Journal Club

July 20, 2021

Jakob Schlör

Universität Tübingen

machine learning in climate science

Jul 27, 2021

Variational Autoencoder Anomaly-Detection of Avalanche Deposits in Satellite SAR Imagery

Saumya Sinha*
University of Colorado
Boulder, USA

Sophie Giffard-Roisin
ISTerre; Univ. Grenoble Alpes, Univ. Savoie Mont Blanc, CNRS, IRD, IFSTTAR
Grenoble, France

Fatima Karbou
CNRM-GAME, Météo-France
Grenoble, France

Michael Deschatres
Irstea, Université Grenoble Alpes
France

Anna Karas
CNRM-GAME, Météo-France
Grenoble, France

Nicolas Eckert
Irstea, Université Grenoble Alpes
France

Cécile Coléou
CNRM-GAME, Météo-France
Grenoble, France

Claire Monteleoni
University of Colorado
Boulder, USA

Climate Informatics (2020)

In short

Question:

Avalanche detection from satellite images with limited labels.

Results:

Unsupervised-learning approach outperforms supervised methods.

Impact:

Automatically identify risk zones, stability of snow pack and improve avalanche forecasting

Motivation

- 4000 avalanches reported in French Alps in winter 2017-2018
- 30 people died in this winter
- Usually reported by local forest offices

Automatic detection of avalanches for remote sensing data:

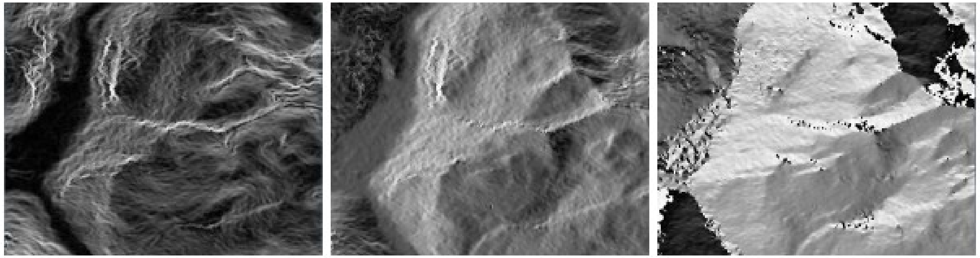
- Stability of snow pack
- Identify avalanche risk zones and time periods

Remote sensing data

- SAR Imagery:
  - Sentinel 1 satellite
  - C-band (4.0 to 8.0 GHz)
  - 6-day repeat cycle
  - 20m resolution
  - Ascending and descending orbit modes
  - Backscatter characteristics over avalanche debris

- topographical feature

Model of a Sentinel 1 [wikipedia]

