SEMANTIC CLOUD SEGMENTATION MODELS FOR SOLAR APPLICATIONS WITH SYNTHETIC DATA

Eduardo Sáez Martínez, Raphael Gut, Felix Häusler, Bijan Nouri, Yann Fabel, David Magiera, Niklas Blum and Luis F. Zarzalejo

25.09.2025

Ciemat

DLR

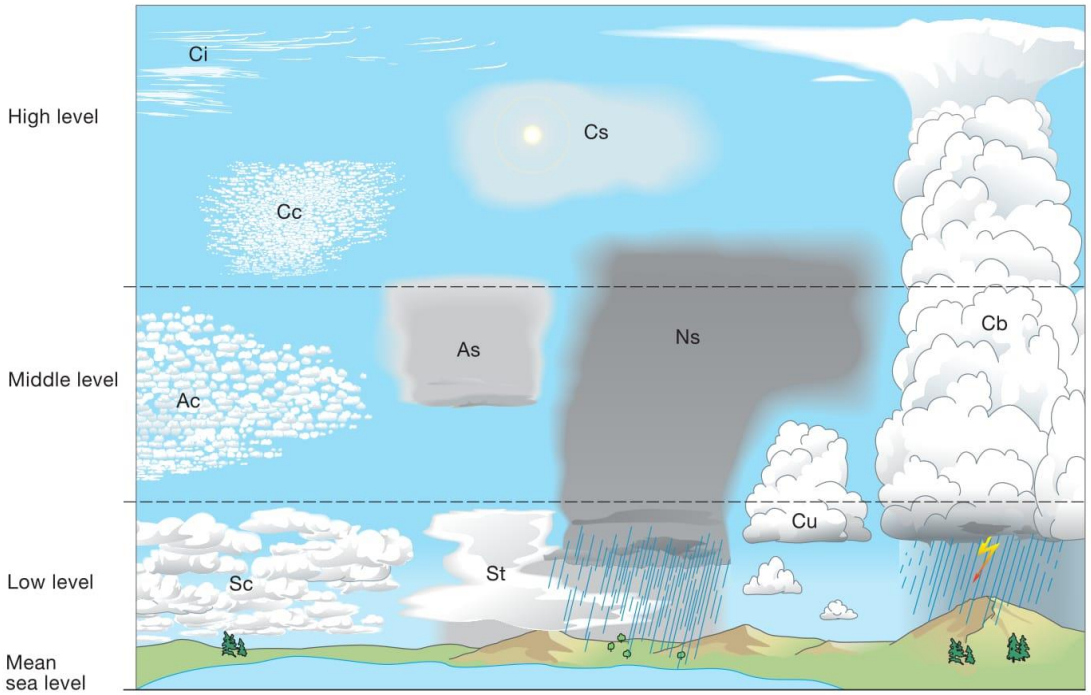
Agenda

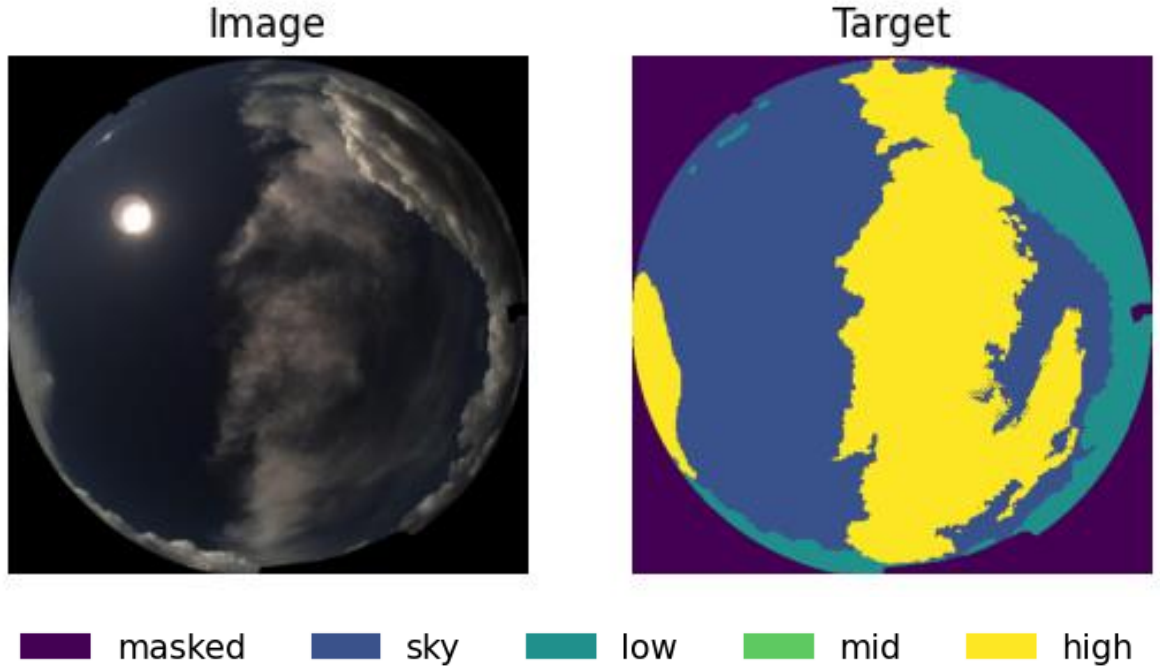
1. **Problem definition**
2. Existing datasets
3. Ground-truth data generation via Semi-supervised Video Segmentation
4. Experimental results
5. Conclusion and outlook

Problem definition

Motivation: Cloud detection for solar irradiance nowcasting systems to optimize operation of solar power plants and electrical grids.

