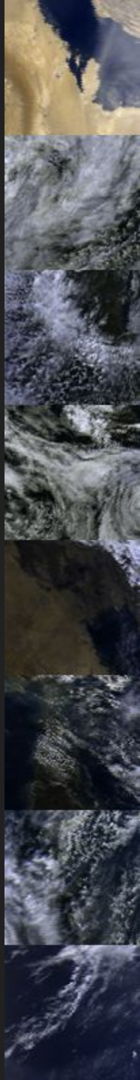
Volumetric Segmentation of Clouds from Multi-Angle Satellite Imagery

Sean Foley, Kirk Knobelspiesse, Judy Hoffman, James Hays



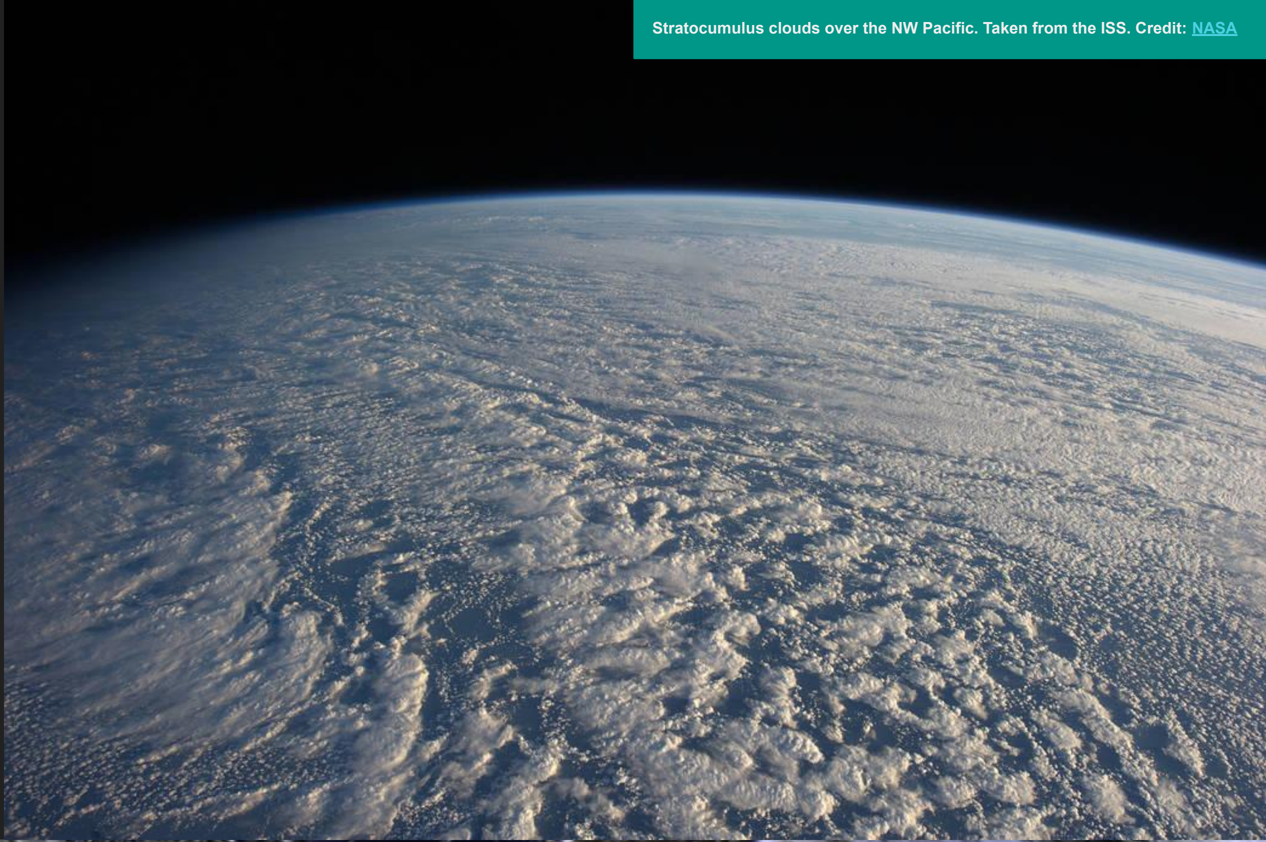
About Me

- ML PhD 5th year, co-advised: James Hays & Judy Hoffman
- Recent Research Interests:
  - applying CV to satellite data
  - volumetric segmentation
  - neural rendering
- Hobbies:
  - piano
  - video games
  - hanging out with my cat

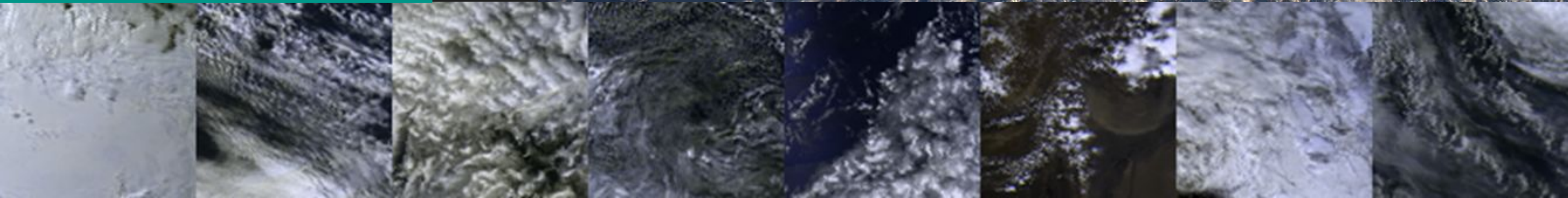
Contact: seanremy@gmail.com

What are clouds?

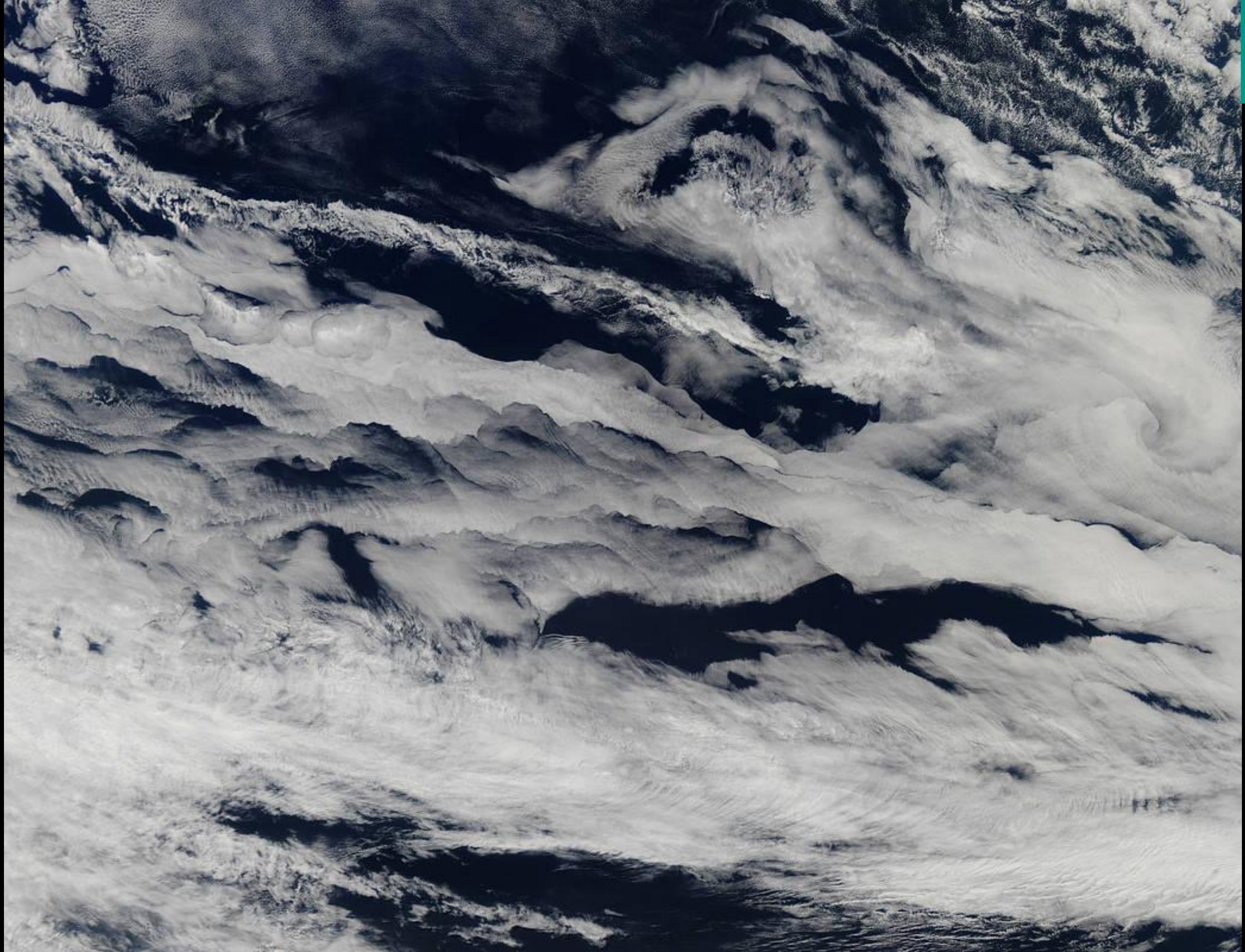
- floating water drops or ice crystals
- condensation nuclei
- many kinds



Stratocumulus clouds over the NW Pacific. Taken from the ISS. Credit: NASA

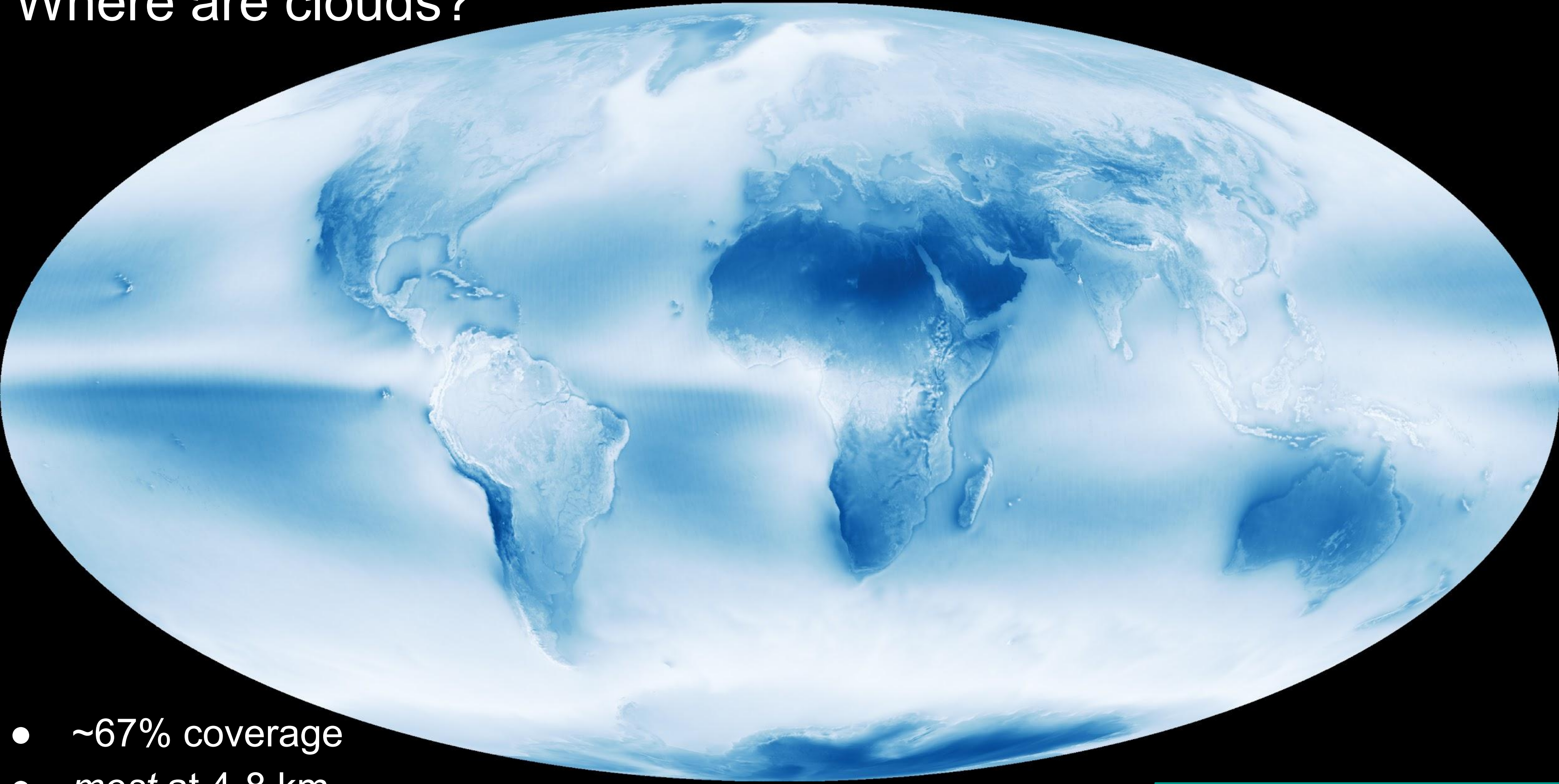


Data from PARASOL / POLDER (CNES). Interpolated false color.



