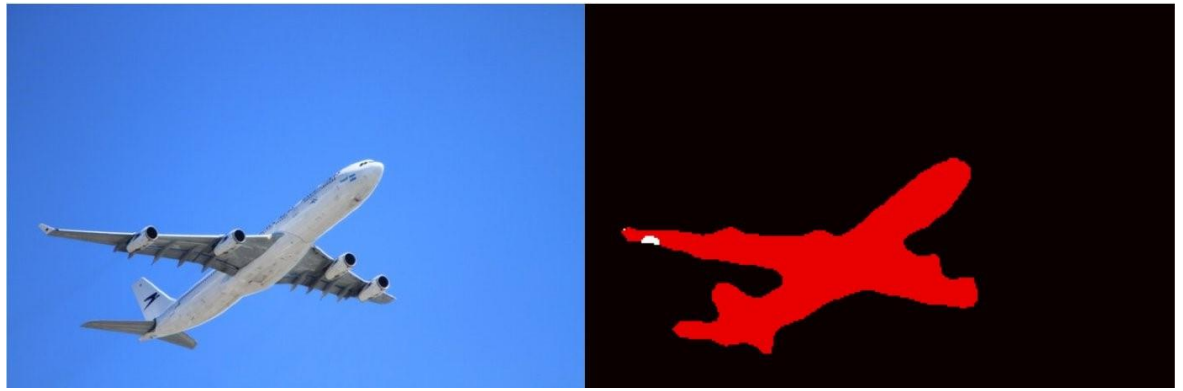


Håkansson et al., 2023

Machine learning setup

- Established procedure:
  1. **Create a binary mask**
  2. Estimate source attributes from mask
- Similar to semantic segmentation
- Why not learn the mask instead?
  - Access to true source catalog
- New problem: generate ideal target mask from source catalog



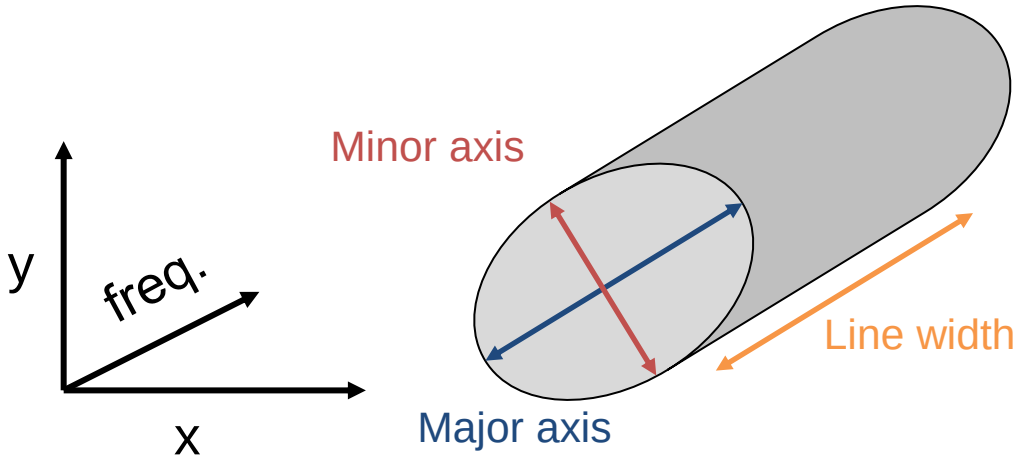


Håkansson et al., 2023

FRAUNHOFER CHALMERS
RESEARCH CENTRE FOR INDUSTRIAL MATHEMATICS

Target mask

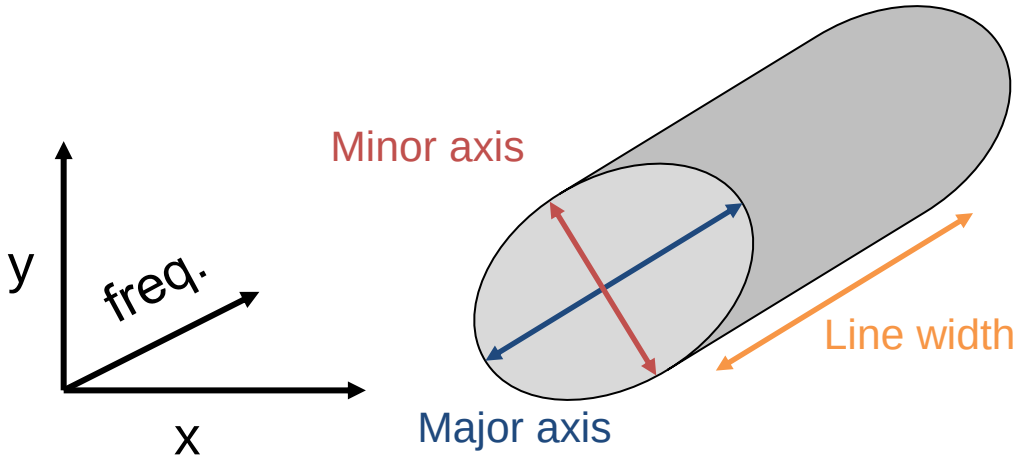
- *Shape attributes* given in the true source catalog
  - Elliptical cylinder
- Covers more than galaxy voxels
- Can masks be tighter and more accurate?

