# *Deep learning and remote sensing*

*beyond classifying pixels*

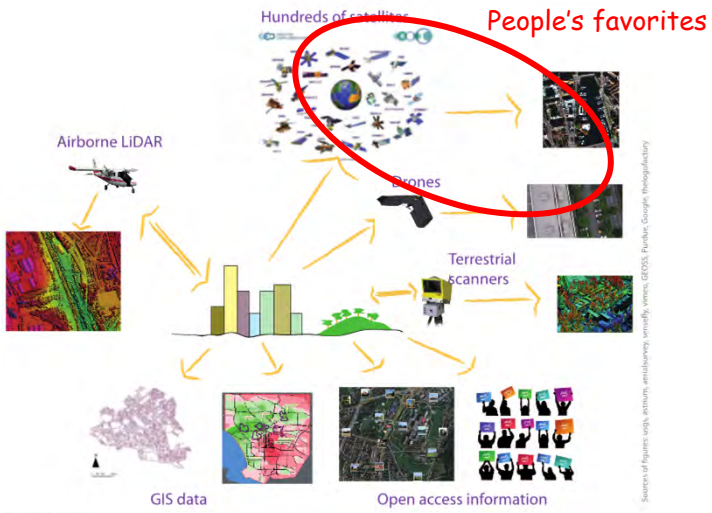
Devis Tuia, *Wageningen University*

EO, Science and Society Symposium, 10 October 2019

WAGENINGEN UNIVERSITY & RESEARCH

1

# Geo-information (data) science is about




WAGENINGEN UNIVERSITY & RESEARCH

2

2

## 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

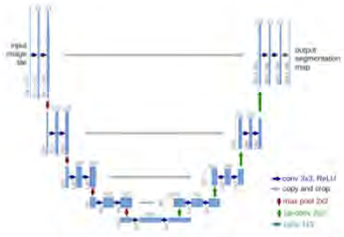
WAGENINGEN UNIVERSITY & RESEARCH

3

3

## 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

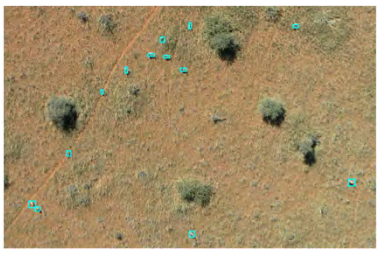
WAGENINGEN UNIVERSITY & RESEARCH

4

4

## The low hanging fruit is a blessing…

- We can advance several applications with this technology from CS

[Kellenberger et al., RSE, 2018]

WAGENINGEN UNIVERSITY & RESEARCH

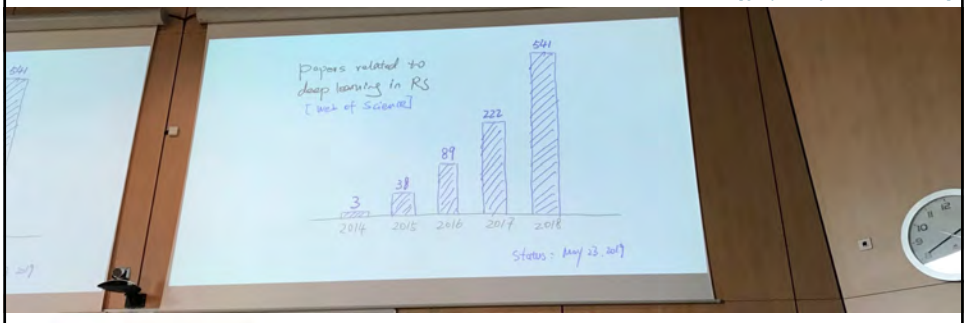
5

5

## 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

[graphic by XX Zhu, 2019]



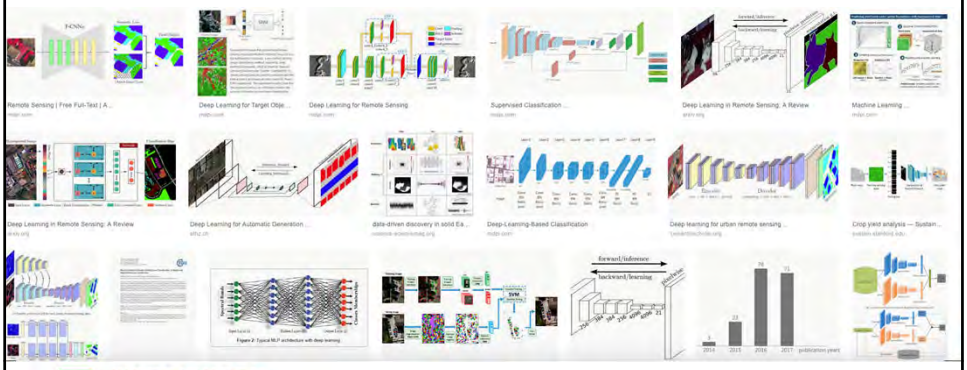

WAGENINGEN UNIVERSITY & RESEARCH

6

6

## 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 get lost into this jungle.



WAGENINGEN UNIVERSITY & RESEARCH

7

7

## How advanced is DL in Geo-info data science?

1. Am I interested only in classifying pixels?

   >> then, it is pretty much advanced.