Goal: Provide a pixel-wise classification (semantic segmentation) of clouds in all-sky images in 3 layers defined by WMO.





Problem definition

Approach:

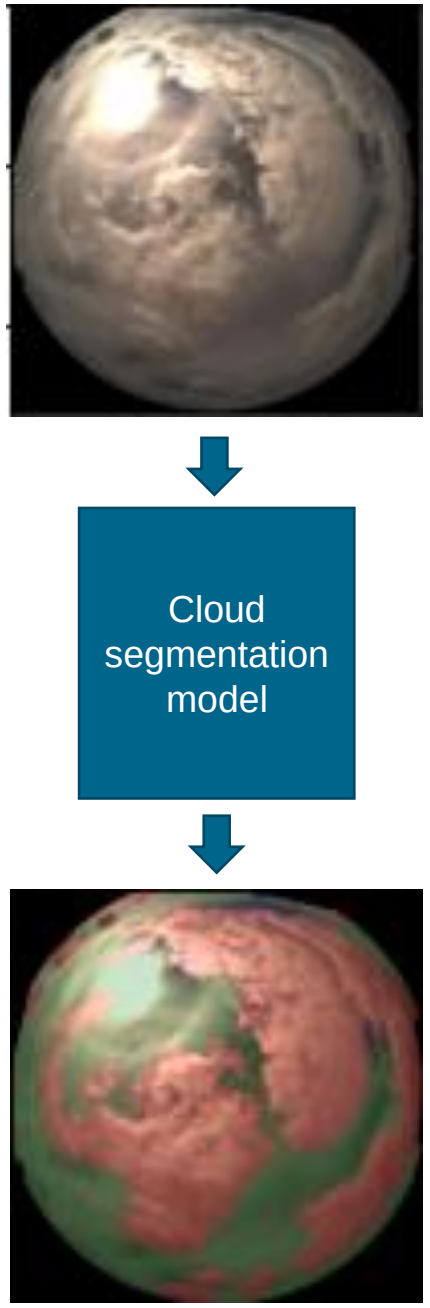
- Deep Learning based methods have been established

Challenges:

- High-quality pixel-level annotated datasets are required
- Lack of a large amount of ground truth readily available
- Expert annotations are time intensive

Solution:

- Automatic extension of manually annotated dataset.



Agenda

1. Problem definition
2. **Existing datasets**
3. Ground-truth data generation via Semi-supervised Video Segmentation
4. Experimental results
5. Conclusion and outlook

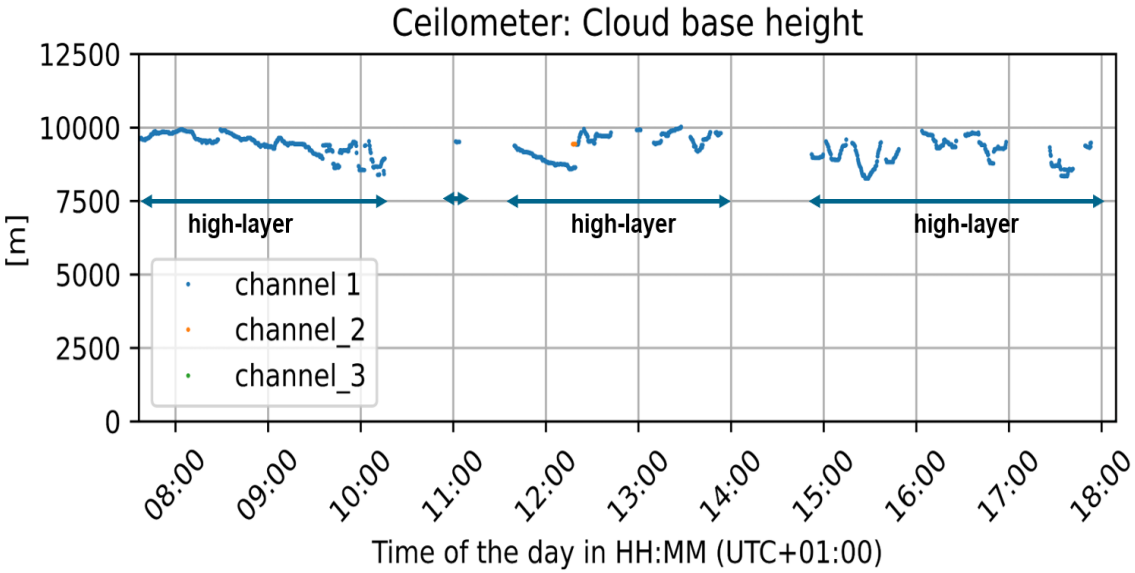
Datasets

Manually annotated dataset [1]**:**

- 770 annotated images

Weakly annotated dataset [2]**:**

- Ceilometer measures cloud height
- Measurements within certain thresholds represent single cloud layers
- Temporal stable single cloud layer conditions are identified
- Assign corresponding cloud layer as image-level weak label to binary cloud segmentations
- 47.000 segmented images