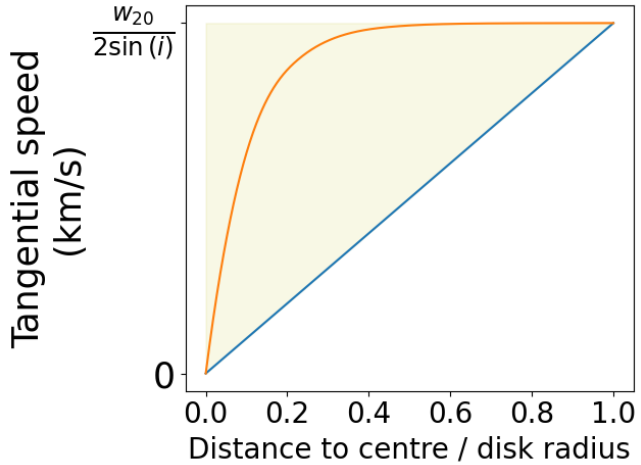


Håkansson et al., 2023

Target mask

- *Shape attributes* given in the true source catalog
  - Elliptical cylinder
- Covers more than galaxy voxels
- Can masks be tighter and more accurate?
- Idea: generate mask from plausible rotation curves
- Target <u>any</u> curve fitting inside the colored area



