ICPR 2020
25th INTERNATIONAL CONFERENCE ON PATTERN RECOGNITION
Milan, Italy 10 | 15 January 2021

DLR

TU Graz

amazon

Aerial Road Segmentation in the Presence of Topological Label Noise

Corentin Henry

Remote Sensing Technology Institute
German Aerospace Center (DLR)
Oberpfaffenhofen, Germany
corentin.henry@dlr.de

Friedrich Fraundorfer

Institute of Computer Graphics and Vision
Graz University of Technology (TUG)
Graz, Austria

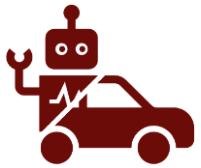
Eleonora Vig

Amazon
Berlin, Germany

Road Segmentation in Remote Sensing

Need for up-to-date road network maps

Navigation Systems



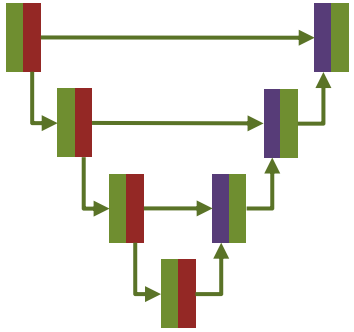
Autonomous Driving

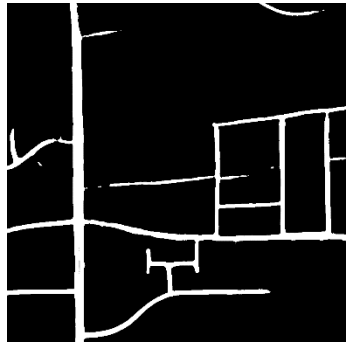
Disaster Relief

Urban Monitoring

Fully-convolutional neural networks

U-Net → Residual U-Net → D-LinkNet [1]





[1] Zhou et al. „D-LinkNet: LinkNet with Pretrained Encoder and Dilated Convolution for High Resolution Satellite Imagery Road Extraction “, CVPR Workshops 2018

Pixel-wise road segmentation datasets

Massachusetts Roads

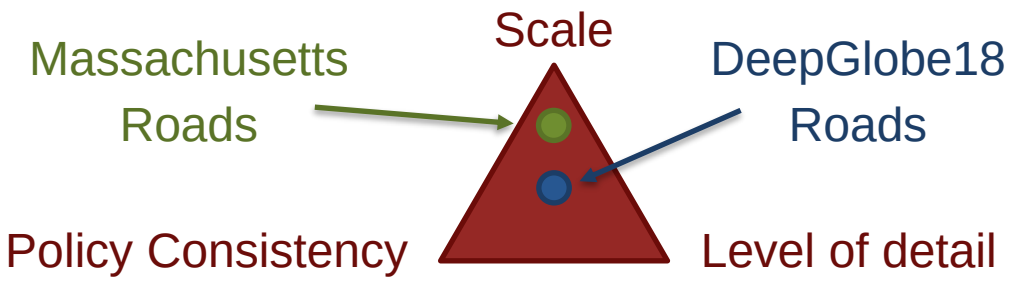
- ~2600 km² at 1m/px
- Rasterized vector annotation from OSM
- Massachusetts, USA

DeepGlobe18 Roads

- ~2200 km² at 50cm/px
- Highly detailed manually drawn annotation
- Thailand, Indonesia, India

Why is Label Noise so Impactful?

