New ways of interacting with remote sensing images: an AI perspective

Devis Tuia, Wageningen University (devis.tuia@wur.nl)

AI4Copernicus day, Namur

WAGENINGEN
UNIVERSITY & RESEARCH

100years
1918 — 2018

17.9.2019

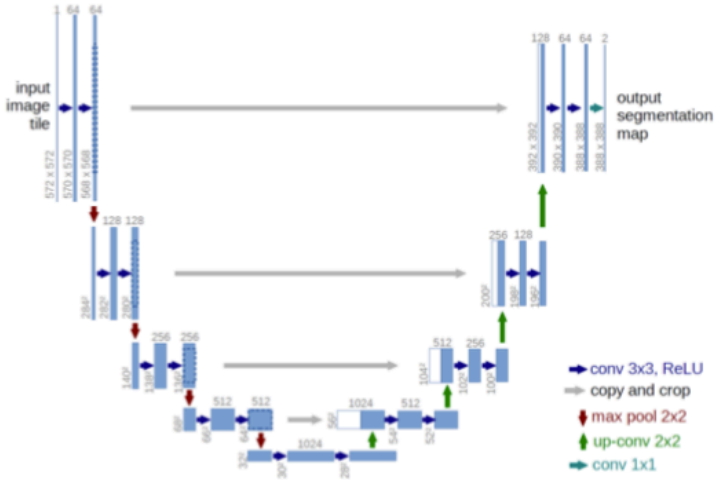
Applying deep learning with optical remote sensing data seems very easy

June 28th 2018: *Bing releases 125 million Building Footprints in the US as Open Data*
How?

Apply ResNet [He et al., 2015] + smart postprocessing

Applying deep learning with optical remote sensing data seems very easy

IGARSS 2018: *Large-scale semantic classification: outcome of the first year of Inria aerial image labeling benchmark* [Huang et al., 2018]
Winner:

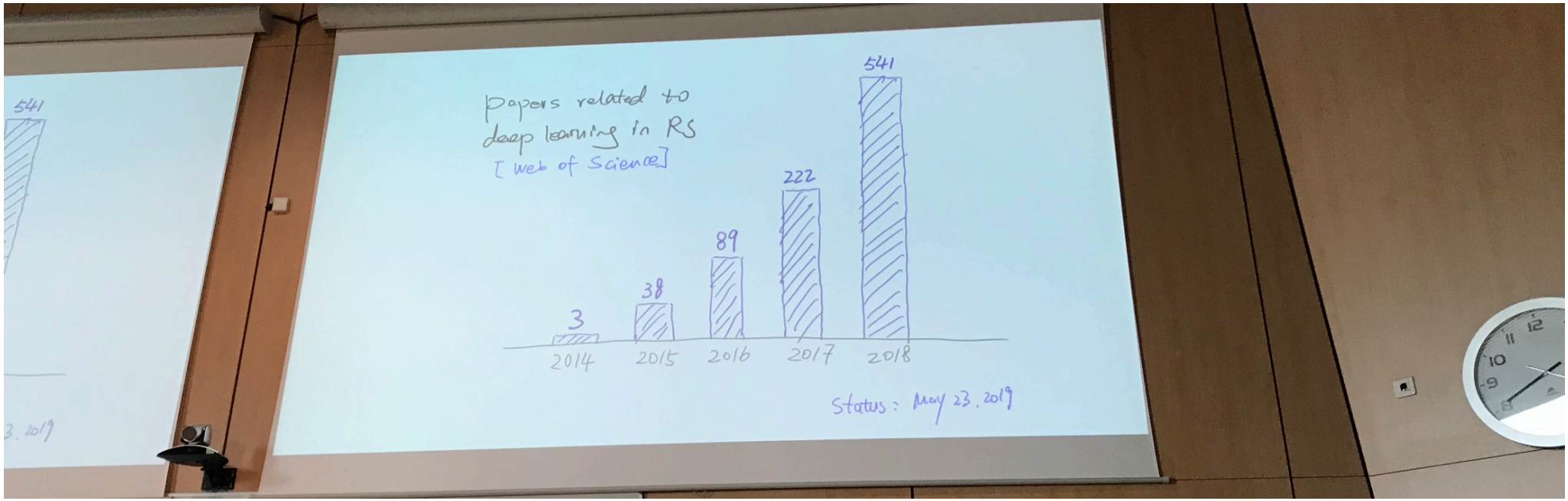


Apply U-Net [Ronneberger et al., 2015] with a modified inference method

The low hanging fruit is a blessing…

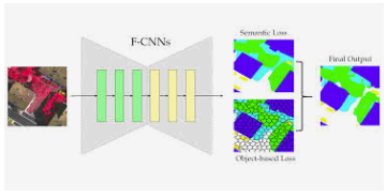
- We can advance several applications with this technology from CS
- Massive increase of "DL-in-RS" papers

[graphic by XX Zhu, 2019]

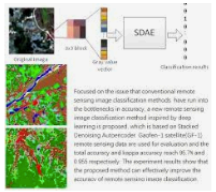


The low hanging fruit is a blessing… in disguise.

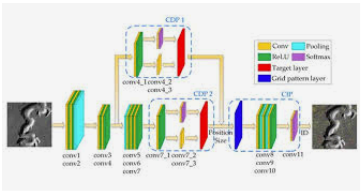
- We can advance several applications with this technology from CS
- Massive increase of “DL-in-RS” papers
- One could be easily lost in all this



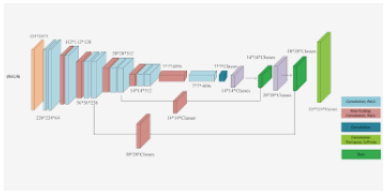
Remote Sensing | Free Full-Text | A ...
mdpi.com

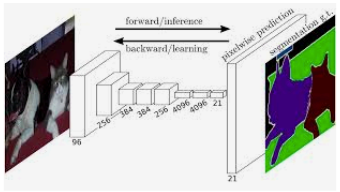


Deep Learning for Target Obje...
mdpi.com

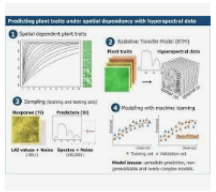


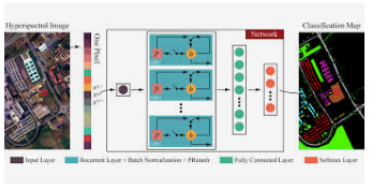
Deep Learning for Remote Sensing
mdpi.com


Supervised Classification ...
mdpi.com

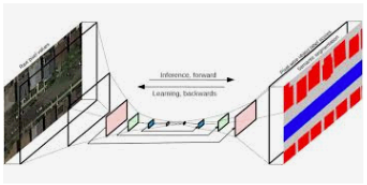


Deep Learning in Remote Sensing: A Review
arxiv.org

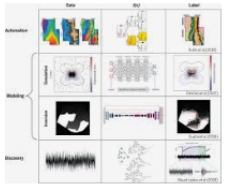

Machine Learning ...
mdpi.com

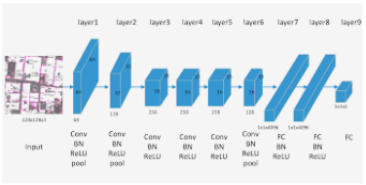


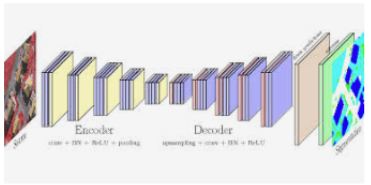
Deep Learning in Remote Sensing: A Review
arxiv.org



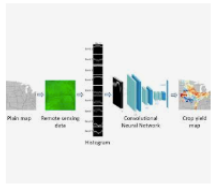
Deep Learning for Automatic Generation ...
ethz.ch

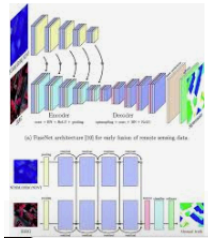

data-driven discovery in solid Ea...
science.sciencemag.org


Deep-Learning-Based Classification ...
mdpi.com



Deep learning for urban remote sensing ...
semanticscholar.org


Crop yield analysis — Sustain...
sustain.stanford.edu



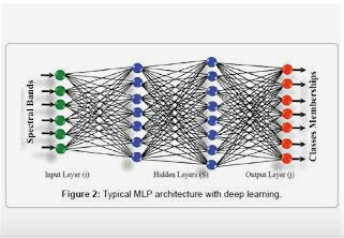
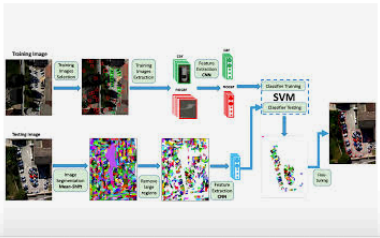
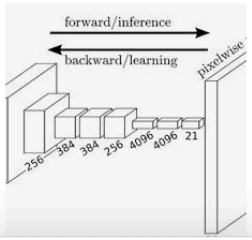
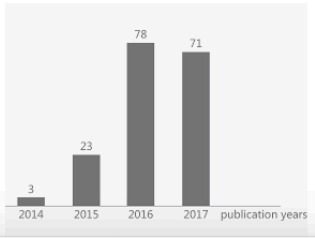
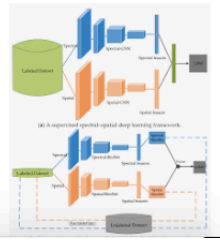


Figure 2: Typical MLP architecture with deep learning.









How advanced is DL in Geo-info data science?

1. Am I interested only in classifying pixels?

   >> then, it is pretty much advanced.

How advanced is DL in Geo-info data science?

1. Am I interested only in classifying pixels?

   >> then, it is pretty much advanced.