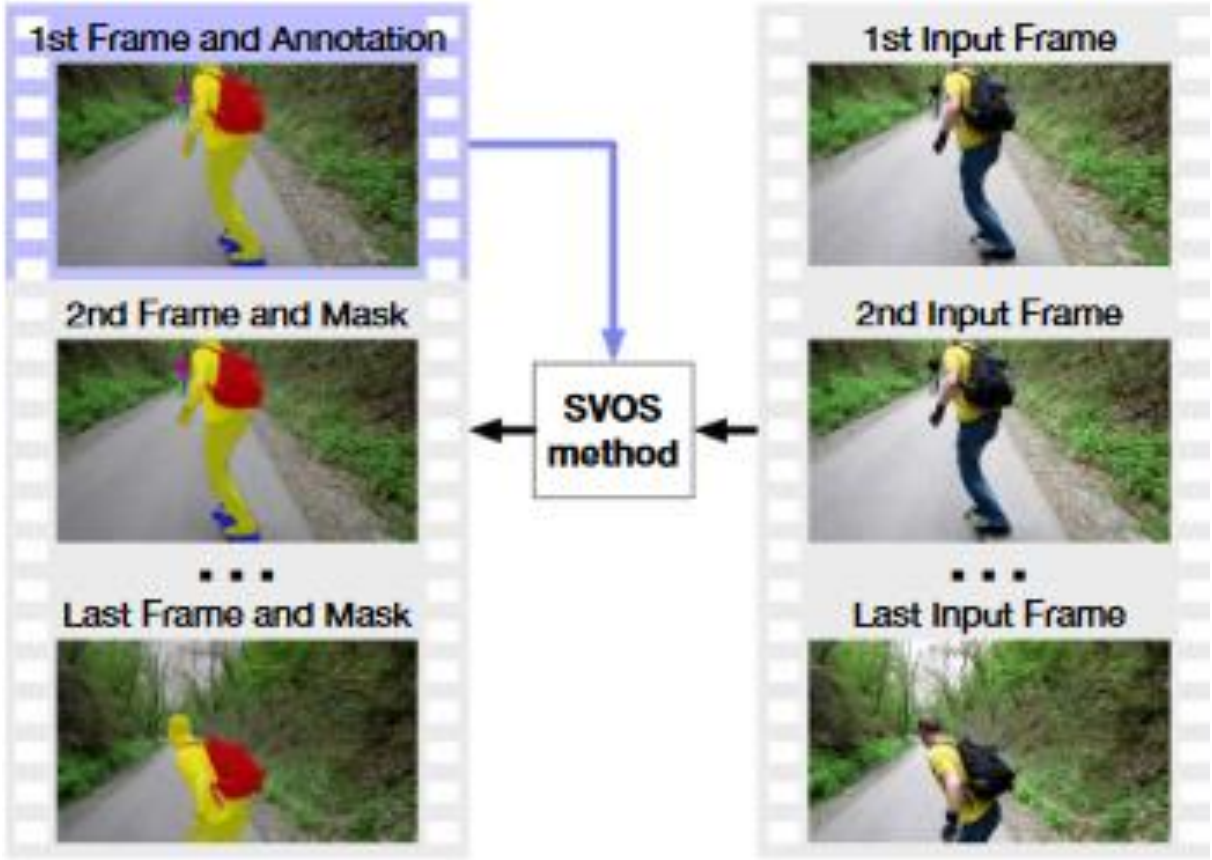


Agenda

1. Problem definition
2. Existing datasets
3. **Ground-truth data generation via Semi-supervised Video Segmentation**
4. Experimental results
5. Conclusion and outlook

