



Towards a data challenge

Guiding questions for discussion



- ◆ Recap: what have we learnt today?
- ◆ What may be useful for our context?
- ◆ How can it become a data challenge?

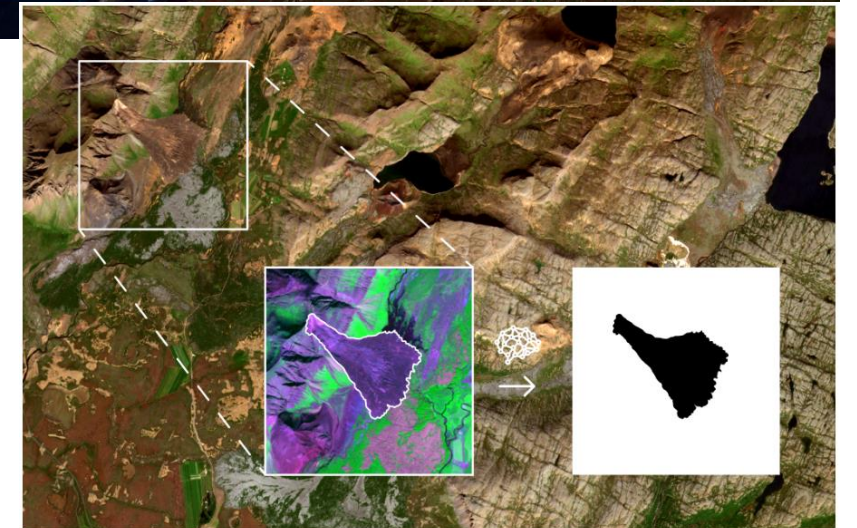
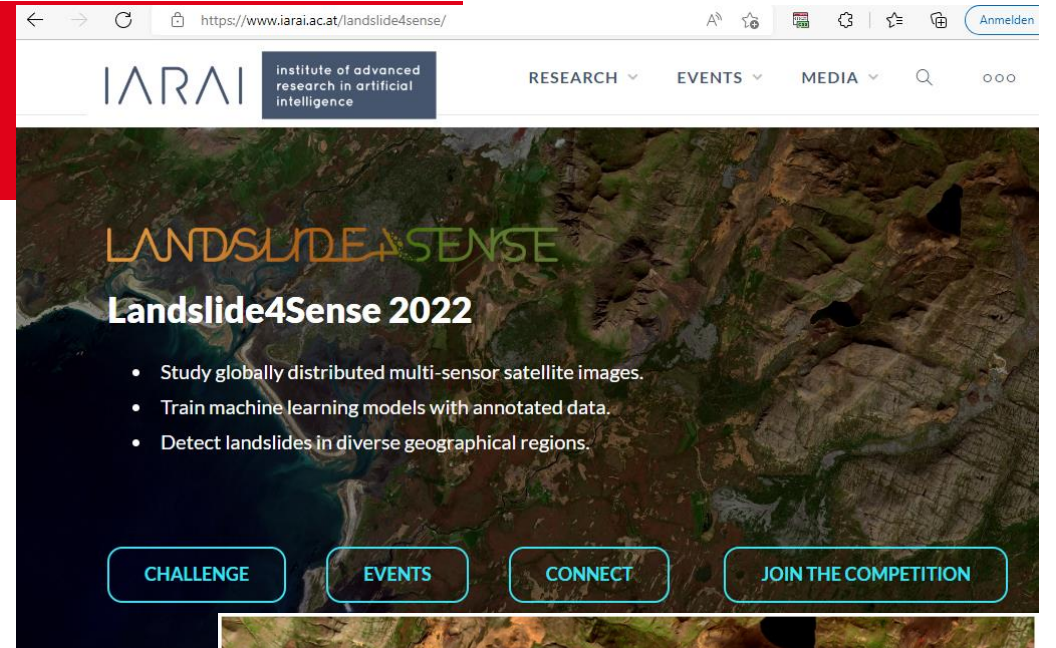
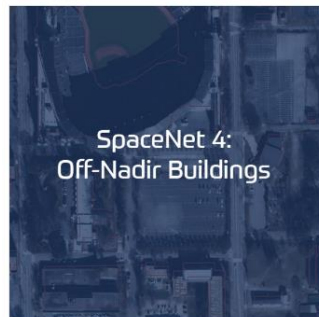
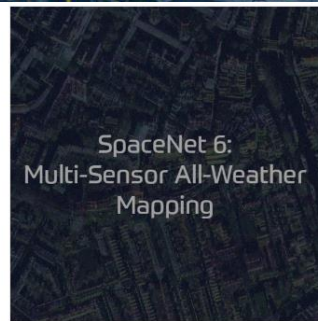
Recap

- ◆ What are the take-home messages from the presentations we heard today? What *is* hybrid AI? Theoretically and practically?

Adaption to humanitarian context

- ◆ How can we use a hybrid AI dwelling extraction for dwelling extraction?
- ◆ What would be the benefits?
- ◆ What would be challenging?

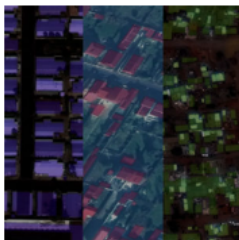
What is a data challenge?



What is a data challenge? Example

SpaceNet 2: Building Detection v2

The Problem



The commercialization of the geospatial industry has led to an explosive amount of data being collected to characterize our changing planet. One area for innovation is the application of computer vision and deep learning to extract information from satellite imagery at scale. CosmiQ Works, Radiant Solutions and NVIDIA have partnered to release the SpaceNet data set to the public to enable developers and data scientists to work with this data.

Today, map features such as roads, building footprints, and points of interest are primarily created through manual techniques. We believe that advancing automated feature extraction techniques will serve important downstream uses of map data including humanitarian and disaster response, as observed by the need to map road networks during the response to recent flooding in Bangladesh and Hurricane Maria in Puerto Rico. Furthermore, we think that solving this challenge is an important stepping stone to unleashing the power of advanced computer vision algorithms applied to a variety of remote sensing data applications in both the public and private sector.

The Data – Over 685,000 footprints across the Five SpaceNet Areas of Interest.

AOI	Area of Raster (Sq. Km)	Building Labels (Polygons)
AOI_2_Vegas	216	151,367
AOI_3_Paris	1,030	23,816
AOI_4_Shanghai	1,000	92,015
AOI_5_Khartoum	765	35,503

The Metric

In SpaceNet Challenge, the metric for ranking entries is based on the Jaccard Index, also called the Intersection-over-Union (IoU). For more information read the full article on [The DownlinQ](#).

Catalog

The data is hosted on AWS as a Public Dataset. It is free to download, but an AWS account is required.

```
aws s3 ls s3://spacenet-  
dataset/spacenet/SN2_buildings/
```

[SpaceNet 2: Building Detection v2](#)

Ideas for a data challenge

- ◆ How could we set up a data challenge to use hybrid AI in for dwelling extraction?
- ◆ What data would be ideal?
- ◆ What would be the minimum?
- ◆ What would be interesting tasks?
- ◆ How do we evaluate the contributions?
- ◆ What are good and bad experiences with Data Challenges?

Suggestive timeline

- ◆ When do we plan to publish the challenge?
- ◆ What has to happen before?
- ◆ When do we meet next?