2. Do I want to use existing DL for my application?

   >> I should be ok.

How advanced is DL in Geo-info data science?

1. Am I interested only in classifying pixels?

   >> then, it is pretty much advanced.

2. Do I want to use existing DL for my application?

   >> I should be ok.

3. Do I want to use the full power of images (beyond RGB)?

   >> oh… this a pre-trained deep net can’t do ☹

How advanced is DL in Geo-info data science?

1. Am I interested only in classifying pixels?

   >> then, it is pretty much advanced.

2. Do I want to use existing DL for my application?

   >> I should be ok.

3. Do I want to use the full power of images (beyond RGB)?

   >> oh… this a pre-trained deep net can’t do ☹

4. Am I forgetting something?

This talk is about human / machine interaction

Users seem to be forgotten in the “AI for remote sensing”.

Without them, no training data, so no models.

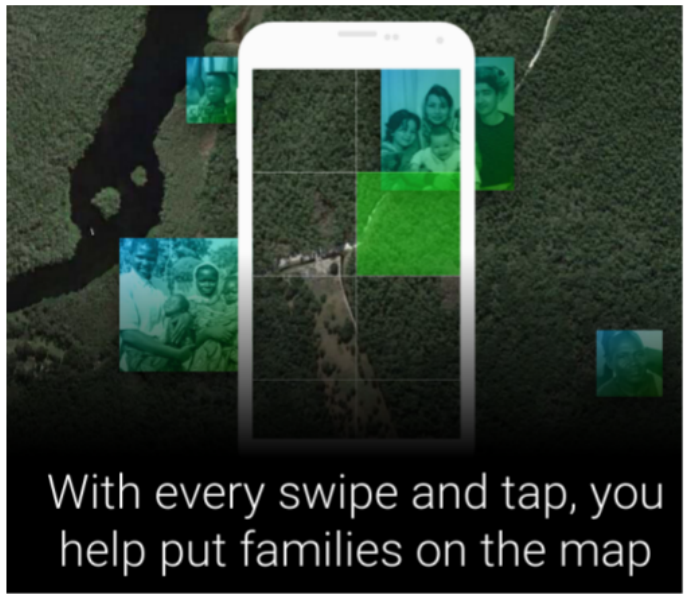
Vargas, Tuia, Falcao. *Supporting digital humanitarians in OpenStreetMap: the role of deep learning and human-machine interaction*. IJGIS, Submitted.

But why are we developing AI, if not for them?

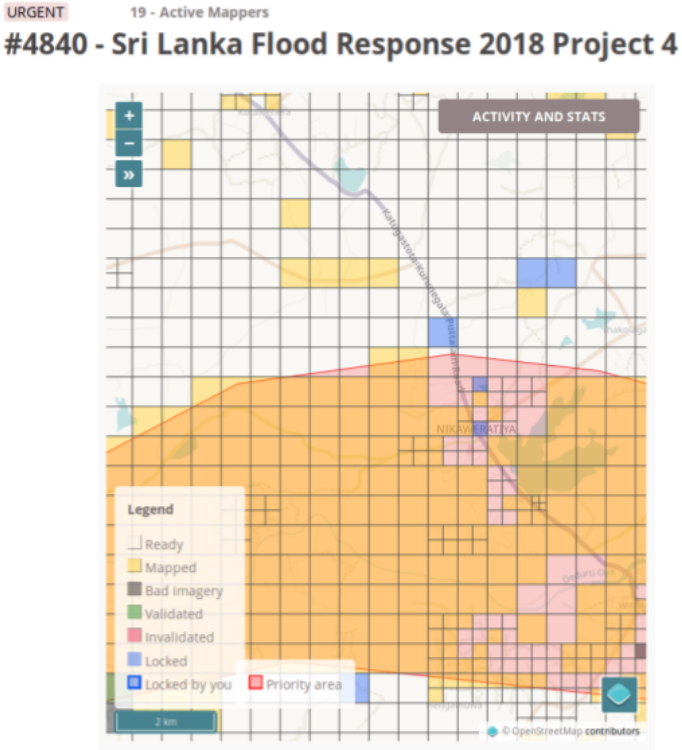
Lobry, Marcos, Murray, Tuia. *Remote Sensing Visual Question Answering*. IGARSS 2019, Yokohama

Why interacting with labelers it relevant?

MapSwipe



Humanitarian OpenStreetMap Team



Supporting volunteers in OpenStreetMap

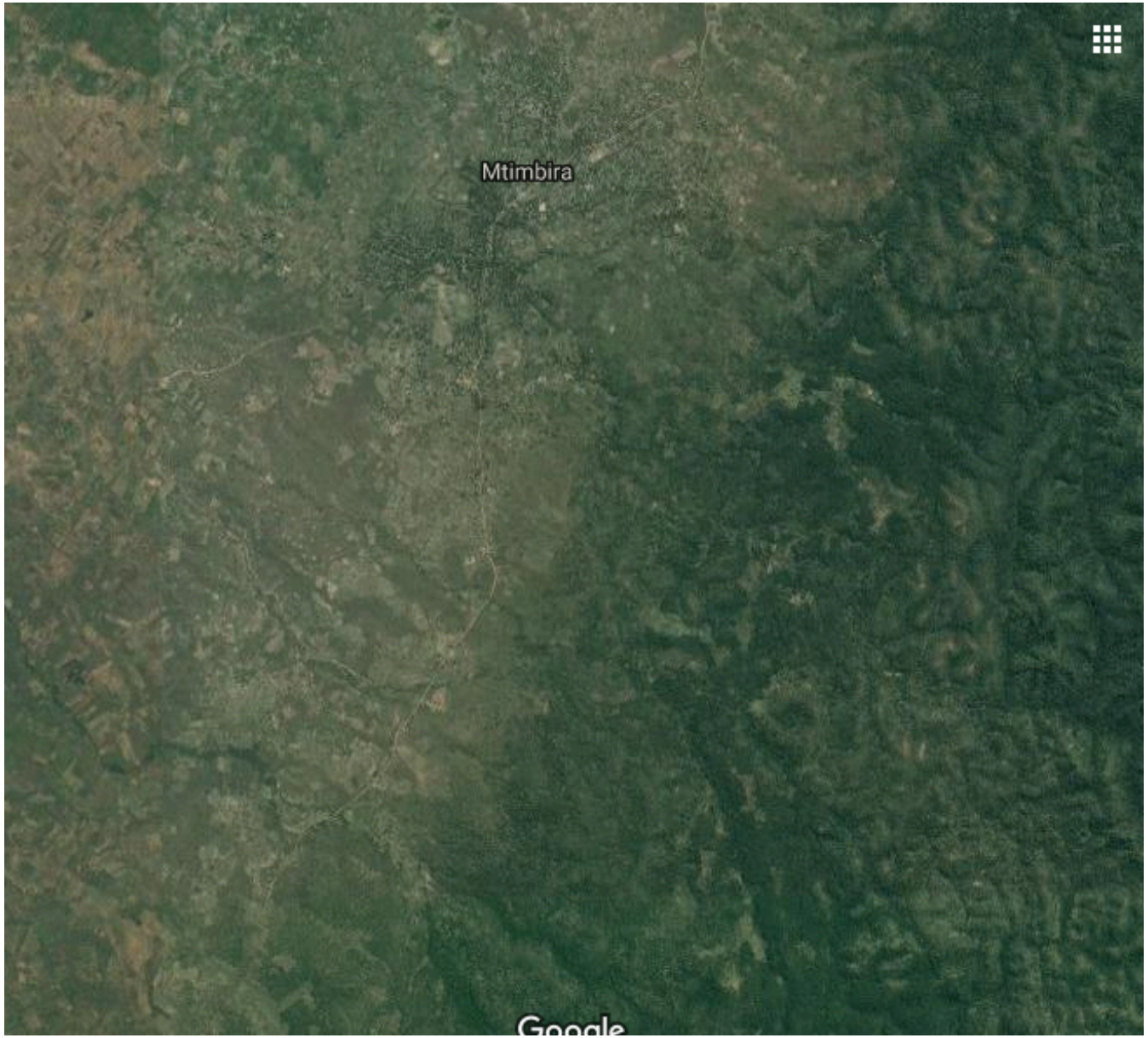
[Vargas et al., IJGIS, under review]

Supporting volunteers in OpenStreetMap

[Vargas et al., IJGIS, under review]

Supporting volunteers in OpenStreetMap

[Vargas et al., IJGIS, under review]



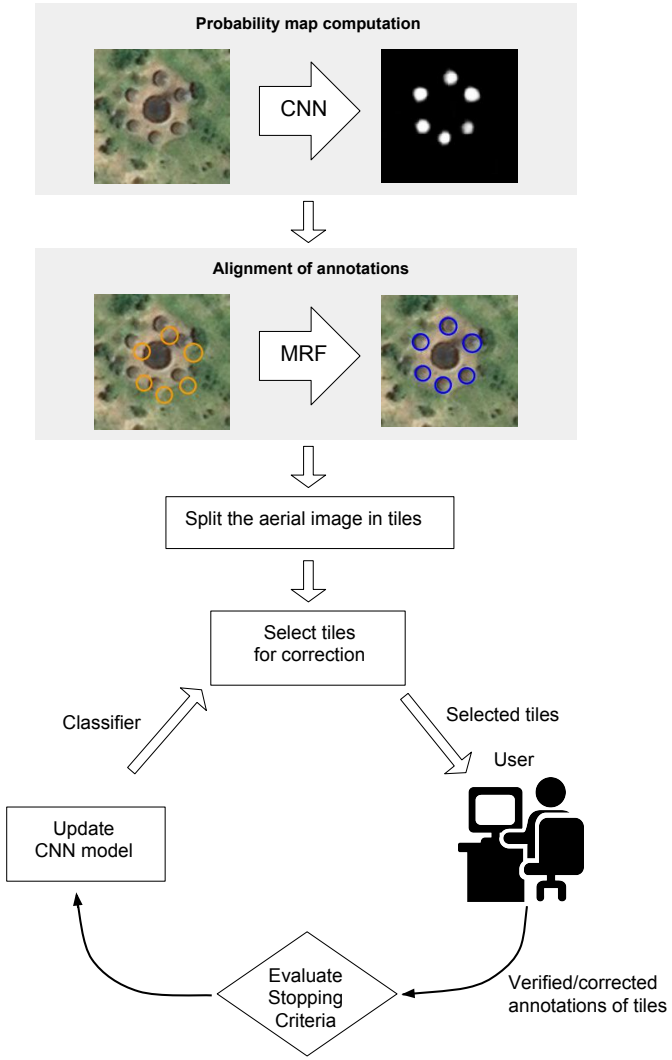
Supporting volunteers in OpenStreetMap

[Vargas et al., IJGIS, under review]