Clouds over the southern Indian Ocean, from MODIS aboard the Aqua satellite. Credit: NASA

Where are clouds?

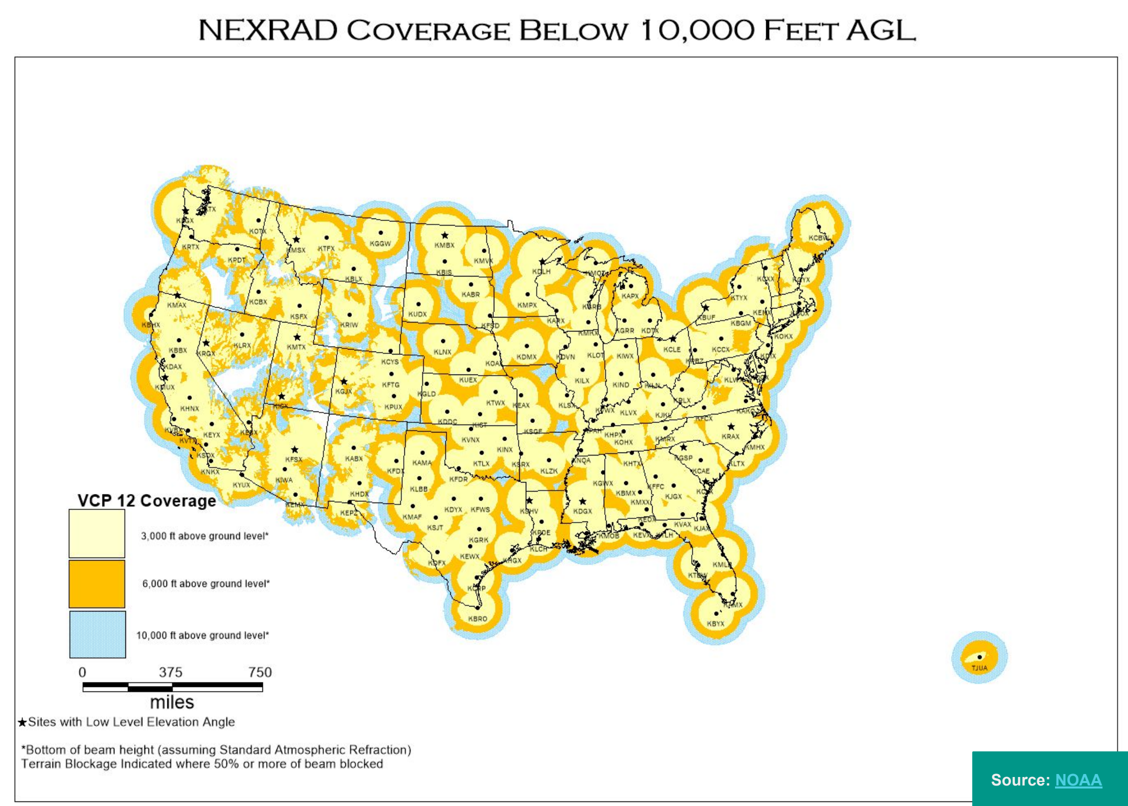


- ~67% coverage
- *most* at 4-8 km

Global cloud fraction, Jul 2002 - Apr 2015, from MODIS data.

Ground-based RADAR

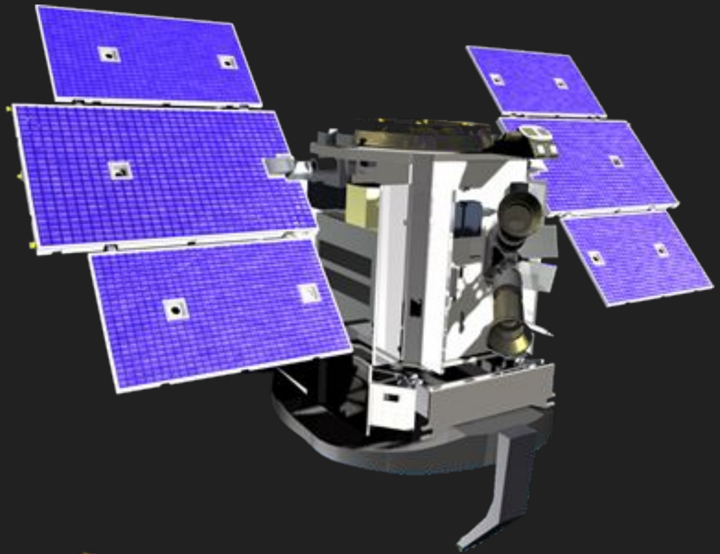
- constant
- real-time
- fixed
- no ocean

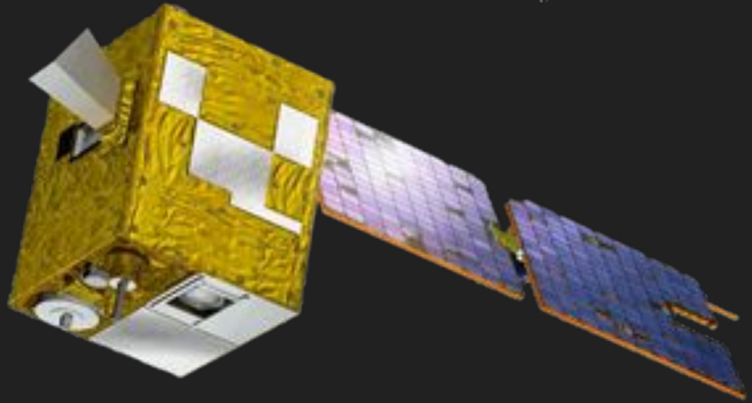


Source: [NOAA]

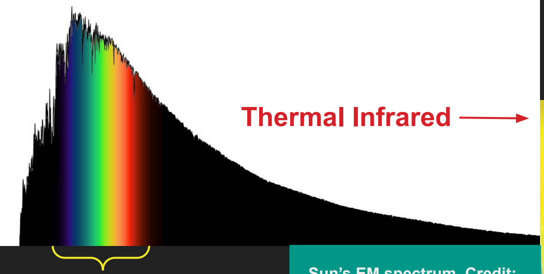
Finding clouds from space

- active sensors
  - RADAR
  - LIDAR
  - accurate!
- passive sensors
  - ‘cameras’ (sort of)
  - multi-angle
  - wide coverage!





Why care about clouds?



Sun's EM spectrum. Credit: Dr. Chris Baird

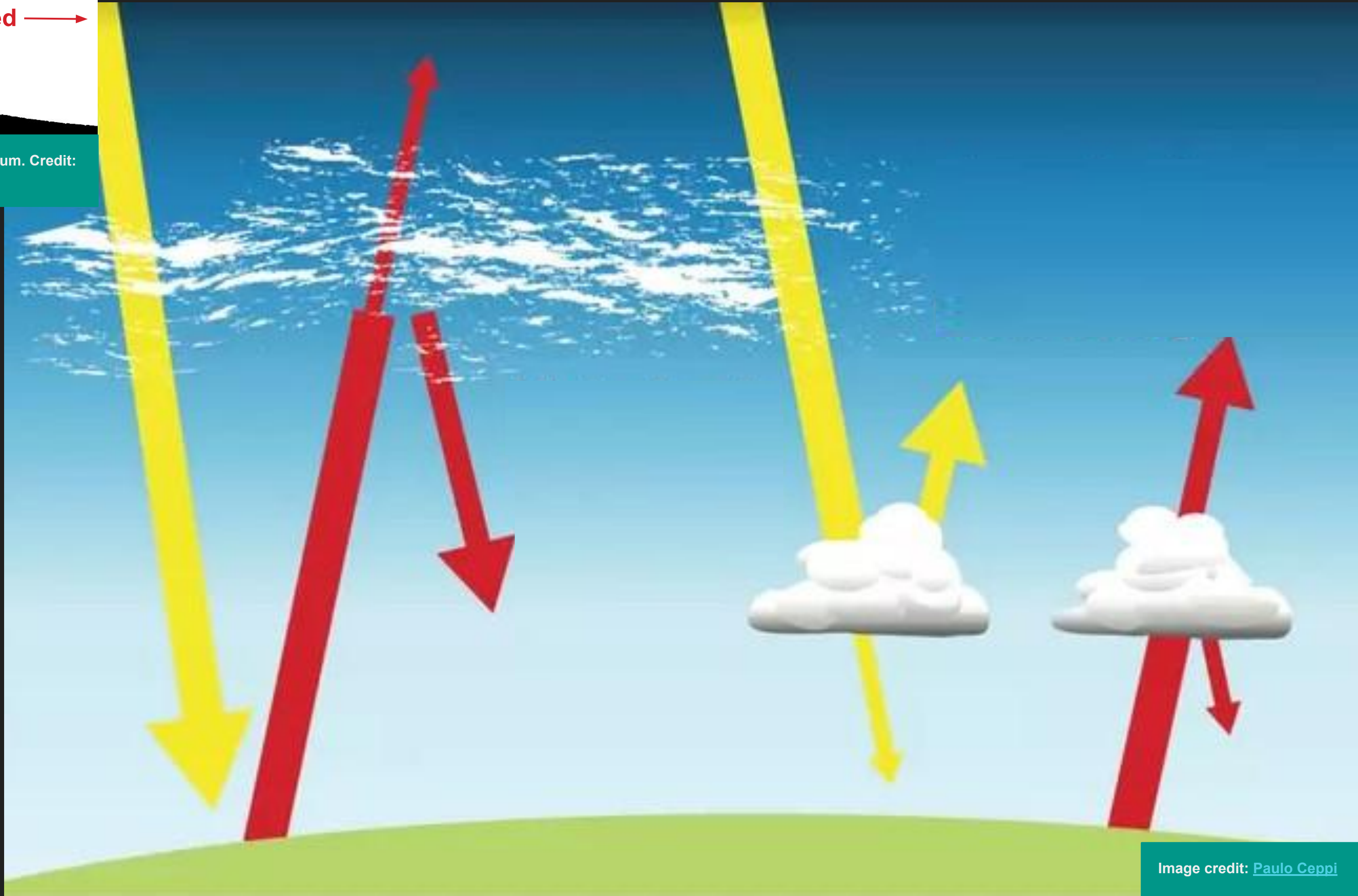
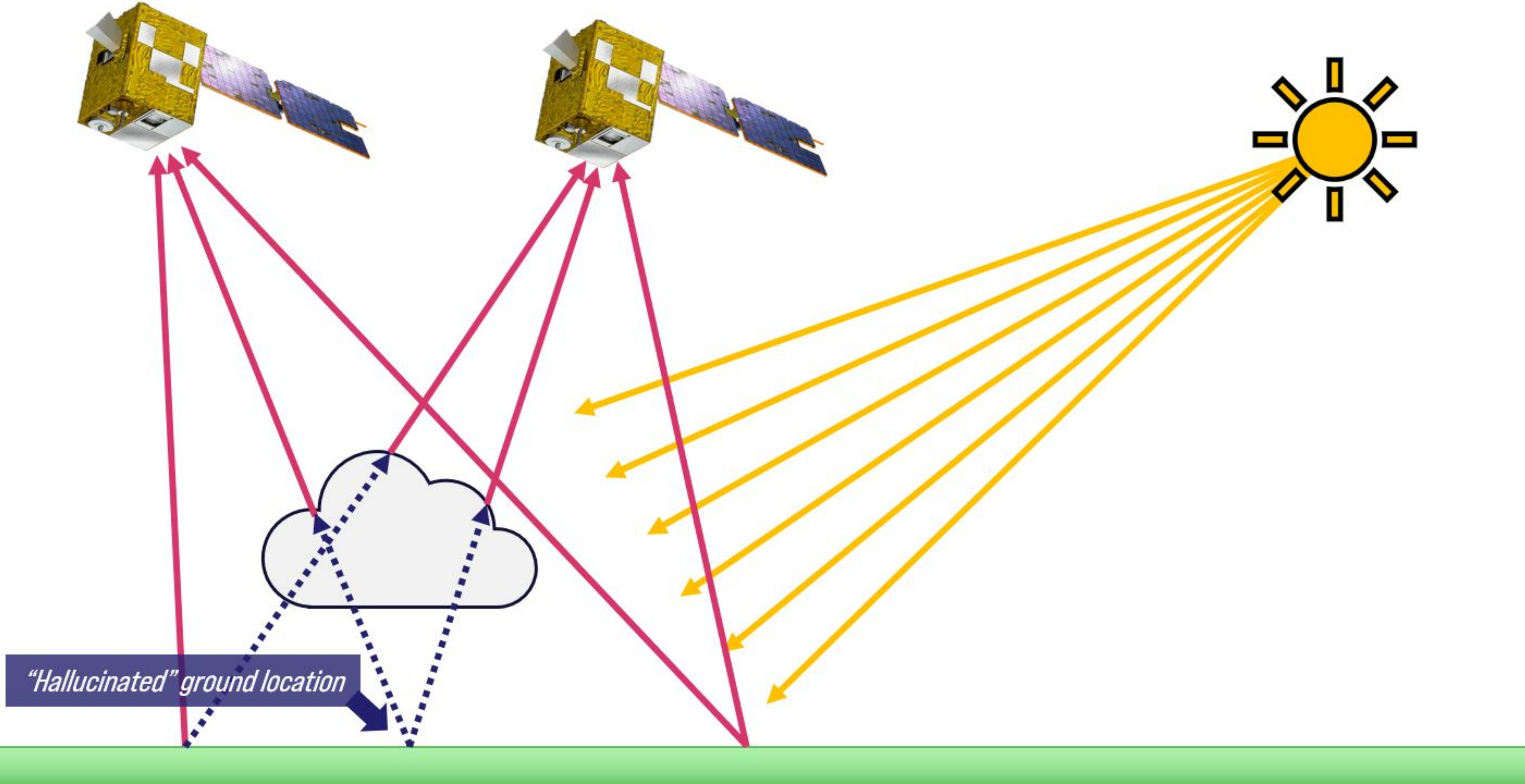
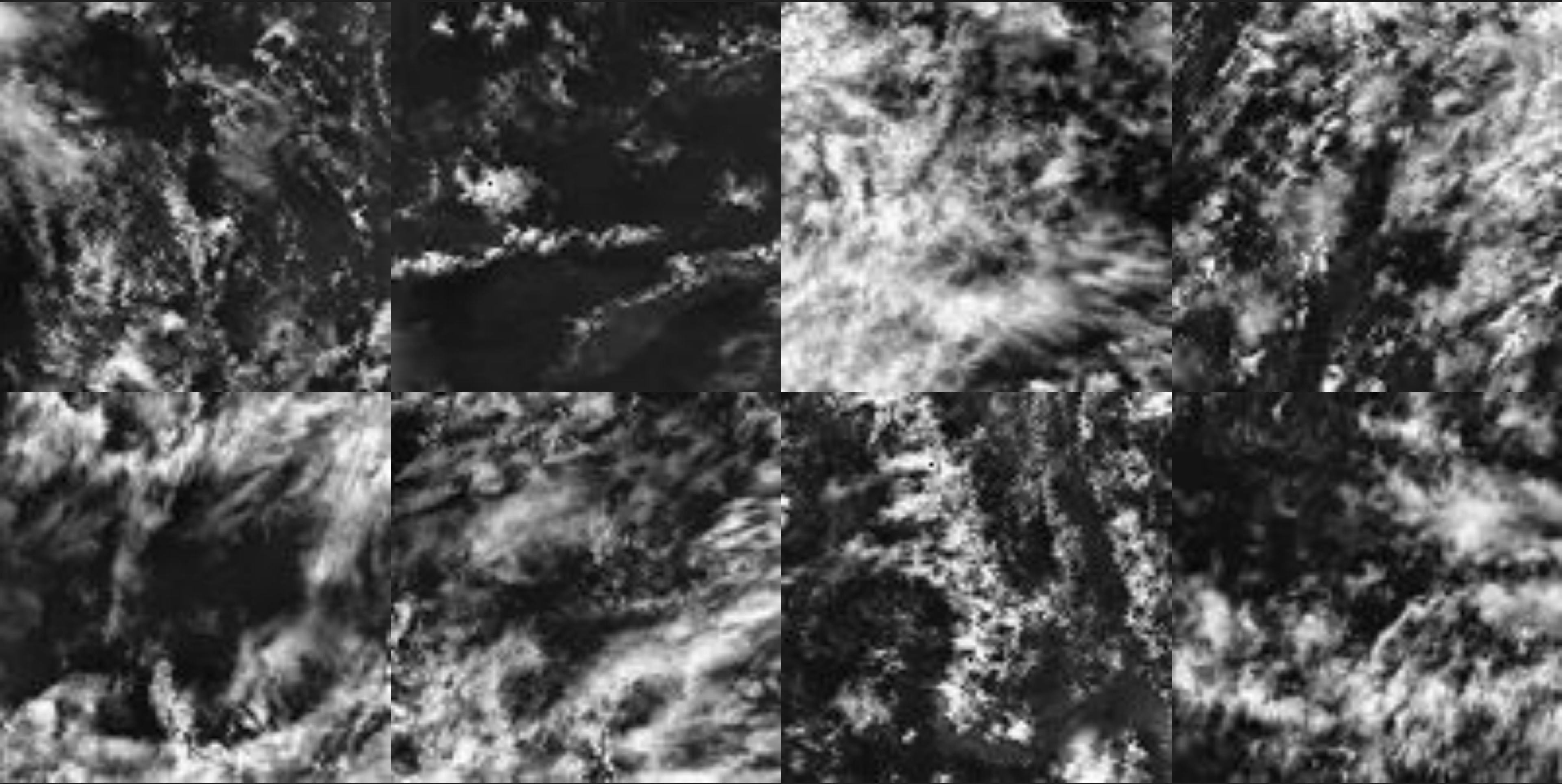