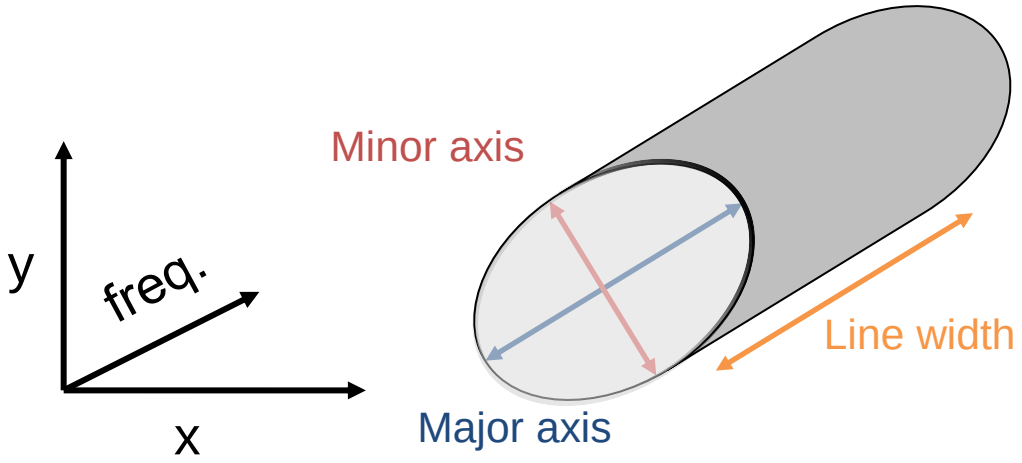


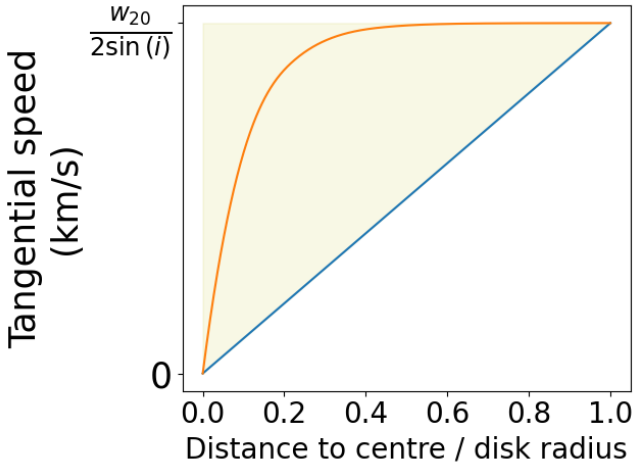
Håkansson et al., 2023

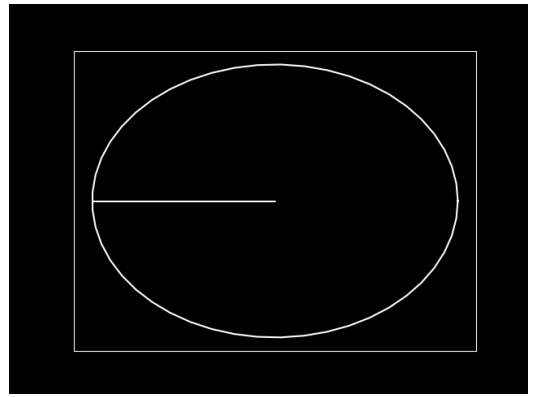
FRAUNHOFER CHALMERS
RESEARCH CENTRE FOR INDUSTRIAL MATHEMATICS

Target mask

- *Shape attributes* given in the true source catalog
  - Elliptical cylinder
- Covers more than galaxy voxels
- Can masks be tighter and more accurate?
- Idea: generate mask from plausible rotation curves
- Target <u>any</u> curve fitting inside the colored area





