ScaDS.AI
DRESDEN LEIPZIG
CENTER FOR SCALABLE DATA ANALYTICS AND ARTIFICIAL INTELLIGENCE

H<CK>THON
SERIES

NHR Summer School Edition:

Unveiling UFOs through HPC and AI

Norman Koch, Apurv Deepak Kulkarni, Christoph Lehmann,
Lalith Manjunath and Elias Werner
Dresden, June 11, 2024

GEFÖRDERT VOM
Bundesministerium für Bildung und Forschung

SACHSEN
Diese Maßnahme wird gefördert durch die Bundesregierung aufgrund eines Beschlusses des Deutschen Bundestages.
Diese Maßnahme wird mitfinanziert durch Steuermittel auf der Grundlage des von den Abgeordneten des Sächsischen Landtags beschlossenen Haushaltes.

Instructors

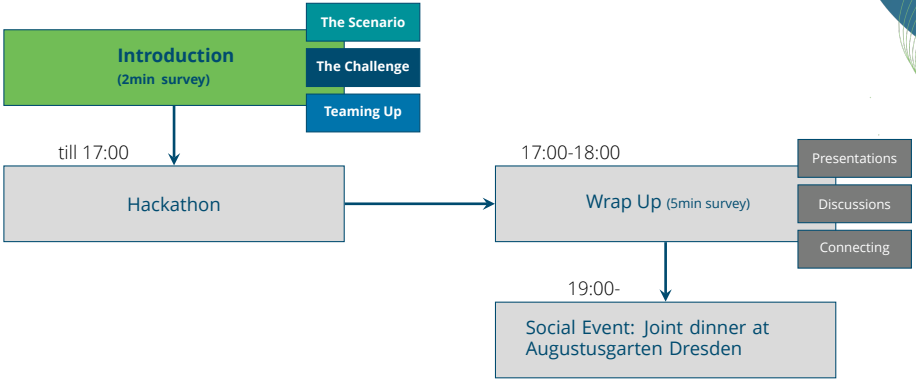
Norman Koch,
AI Application Engineer, ScaDS.AI
Philosophie, Perl, Machine Learning

Apurv Deepak Kulkarni,
Research Associate, ScaDS.AI
Big Data Analytics and Distributed Computing, Living Lab

Christoph Lehmann,
Senior Researcher, ScaDS.AI
Statistics, Deep Learning, HPC

Lalith Manjunath,
Research Associate, ScaDS.AI
Empirical Analysis. Language models. Scale Optimist.

Elias Werner,
Research Associate, ScaDS.AI
Parallelization and performance analysis of data-intensive applications


Introduction
(2min survey)
The Scenario
The Challenge
Teaming Up
till 17:00
Hackathon
17:00-18:00
Wrap Up (5min survey)
Presentations
Discussions
Connecting
19:00-
Social Event: Joint dinner at Augustusgarten Dresden

Scenario: Is it a bird? Is it a plane?

- **U**nidentified **F**lying **O**bjects always attract attention

Scenario: Is it a bird? Is it a plane?

- **U**nidentified **F**lying **O**bjects always attract attention
- Everything that can be in the sky can become an UFO:
  - Flying tents
  - rocket exhausts
  - smoke rings

Scenario: Is it a bird? Is it a plane?

- **U**nidentified **F**lying **O**bjects always attract attention
- Everything that can be in the sky can become an UFO:
  - Flying tents
  - rocket exhausts
  - smoke rings
- UFO sightings databases are publicly available (e.g. UFOStalker, NUFORC)
- Most sightings are easily explainable when you know what you are seeeing

Scenario: Is it a bird? Is it a plane?

- **U**nidentified **F**lying **O**bjects always attract attention
- Everything that can be in the sky can become an UFO:
  - Flying tents
  - rocket exhausts
  - smoke rings
- UFO sightings databases are publicly available (e.g. UFOStalker, NUFORC)
- Most sightings are easily explainable when you know what you are seeeing
- **Fun Fact:** UFO sightings databases are anonymous: → no fame, no money → not many fakes
- Contrary: a lot of fake videos on Youtube

The UFO phenomenon

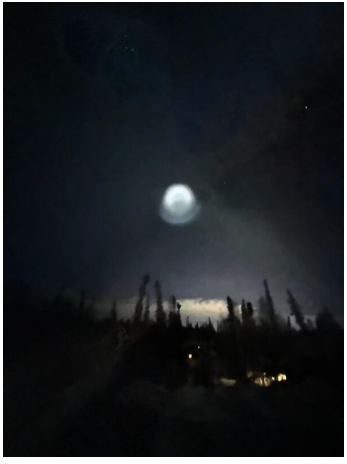
The UFO phenomenon

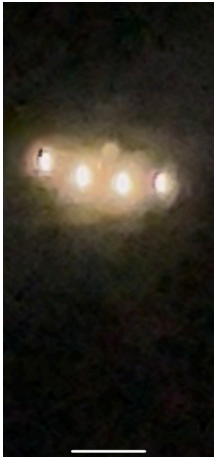
The UFO phenomenon

The UFO phenomenon

The UFO phenomenon – Unexplained Ones

The UFO phenomenon – Unexplained Ones





Conclusions

- Most, maybe all, UFO-sightings are misidentifications of natural or human-made phenomenon
- **But**, interesting phenomena around (rocket spirals,...)
- No evidence that we are currently visited by Alien