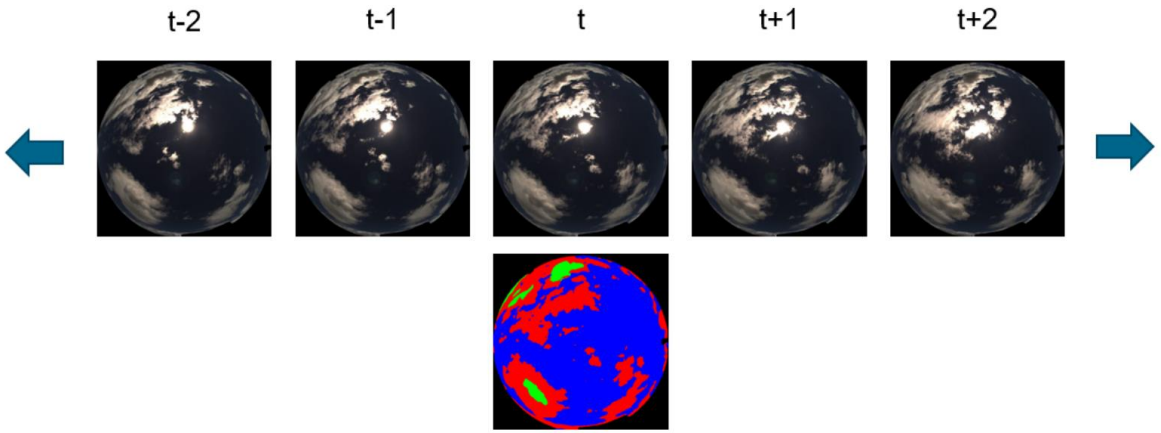
Ground-truth data generation via Semi-supervised Video Segmentation: Overview

- VOS (Video Object Segmentation) is an AI technique that segments and tracks objects frame by frame in a video.
- In Semi-Supervised VOS (SVOS) an initial mask of the object to be segmented in the first frame is given.



Ground-truth data generation via Semi-supervised Video Segmentation: Strategy

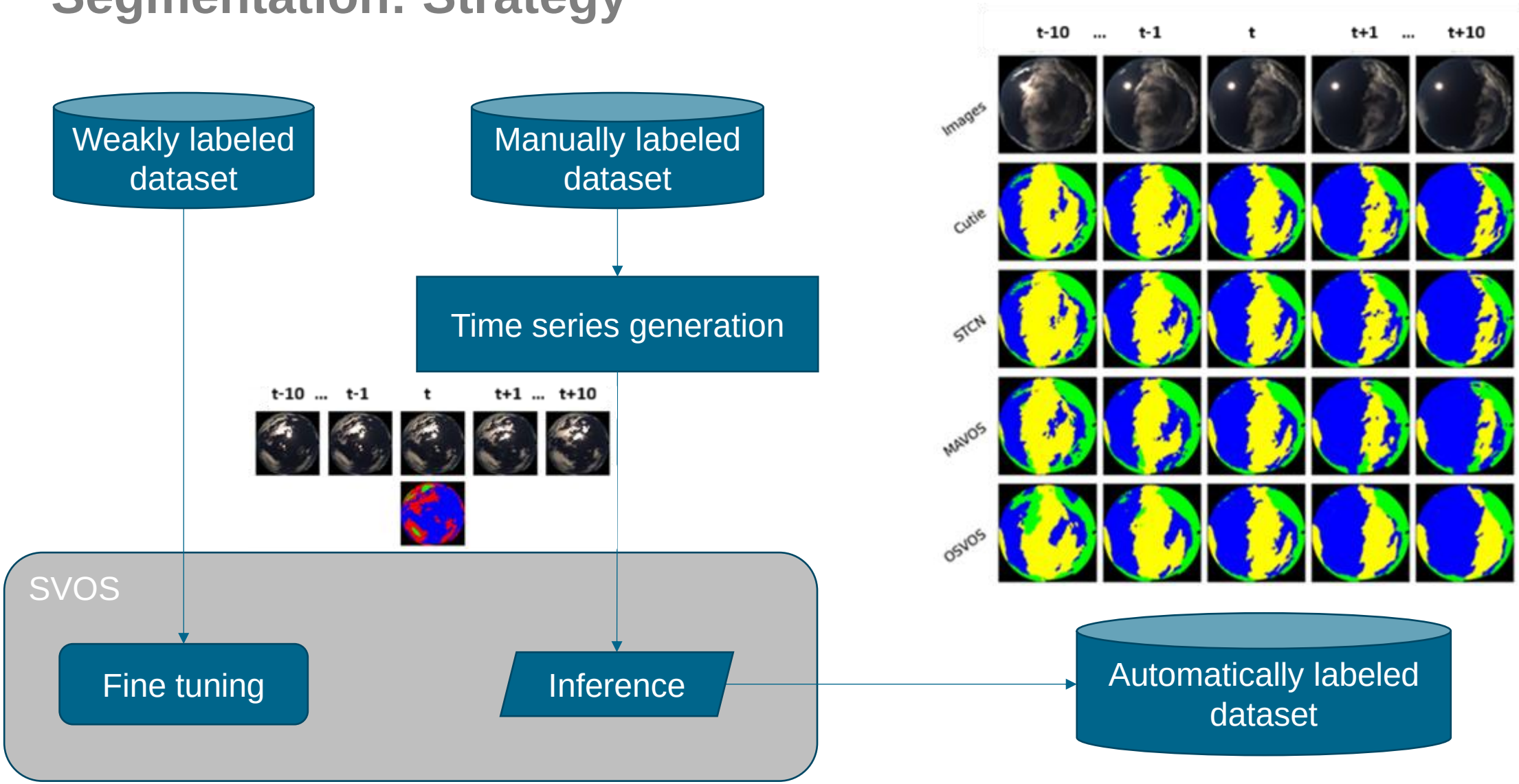
- **Goal**:
  - Expand the existing Manually annotated dataset
  - Study the degradation of quality in label propagation