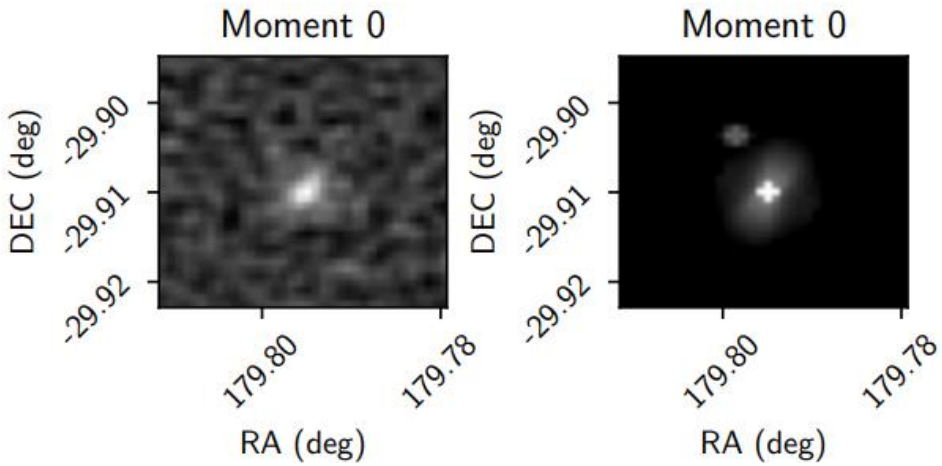


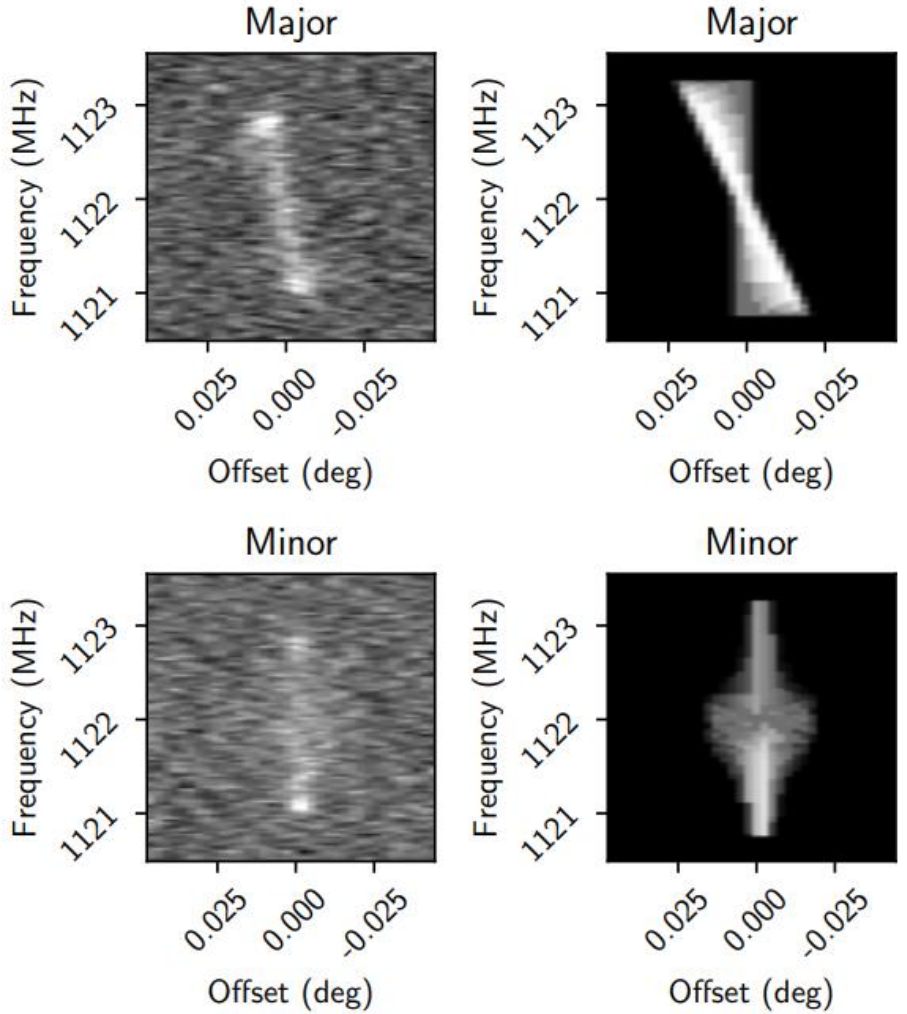
Håkansson et al., 2023

FRAUNHOFER CHALMERS
RESEARCH CENTRE FOR INDUSTRIAL MATHEMATICS

Deep learning training with mask

- Mask resembles occupation with margin



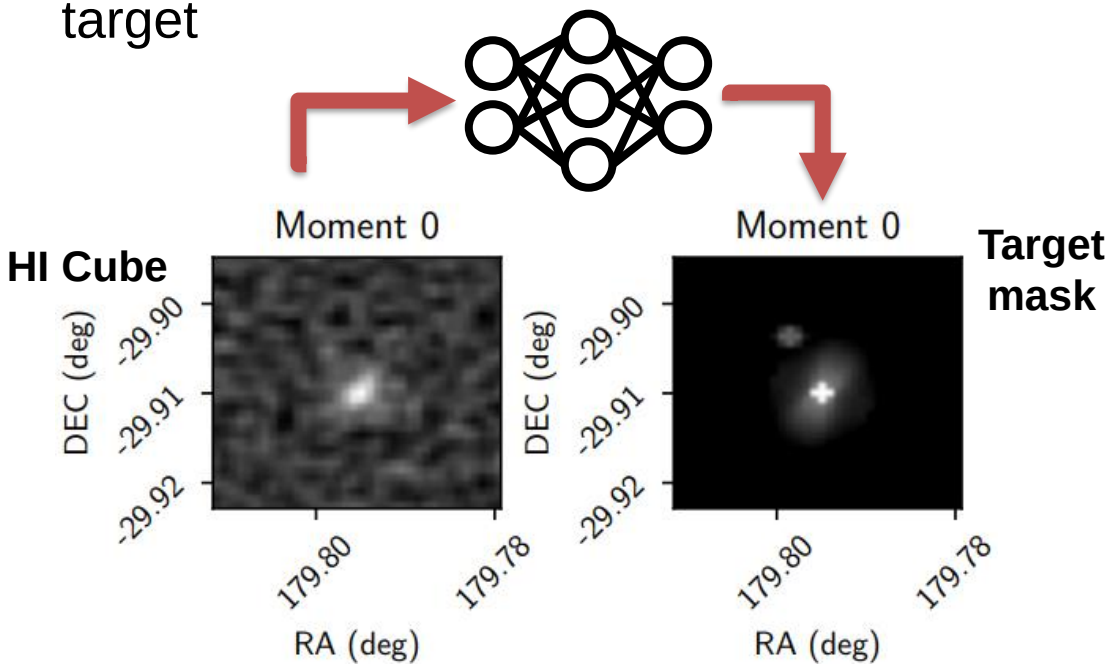


