Application: United Nations Satellite Centre (UNOSAT)

Edoardo Nemni, Data Scientist

Taoyuan Liu, ML Trainee

10/03/2023, CERN IT Machine Learning Infrastructure Workshop

Background of damage assessment from satellite imageries

- Accurate information about the extent of building damage is essential for **humanitarian relief** and **disaster response**
- **Application**: urban planning, population and growth estimation, damage assessment, etc…
- Multi-temporal (pre- and post-) high-resolution satellite images can be used but there are complex challenges:

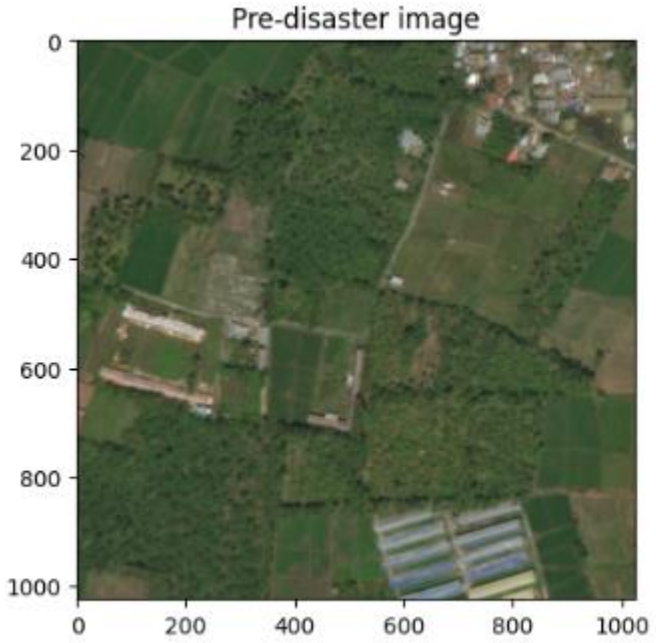
Earthquake

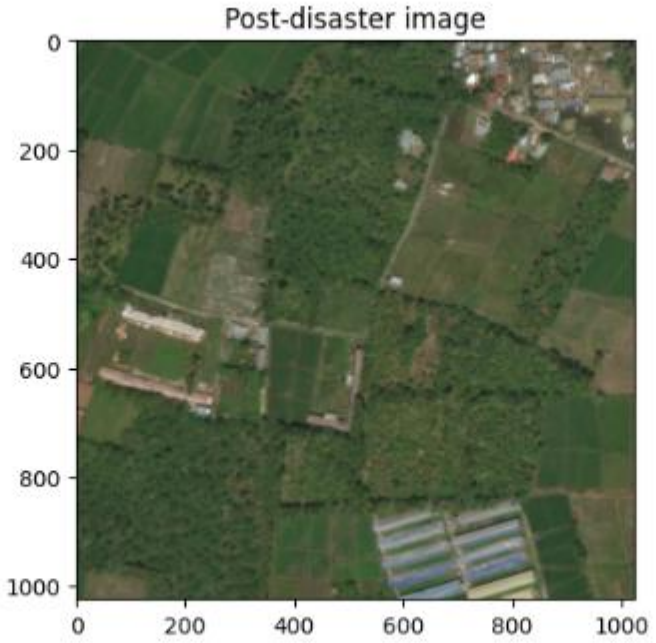
Flood

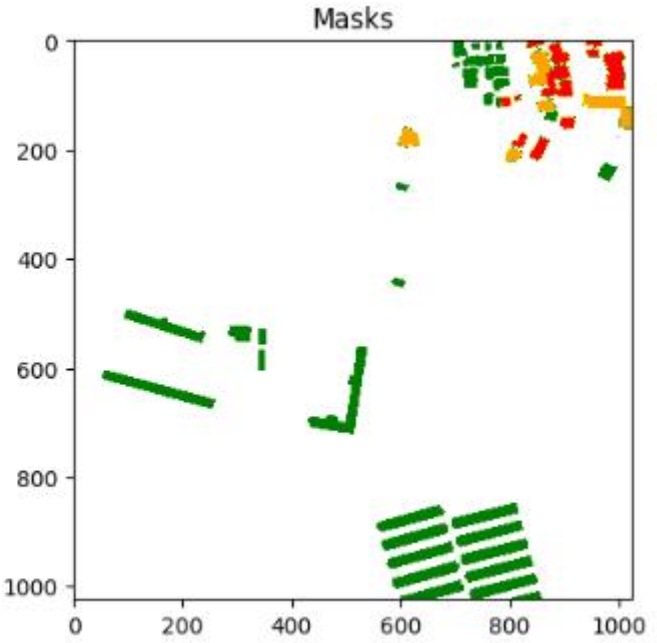
Tsunami

Wildfire

Data

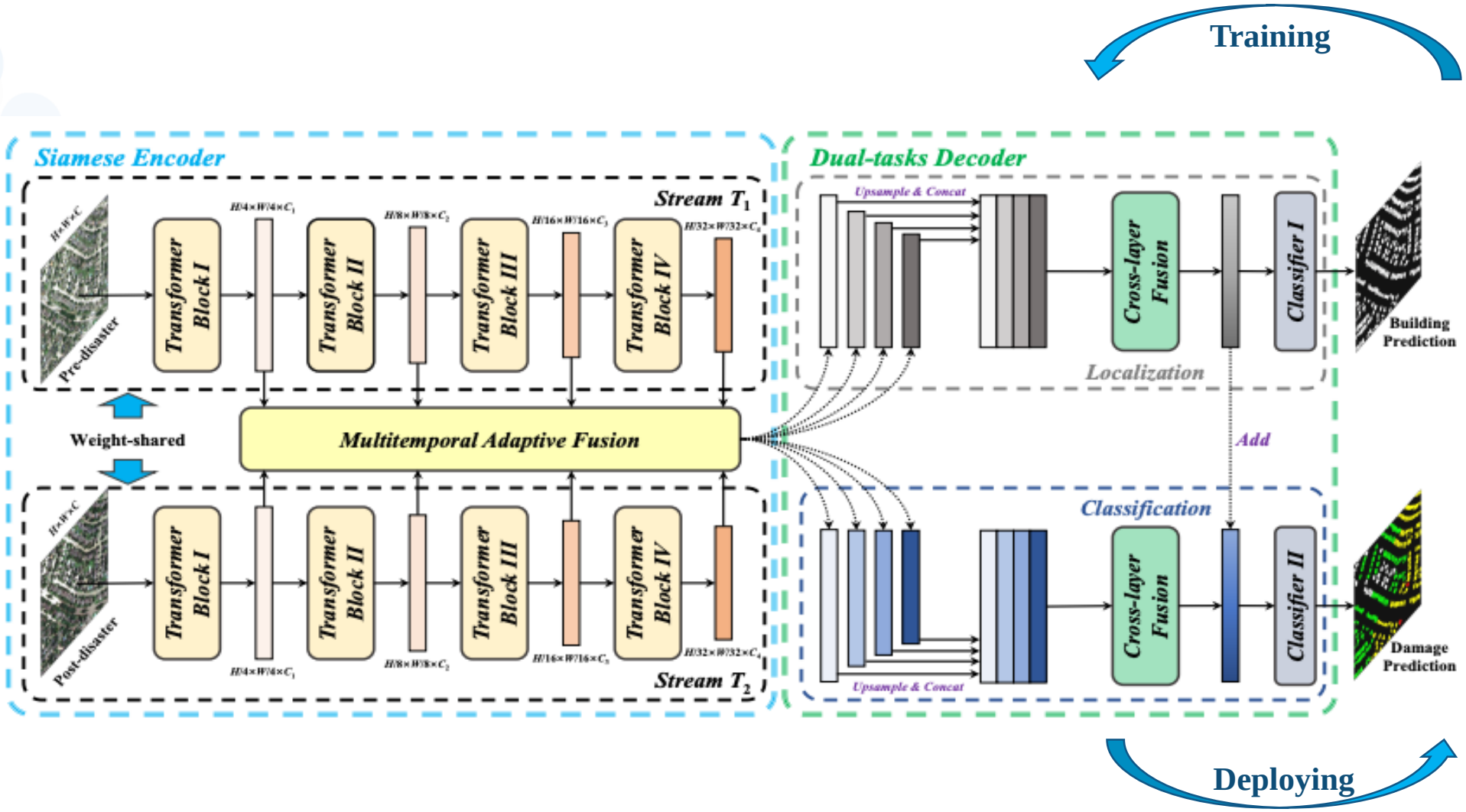


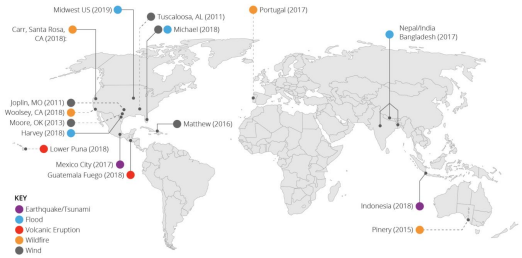




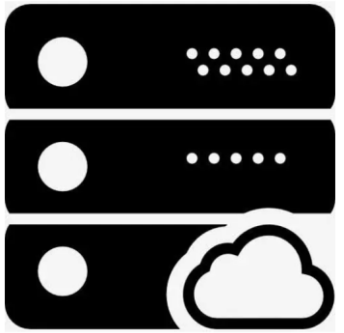
Score	Label	Visual Description of the Structure
0	No damage	Undisturbed. No sign of water, structural damage, shingle damage, or burn marks.
1	Minor damage	Building partially burnt, water surrounding the structure, volcanic flow nearby, roof elements missing, or visible cracks.
2	Major damage	Partial wall or roof collapse, encroaching volcanic flow, or the structure is surrounded by water or mud.