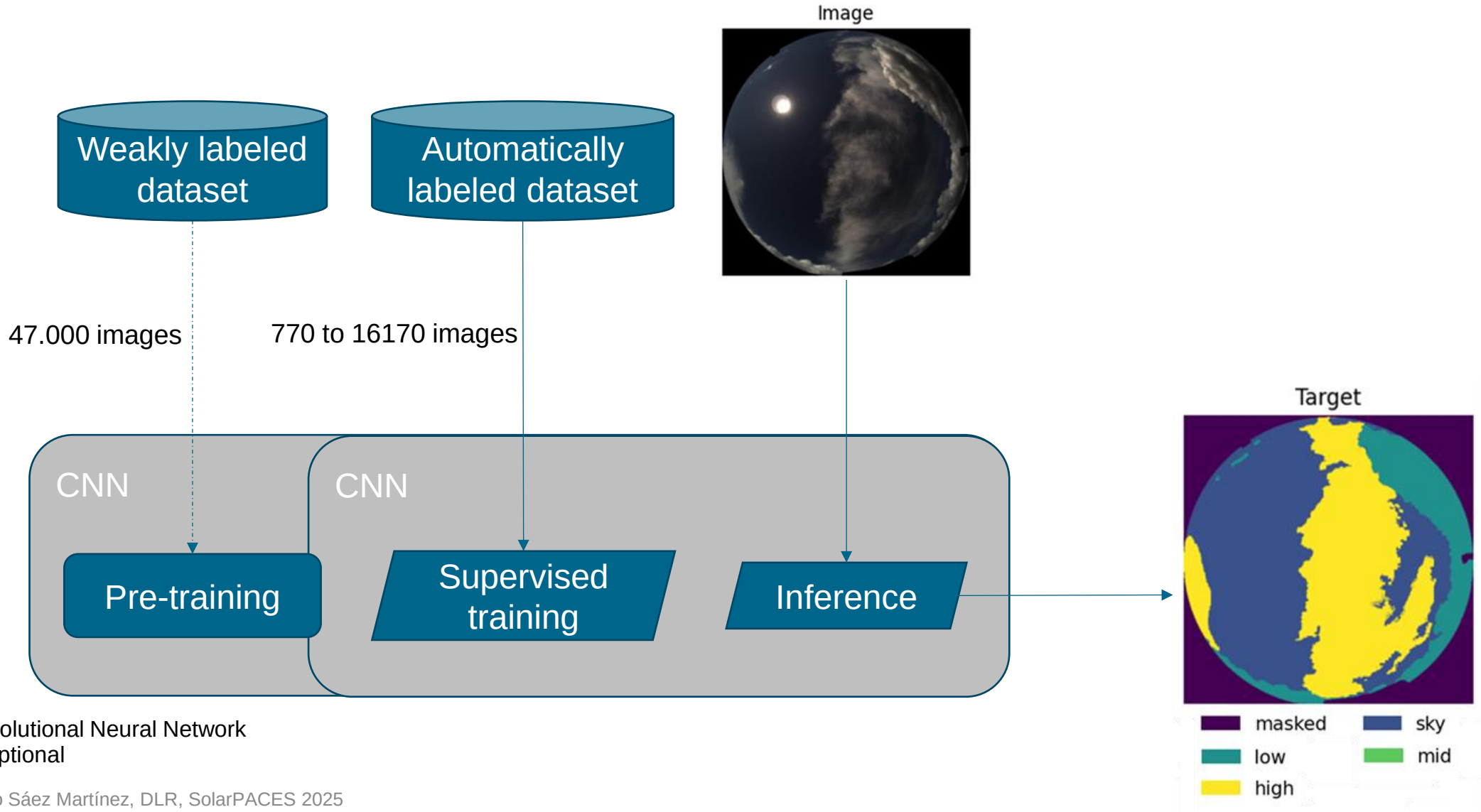
- Test 4 models:
  - OSVOS (2017) [3]
  - STCN (2020) [4]
  - Cutie (2024) [5]
  - MAVOS (2025) [6]