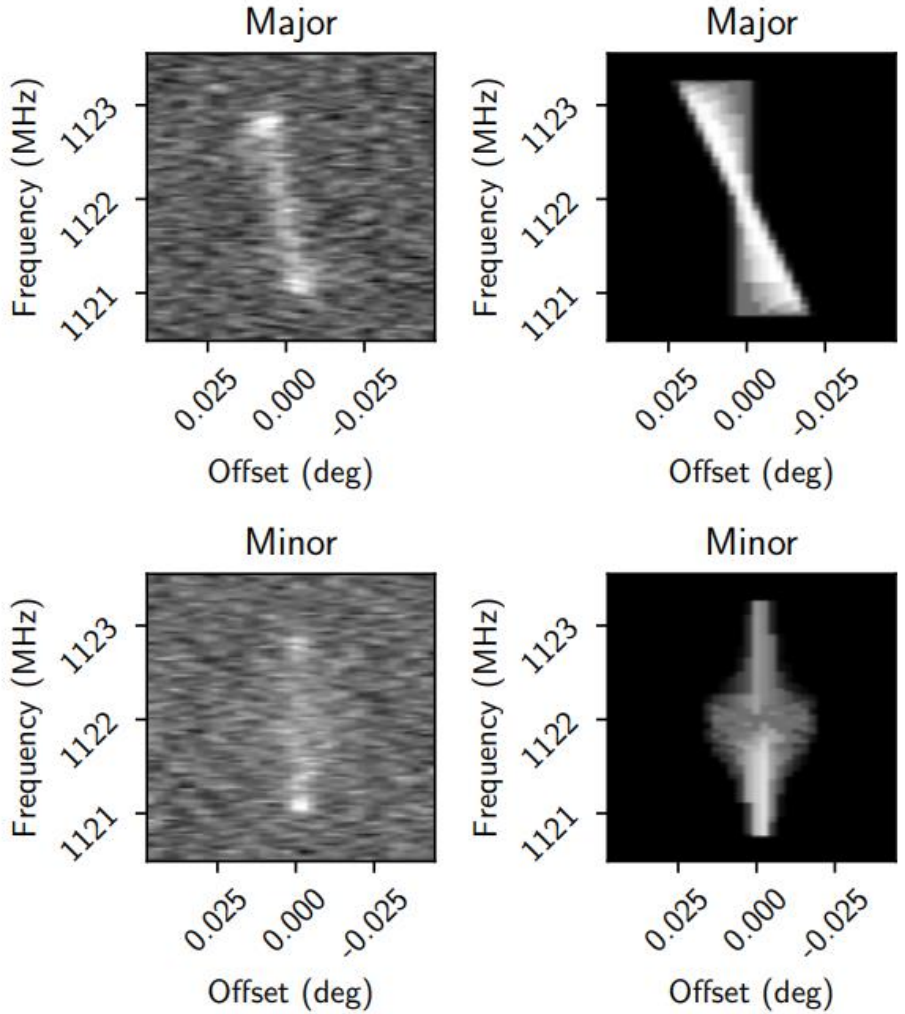
Håkansson et al., 2023

FRAUNHOFER CHALMERS
RESEARCH CENTRE FOR INDUSTRIAL MATHEMATICS

Deep learning training with mask

- Mask resembles occupation with margin
- Train a convolutional network (deep learning segmentation model) to output target