Supporting volunteers in OpenStreetMap

[Vargas et al., IJGIS, under review]

The interactive system

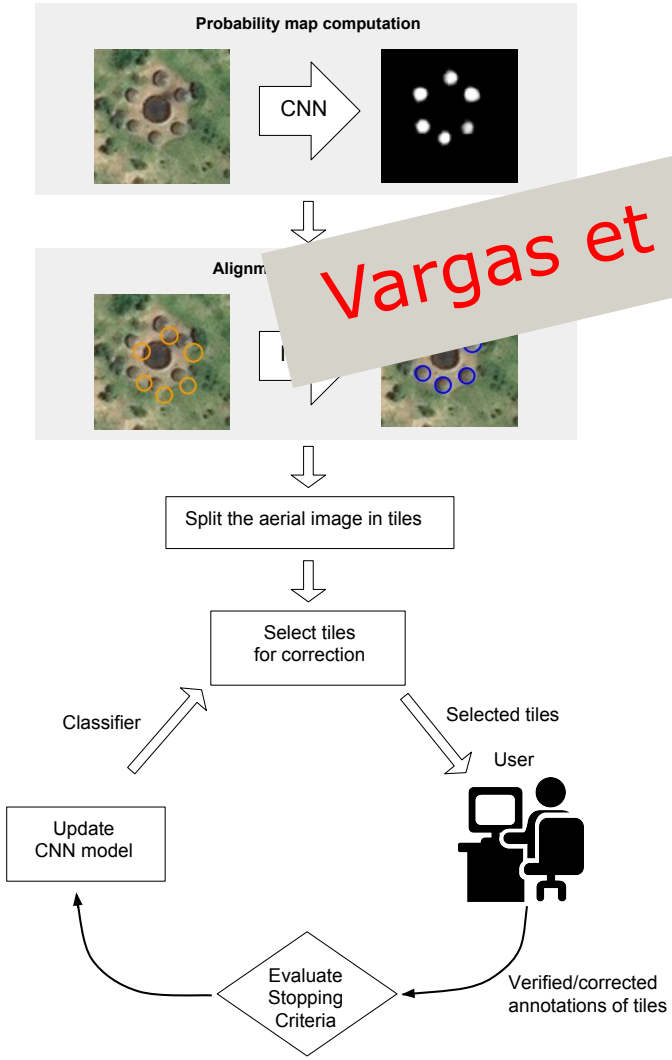


1. Predict with Convolutional network

OFFLINE

2. Align existing OSM footprints

The interactive system



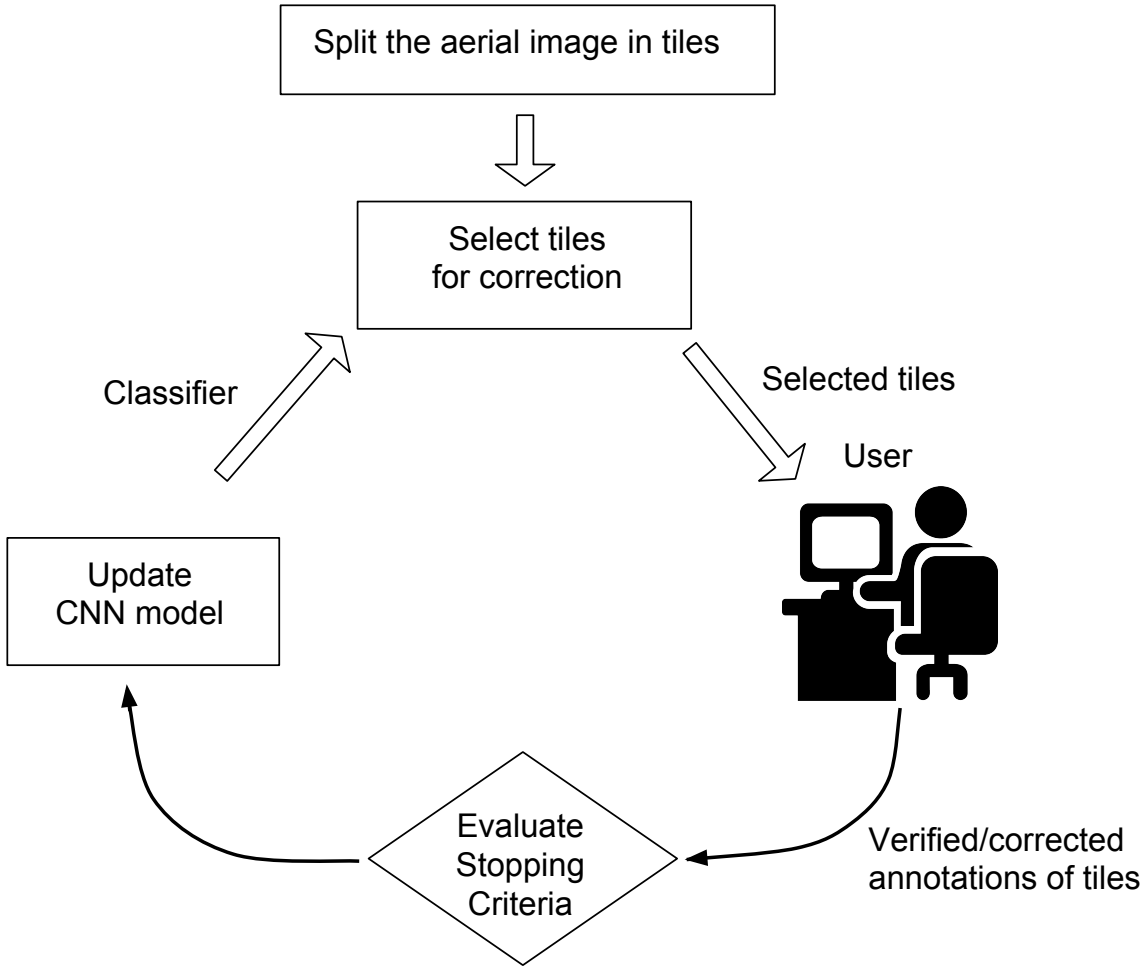
1. Predict wi… …k

Vargas et al., J. ISPRS, 2019

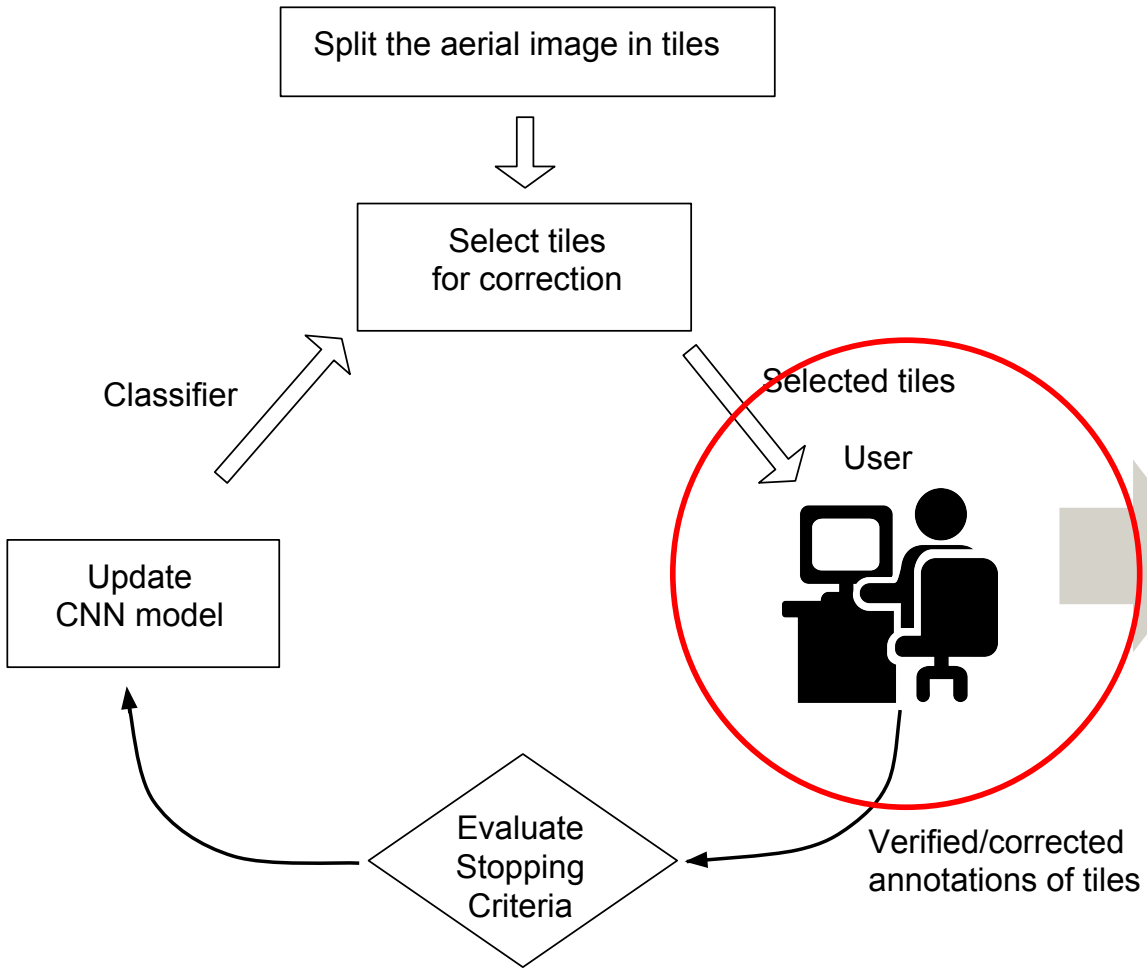
OFFLINE

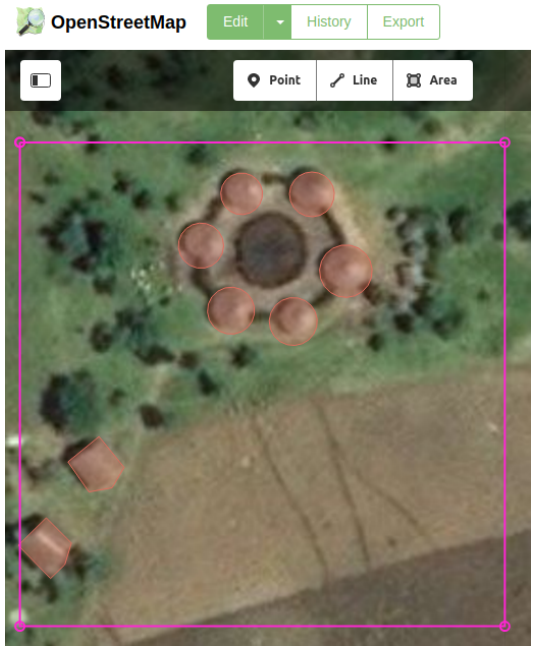
2. Align existing OSM footprints