All annotations are flawed to an extent



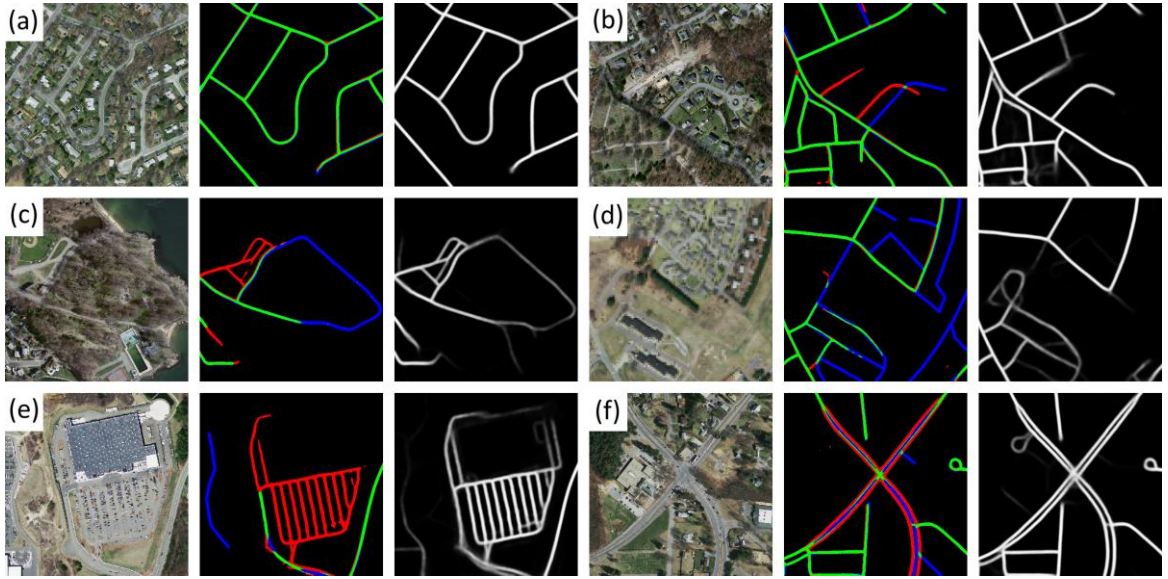
Main types of topological noise

- **Omission:** Missing label (e, g-i)
- **Registration:** Offset label (b)
- **Geometry:** Coarsely or wrongly shaped label (f)
- **Inconsistency:** Same objects, different labels (c)

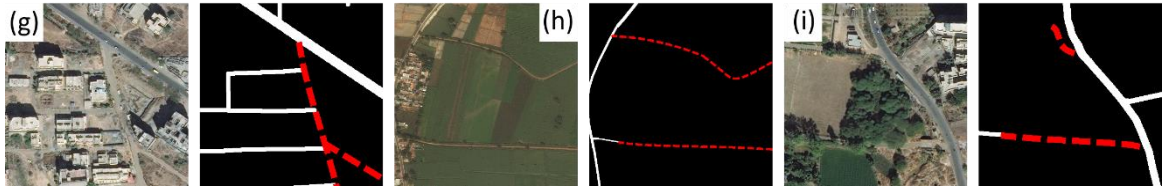
Consequences:

- **Difficult training**
- **Unreliable evaluation!**

Confused road extraction models



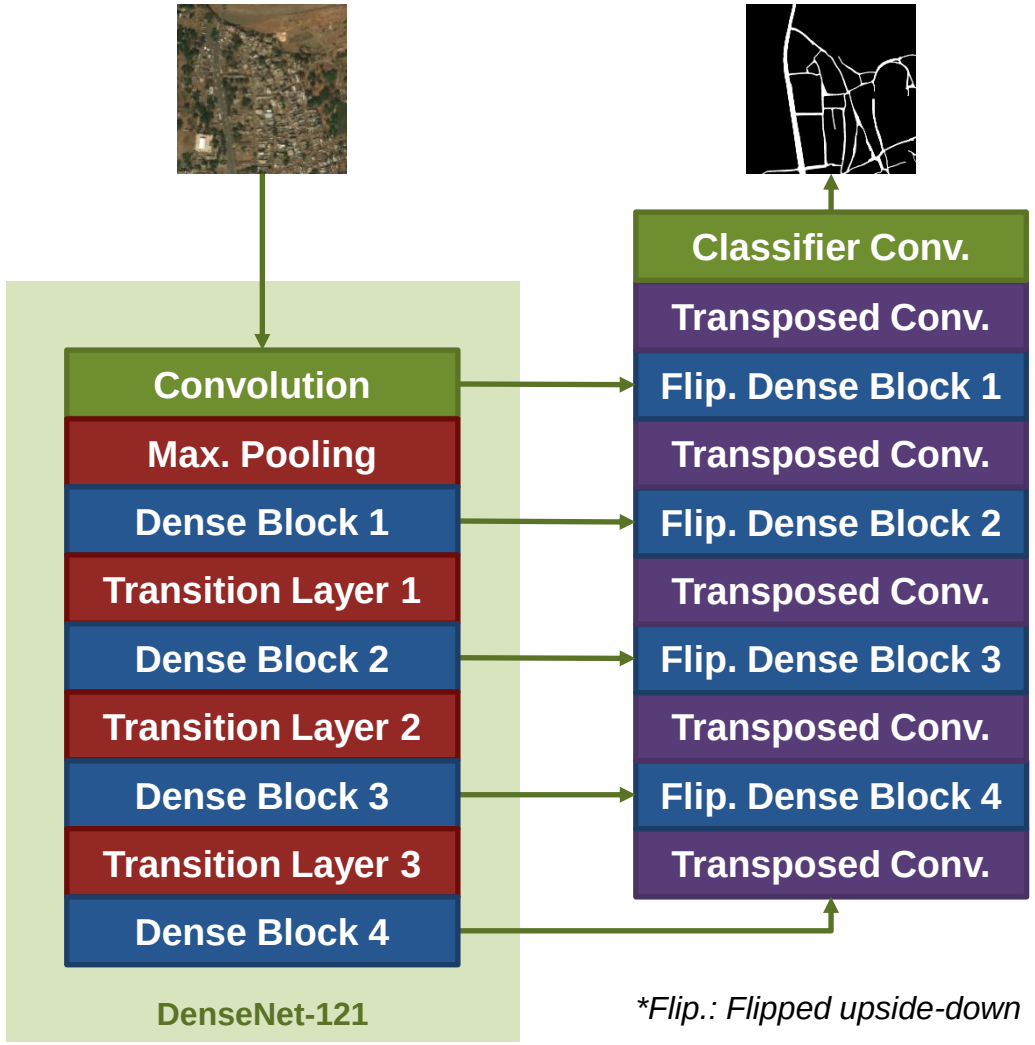
True positives
False negatives
False positives
Predicted road probability (0-100%)