Image credit: Paulo Ceppi


"Hallucinated" ground location

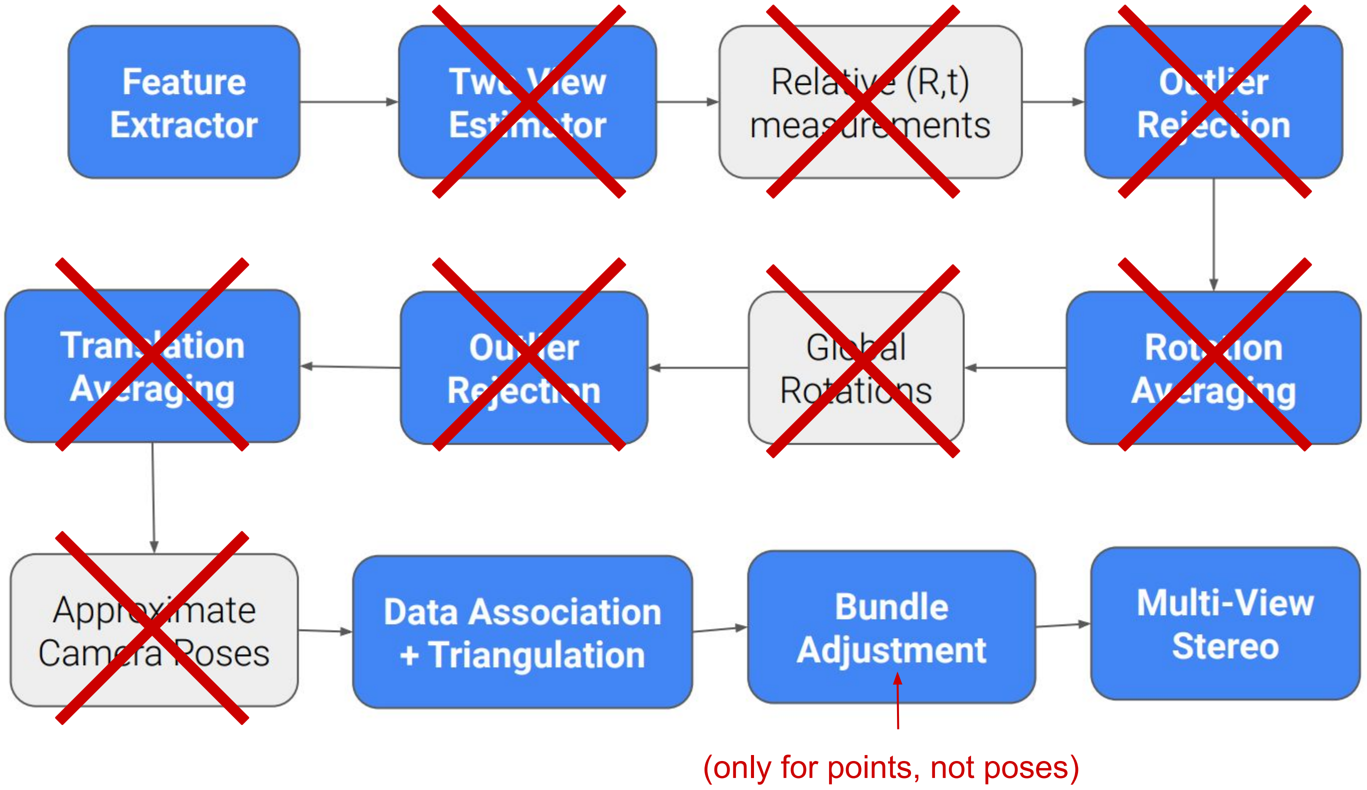
Warning: next slide has some flickering



Goal:
3D cloud segmentation in passive sensor data

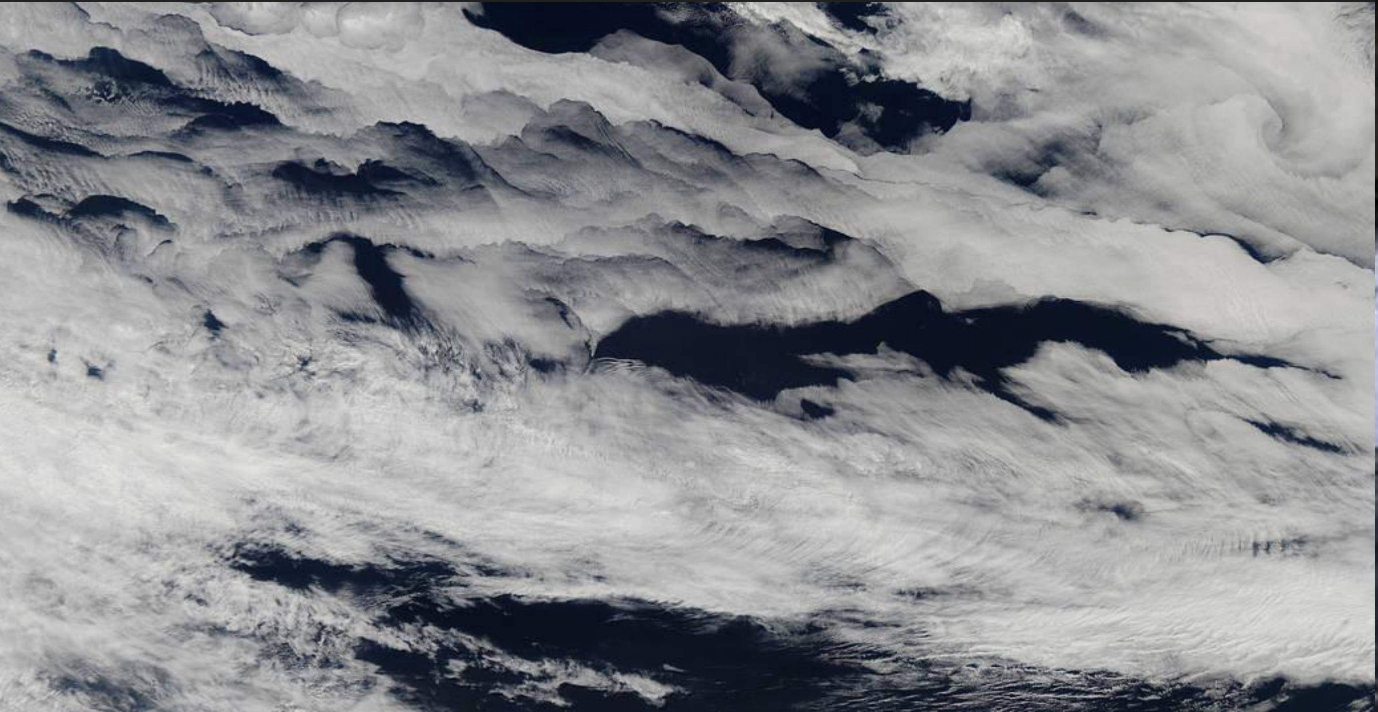
Questions?

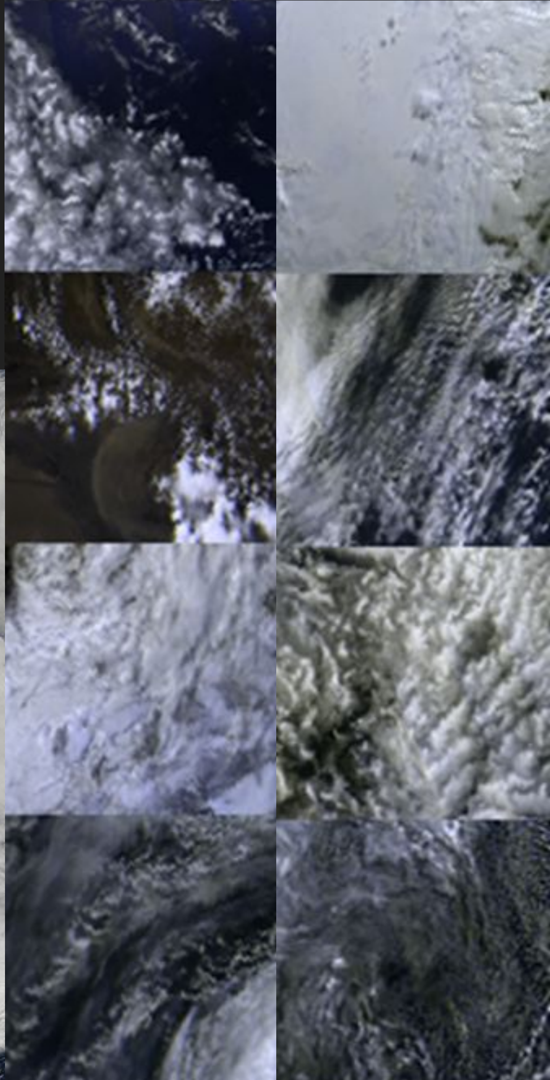
What if we already know camera poses?