8

8

## 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.

9

9

## 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 ☹

10

10

## 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)?

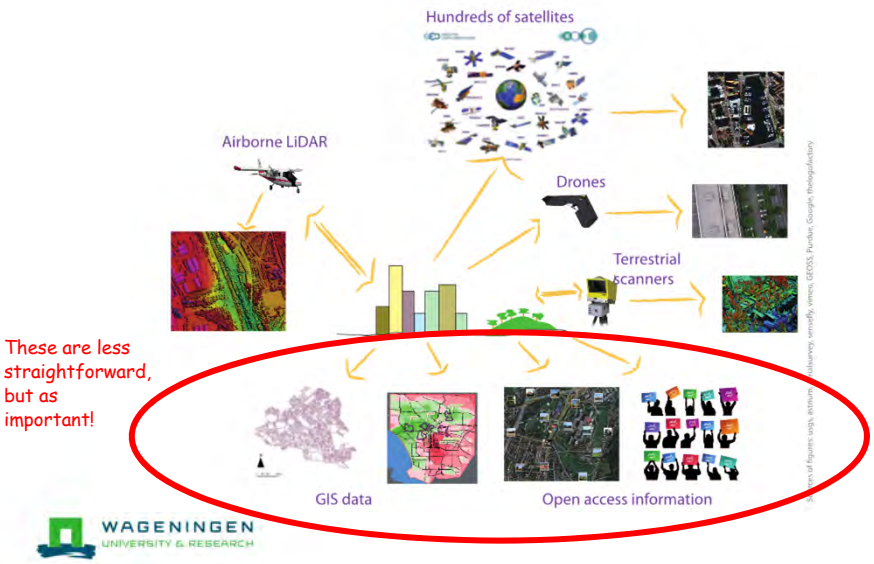
   >> oh… this a pre-trained deep net can't do ☹

4. Am I forgetting something?

11

11

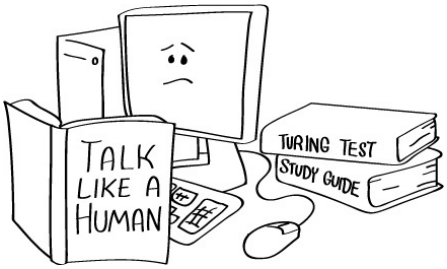
## Geo-information (data) science is ALSO about




WAGENINGEN UNIVERSITY & RESEARCH

12

12




## Let's dialogue with images

Based on:

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

WAGENINGEN UNIVERSITY & RESEARCH



13

## We are good at solving single tasks

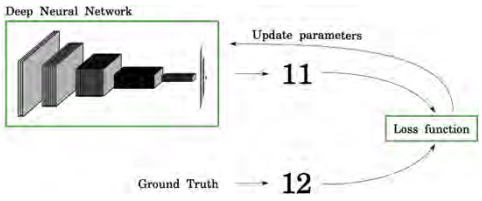
WAGENINGEN UNIVERSITY & RESEARCH



14

## We are good at solving single tasks

How many cars?




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

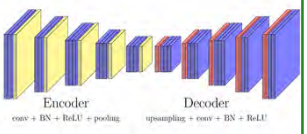
WAGENINGEN UNIVERSITY & RESEARCH

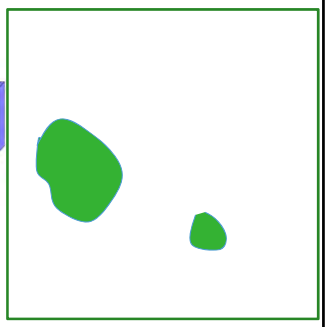
15

15

## We are good at solving single tasks

Are there trees?