Missing road labels in DeepGlobe

Training Noise-Resilient Models

Using a Densely-Connected U-Net



**Flip.: Flipped upside-down*

Key hyper-parameters:

- Pre-training: ImageNet
- Optimizer: ADAM
- Learning rate: fixed to 1e-4
- # Training epochs: 40

Augmentations:

- Random horizontal flip
- Random 90°-wise rotation

Other architectures benchmarked:

- DeepLabv3+
- DenseASPP
- Residual U-Net
- D-LinkNet

Training Noise-Resilient Models

Using noise-aware losses

Binary cross-entropy (BCE)

$$-\sum_{k}^{C}\sum_{i}^{N} \mathbf{y}_{ik} \log(\mathbf{p}_{ik})$$

Most widely used

✓ General purpose ✗ For thin objects

Dice coefficient

$$1 - \frac{1 + \sum_i^N 2\mathbf{y}_i\mathbf{p}_i}{1 + \sum_i^N (\mathbf{y}_i{}^2 + \mathbf{p}_i{}^2)}$$

Helps maximizing dice metric

✓ Data agnostic ✗ Limited noise-awaress

Noise-aware sigmoid

$$\frac{1}{N}\sum_{i}^{N} \mathrm{Sigmoid}(-\boldsymbol{\beta}\mathbf{y}_i\mathbf{p}_i)$$

Controllable level of trust in label

✓ Implicit resilience ✗ Not standalone

Bootstrapped BCE

$$-\sum_{k}^{C}\sum_{i}^{N} [\boldsymbol{\beta}\mathbf{y}_{ik} + (1-\boldsymbol{\beta})\mathbf{p}_{ik}] \log(\mathbf{p}_{ik})$$

Bootstrapped dice coefficient

$$1 - \frac{1 + \sum_i^N 2[\boldsymbol{\beta}\mathbf{y}_i + (1-\boldsymbol{\beta})\mathbf{p}_i]\mathbf{p}_i}{1 + \sum_i^N [\boldsymbol{\beta}\mathbf{y}_i + (1-\boldsymbol{\beta})\mathbf{p}_i]^2 + \mathbf{p}_i{}^2}$$

The less the label is trusted, the more the predictions are trusted

✓ Explicit resilience ✗ Too aggressive

$\mathbf{y}_{ik}$ Label (pixel i, class k)

$\mathbf{p}_{ik}$ Predicted probability (pixel i, class k)

$\boldsymbol{\beta}$ Label trust coefficient

Training Noise-Resilient Models

Using synthetic noise augmentations

Noise-aware losses are still sensitive, because:

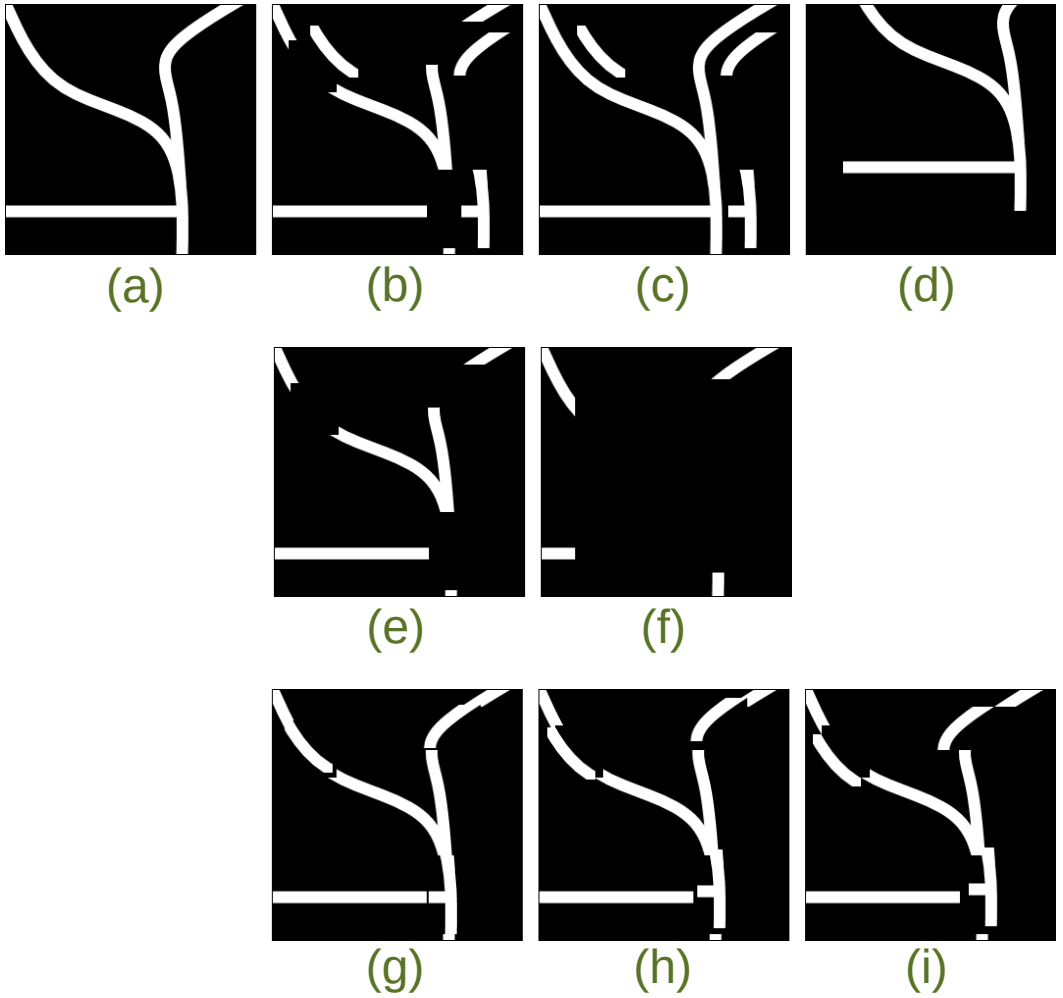
- Label noise is sparse
- Noise types are unequally represented

Noise augmentation during training:

- Uniform frequency
- Random amplitude (g-i)

Synthetic noise types:

(a) Original ground truth
(b) Registration - segment offset
(c) Registration - segment duplication
(d) Registration - area offset
(e) Omission - segment
(f) Omission - area



Evaluating the Training Strategy

Noise-awareness improves the performance

Synth. Noise Type (Amplitude)	Loss	Massach. Custom Test IoU	F1	Qual.	DeepGlobe Custom Valid IoU	F1	Qual.
Registration (None) Omission (None)	BCE	57.12	73.03	70.06	65.13	79.19	72.43
	Boot. BCE	57.87	**73.53**	70.02	**65.87**	**79.58**	**73.28**
	Boot. Dice	**57.91**	73.30	**70.22**	64.88	79.00	71.69