Problem:

Clouds lack good keypoints!





Problem:

Thin clouds can be *translucent*

A pixel may correspond to *multiple* 3D locations

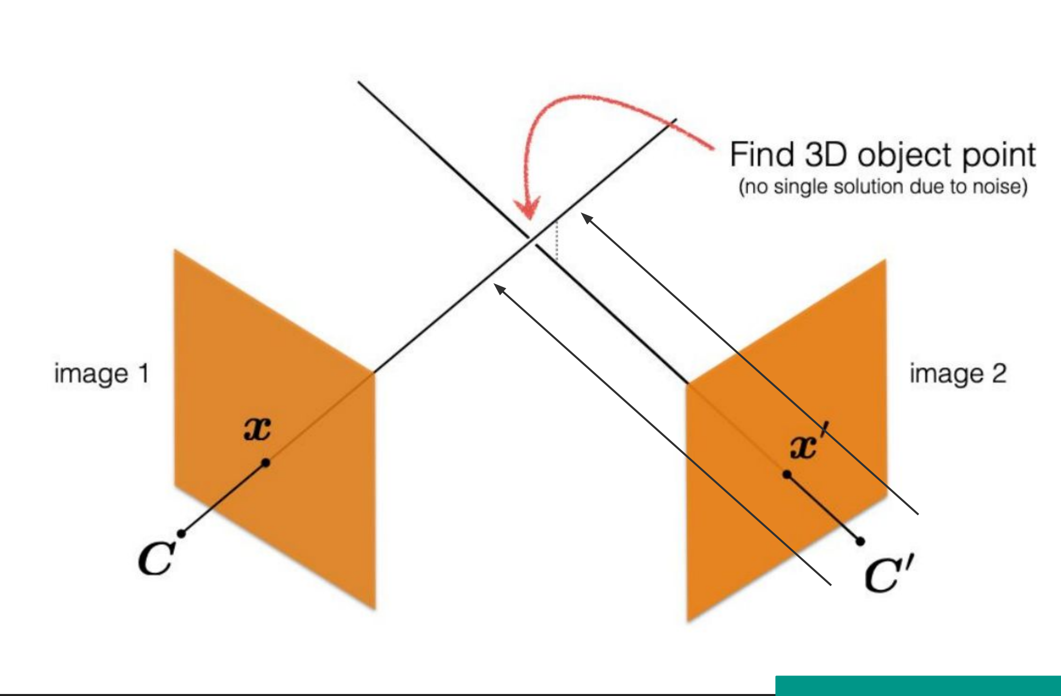


Image credit: [Dr. John Lambert]

So, now what?

Throw deep learning at the problem!

What about training data?

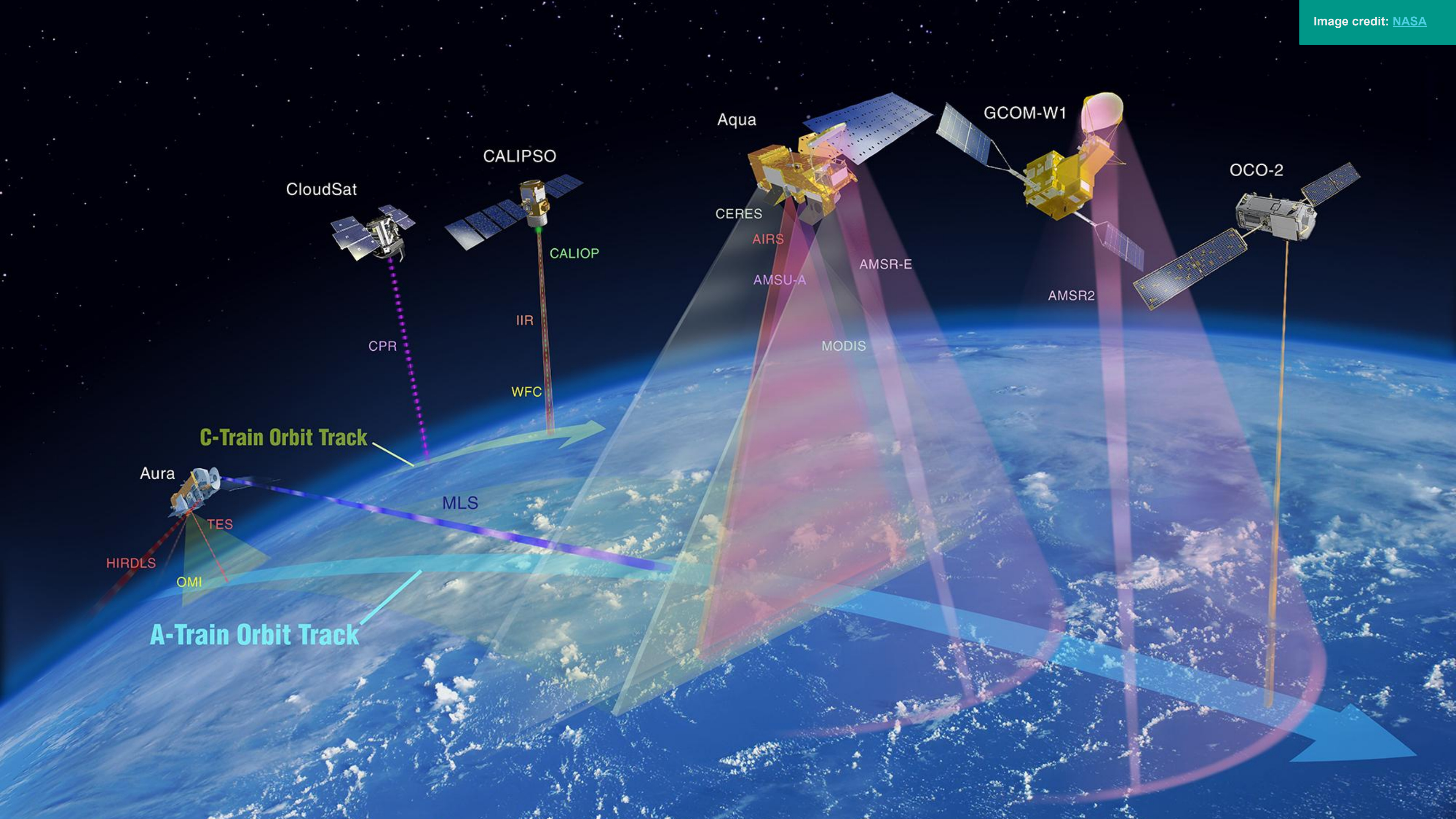
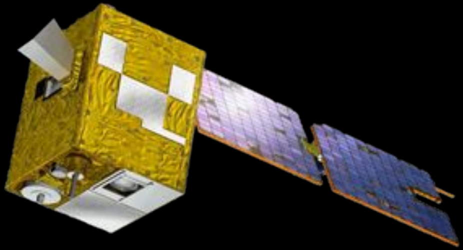
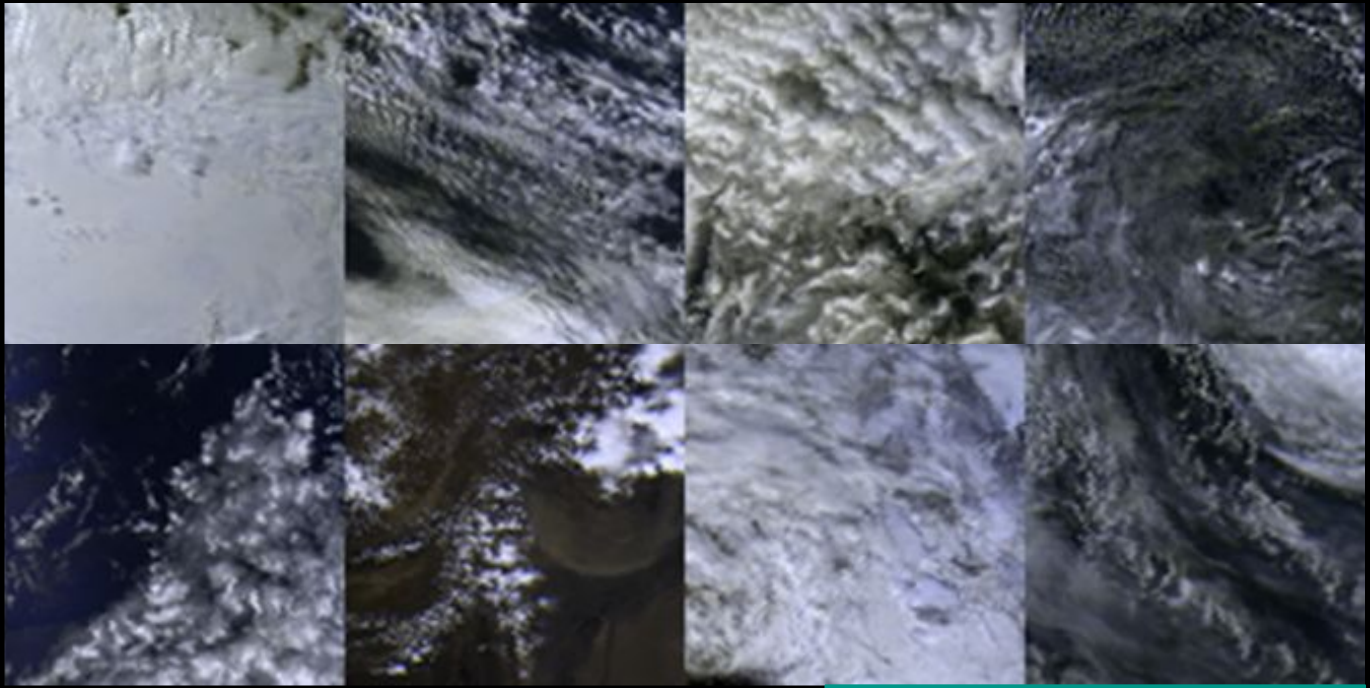


Image credit: NASA

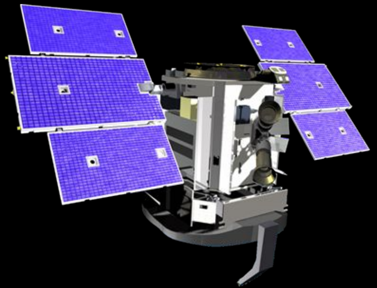
PARASOL / POLDER



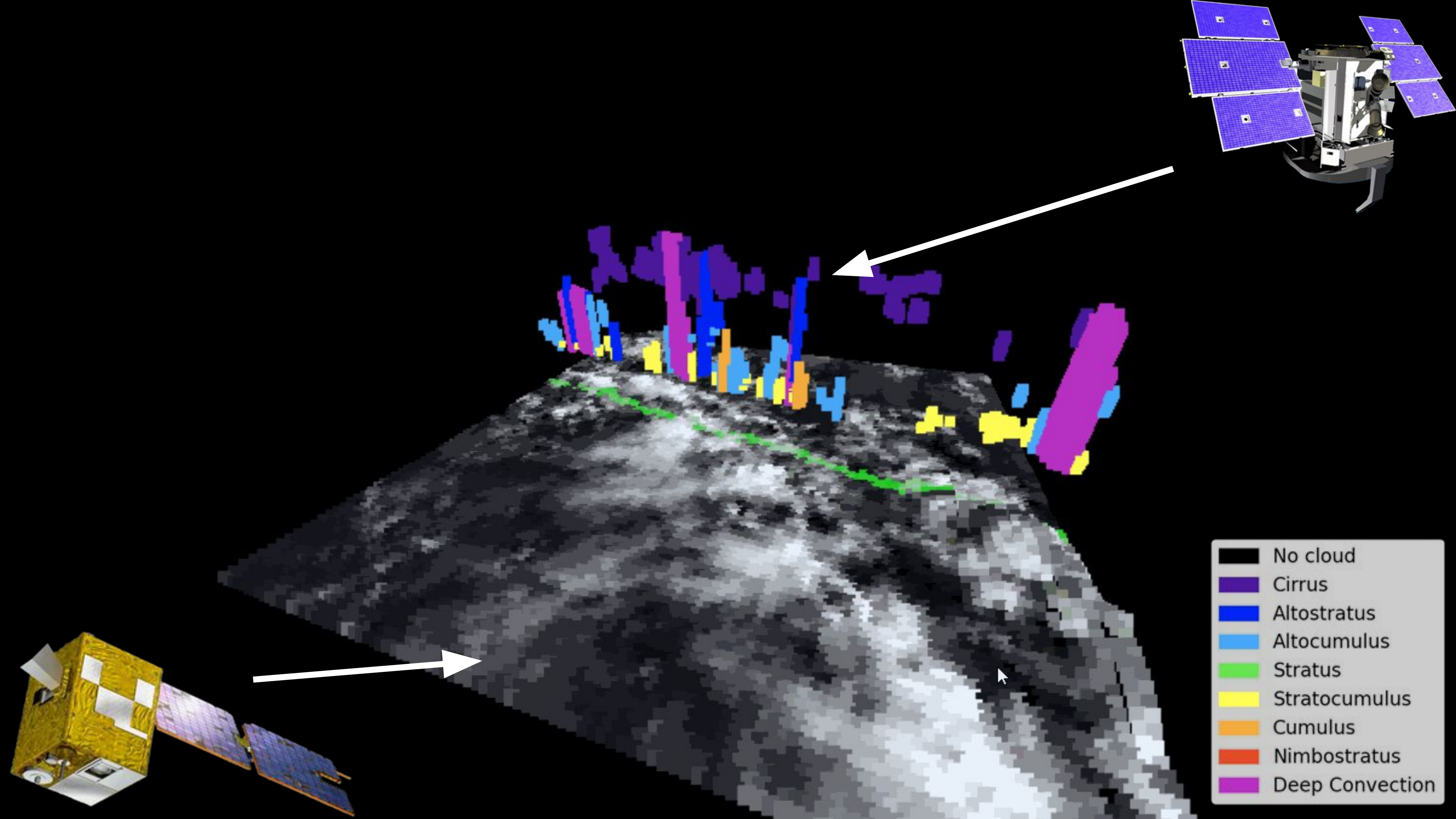


Data from PARASOL / POLDER (CNES). Interpolated false color.