- Maximum of +/- 10 frames is evaluated

Ground-truth data generation via Semi-supervised Video Segmentation: Strategy



Ground-truth data generation via Semi-supervised Video Segmentation: Strategy



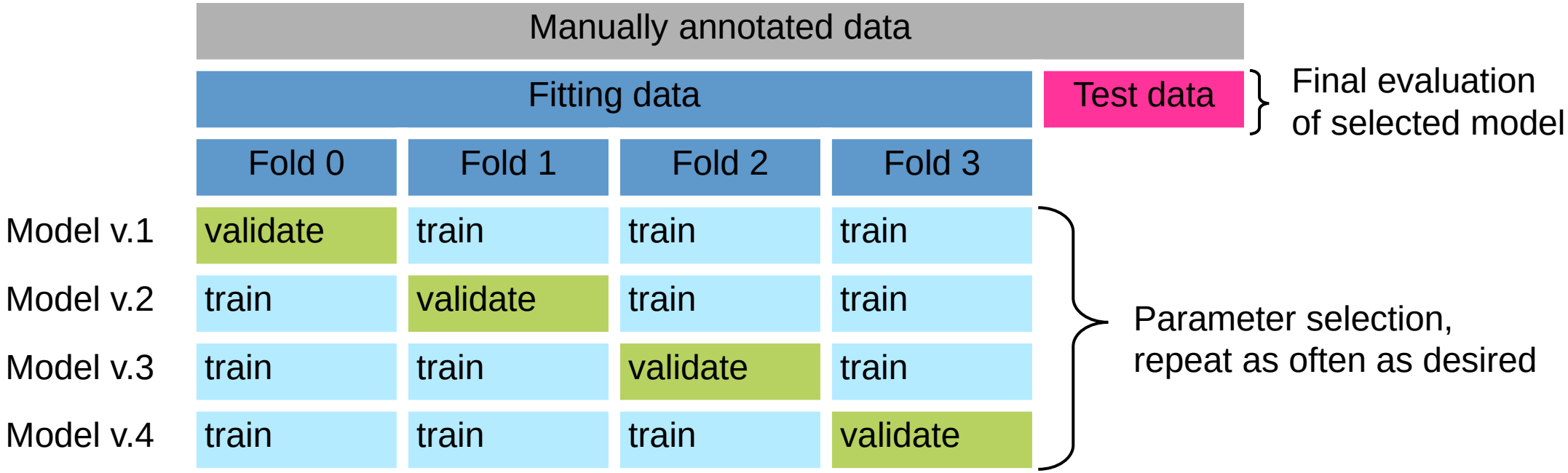
CNN: Convolutional Neural Network

Optional

Agenda

1. Problem definition
2. Existing datasets
3. Ground-truth data generation via Semi-supervised Video Segmentation
4. **Experimental results**
5. Conclusion and outlook