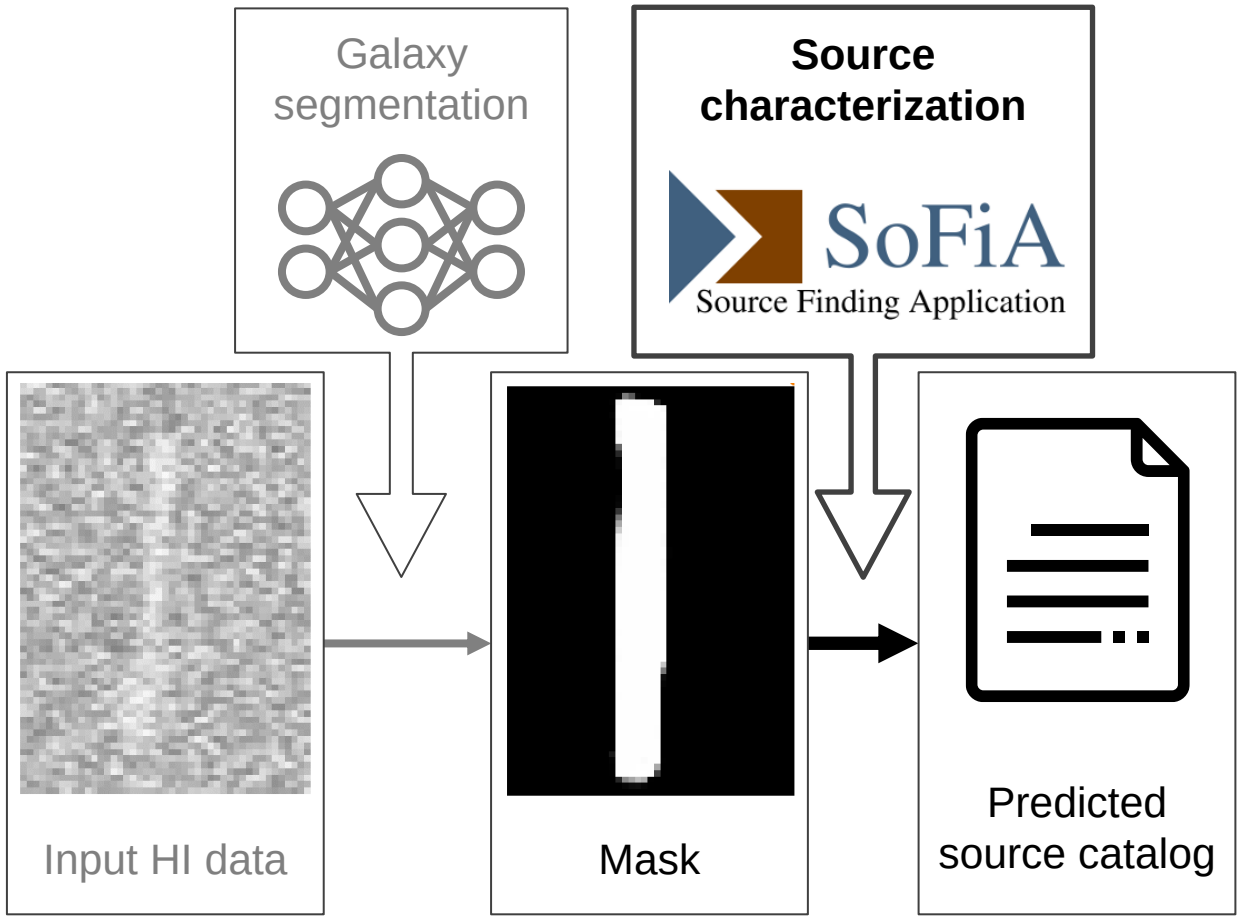


Håkansson et al., 2023

FRAUNHOFER CHALMERS
RESEARCH CENTRE FOR INDUSTRIAL MATHEMATICS

Source characterization

- 1st step: Trained deep learning model for creating mask
- 2nd step: create source catalog from mask
- Subset of SoFiA modules used
- Merging coherent and almost-coherent segments
- Compute characteristics: disk radius, line flux integral, inclination, etc.
- Next problem: How to configure SoFiA?