2B-CLDCLASS



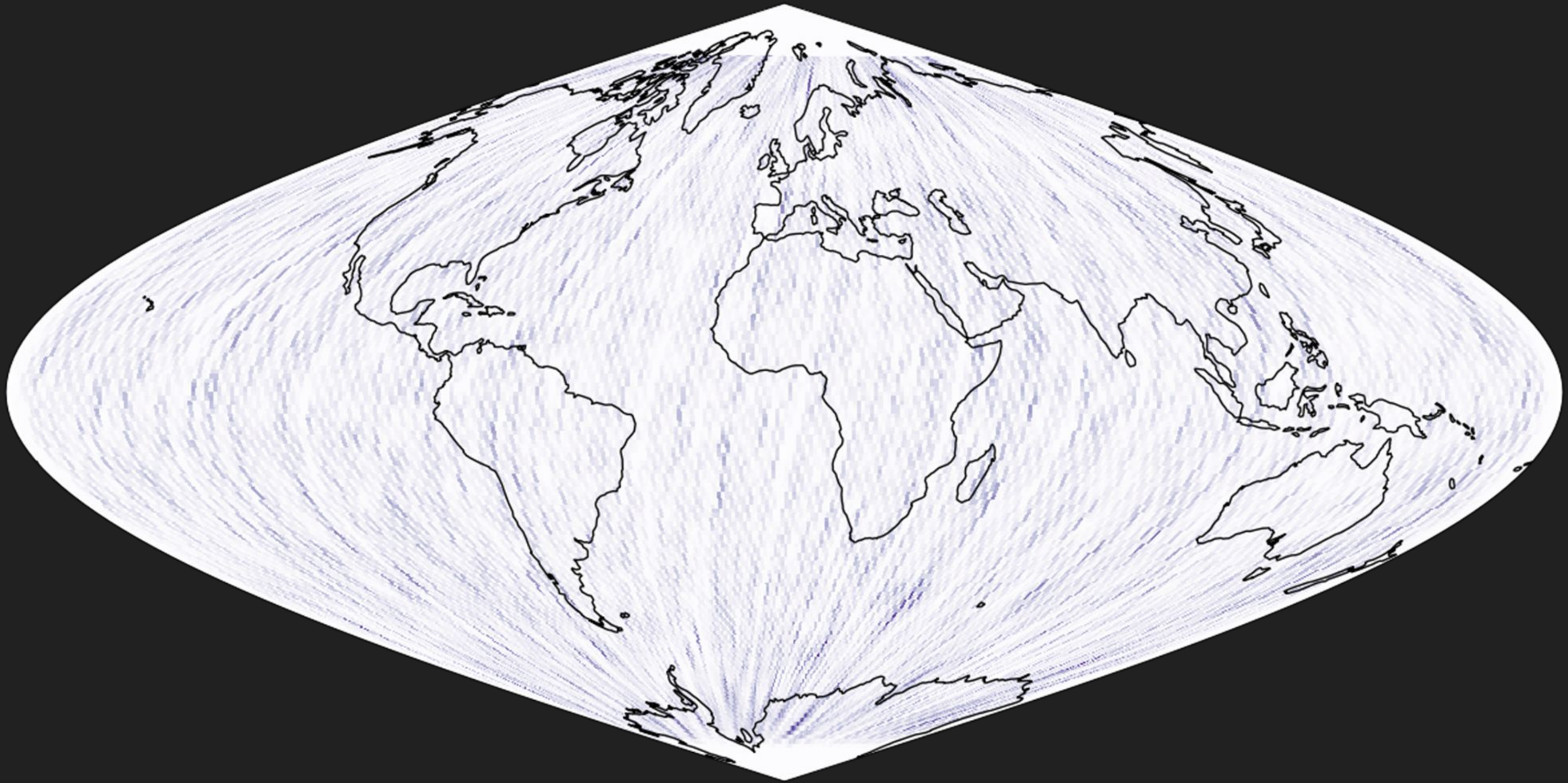

2009 Aug 19 (231) | 2B-CLDCLASS | Granule 17611
18
17:25:55 17:22:44 UTC
Cloud Type:
Deep Ns Cu Sc St Ac As High Missing
CIRA CloudSat DPC


No cloud
Cirrus
Altostratus
Altocumulus
Stratus
Stratocumulus
Cumulus
Nimbostratus
Deep Convection

A-Train Cloud Segmentation Dataset (ATCS)

- 20k train/val instances
- Python API
- Pytorch training/evaluation code

https://github.com/seanremy/atrain-cloudseg

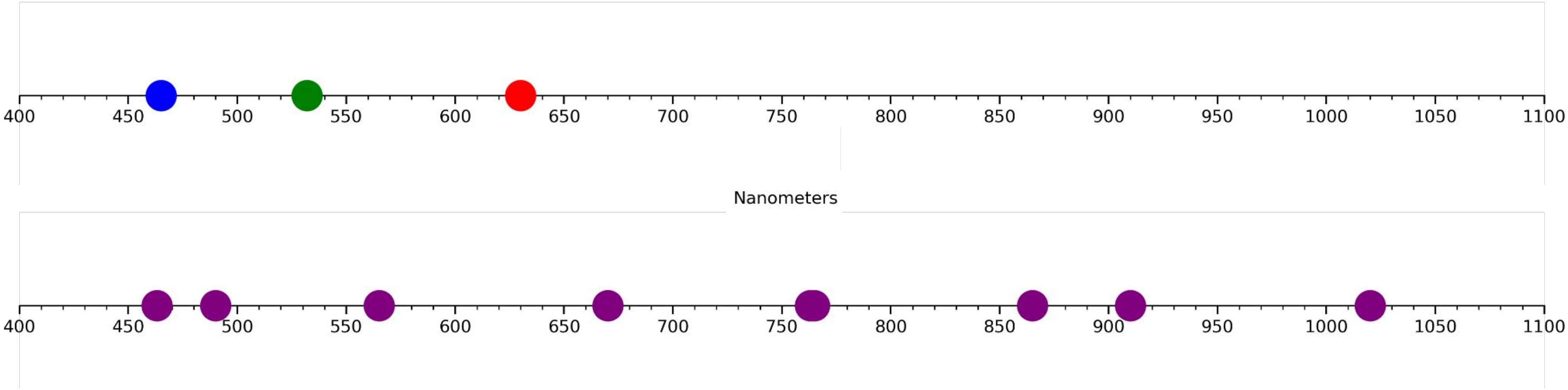



2007
2008
2009
M
W
F
S
Jan
Feb
Mar
Apr
May
Jun
Jul
Aug
Sep
Oct
Nov
Dec

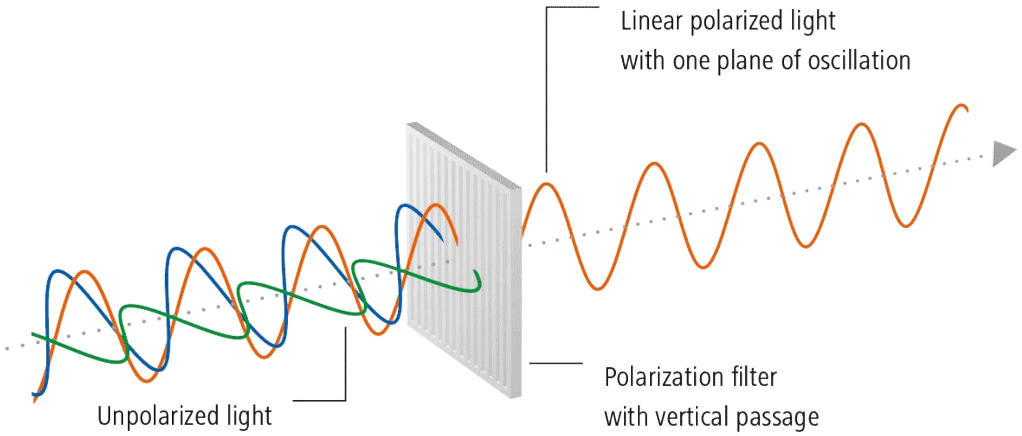
Multi-angular imagery

- pixel -> surface location (latitude longitude)
- 16 angled cameras
- up to 13 for one pixel

POLDER's spectra



Polarization



Why?

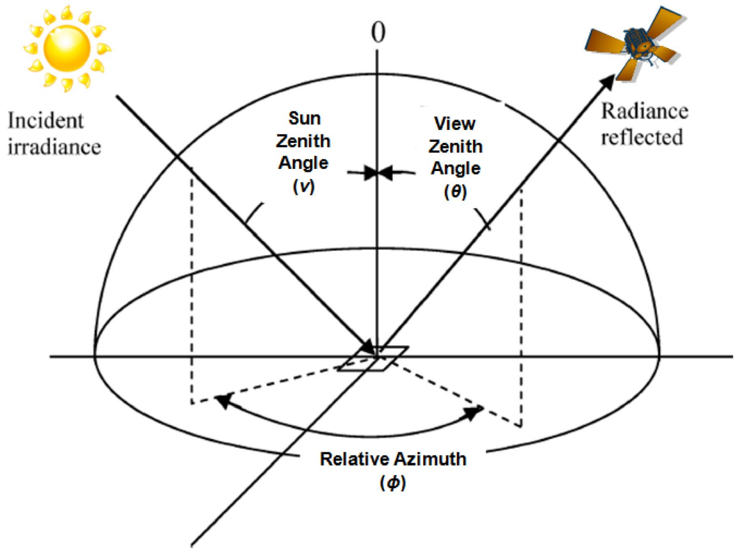
- clouds vs bright surfaces
- clouds vs icy clouds

Geometry

for each pixel:

- solar azimuth angle
- for each angle:
    - solar zenith angle
    - view zenith angle
    - view azimuth angle

Azimuth: 8 overlapping bins



Questions?