Interactive module



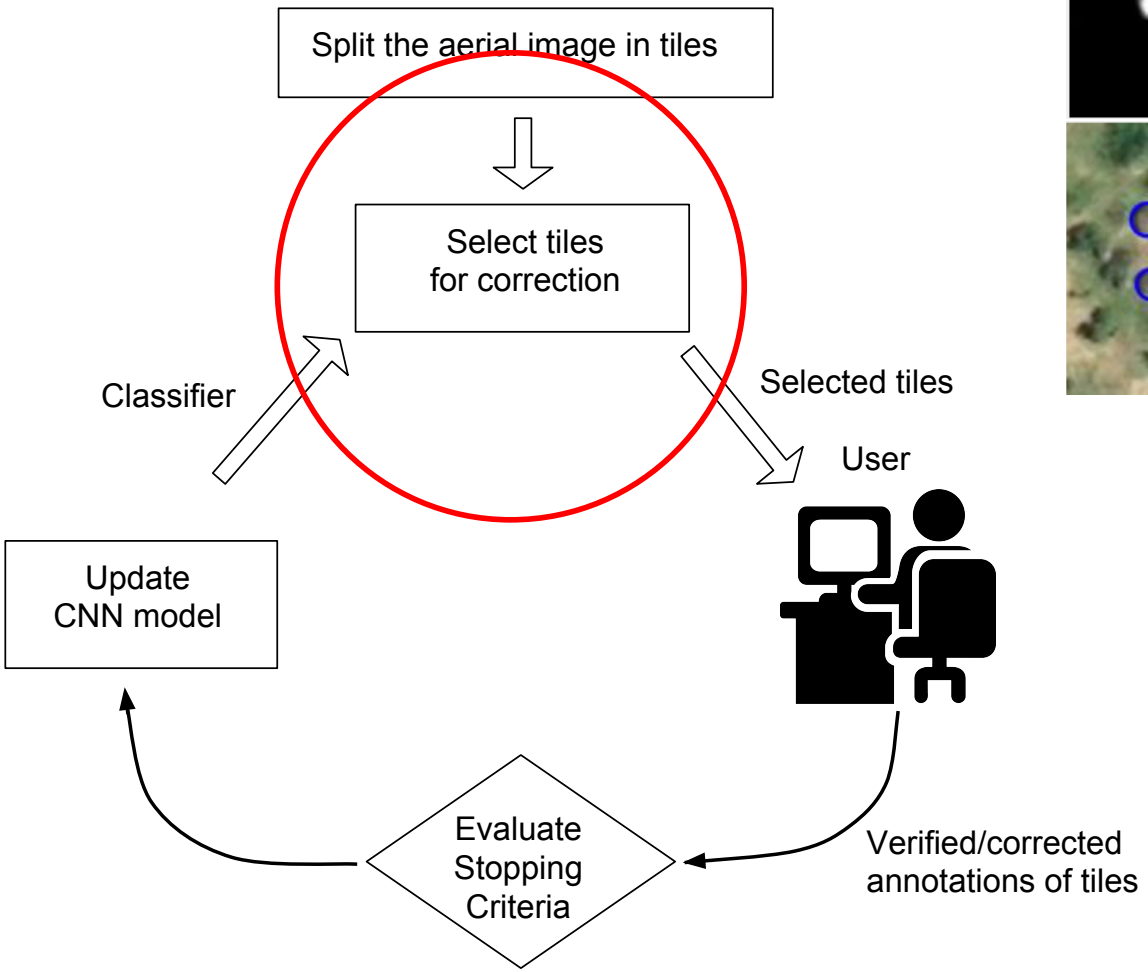
Interactive module

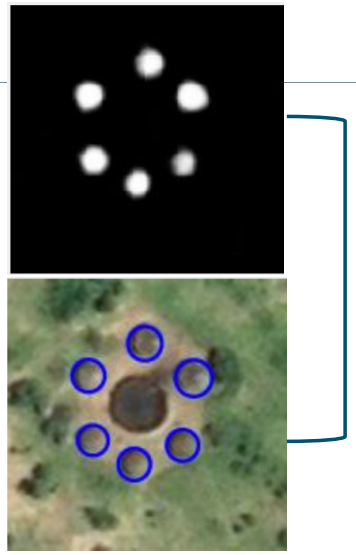




Which one to chose
when you have millions?

Interactive module





- Compare
- Estimate number of edits

The Tanzania dataset

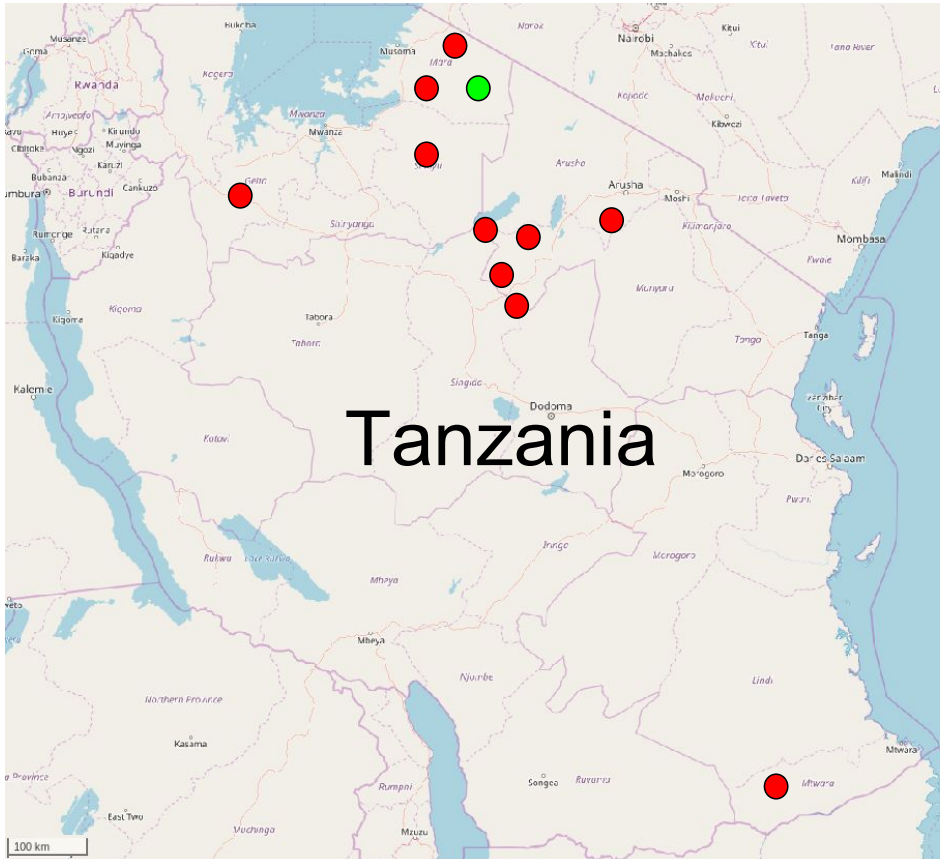
23.75 km2 of Bing imagery

- RGB, 30cm resolution

- each dot is multiple images

Training area: 3134 OSM footprints

Test area: 1392 OSM footprints.