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

WAGENINGEN UNIVERSITY & RESEARCH

16

16

## 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** ←

WAGENINGEN UNIVERSITY & RESEARCH

17

17

## 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**.

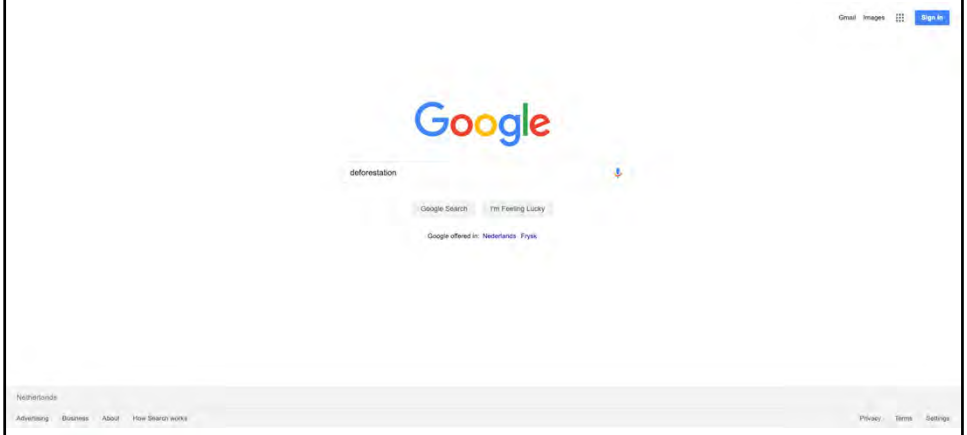
WAGENINGEN UNIVERSITY & RESEARCH

18

18

## What do we need?

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



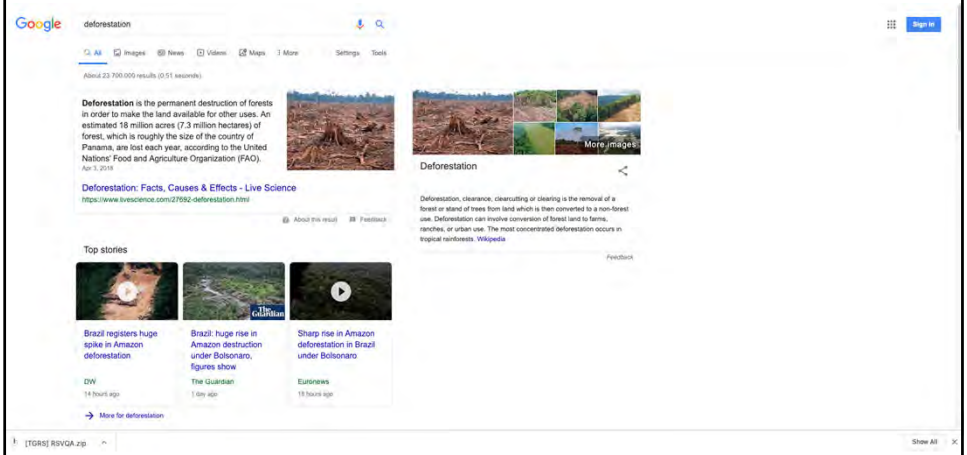

WAGENINGEN UNIVERSITY & RESEARCH



19

## What do we need?

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




WAGENINGEN UNIVERSITY & RESEARCH

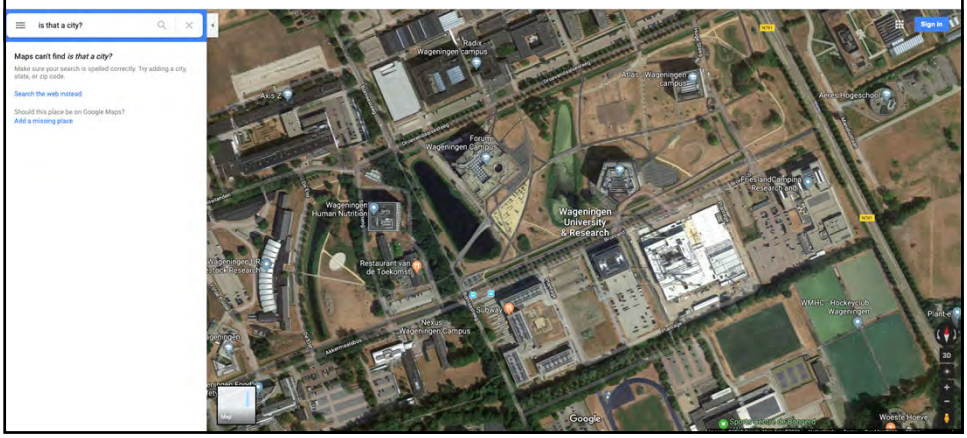


20

## What do we need?

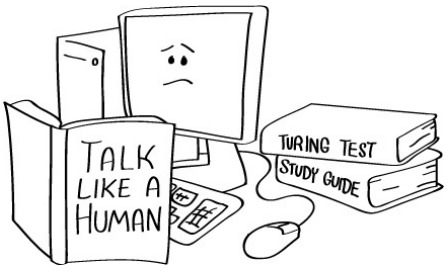
- For web-search it works a bit like that.
- With satellite images it just doesn't work

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




21

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



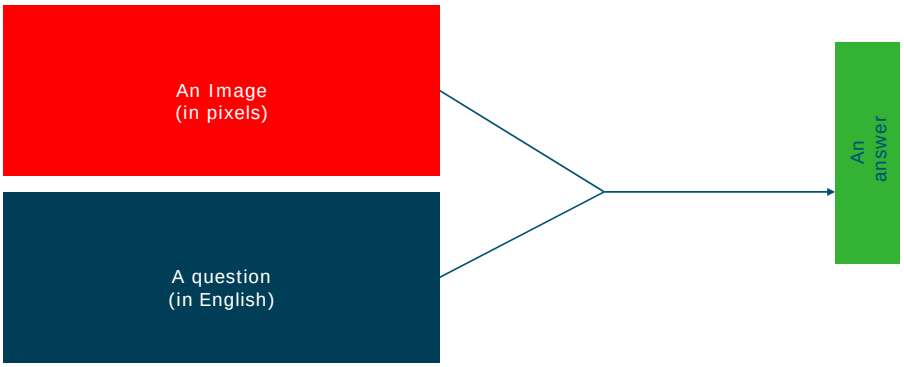

Source: CS unplugged.