Experiment Results: Dataset – 4-Fold Cross Validation

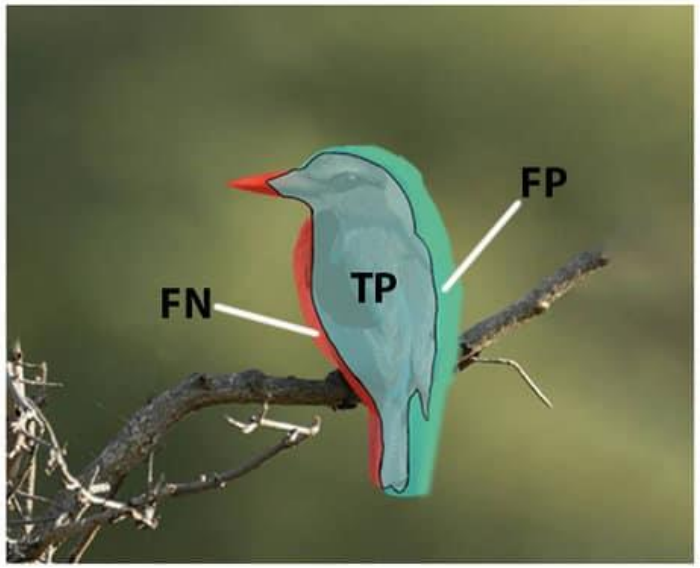


Analyze variance, avoids overfitting, uses all data

Experiment Results: Intersection over Union

Ground Truth Mask

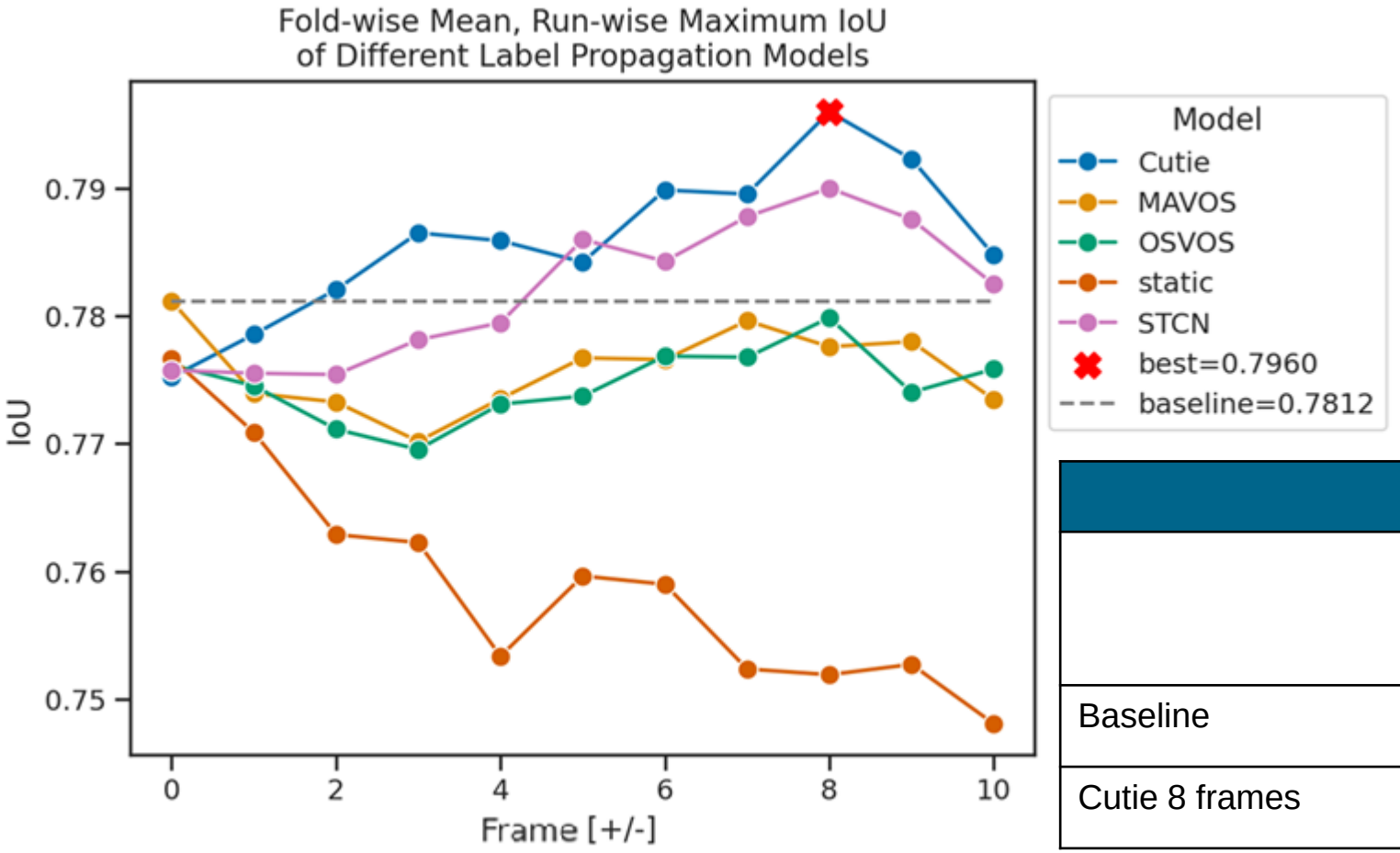
Predicted Mask



https://learnopencv.com/wp-content/uploads/2022/06/5-segmentation-iou.jpg

$$mean\ IoU = \frac{1}{|Classes|} \sum_{c \in Classes} \frac{Target_c \cap Prediction_c}{Target_c \cup Prediction_c} = \frac{1}{|Classes|} \sum_{c \in Classes} \frac{TP_c}{TP_c \cup FP_c \cup FN_c}$$