The Tanzania dataset

23.75 km2 of Bing imagery

- RGB, 30cm resolution

- each dot is multiple images

Training area: 3134 OSM footprints

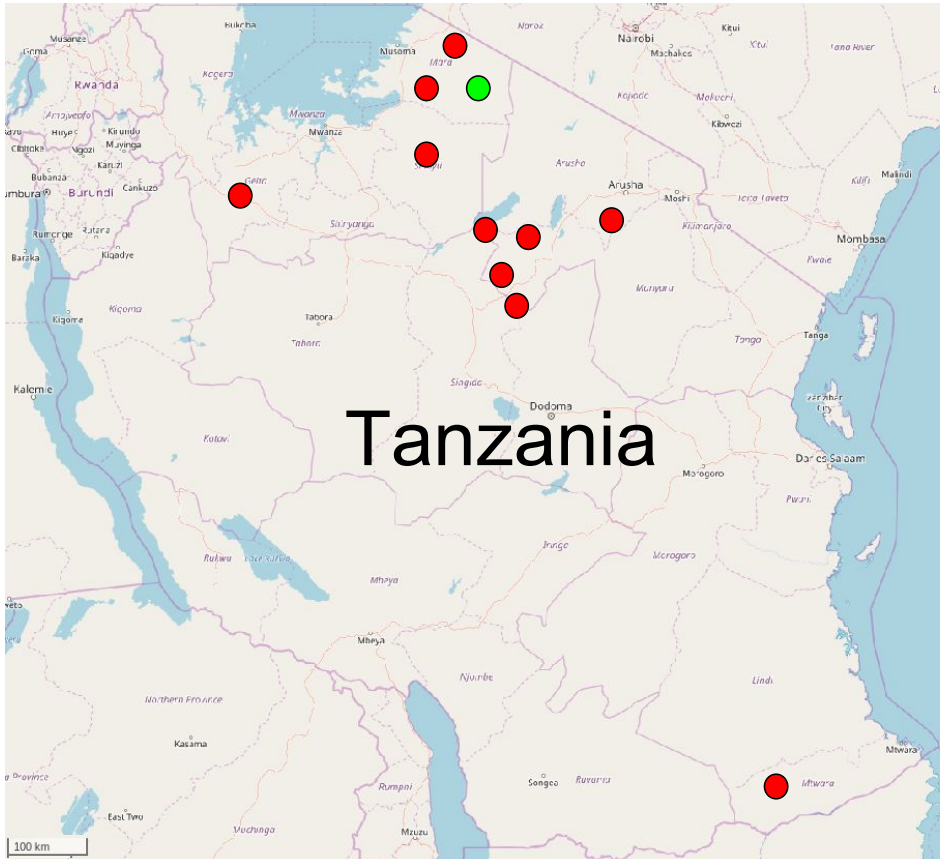
Test area: 1392 OSM footprints.

BUT:

- missing buildings

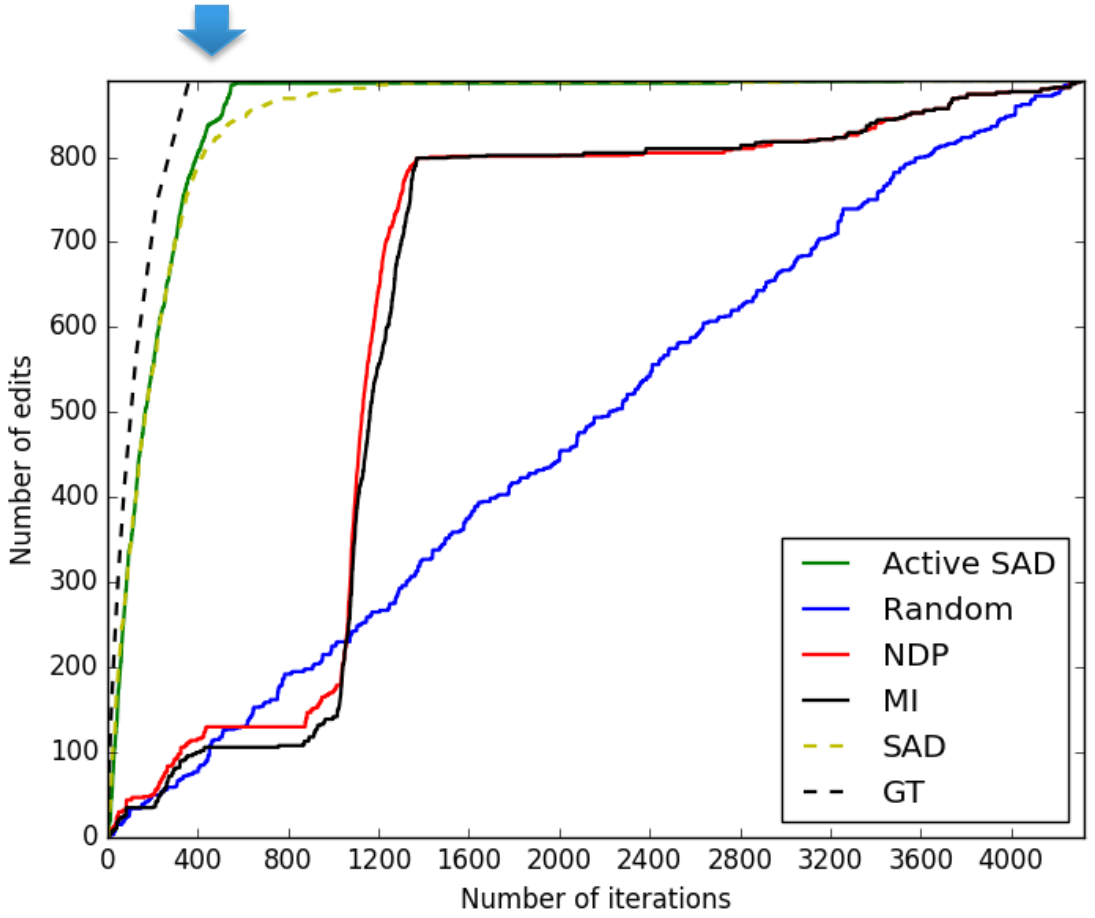
- annotation errors

Correction needs approx. 900 edits!



Results for real interaction [Vargas et al., IJGIS, under review]

The problematic images are found faster!



This talk is about human / machine interaction

User seem to be forgotten in the “AI for remote sensing”.

Without them, no training data, so no models.

Vargas, Tuia, Falcao. *Supporting digital humanitarians in OpenStreetMap: the role of deep learning and human-machine interaction*. IJGIS, Submitted.

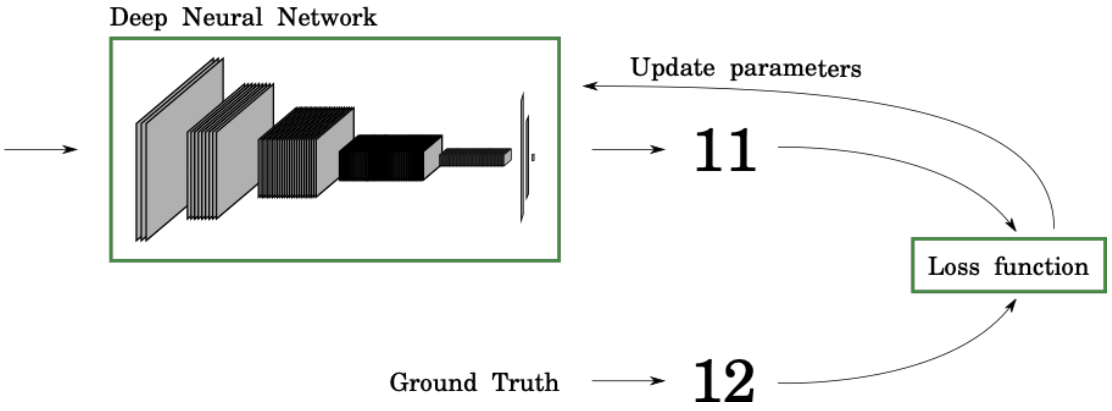
But why are we developing AI, if not for them?

Lobry, Marcos, Murray, Tuia. *Remote Sensing Visual Question Answering*. IGARSS 2019, Yokohama

We are pretty good at solving single tasks

We are pretty good at solving single tasks

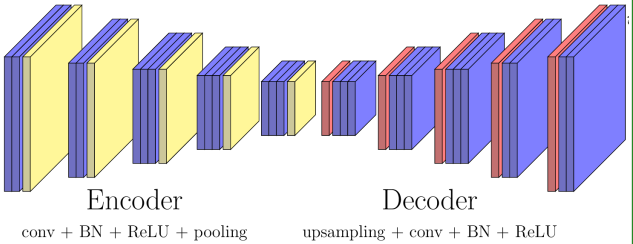
How many cars?



[Lobry and Tuia, JURSE 2019; Lang et al., LPS 2019]

We are pretty good at solving single tasks

Are there trees?



[Audebert et al., Maggiori et al.; Volpi and Tuia; …]

We are not very good at reacting to unforeseen questions

What is the road %?

Is it Paris?

Are there trees?

Am I in a city?

How many cars?

