Heidelberg Institute for Theoretical Studies
HITS

University of Copenhagen
Department of Computer Science

Automatic Classification of Galaxies via Machine Learning Techniques

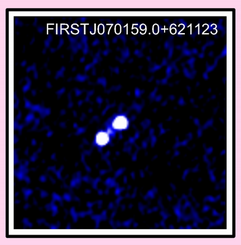
Parallelized rotation/flipping INvariant Kohonen map (PINK)

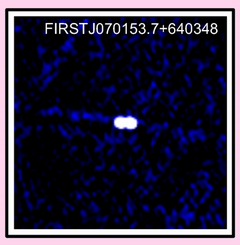
Kai Lars Polsterer[1] Kai.Polsterer@h-its.org
Fabian Gieseke[2] Fabian.Gieseke@di.ku.dk
Christian Igel[2] Igel@di.ku.dk

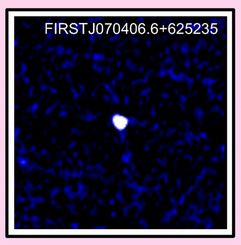
[1]Heidelberg Institute for Theoretical Studies, Astroinformatics, Heidelberg, Germany [2]University of Copenhagen, Department of Computer Science, Copenhagen, Denmark

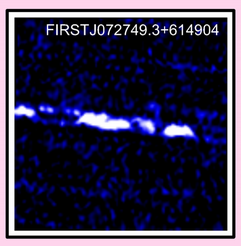
Abstract

In the last decades more and more all-sky-surveys created an enormous amount of data which is publicly available on the Internet. Crowd-sourcing projects like Galaxy-Zoo or Radio-Galaxy-Zoo used encouraged users from all over the world to manually conduct various classification tasks. The combination of the pattern-recognition capabilities of thousands of volunteers enabled scientists to do the data-analysis within acceptable time. For upcoming surveys with billions of sources, however, this approach is not feasible anymore. In this work, we present an unsupervised method that can automatically process large amounts of galaxy data and generates a set of prototypes. This resulting modle can be used to both, visualize the given galaxy data as well as to classify new, yet unseen images.







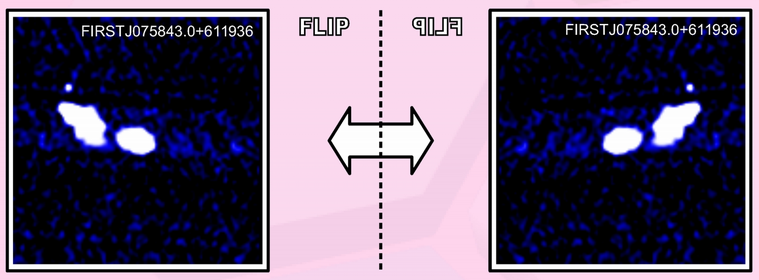


Task

The goal is to enable astronomers to efficiently perform a morphological analysis on huge amounts of data, e.g. images or radio-synthesis data. Thereby we want to make full use of the human pattern-recognition capabilities. The data-sets in astronomy are rapidly growing which therefore renders a manual inspection of all objects by an expert impossible. Currently the researchers are limited by the number of Zooniverse projects. Setting up and starting new projects is the only solution to deal with new data-sets and/or astrophysical questions. By using computers to pre-process and pre-analyze the data we try to assist astronomers to conduct such tasks in a semi-automatic manner instead of a fully manual analysis via crowd-sourcing projects. Similar and frequent objects can be combined / sorted by machine learning models which yield only a single representative that needs manual inspection by the scientist.

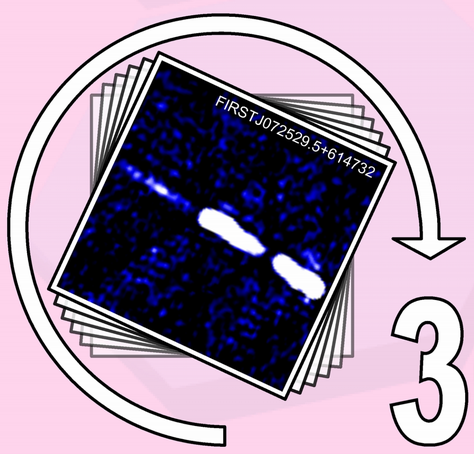
Approach

We promote a framework which generates a classification scheme in an unsupervised way and, thus, permits a semi-automatic analysis of the data by the user. In particular, we make use of so-called Kohonen-maps that can be seen as a simple yet effective dimensionality reduction which projects data to a two-dimensional map. This projections preserves the similarity in high dimensions and groups similar objects together. For the imaging data at hand a rotation and flipping invariant similarity measure is essential to achieve satisfying results. When comparing images, the human brain automatically performs this kind of alignment. For a given image, our approach basically calculates the Euclidean distances for all possible rotations/flipped/un-flipped objects in the map to determine the best match. Since these brute-force comparisons are computationally very demanding, this task depicts an ideal use case for massively parallel implementations. The current version makes use of modern multi-core systems while a new GPU based version is under development.