WAGENINGEN UNIVERSITY & RESEARCH

22

22

## Remote sensing visual question answering (RSVQA)

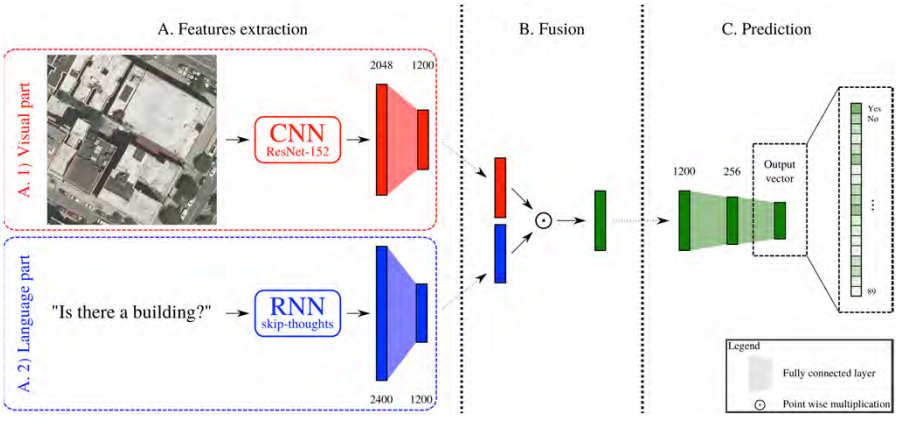



WAGENINGEN UNIVERSITY & RESEARCH

[Lobry, Marcos, Murray, Tuia, IGARSS 2019] 23

23

## Remote sensing visual question answering (RSVQA)




WAGENINGEN UNIVERSITY & RESEARCH

[Lobry, Marcos, Murray, Tuia, IGARSS 2019] 24

24

## Remote sensing visual question answering (RSVQA)

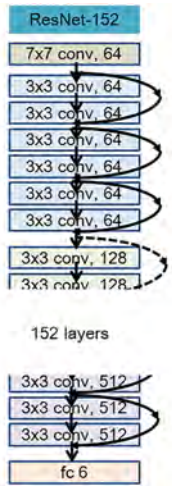

WAGENINGEN UNIVERSITY & RESEARCH

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



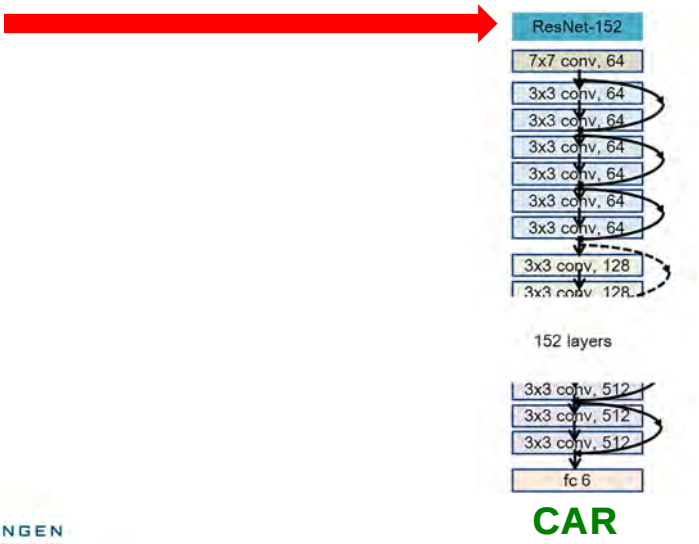


## Visual part: ResNet-152 (He et al., CVPR 2016)




WAGENINGEN UNIVERSITY & RESEARCH

## Visual part: ResNet-152 (He et al., CVPR 2016)

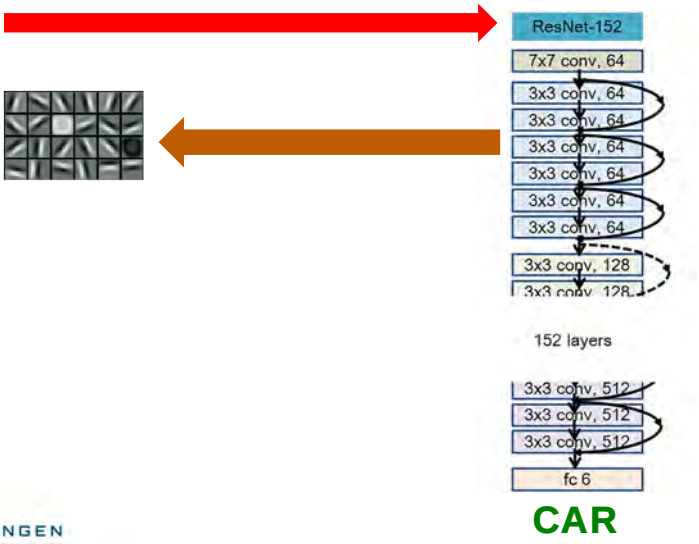



WAGENINGEN UNIVERSITY & RESEARCH

27

27

## Visual part: ResNet-152 (He et al., CVPR 2016)

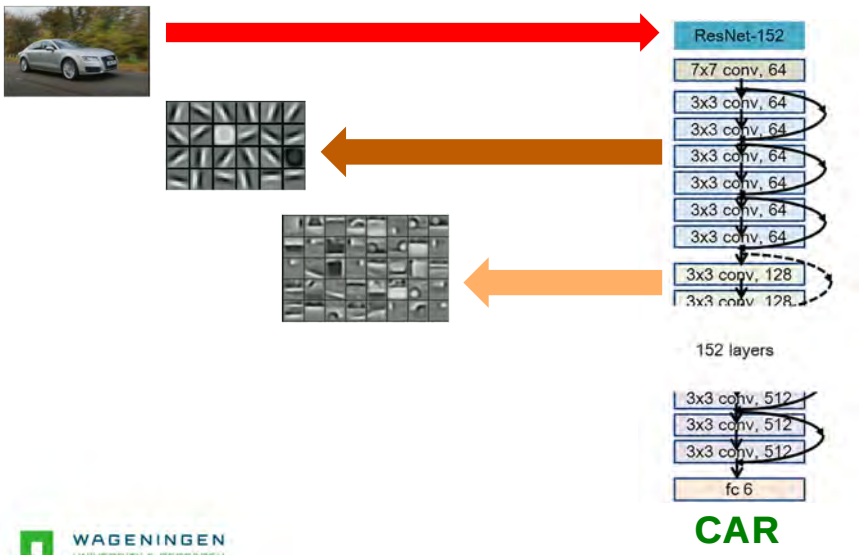



WAGENINGEN UNIVERSITY & RESEARCH

28

28

## Visual part: ResNet-152 (He et al., CVPR 2016)

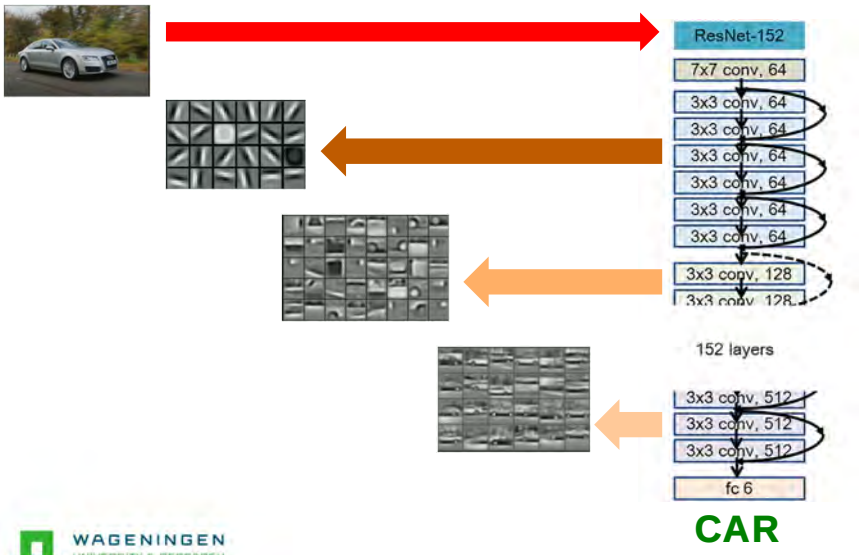



WAGENINGEN UNIVERSITY & RESEARCH

29

29

## Visual part: ResNet-152 (He et al., CVPR 2016)

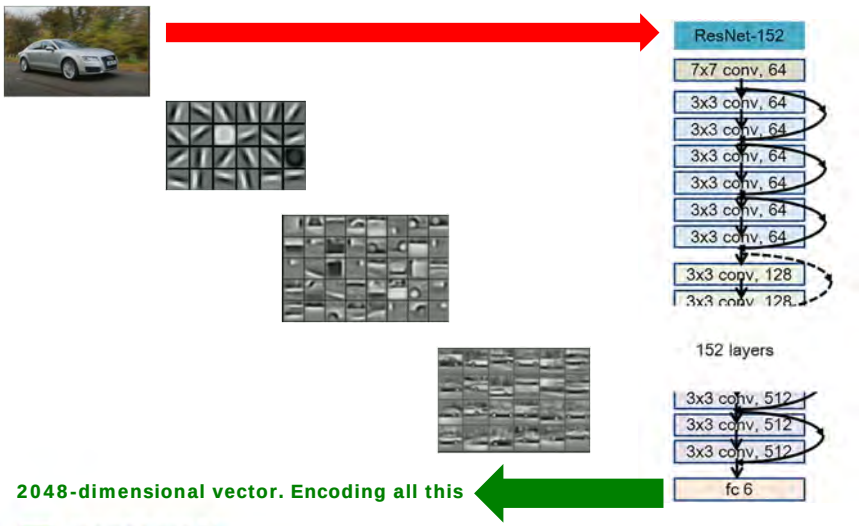