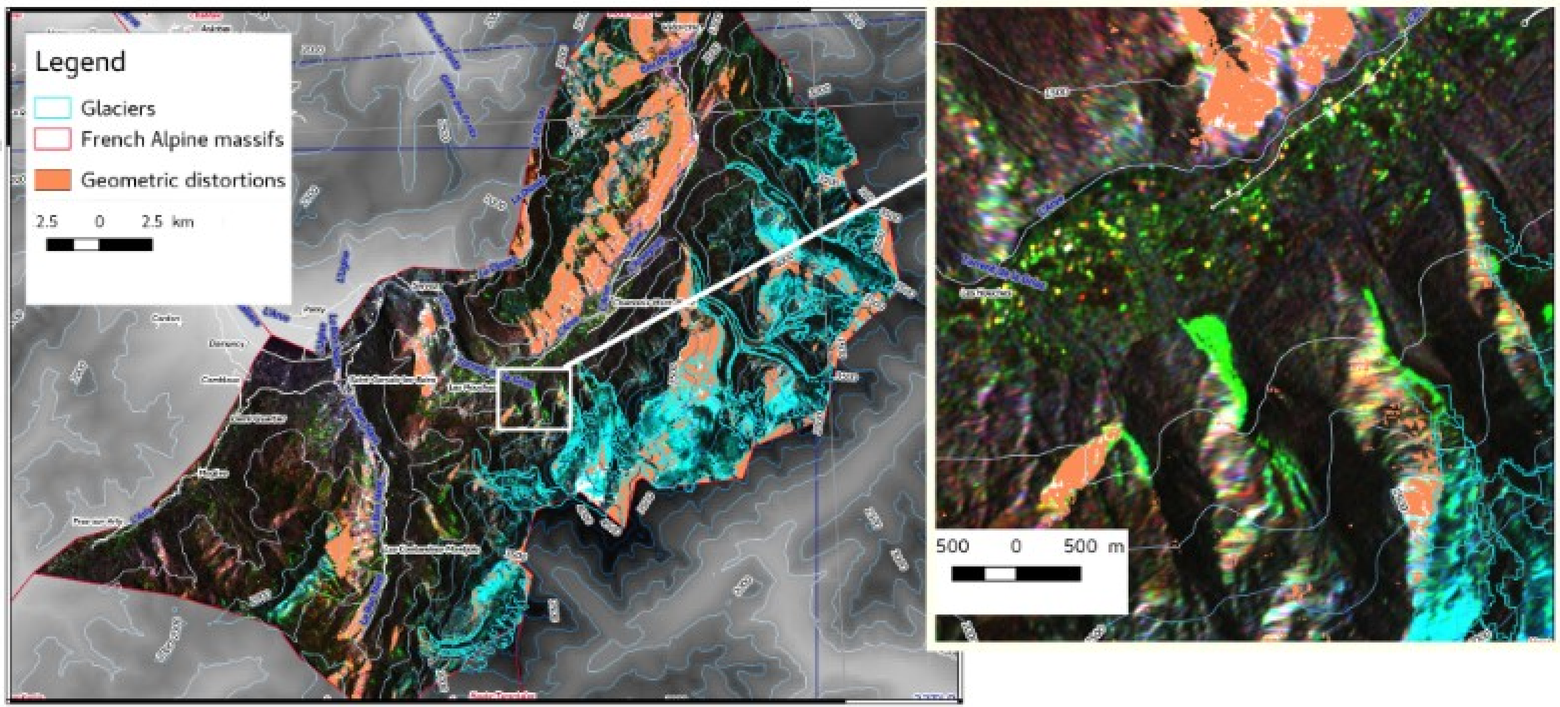


slope angle aspect

Remote sensing data

Model of a Sentinel 1 [wikipedia]

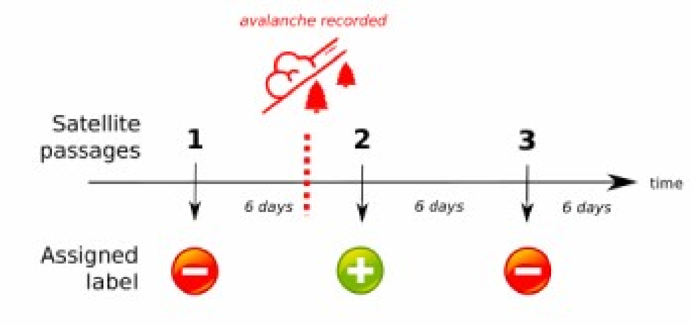
SAR Imagery:

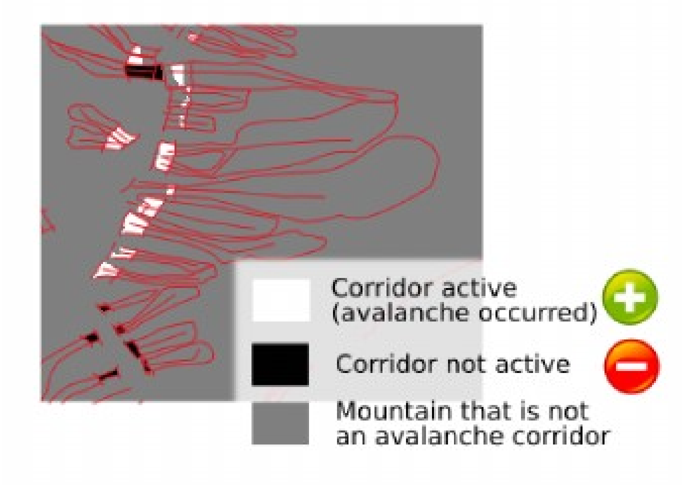


Project Summary

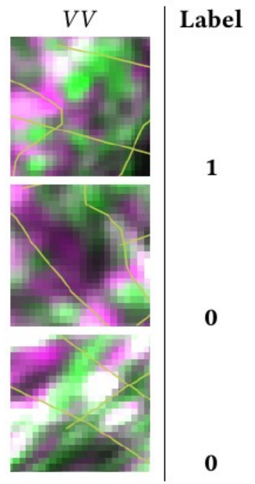
Jakob Schlör

Preprocessing of data

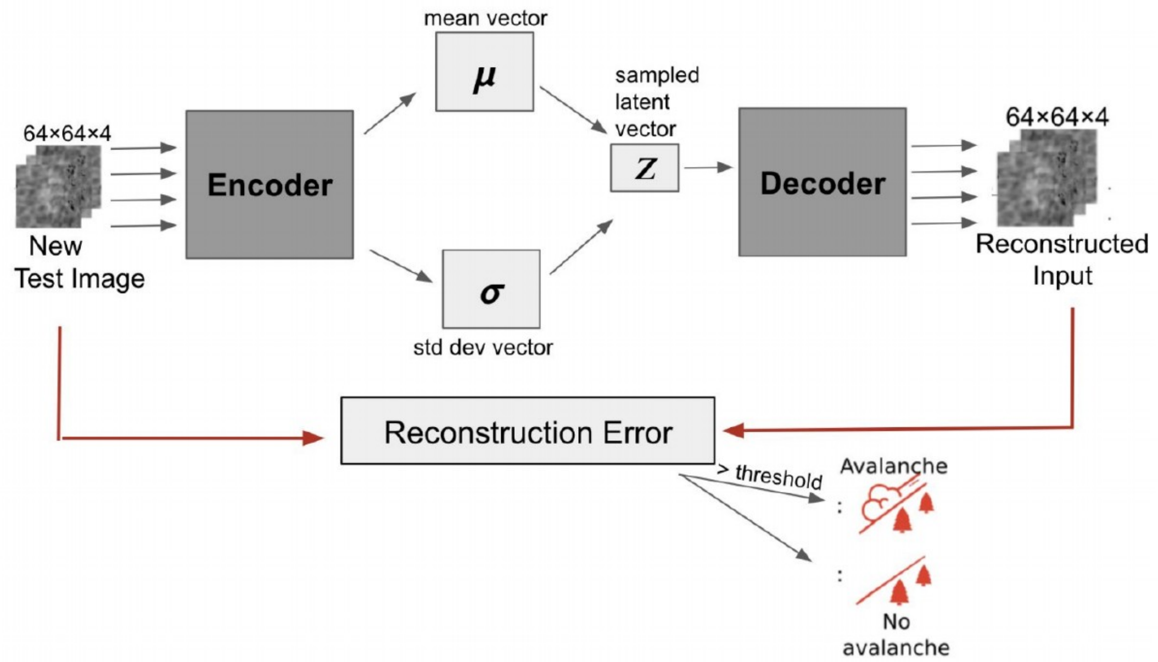




- Consider only 3000 predefined avalanche paths
- Take difference to snow-free summer images
- Take difference to previous satellite image
- Only small amount of data is labeled
- ~1 % of the data includes avalanches



Method



Anomaly score:

1) Encode new test image
2) Draw L samples from p(z|x) and decode
3) If mean reconstruction error > optimal threshold

$$\texttt{AnomalyScore}(x^i) = -\frac{1}{L}\sum_{j=1}^{L}[log(p(x^i|z^j;\theta)]$$

Method

1) Training the VAE

Unsupervised - VAE	Train VAE on unlabeled training data
Semi-supervised - VAE	Train VAE only on image patches showing no avalanches

2) Tune threshold using labeled validation data set

Model evaluation

	All Alps		Haute Maurienne	
	Balanced Accuracy	F1-score	Balanced Accuracy	F1-score
Baseline	0.58	0.05	0.58	0.12
Supervised - CNN	0.53	0.10	0.53	0.12
Semi-supervised - VAE	0.59	0.11	0.6	0.23
Unsupervised - VAE	**0.69**	**0.14**	**0.68**	**0.26**

Recall:

$$\text{Balanced accuracy: } BA = \frac{TPR + TNR}{2}$$

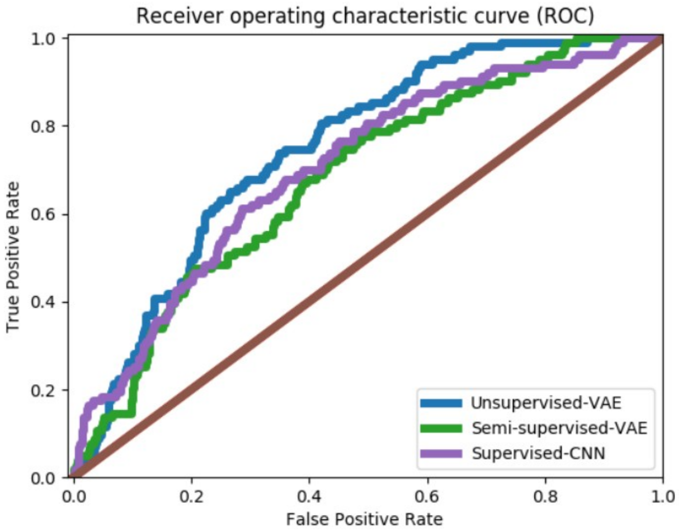
$$\text{F1-score: } F1 = 2 \cdot \frac{PPV \cdot TPR}{PPV + TPR}$$

$$\text{sensitivity: } TPR = \frac{TP}{TP + FN}$$

$$\text{specificity: } TNR = \frac{TN}{TN + FP}$$

$$\text{precision: } PPV = \frac{TP}{TP + FP}$$

Model evaluation



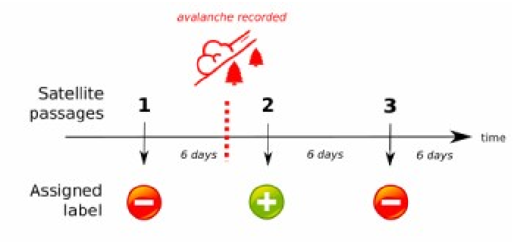
Method	AUC ROC
Supervised - CNN	70.7
Semi-supervised - VAE	68.3
Unsupervised - VAE	**75**

Summary & Discussion

- Propose a semi-supervised approach for anomaly detection with limited amount of labels

- Unsupervised method outperforms semi-supervised VAE,
  - More trainings data
  - Noisier data

- Including one-month-old avalanches improved detection ability



Open Questions

- Is there an upper limit in performance due to anomalies which do not correspond to avalanches?

- Which other variables would improve detection ability?