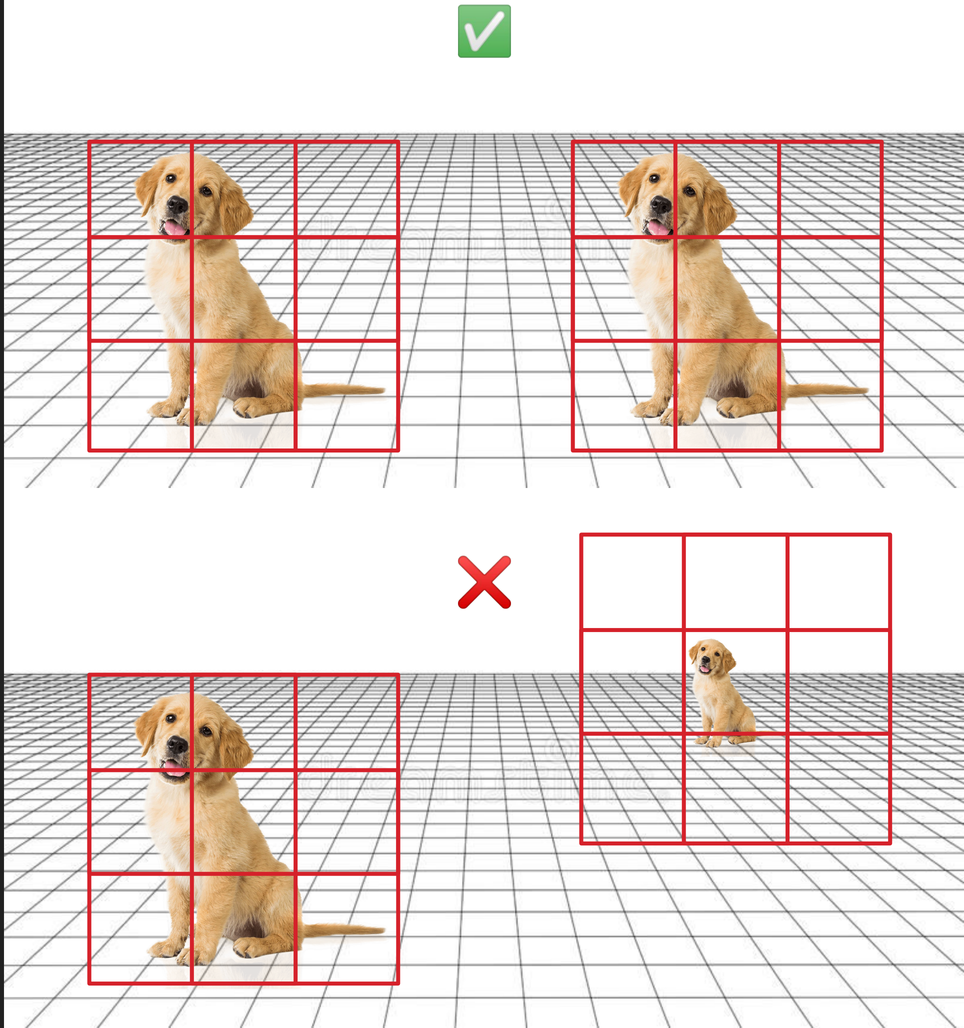
RGB: 3

ATCS: 289

That’s alotta channels!

Translational Invariance

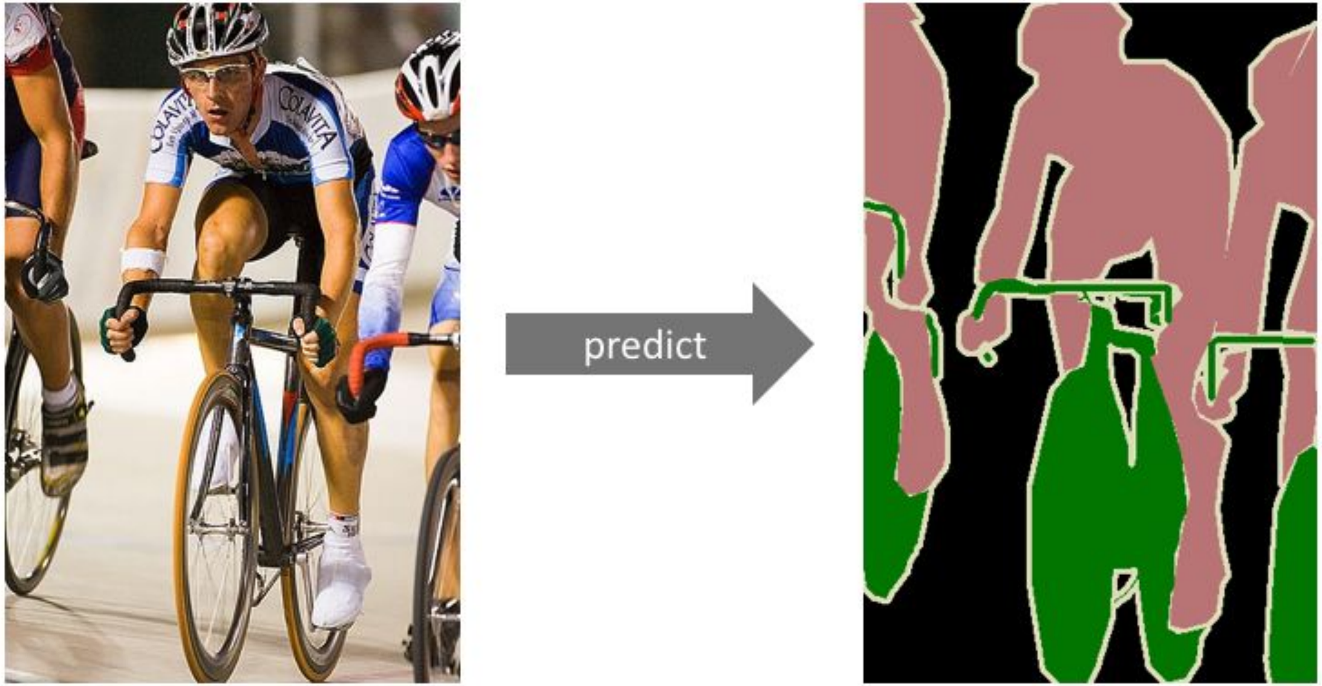
- Motivation for convolutions
- things that break it:
    - scale (distance)
    - lighting
    - rotation



What output representation?

Source: PASCAL-VOC

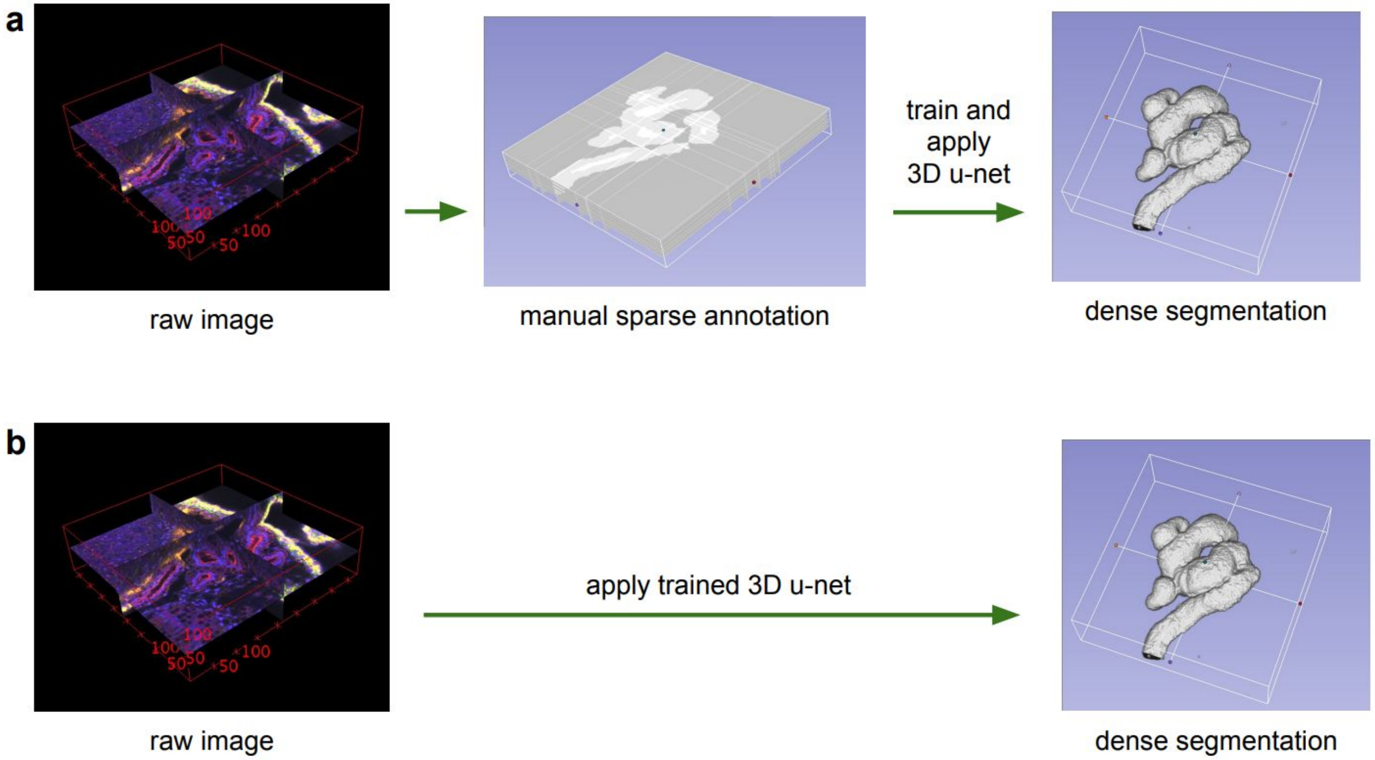
Semantic Segmentation



Person
Bicycle
Background

Image credit: 3D U-Net

Volumetric Segmentation



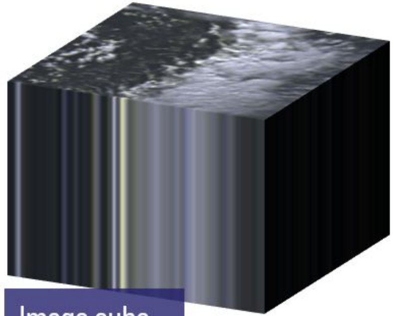
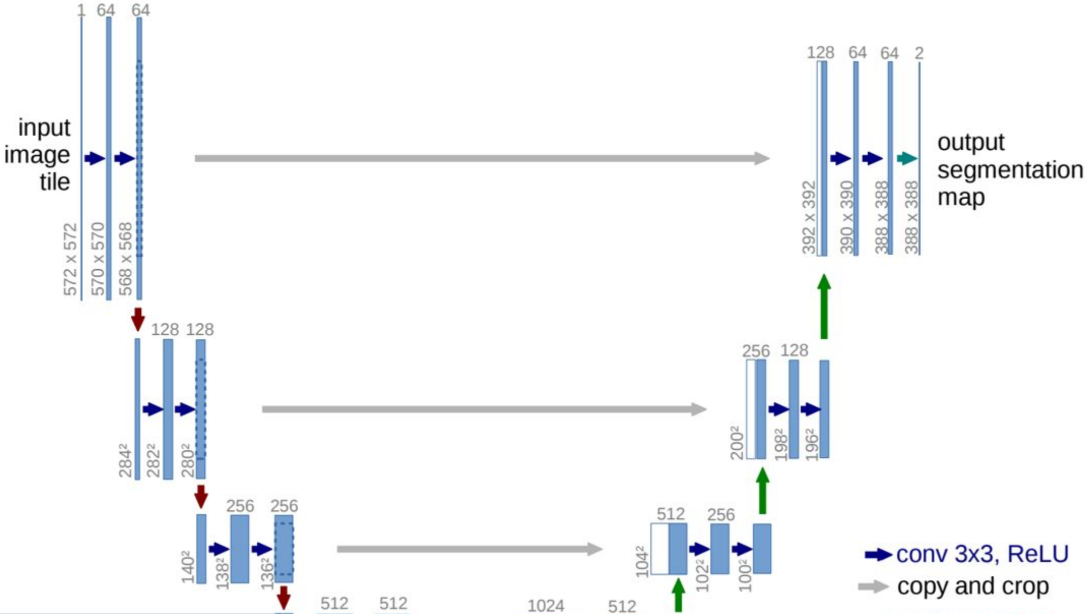
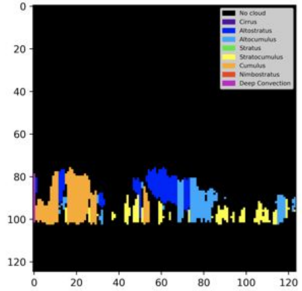