Synthetic noise can boost the performance

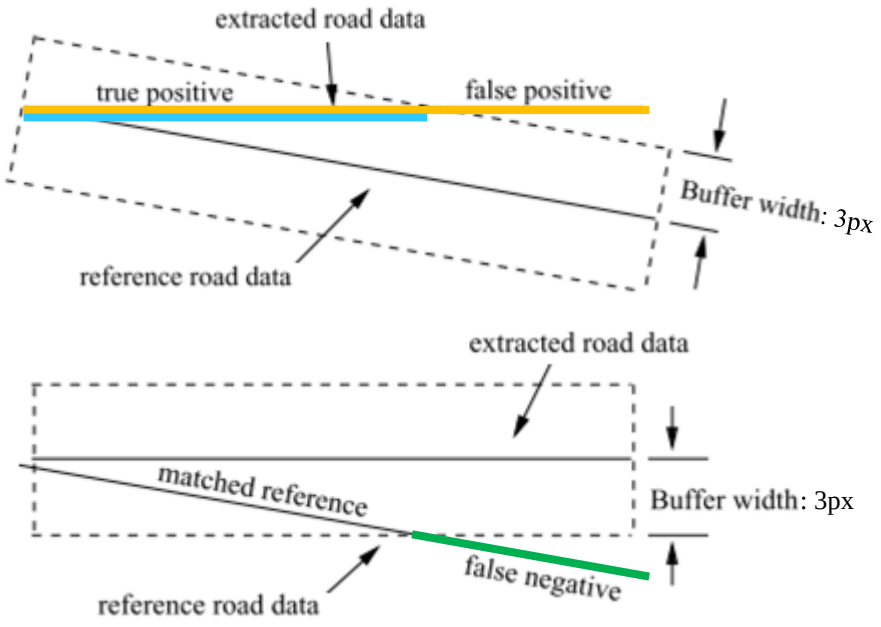
Synth. Noise Type (Amplitude)	Loss	DeepGlobe Custom Valid IoU	F1	Qual.
Registration (Low) Omission (None)	Boot. BCE	66.36	79.85	72.94
	Boot. Dice	**68.03**	**81.13**	**74.77**
Registration (Medium) Omission (None)	Boot. BCE	66.03	79.61	72.66
	Boot. Dice	**67.72**	**80.91**	**74.89**

Training can recover from extreme noise

Synth. Noise Type (Amplitude)	Loss	Massach. Custom Test IoU	F1	Qual.	DeepGlobe Custom Valid IoU	F1	Qual.
Registration (High) Omission (None)	BCE	4.18	8.25	5.95	34.28	51.10	27.47
	Boot. BCE	12.16	22.58	12.43	41.80	58.45	34.48
	Boot. Dice	**23.24**	**39.39**	**20.83**	**42.58**	**59.71**	**42.76**
Registration (None) Omission (High)	BCE	0.00	0.00	0.01	0.03	0.06	0.41
	Boot. BCE	38.34	55.35	63.47	45.90	63.04	60.87
	Boot. Dice	**57.11**	**70.11**	**72.92**	**64.41**	**78.73**	**71.31**

Measuring the road quality metric [2]

Computed on skeletonized GT and predictions!



$$Road\_Quality = \frac{|matched\_extraction|}{|extraction| + |unmatched\_reference|}$$

[2] Wiedemann et al. „Empirical Evaluation Of Automatically Extracted Road Axes“, Empirical Evaluation Techniques in Computer Vision, 1998

Annotation Consistency is Critical

Consistency in annotation

Test annotations are precise if:

- No road is wrongly identified
- Labels do not overshoot

Test annotations are complete if:

- All roads are idenfied
- Labels cover entire drivable area

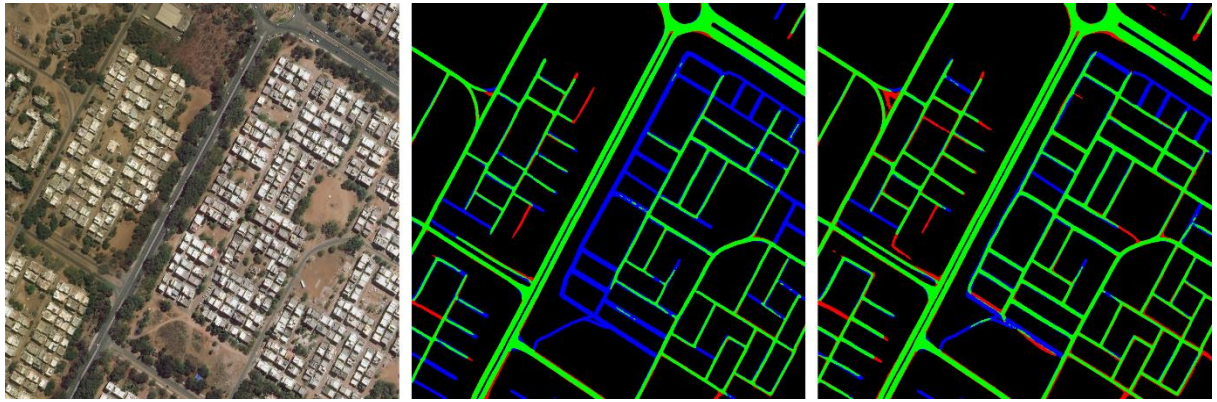
Both datasets have noisy test labels!
It affects performance benchmarks

Actual performance improvements?

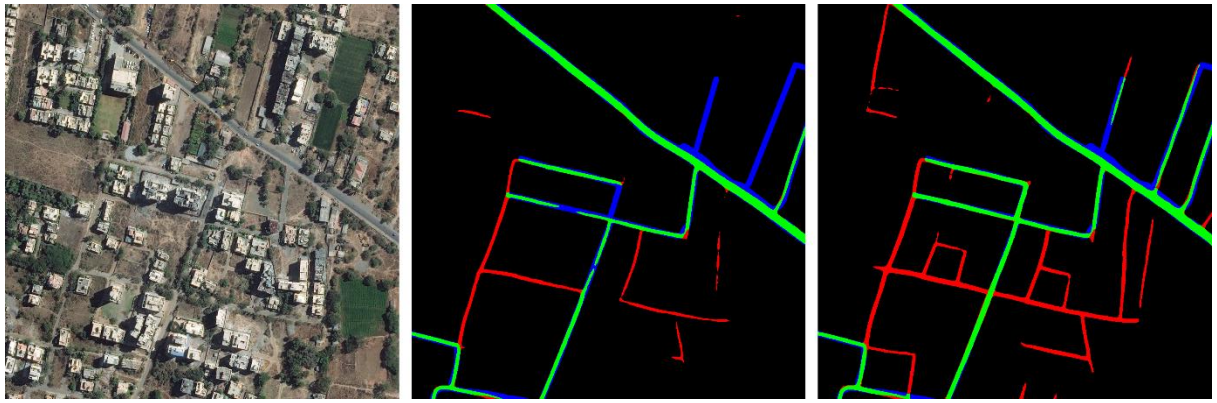
- Qualitative: Yes, as shown on the right
- Quantitative: Yes, but underestimated

Effect on DeepGlobe images: before and after

Fewer roads are missed:



Additional roads are detected:



What Comes Next?

Achievements so far

Label noise training counter-measures:

- Noise-aware losses are effective
- Synthetic noise augmentation is effective
- Most effective when both are combined

Areas of improvements

Using more advanced architectures:

- *Bastani et al. „RoadTracer: Automatic Extraction of Road Networks from Aerial Images“, CVPR18*
- *He et al. „Sat2Graph: Road Graph Extraction through Graph-Tensor Encoding“, ECCV20*

Using more advanced metrics:

- *Citraro et al. “Towards Reliable Evaluation of Algorithms for Road Network Reconstruction from Aerial Images”, ECCV20*

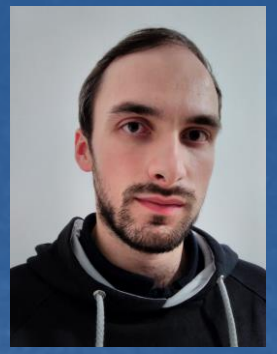
A most critical next step!

Creating reliable benchmark datasets [3]:

- Large-scale
- High level of detail annotation
- Thorough annotation quality check

[3] Azimi et al. „SkyScapes - Fine-Grained Semantic Understanding of Aerial Scenes“, ICCV19

We thank you for your attention!



Corentin Henry
German Aerospace Center (DLR)
Oberpfaffenhofen, Germany
corentin.henry@dlr.de

Friedrich Fraundorfer
Graz University of Technology (TUG)
Graz, Austria

Eleonora Vig
Amazon
Berlin, Germany