WAGENINGEN UNIVERSITY & RESEARCH

30

30

## Visual part: ResNet-152 (He et al., CVPR 2016)




WAGENINGEN UNIVERSITY & RESEARCH

31

31

## Remote sensing visual question answering (RSVQA)

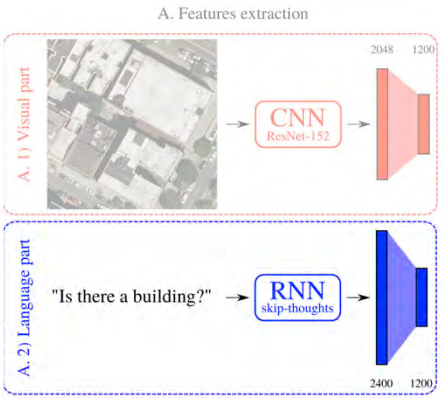

WAGENINGEN UNIVERSITY & RESEARCH

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

32

32

# Introducing remote sensing visual question answering (RSVQA)




WAGENINGEN UNIVERSITY & RESEARCH

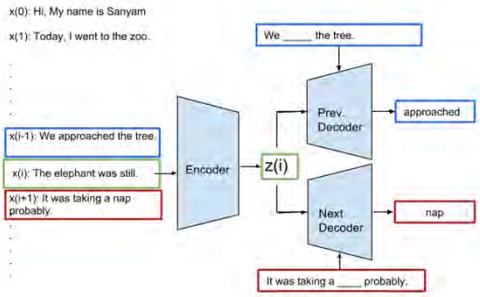
[Lobry, Marcos, Murray, Tuia, IGARSS 2019]





# Language part: skip-thoughts (Kiros et al., NIPS)

- We want to encode a sentence in a fixed length vector representation
- We use a method from the literature, the skip-thoughts model
- Given a sentence, it predicts the previous and following one in a text

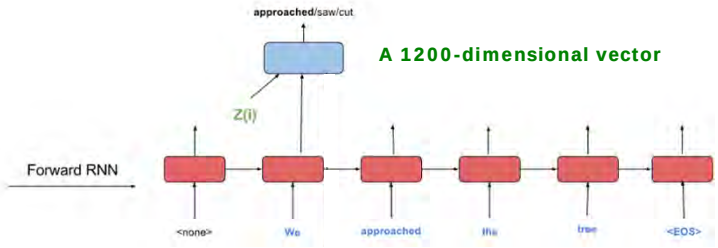



WAGENINGEN UNIVERSITY & RESEARCH

[http://sanyam5.github.io/my-thoughts-on-skip-thoughts/]

# Language part: skip-thoughts (Kiros et al., NIPS)

- We want to encode a sentence in a fixed length vector representation
- We use a method from the literature, the skip-thoughts model
- Given a sentence, it predicts the previous and following one in a text
- It is based on recurrent nets