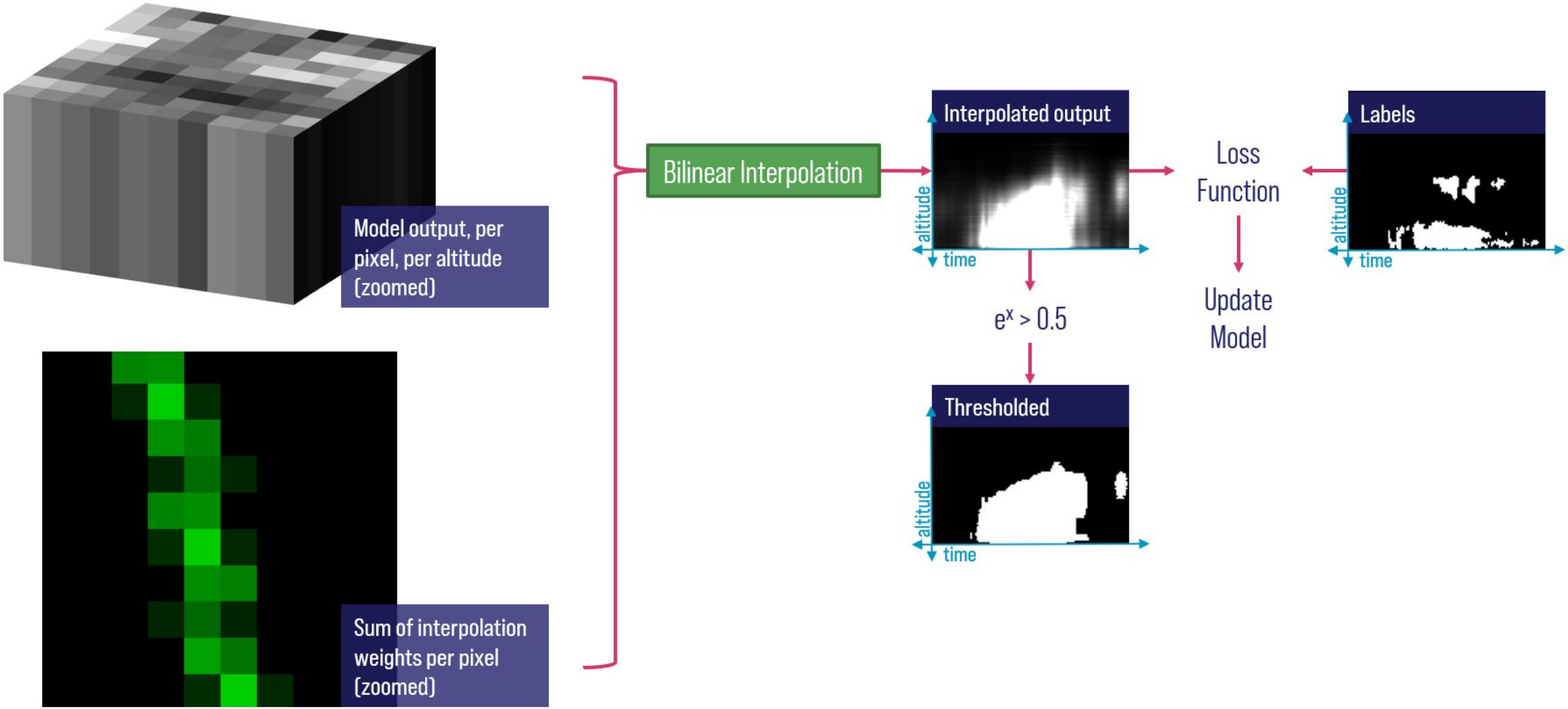
Image cube



Architecture diagram of a U-Net. (Ronneberger et al. 15)

Interpolation & thresholding





Model output, per pixel, per altitude (zoomed)
Sum of interpolation weights per pixel (zoomed)
Bilinear Interpolation
Interpolated output
altitude
time
$e^x > 0.5$
Thresholded
altitude
time
Loss Function
Update Model
Labels
altitude
time

Training Details

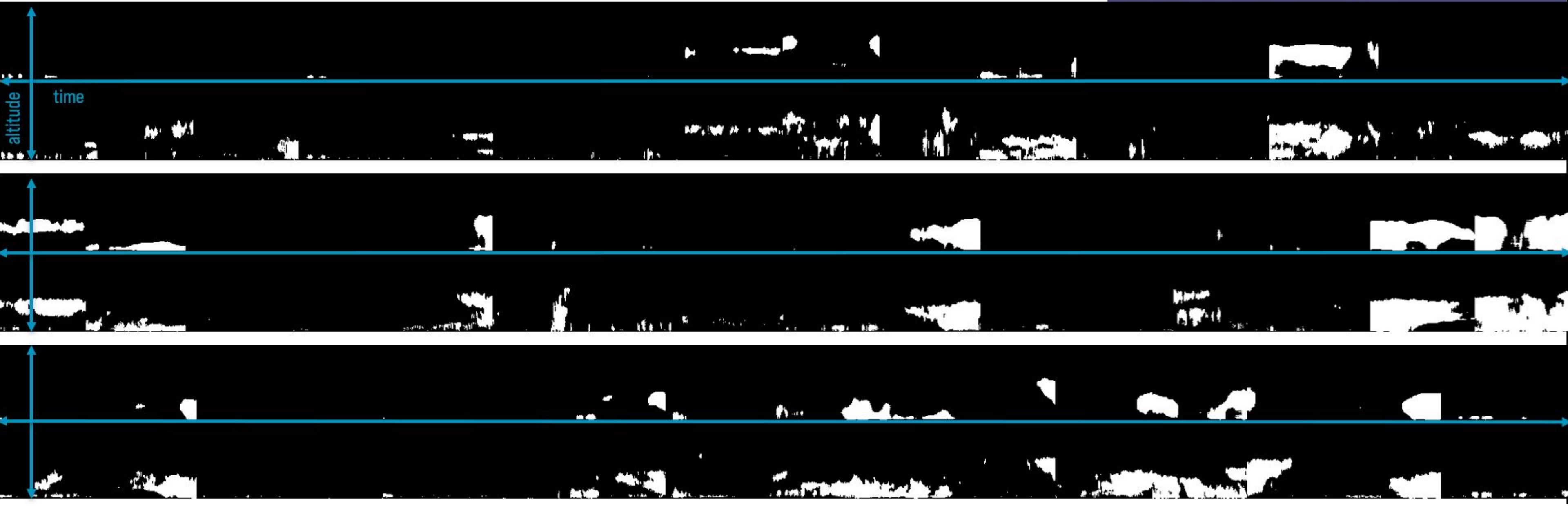
- U-Net
- 50 epochs
- Binary Cross Entropy
- Adam

Binary Cross Entropy

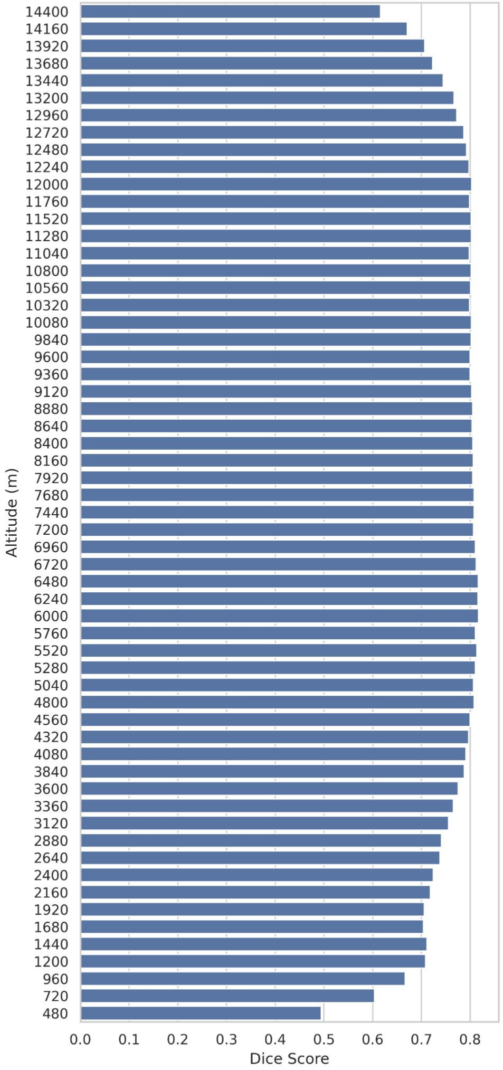
$$\ell(x, y) = L = \{l_1, \dots, l_N\}^\top, \quad l_n = -w_n \left[y_n \cdot \log x_n + (1 - y_n) \cdot \log(1 - x_n) \right]$$

Questions?

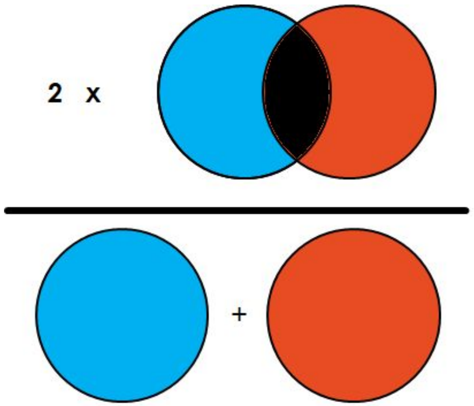
Qualitative results on multiple patches. Top of each row: predicted. Bottom: true profile.



<3D qualitative example>


14400
14160
13920
13680
13440
13200
12960
12720
12480
12240
12000
11760
11520
11280
11040
10800
10560
10320
10080
9840
9600
9360
9120
8880
8640
8400
8160
7920
7680
7440
7200
6960
6720
6480
6240
6000
5760
5520
5280
5040
4800
4560
4320
4080
3840
3600
3360
3120
2880
2640
2400
2160
1920
1680
1440
1200
960
720
480
Altitude (m)
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
Dice Score

$$Dice = \frac{2 \times TP}{(TP + FP) + (TP + FN)}$$



Architecture	# Params.	Dice Score
Single-Pixel	1.3E5	75.8
Simple ConvNet	6.88E5	77.5
U-Net	2.76E8	78.1

ϕ_{sol}	ϕ_{rel}	θ_{sol}	θ_{sen}	Dice Score
				58.4
✓	✓	✓	✓	78.1

Some other findings

- more angles = better
  - diminishing returns
- 3D convs don’t work… yet
- *some* geometry essential

Can we do better?

Directions

- different architecture

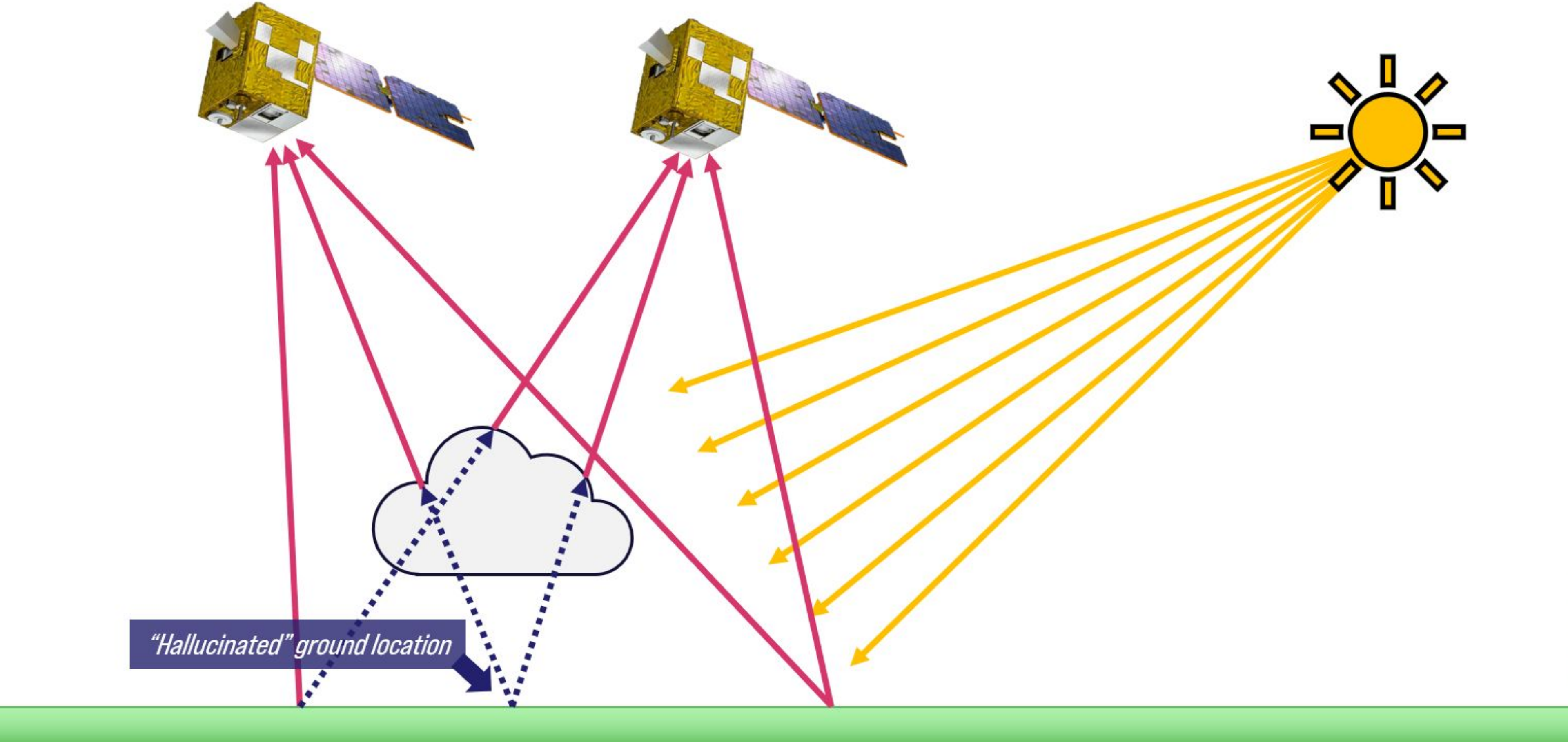
Directions

- ~~different architecture~~

Architecture	# Params.	Dice Score
Single-Pixel	1.3E5	75.8
Simple ConvNet	6.88E5	77.5
U-Net	2.76E8	78.1