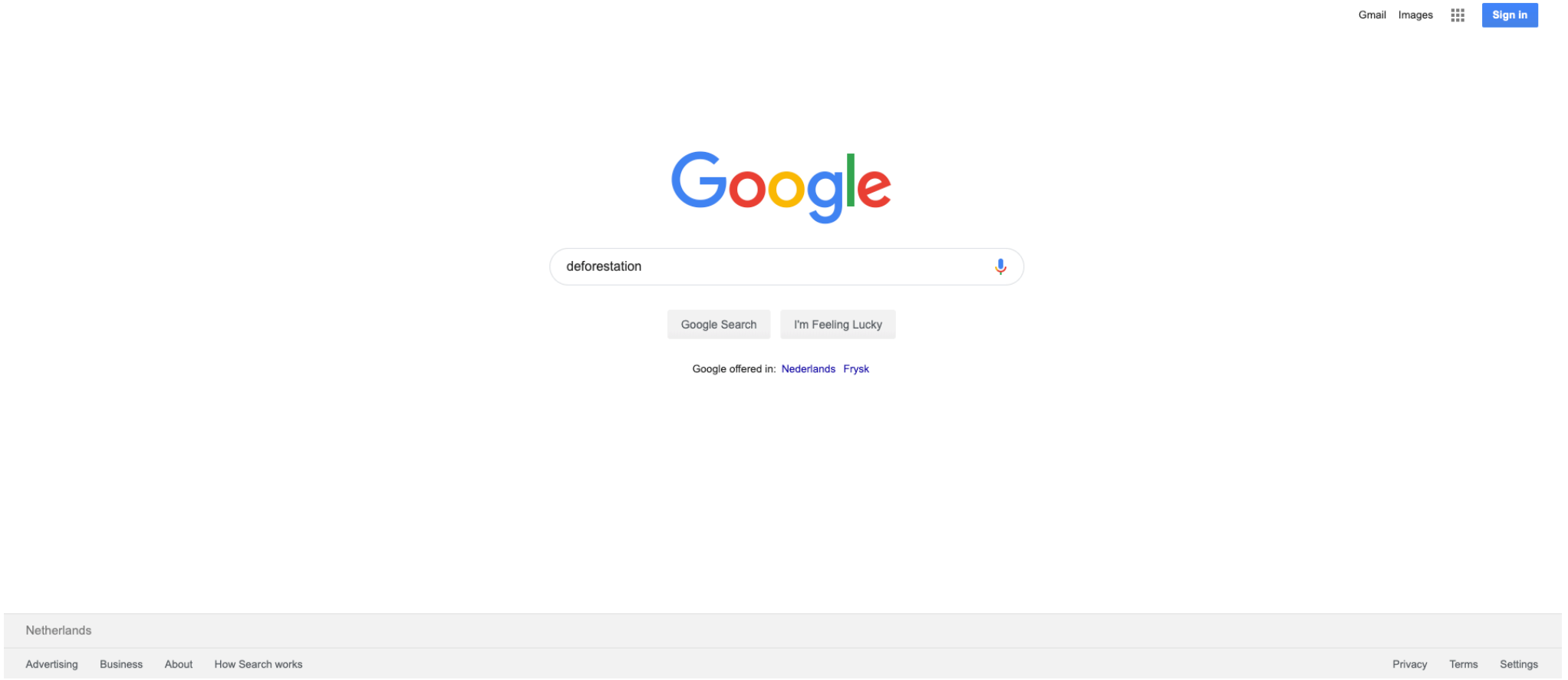
ANSWER

But this has great potential.

- Non-experts are … **non technical** experts.
- Non-experts want answer to **specific questions**.
- Non-experts want to formulate questions as **sentences**.

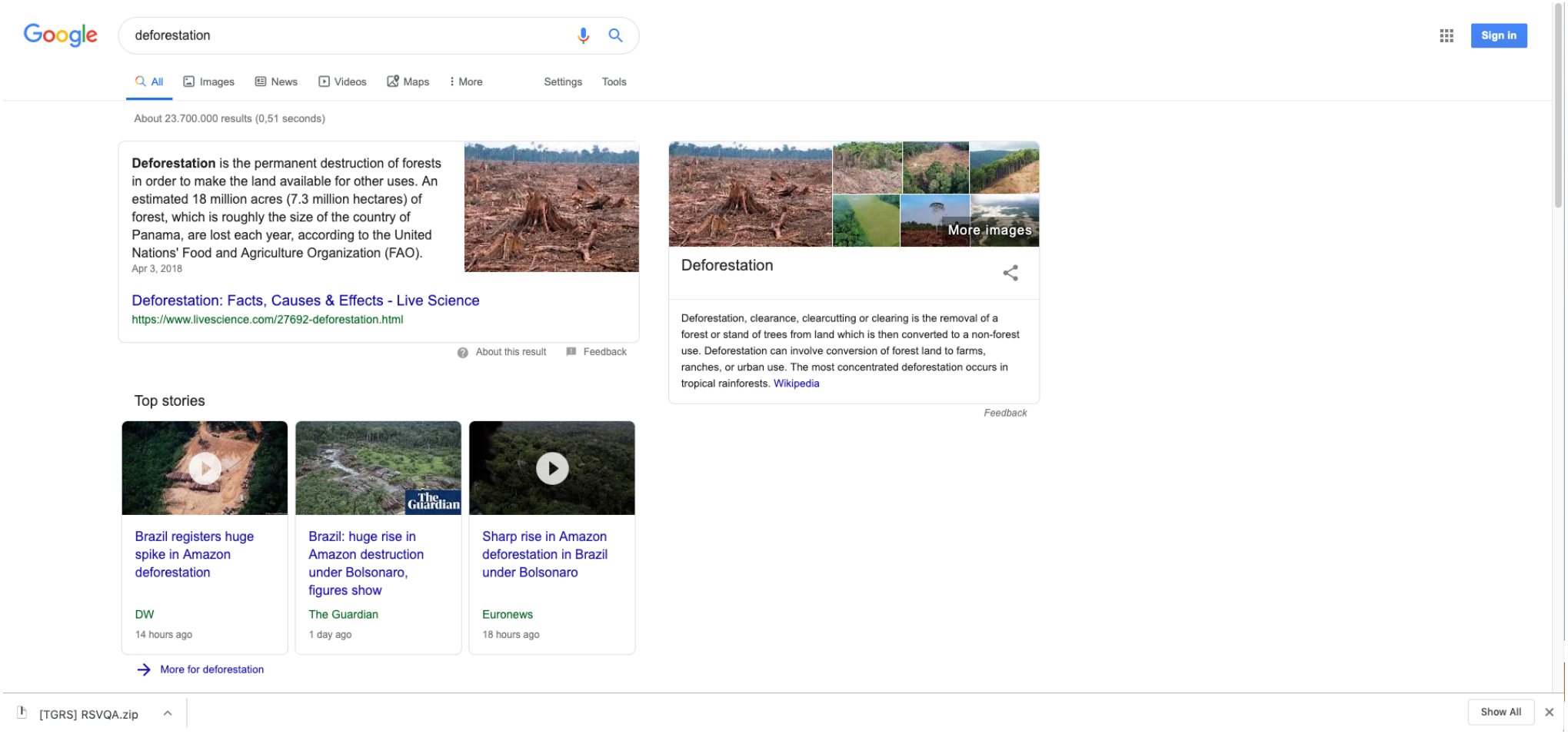
What do we need?

- For web-search it works a bit like that.



What do we need?

- For web-search it works a bit like that.

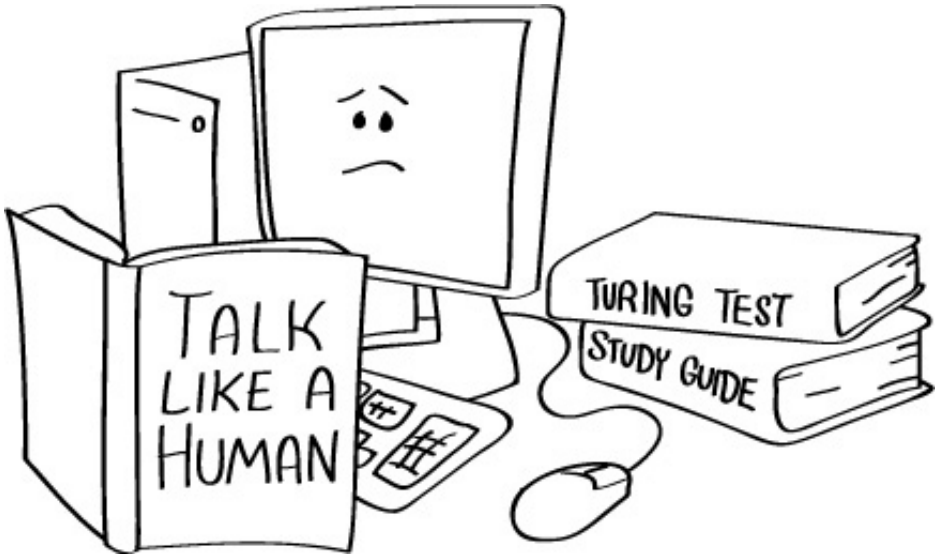


What do we need?

- For web-search it works a bit like that.
- With satellite images it just doesn’t work