3	Destroyed	Structure is scorched, completely collapsed, partially or completely covered with water or mud, or no longer present.

UNOSAT DamFormer Architecture





xBD Dataset



CERN Server

https://arxiv.org/pdf/2201.10953.pdf

Requirement

- Batch Size: 8
- GPU memory usage: 28 GB
- GPU required: 2*T4 or 1* V100s
- Data transfer worker: 4
- Training time required per epoch: 1 hour
- Epochs required for global best convergence: 100+

ml.cern.ch positive experience

- High availability: Kubernetes supports high availability and self-healing. If a container or node fails, the system can automatically restart or migrate containers to keep the application available.

- Elastic scaling: Using a Kubernetes cluster makes it easy to scale compute resources for training tasks to meet training needs of different sizes without having to manually manage resources.

- Unified management: Using a Kubernetes cluster allows you to unify the management of different types of containers and applications, thereby improving management efficiency and reducing complexity.

ml.cern.ch challenges

- **Availability** of ml.cern.ch
- Max **idle time** < 24h
- **Network communication problems**: Network communication problems can arise when the cluster suddenly loses connection to /eos, leading to data read errors and program interruptions. To mitigate this issue, we have implemented a try/except method to read data that allows for a buffer margin in the event of a reading failure.
- **Pods communication problems**: For a multi-pods tasks, if a pod experiences an error, it can cause all other pods to pause and wait for the faulty pod to reconnect and resume training. However, the current distributed training initialization method of NCCL+:/env (the default method used in the CERN cluster) can cause the error pod to be unable to determine its own rank number after reconnection. This is because the rank number is randomly assigned during initialization. To address this issue, we suggest recording the rank number of initialized pods or modifying the initialization method in the code.

@UNOSAT

@UNITAR.unosat

/UNOSAT

unitar
United Nations Institute for Training and Research

UNOSAT
United Nations Satellite Centre

UNOSAT, United Nations Institute for Training and Research (UNITAR)
7 bis, Avenue de la Paix, CH-1202 Geneva 2, Switzerland

T +41 022 917 4720
E unosat@unitar.org
www.unosat.org