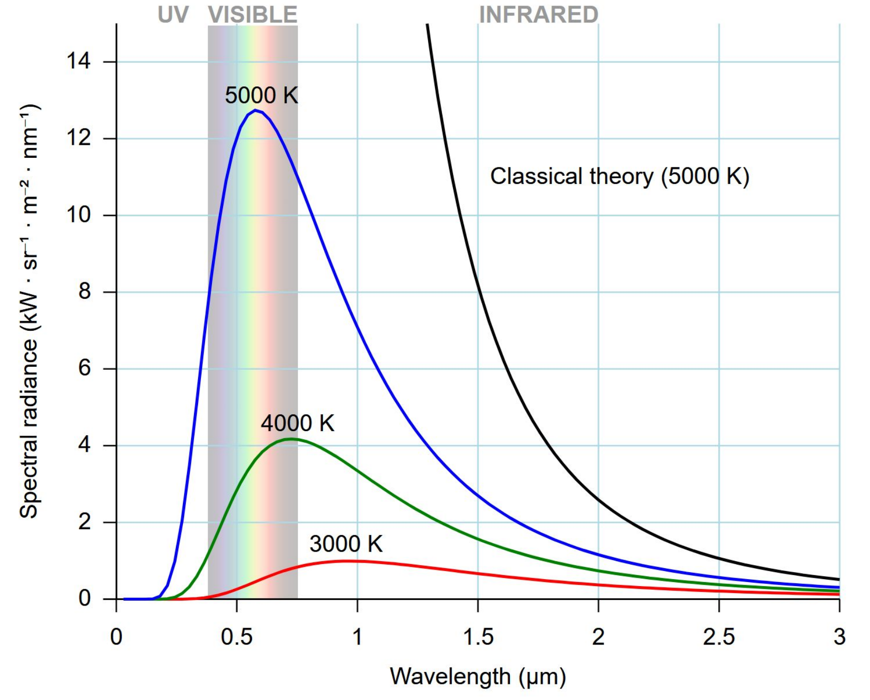
Directions

- ~~different architecture~~
- different representation

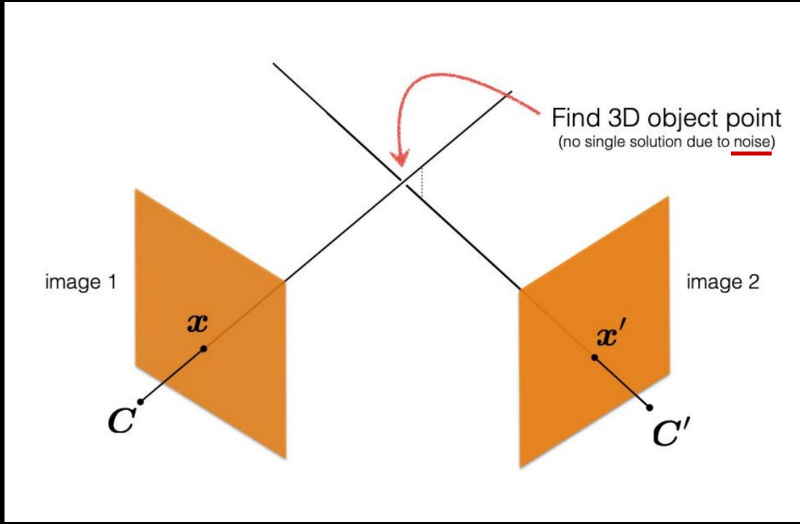

"Hallucinated" ground location

Radiative Transfer

$$\frac{1}{c}\frac{\partial}{\partial t} I_\nu + \hat{\Omega}\cdot\nabla I_\nu + \underbrace{(k_{\nu,s}+k_{\nu,a})\rho I_\nu}_{\text{absorption}} = \underbrace{j_\nu \rho}_{\text{emission}} + \underbrace{\frac{1}{4\pi} k_{\nu,s}\rho \int_\Omega I_\nu \, d\Omega}_{\text{scattering}}$$

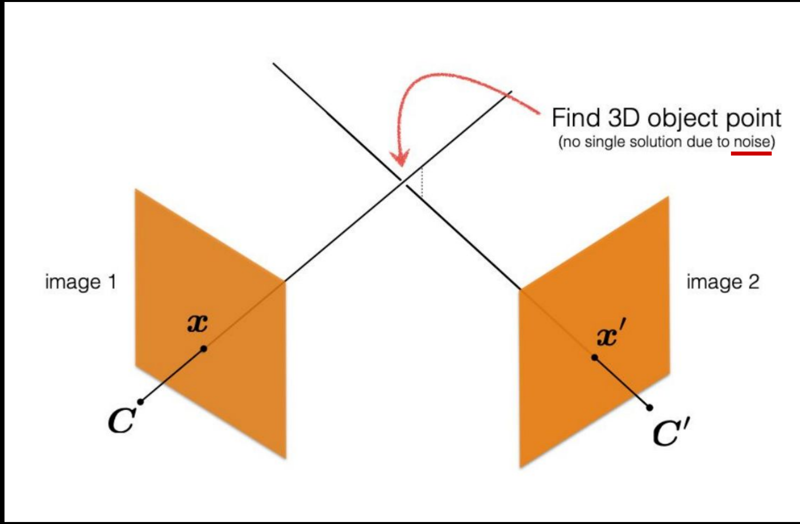


$$\frac{1}{c}\frac{\partial}{\partial t}I_\nu + \hat{\Omega}\cdot\nabla I_\nu + \underbrace{(k_{\nu,s}+k_{\nu,a})\rho I_\nu}_{\text{absorption}} = \cancel{\underbrace{j_\nu\rho}_{\text{emission}}} + \underbrace{\frac{1}{4\pi}k_{\nu,s}\rho\int_\Omega I_\nu d\Omega}_{\text{scattering}}$$



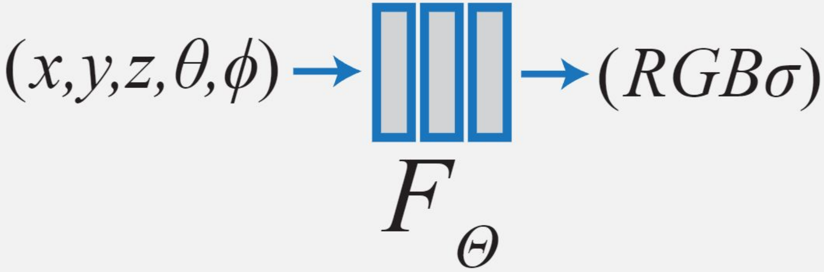
matching & triangulation assumes no partial absorption

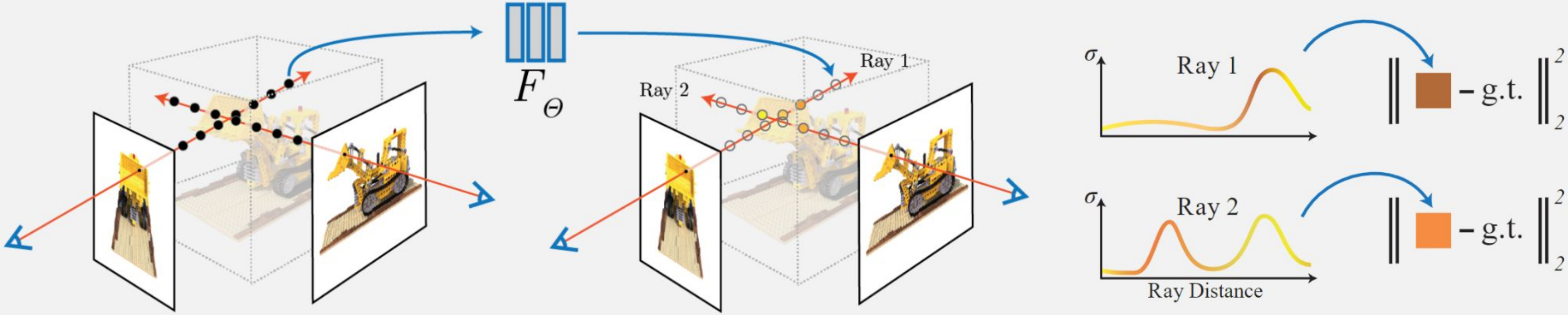
$$\frac{1}{c}\frac{\partial}{\partial t}I_\nu + \hat{\Omega}\cdot\nabla I_\nu + \underbrace{(k_{\nu,s} + k_{\nu,a})\rho I_\nu}_{\text{absorption}} = \underbrace{\cancel{j_\nu \rho}}_{\text{emission}} + \underbrace{\frac{1}{4\pi}k_{\nu,s}\rho\int_\Omega I_\nu d\Omega}_{\text{scattering}}$$



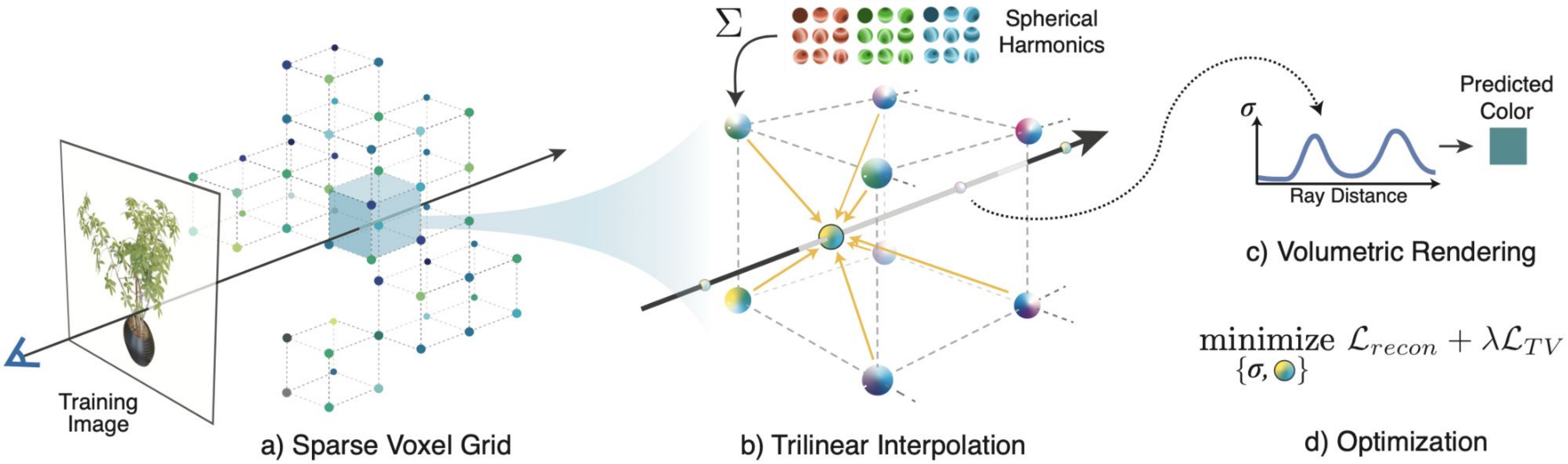
matching & triangulation assumes mostly direction-invariant scattering

$$\frac{1}{c}\frac{\partial}{\partial t}I_\nu + \hat{\Omega}\cdot\nabla I_\nu + \underbrace{(k_{\nu,s}+k_{\nu,a})\rho I_\nu}_{\text{absorption}} = \cancel{\underbrace{j_\nu\rho}_{\text{emission}}} + \underbrace{\frac{1}{4\pi}k_{\nu,s}\rho\int_\Omega I_\nu d\Omega}_{\text{scattering}}$$


(x,y,z,θ,ϕ)
$(RGB\sigma)$
F_Θ


F_Θ
Ray 1
Ray 2
σ
Ray 1
σ
Ray 2
Ray Distance
– g.t.
– g.t.

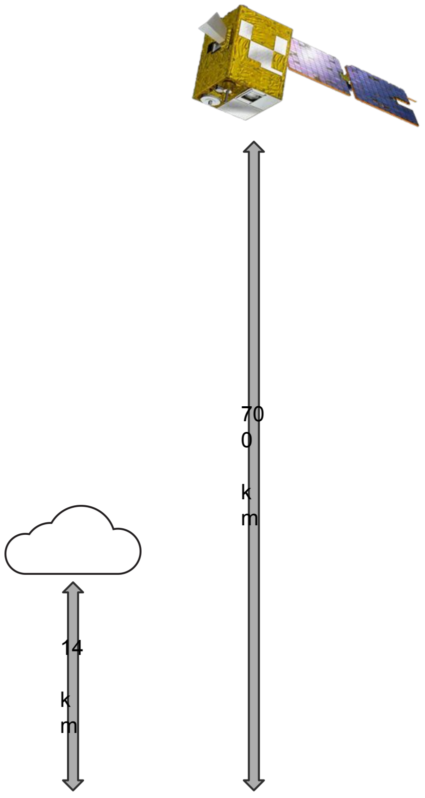
NeRF



Spherical
Harmonics
Predicted
Color
σ
Ray Distance
Training
Image
a) Sparse Voxel Grid
b) Trilinear Interpolation
c) Volumetric Rendering
minimize$_{\{\sigma,\ \cdot\}}$ $\mathcal{L}_{recon} + \lambda\mathcal{L}_{TV}$
d) Optimization

Neural rendering: Challenges

- low spatial resolution
- high spectral resolution
- distant observation
- incorporate supervision



Questions?

- email: seanremy@gmail.com (happy to answer any questions)
- check out the dataset: https://github.com/seanremy/atrain-cloudseg

Thank you!