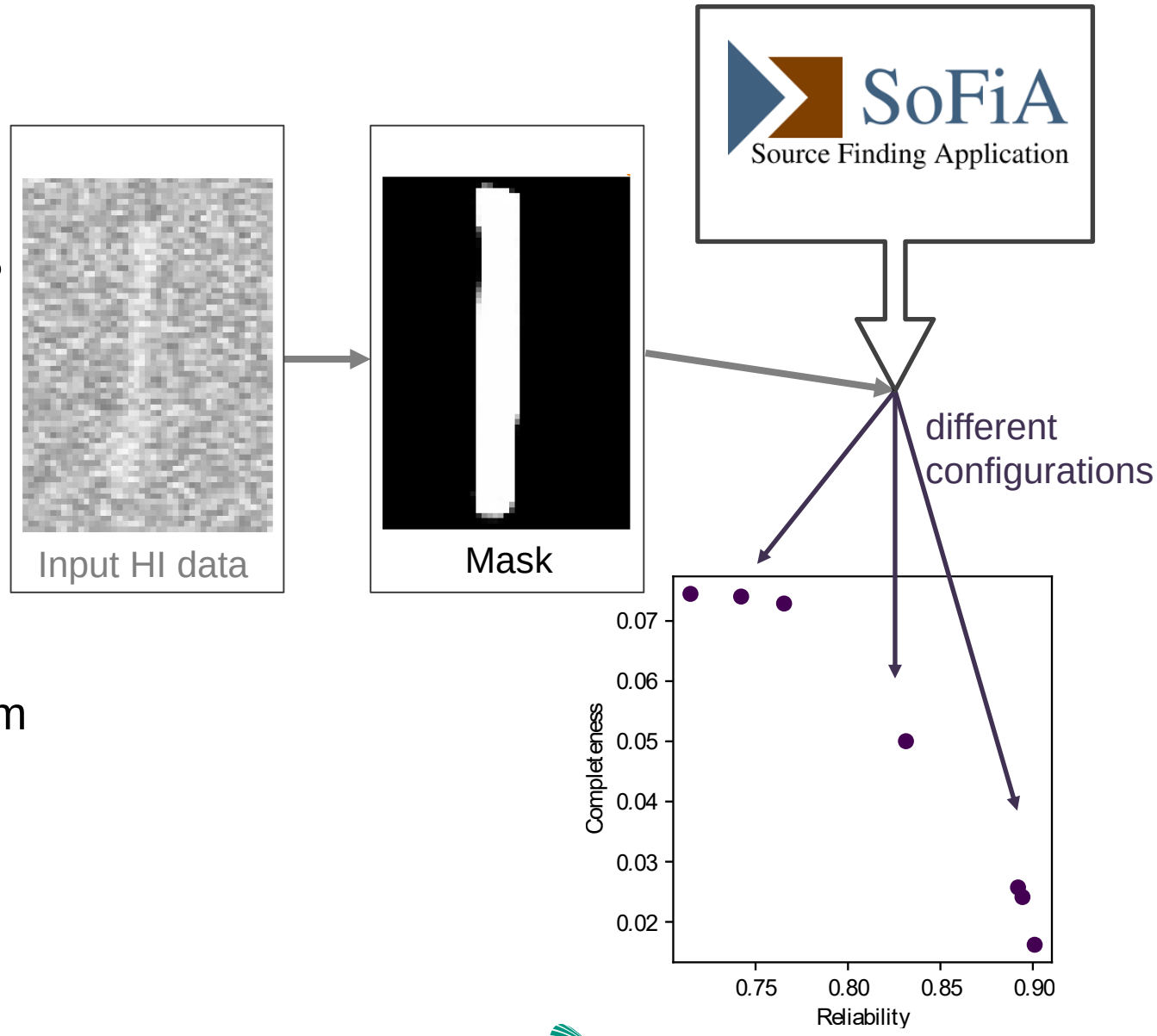


Håkansson et al., 2023

FRAUNHOFER CHALMERS
RESEARCH CENTRE FOR INDUSTRIAL MATHEMATICS

SoFiA configuration

- What do we value most from the pipeline?
  - Reliability or completeness?
  - We can not have both ☹
- Vary only SoFiA configuration → tradeoff!
  - Mask always the same
- How to configure SoFiA?
  - Depends on what you want
- SKA Data Challenge 2 score weighted sum reliability and completeness
  - Not general use case