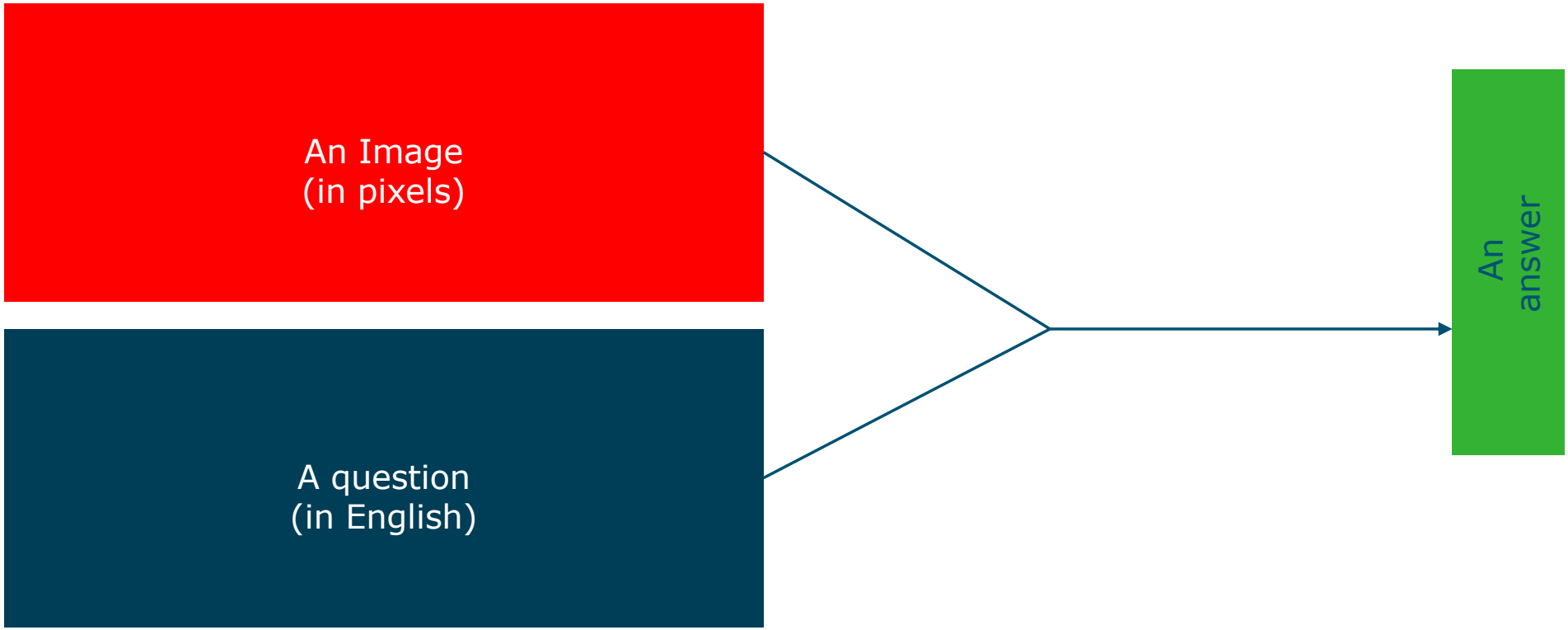
  (it’s normal. It wasn’t built for that)

But what if you could… ask questions to remote sensing images?



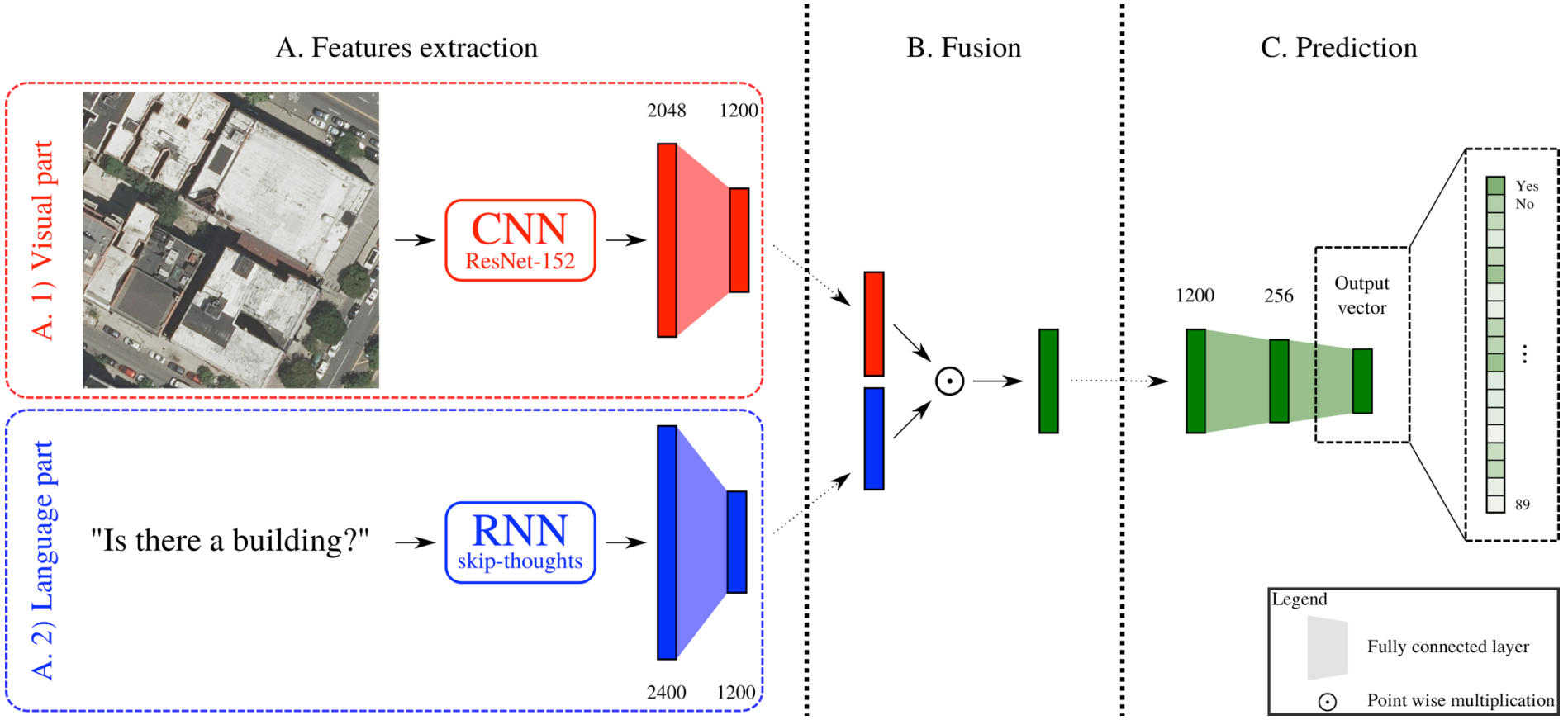
Source: CS unplugged.

Remote sensing visual question answering (RSVQA)



[Lobry, Marcos, Murray, Tuia, IGARSS 2019]

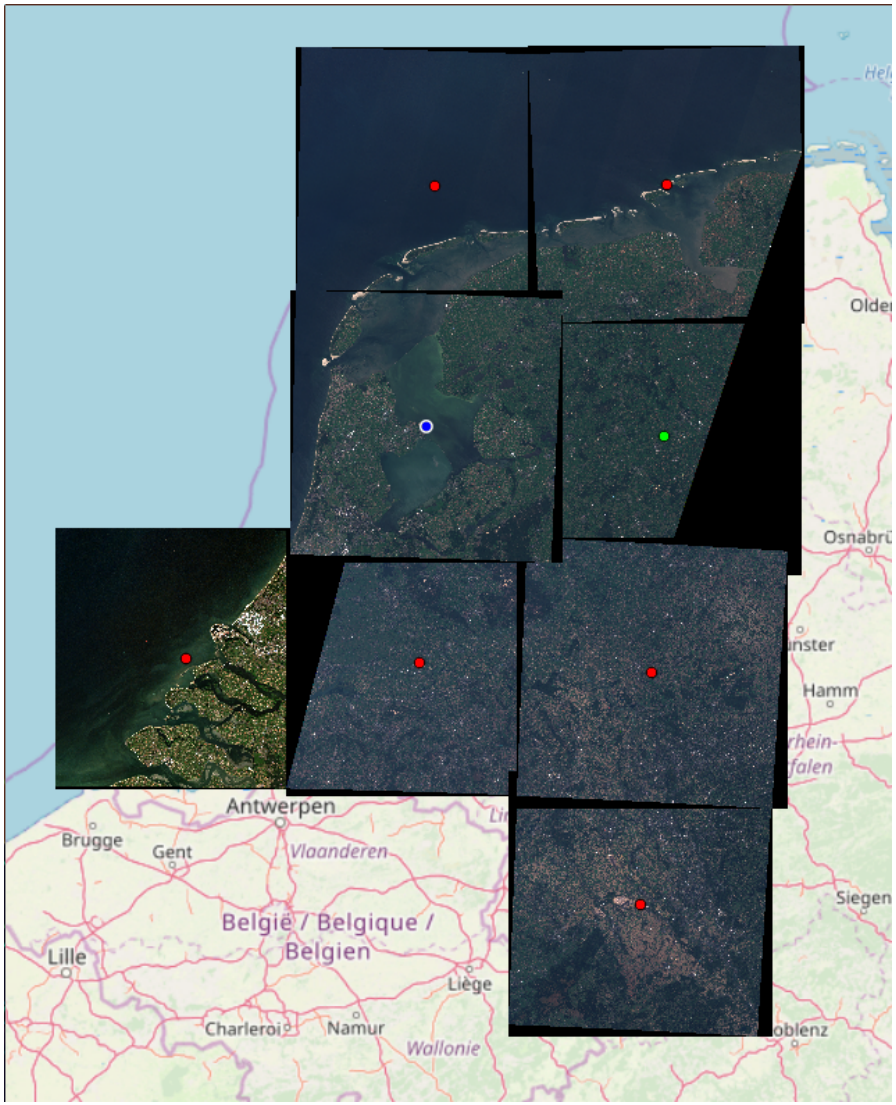
Remote sensing visual question answering (RSVQA)



[Lobry, Marcos, Murray, Tuia, IGARSS 2019]

Asking questions about the Netherlands

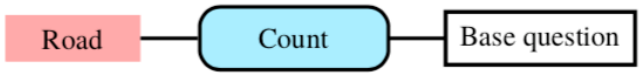
- We created a dataset of
  - Sentinel-2 images (RGB)
    - 9 scenes
    - 772 tiles (256 x 256)
  - 77’232 {image-question-answer} triplets using OSM vector data
  - Covers the whole Netherlands



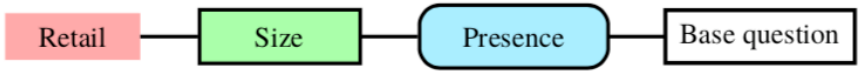
How do we train this monster?