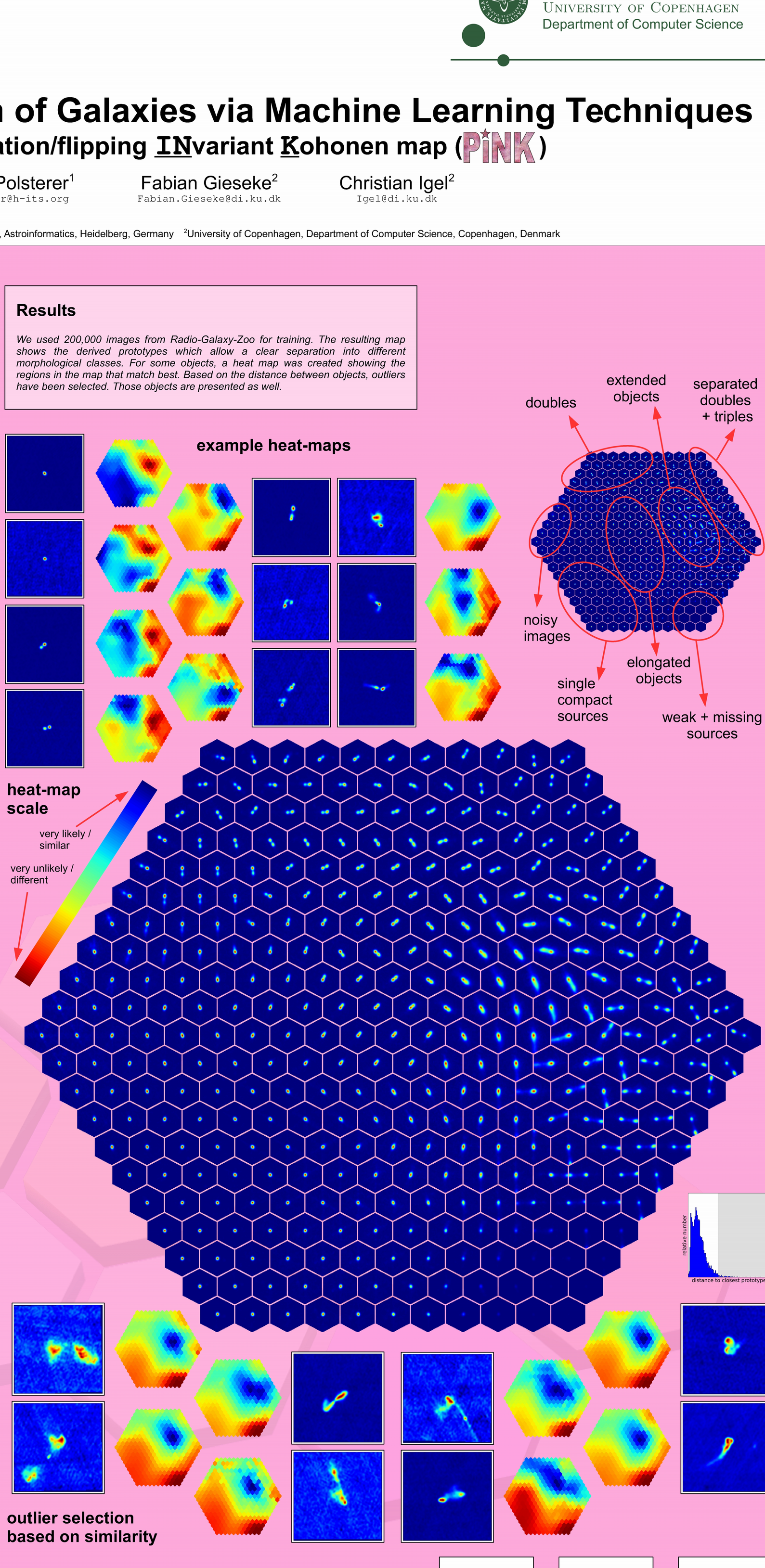
flipping invariant:
by using the original
and the flipped data

rotation invariant:
by rotating the data
in 1° steps



Results

We used 200,000 images from Radio-Galaxy-Zoo for training. The resulting map shows the derived prototypes which allow a clear separation into different morphological classes. For some objects, a heat map was created showing the regions in the map that match best. Based on the distance between objects, outliers have been selected. Those objects are presented as well.



Kai Lars Polsterer

www.h-its.org/english/research/ain/

@AstroInformatix

ADASS 2014, Calgary, Canada, October 5 – 9, 2014