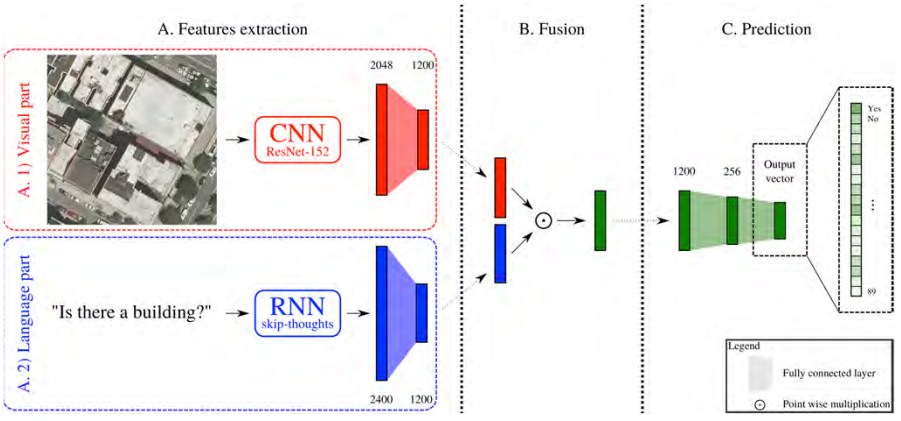

WAGENINGEN UNIVERSITY & RESEARCH

[http://sanyam5.github.io/my-thoughts-on-skip-thoughts/]

35

35

# Remote sensing visual question answering (RSVQA)




WAGENINGEN UNIVERSITY & RESEARCH

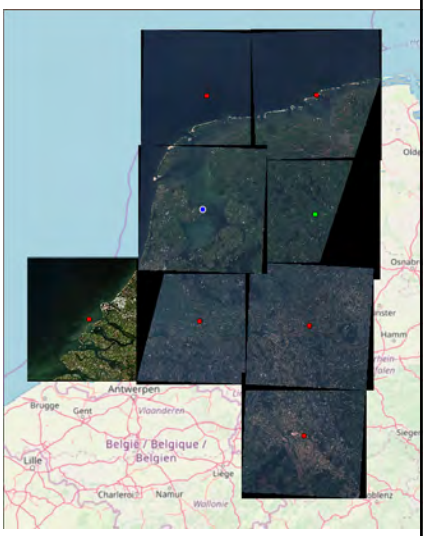
[Lobry, Marcos, Murray, Tuia, IGARSS 2019]

36

36

## How do we train this monster?

- We created a dataset of
  - Sentinel-2 images (RGB)
    - 9 scenes
    - 772 tiles (256 x 256)
  - OpenStreetMap layers
  - Covers the whole Netherlands




WAGENINGEN UNIVERSITY & RESEARCH

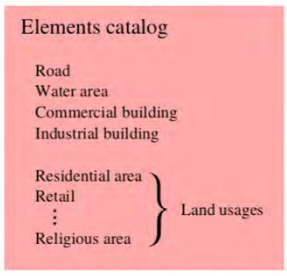
37

37

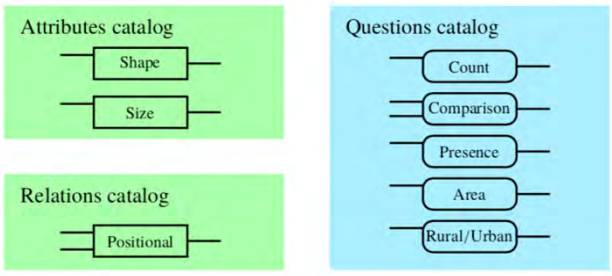
## How do we train this monster?

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

Check on OSM




Random generator




WAGENINGEN UNIVERSITY & RESEARCH

38

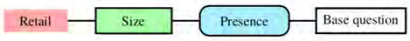
38

# 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?"

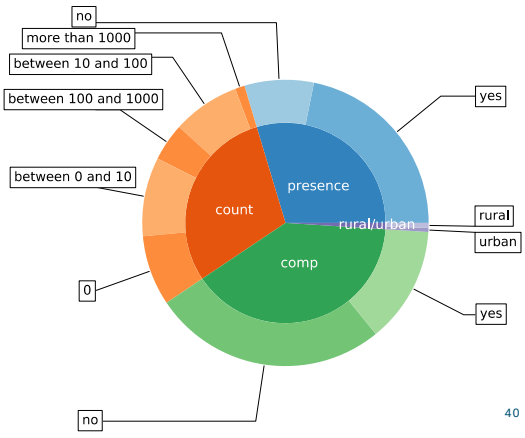

WAGENINGEN UNIVERSITY & RESEARCH

39

39

# How do we train this monster?

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

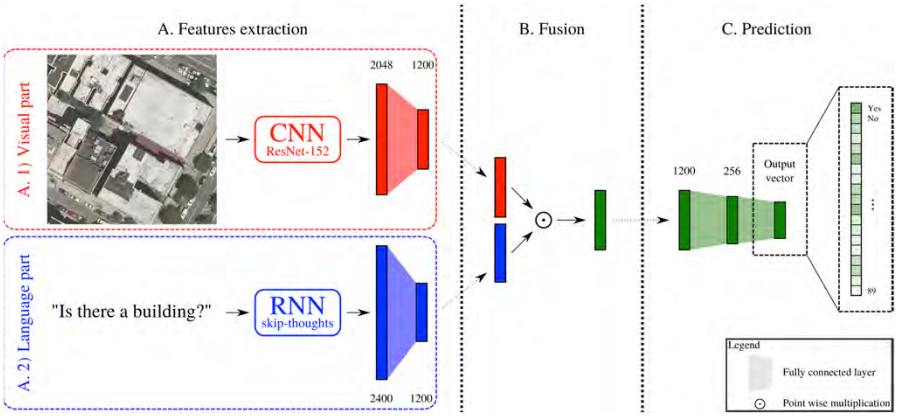



WAGENINGEN UNIVERSITY & RESEARCH

40

40

## Remote sensing visual question answering (RSVQA)




WAGENINGEN UNIVERSITY & RESEARCH

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

41

41

## Results – Sentinel 2

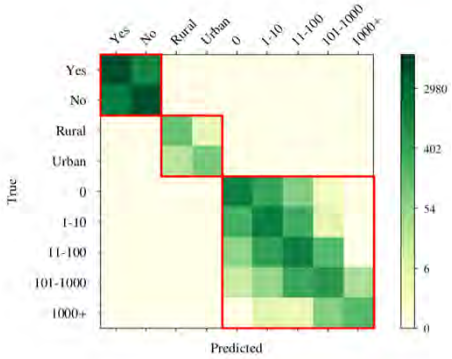
79% overall accuracy!