- We generated **77’232** {image, **question**, answer} triplets

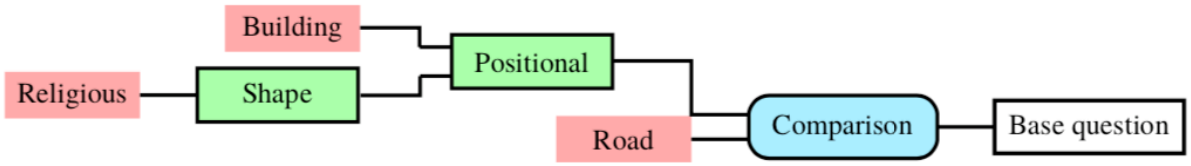
"How many roads are present in the image?"



"Is there a small retail place?"



"Is there more buildings at the top of a circular religious place than roads in the image?"



Results – Sentinel 2

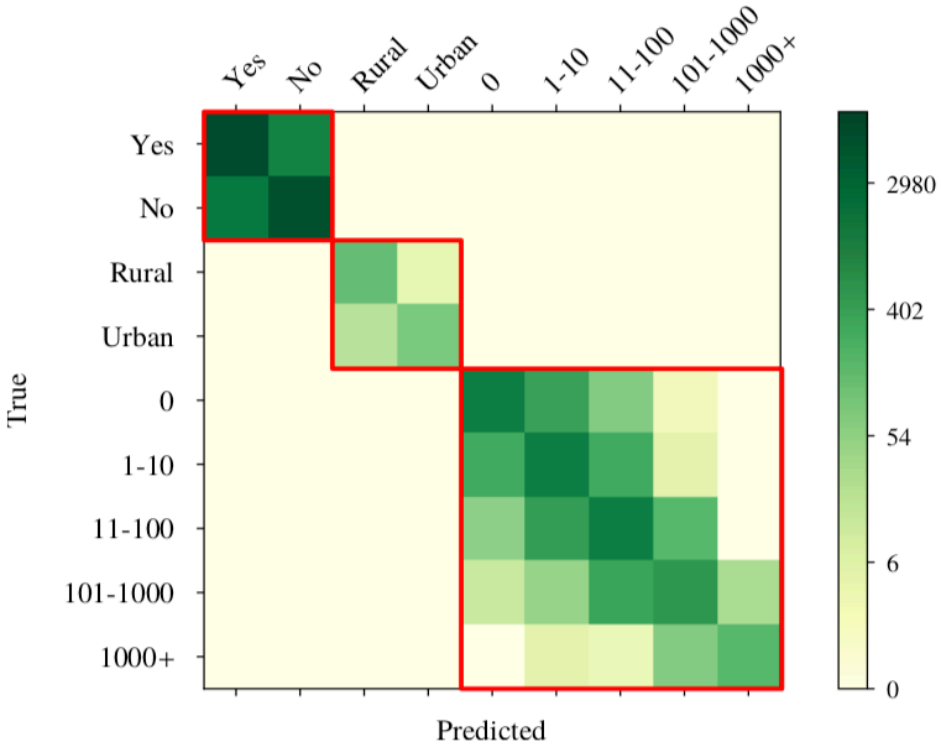
79% overall accuracy!

73% if randomizing the image part

Count questions less accurate

Type	Accuracy
Count	67.01% (0.59%)
Presence	87.46% (0.06%)
Comparison	81.50% (0.03%)
Rural/Urban	90.00% (1.41%)
AA	81.49% (0.49%)
OA	79.08% (0.20%)

The model can make a good distinction between types of questions



Results – Sentinel 2

Is it a rural or an urban area?	
Ground truth Rural	Prediction Rural

(g) LR, test set

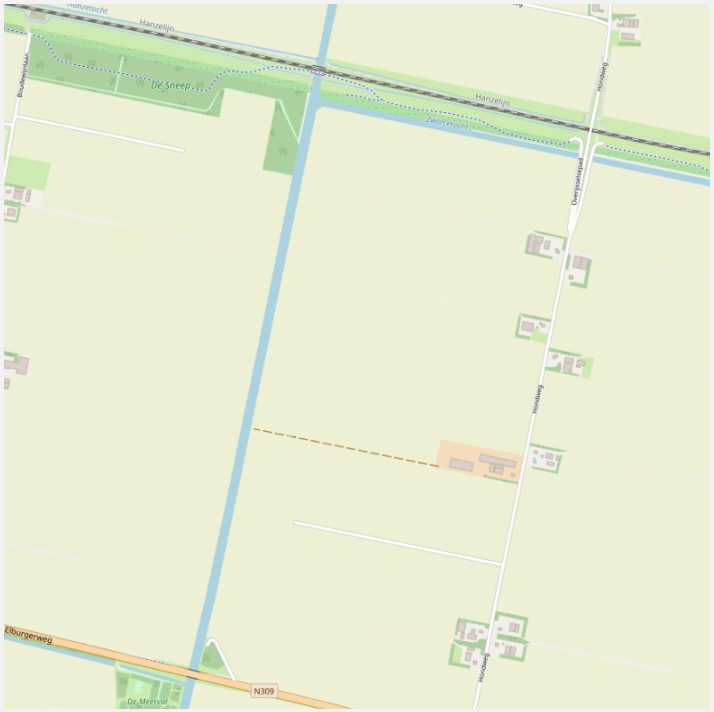
Is it a rural or an urban area?	
Ground truth Urban	Prediction Urban

Results – Sentinel 2

Are there more water areas than commercial buildings?	
Ground truth Yes	Prediction No

(i) LR, test set

Results – Sentinel 2

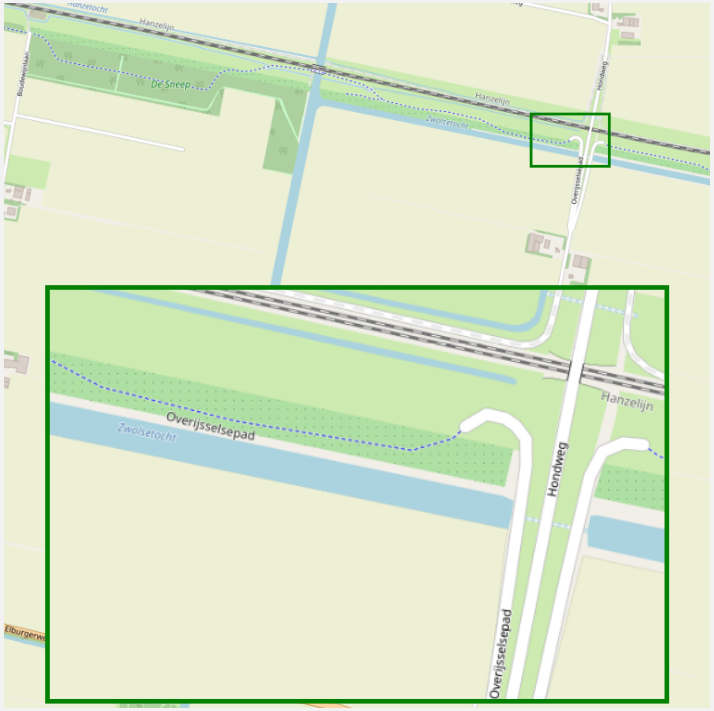


Are there more water areas than commercial buildings?	
Ground truth Yes	Prediction No

Are there more water areas than commercial buildings?	
Ground truth Yes	Prediction No

(i) LR, test set

Results – Sentinel 2

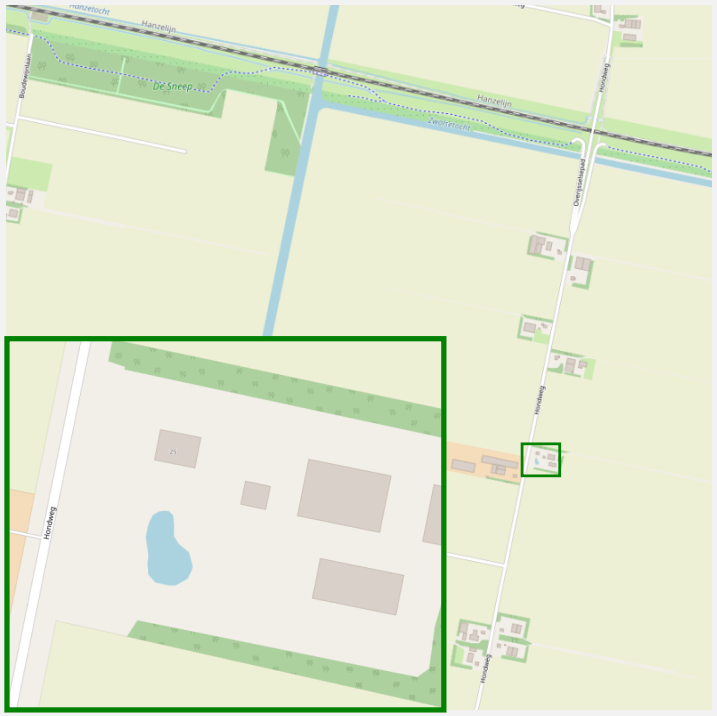


Are there more water areas than commercial buildings?	
Ground truth	Prediction
Yes	No

Are there more water areas than commercial buildings?	
Ground truth	Prediction
Yes	No

(i) LR, test set

Results – Sentinel 2



Are there more water areas than commercial buildings?	
Ground truth Yes	Prediction No

Are there more water areas than commercial buildings?	
Ground truth Yes	Prediction No

(i) LR, test set

Visual Question Answering (VQA)

[Lobry, Marcos, Murray, Tuia, IGARSS 2019]

- Joins image recognition and natural language processing deep models
- Opens use of EO image data to the laymen
- Towards an EO search engine

- A project in collaboration with:

Take home messages

- CNNs are beautiful, indeed
- But they do not do everything.

- We have important problems to solve, societally relevant
- They are not ONLY about classifying pixels!
- They involve human operators and decision makers!

- Deep learning brings us new potential, let’s explore uncharted territory!

Thanks!







Contact me!

devis.tuia.googlepages.com (with links to codes!)

devis.tuia@wur.nl