Experiment Results: VOS Comparison

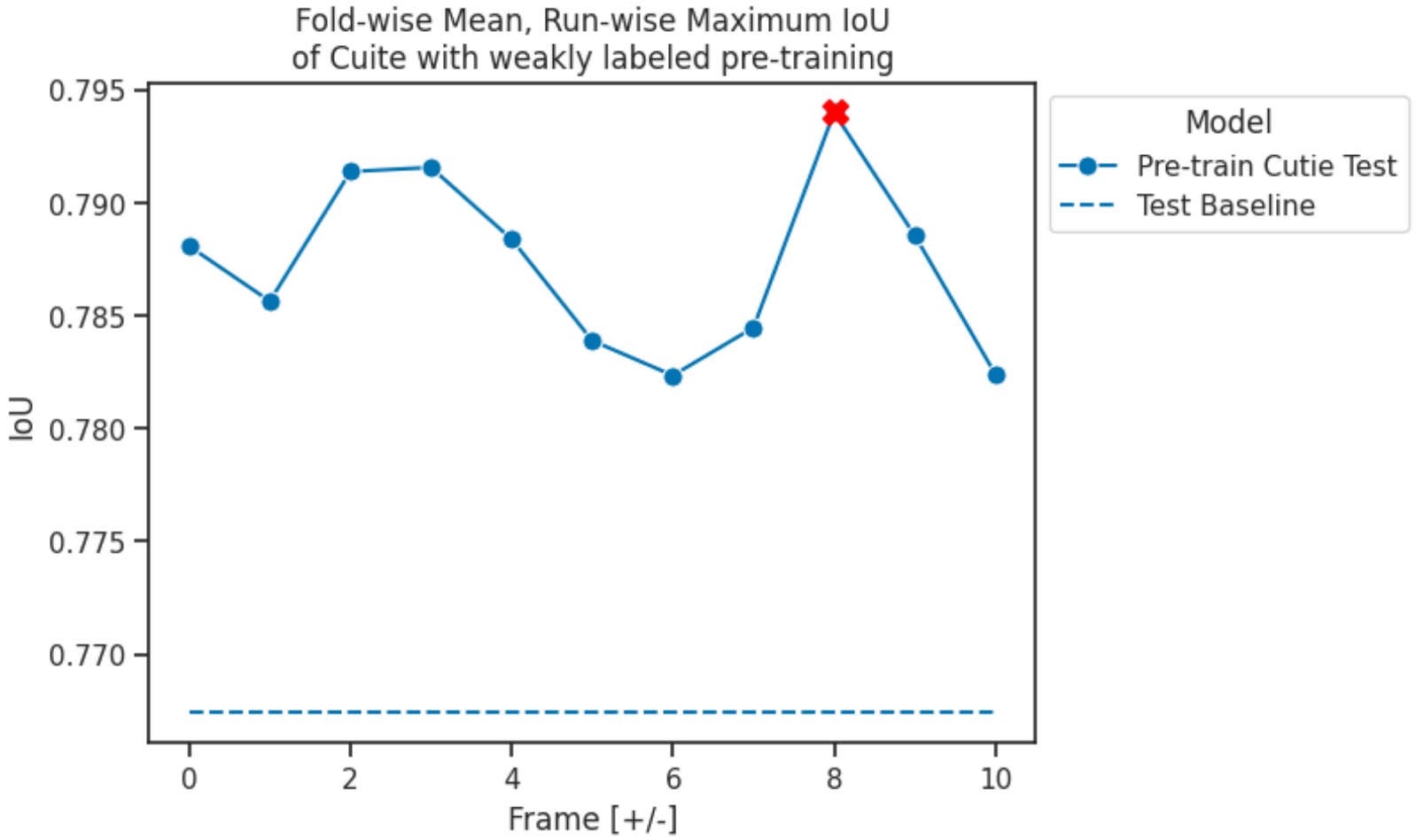


Results				
	Val. IoU	Val. IoU over baseline [%]	Test IoU	Test IoU over baseline [%]
Baseline	0.781	-	0.746	-
Cutie 8 frames	**0.796**	**1.9**	**0.767**	**2.8**

Pre-train: None.

Training: 4k iterations on the automatically labeled dataset.

Experiment Results: Pre-train Cutie



Pre-train: 4k iterations on the weakly labeled dataset (sun elevation above 10º).

Training: 4k iterations on the automatically labeled dataset.

Experiment Results

Results		
	Test IoU	Test IoU over baseline [%]
Baseline	0.746	-
Cutie 8 frames	0.767	2.8
Pre-train Cutie 8 frames (Short Pre-training)	0.776	4.0
Pre-train Cutie 8 frames (Extended Pre-training)	**0.794**	**6.4**

Short Pre-training:

- 4k iterations on the weakly labeled dataset (sun elevation above 20º).

Extended Pre-training:

- 4k iterations on the weakly labeled dataset (sun elevation above 10º).

Agenda

1. Problem definition
2. Existing datasets
3. Ground-truth data generation via Semi-supervised Video Segmentation
4. Experimental results
5. **Conclusion and outlook**

Conclusion and Outlook

Conclusion:

- SVOS are promising for segmentation tasks with limited labeled data as significant improvements are achieved against the baseline.
- Further improvement achieved by incorporating weakly labeled data in pre-training step.

Outlook:

- Use time series information
  - Allows more views on the same clouds without extra hardware
- Introduce confidence weighting (e.g. decreasing weights for label propagation)
- Introduce label relaxation to smooth class boundaries of propagated targets

References

[1] Y. Fabel, “Cloud Segmentation and Classification from All-Sky Images Using Deep Learning,” 2020.

[2] D. Magiera, “Semi-supervised Learning for Probabilistic Cloud Detection in Ground-based Imagery,” 2024.

[3] S. Caelles, K.-K. Maninis, J. Pont-Tuset, L. Leal-Taixé, D. Cremers, and L. Van Gool, “One-Shot Video Object Segmentation,” 2017.

[4] H. K. Cheng, Y.-W. Tai, and C.-K. Tang, “Rethinking Space-Time Networks with Improved Memory Coverage for Efficient Video Object Segmentation,” 2021.

[5] H. K. Cheng, S. W. Oh, B. Price, J.-Y. Lee, and A. Schwing, “Putting the Object Back into Video Object Segmentation,” 2023.

[6] A. Shaker, S. T. Wasim, M. Danelljan, S. Khan, M.-H. Yang, and F. S. Khan, “Efficient Video Object Segmentation via Modulated Cross-Attention Memory,” in Proceedings of the Winter Conference on Applications of Computer Vision (WACV), Feb. 2025, pp. 8670–8679.

Thank you for your attention!

bijan.nouri@dlr.de

eduardo.saezmartinez@dlr.de