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

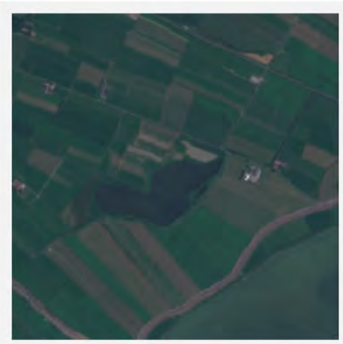



WAGENINGEN UNIVERSITY & RESEARCH

42

42

## Results – Sentinel 2



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

(g) LR, test set

| Are there less buildings than water areas? | |
|---|---|
| Ground truth: No | Prediction: No |

(j) LR, test set

WAGENINGEN UNIVERSITY & RESEARCH



43

## Results – Sentinel 2

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

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

(i) LR, test set

WAGENINGEN UNIVERSITY & RESEARCH



44

## 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

WAGENINGEN UNIVERSITY & RESEARCH

45

45

## 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

WAGENINGEN UNIVERSITY & RESEARCH

46

46

## Results – Sentinel 2



(i) LR, test set

WAGENINGEN UNIVERSITY & RESEARCH

47

47

## RSVQA - Summary

- 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:

Université Bretagne Sud ubs : cnes CENTRE NATIONAL D'ÉTUDES SPATIALES

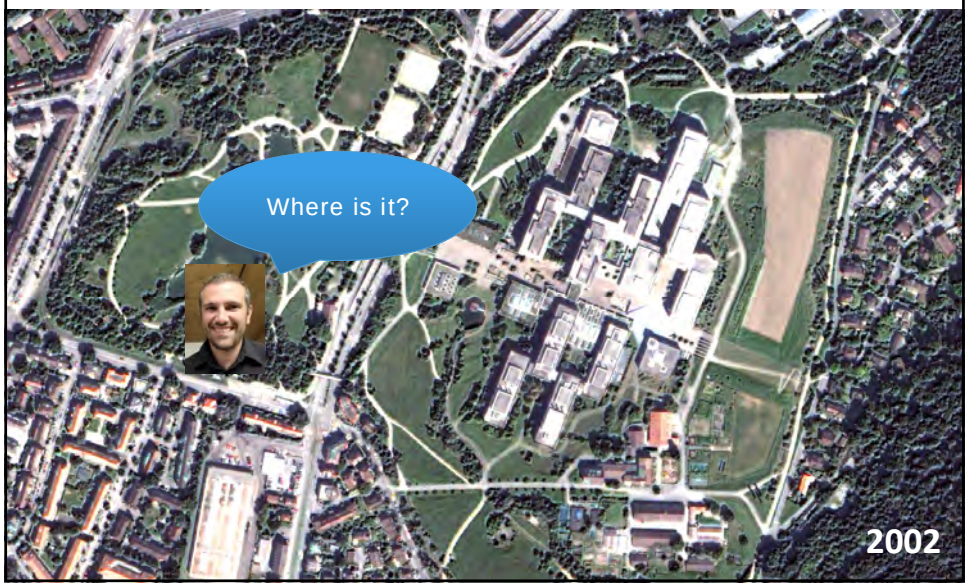
WAGENINGEN UNIVERSITY & RESEARCH

48

48


49

## One model to answer them all!

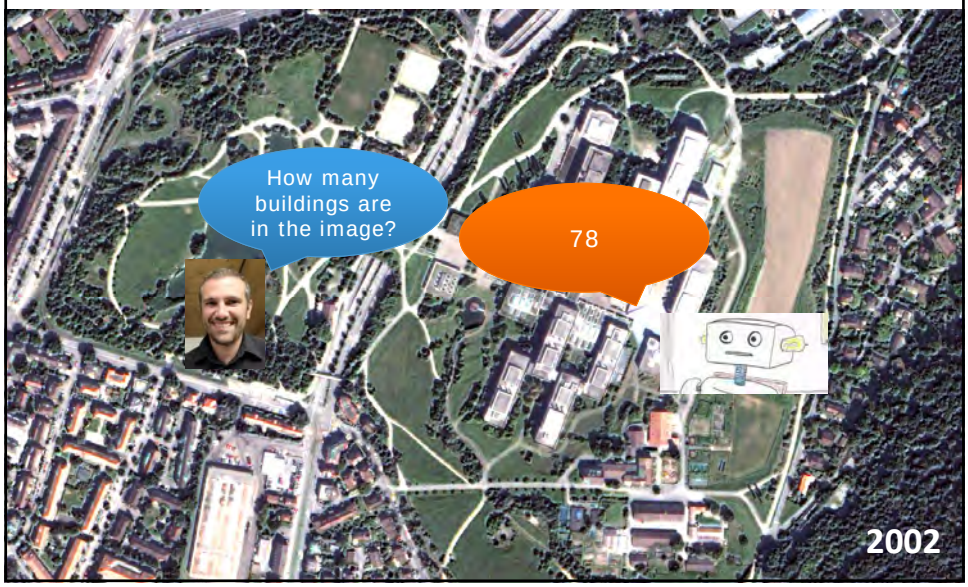



50