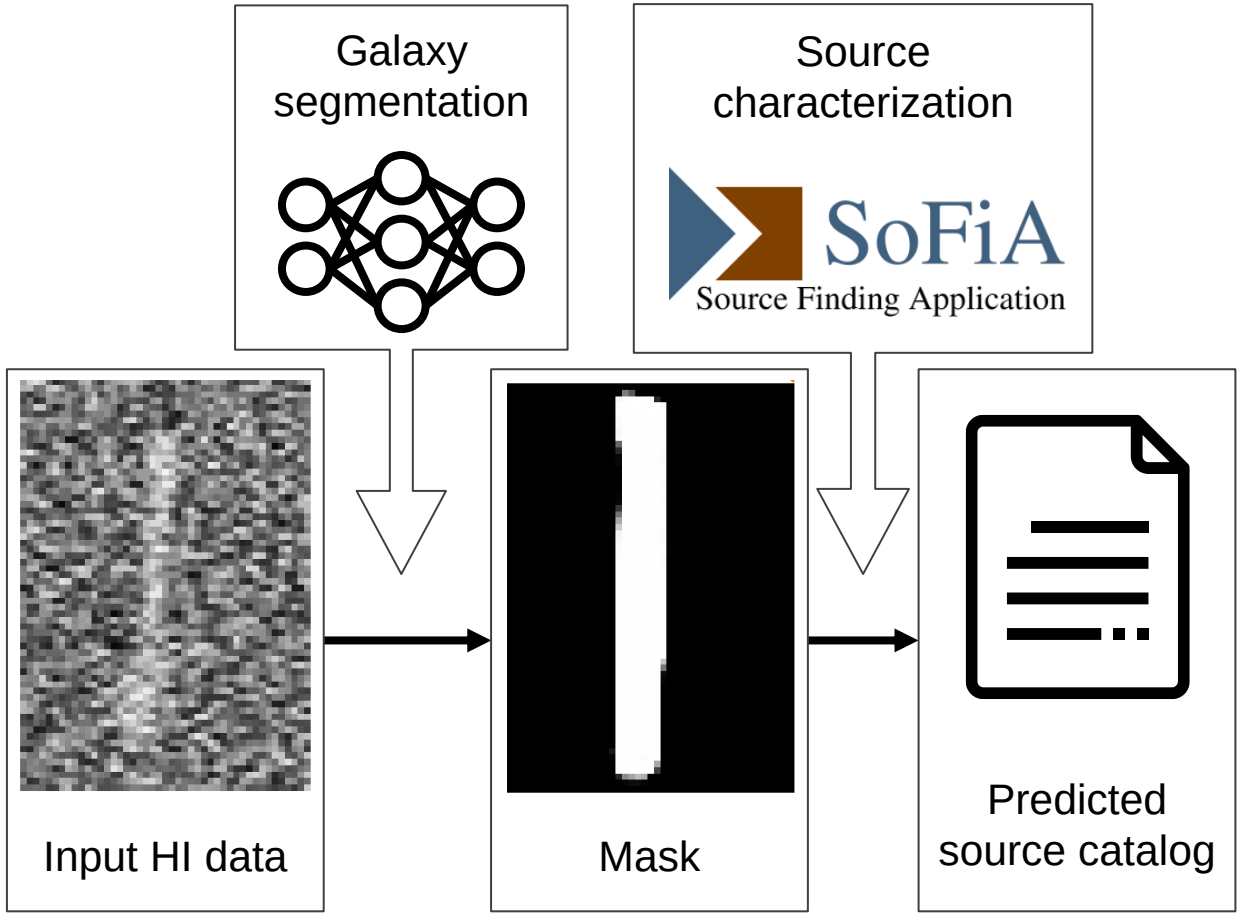


Håkansson et al., 2023

FRAUNHOFER CHALMERS
RESEARCH CENTRE FOR INDUSTRIAL MATHEMATICS

Summary

- Tools from modern computer vision prominent for source finding
- Astronomical knowledge beneficial for algorithm development
- Existing routines can be combined with machine learning

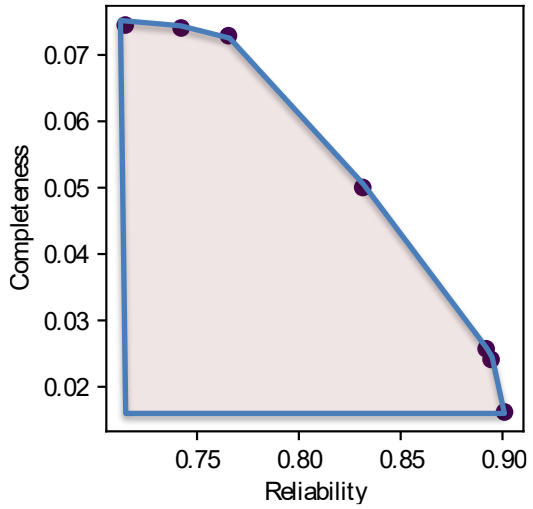


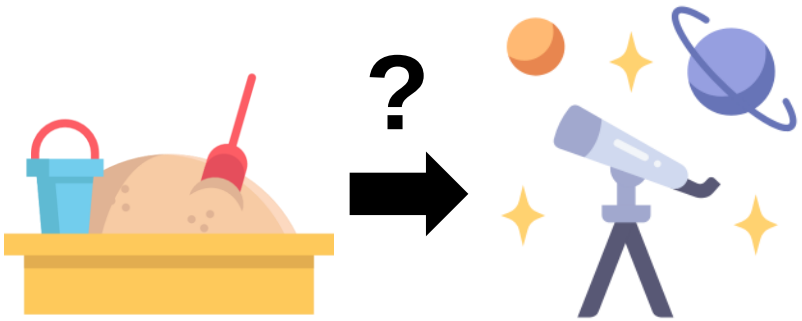
Håkansson et al., 2023

FRAUNHOFER CHALMERS
RESEARCH CENTRE FOR INDUSTRIAL MATHEMATICS

Outlook for machine learning in source finding

- General purpose evaluation metrics
  - Tradeoff common in machine learning
  - Average precision: area under curve
  - Needs domain-specific tuning
- Validating on real data
  - Trained on simulation, validated on simulation
  - How much can we trust the machine learning model?





Håkansson et al., 2023