- Spectacular UFO images are dubious in origin or published in sources where submitter can gain fame
- Fakes can be avoided by using anonymous sources

Conclusions

- Most, maybe all, UFO-sightings are misidentifications of natural or human-made phenomenon
- **But**, interesting phenomena around (rocket spirals,...)
- No evidence that we are currently visited by Alien
- Spectacular UFO images are dubious in origin or published in sources where submitter can gain fame
- Fakes can be avoided by using anonymous sources
- **Absence of evidence is not evidence of absence!**

Challenge: Build your own UFO Identifier

Tasks:

1. Set up your developing environment with HPC
2. Build your own AI-based UFO detector
3. Present your pipeline and findings

We provide:

- Access to the HPC via JupyterHub
- Prepared Jupyter notebooks with a simple ML pipeline
- Hints and advice about ML and HPC basics

Hint

For the competition, we provide an unlabeled data set. We measure how well your AI classifies the images in that data set. You can upload the data directly from JupyterHub and visualize the scoreboard. Or use our codabench server: https://hackathon.scads.ai/

A Perfect Opportunity

> **Hint**
>
> It's not only about solving the challenge!

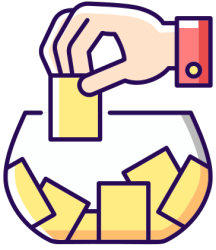
During the hackathon you can:

- Get in touch with HPC and learn about HPC principles
- Develop your own AI pipeline
- Get support if you have questions
- Socialise and connect to people
- Learn about aliens and UFOs
- ...

It's your turn

Teaming Up

- Work in teams of 2 people
- Drawing lots

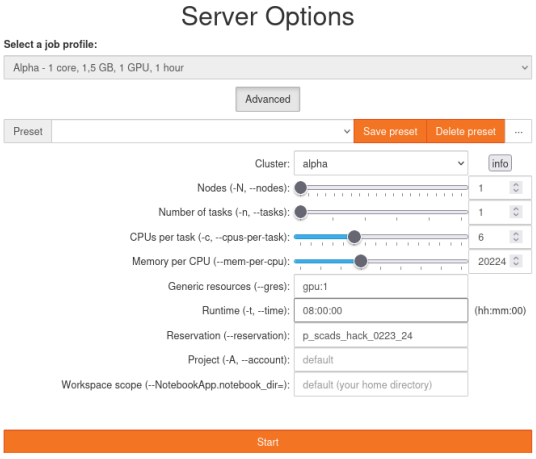


In a nutshell

1.) Login to JupyterHub:
https://jupyterhub.hpc.tu-dresden.de

2.) Spawn Jupyter Server:
Use the JH with appropriate spawner options (see image)

> **Hint**
>
> Please use the following reservation: p_nhr_summerschool_54



In a nutshell

3.) Start hacking:

- **First**, open a Terminal in JupyterHub and copy the code: `cp /data/horse/ws/1/s4122485-nhr_hackathon/code . -R`
- **Second**, continue with the Uni-Note.ipynb that is located in the `code` directory

- Datasets:
  - Read the data from: `/data/horse/ws/1/s4122485-nhr_hackathon/datasets/<DATASET>`
  - `tiny` - used for testing if everything runs correctly
  - `selected` - a selection of the full dataset
  - `full` - the full dataset

- Develop approaches and discuss in your team
- Check the documentations:
  - HPC: https://compendium.hpc.tu-dresden.de/
  - YoloV5: https://github.com/ultralytics/yolov5

Hint

In case of any questions, do not hesitate to ask us.

Ideas

What happens if you...

- ...change/tune hyperparameters, e.g. image size, structure of the neural network?
- ...use more data (i.e. incorporate `full` dataset)?
- ...weight the samples?
- ...add additional pre-processing steps (normalization, standardization, grey-scaling, zero-component analysis)?
- ...increase the model depth and add skip connections?
- ...apply curriculum learning - progressive training of images based on how hard (i.e. high/low loss) it is for the model to learn?

Troubleshooting

- Problems when spawning:
  - delete cookies
  - delete access tokens
  - use incognito mode
  - check whether your spawner is still running via `squeue --me` command in CLI (ssh connection)

- General:
- check the `jupyter-session-XXXXXXX.log` file in your HOME directory
- it gives a lot of insights of your notebook servers behaviour (e.g. if module (pre-)load was successful)
- Pika Job Monitoring tool shows you utilization statistics of your notebook server

- ssh connection to the node with the notebook server software (get name of the node via `squeue --me`), then `ssh taurusi_nodeID`, lets you investigate your computing node(s) further

Hint

Having a look in the compendium is always useful: https://compendium.hpc.tu-dresden.de/

Wrap Up

Please take part in our 2min survey:
https://tud.link/2tdv9k

Wrap Up

Please take part in our 2min survey:
https://tud.link/2tdv9k

Note

Group presentations - tell us in 3 minutes about your experiences, ideas, and lessons learned.