## One model to answer them all!

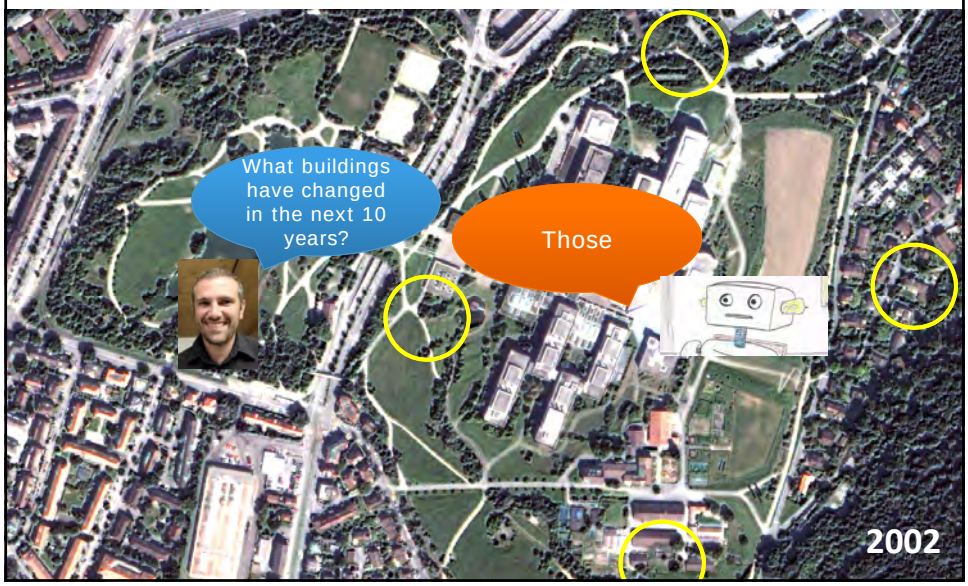

51

## One model to answer them all!



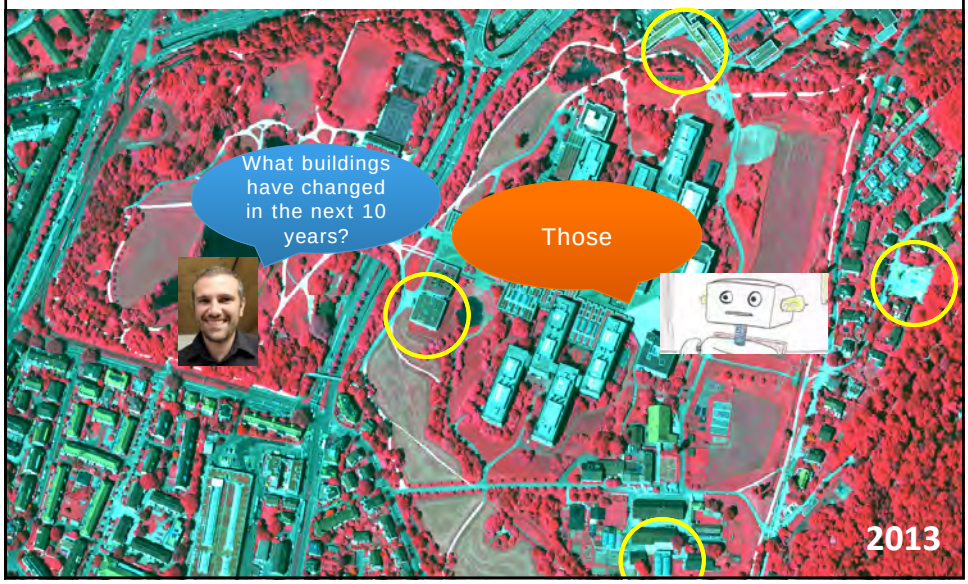

52

## One model to answer them all!




53

## One model to answer them all!




54

# RSVQA - Summary

- RSQA is a new task. We could not do it before deep learning.
- It is far from solved.
- We are putting data and codes online soon.


55

# Wrapping up!

- Machine / deep learning are great tools for remote sensing.
- They allow pushing the state of art!
- But…

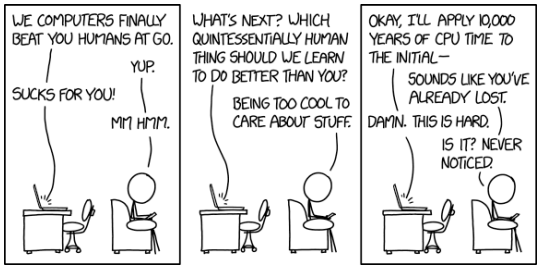

WAGENINGEN UNIVERSITY & RESEARCH

56

56

# Wrapping up!

- Machine / deep learning are great tools for remote sensing.
- They allow pushing the state of art!
- Most importantly: they **allow to imagine new uses of geodata**
- **It's hard, but this is where we can make a difference.**

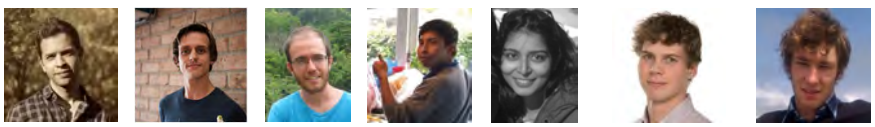



WAGENINGEN UNIVERSITY & RESEARCH

57

57

# Thanks!



**Contact me!**

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

**devis.tuia@wur.nl**

**We are organizing IGARSS 2021 in the lowlands!**

- Want to be an exhibitor?
- Want to sponsor?
- Want to participate?

igarss BRUSSELS 2021

WAGENINGEN UNIVERSITY & RESEARCH

58

58