



Finanziato
dall'Unione europea
NextGenerationEU



Ministero
dell'Università
e della Ricerca



Italiadomani

PIANO NAZIONALE
DI RIPRESA E RESILIENZA



Centro Nazionale di Ricerca in HPC,
Big Data and Quantum Computing

HaMMon: Segmentation and Classification of aerial images

Nicoletta Sanvitale (IRA – INAF)

Spoke 3 II Technical Workshop, Perugia, May 26-29, 2025

HaMMon (Hazard Mapping and vulnerability Monitoring) : WP 4

WP4 objective:

Developing vulnerability curves for flood, windstorm, hail, and seismic risks, utilizing the features of buildings that can be automatically identified from images through machine learning algorithms.



AI solutions for automatic extraction of building features



INAF focus on WP4



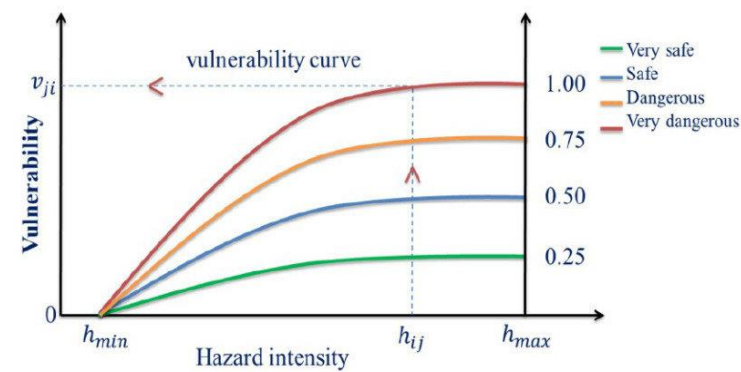
satellite



airborne



Street view images



On-going tasks

Semantic Segmentation using the hybrid architecture TransUnet (Unet architecture + Vision Transformers)

Building Footprint



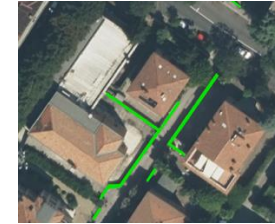
Solar Panels



Skylight

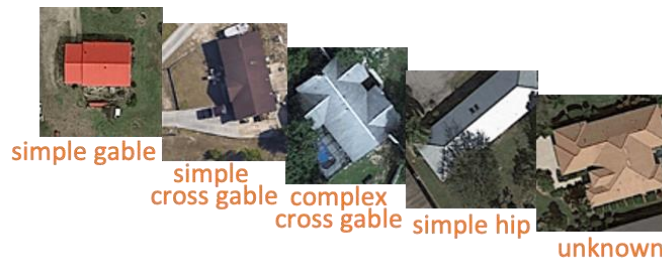


Fence



Classification using BEitV2 (Bidirectional Encoder representation from Image Transformers)
and CNN-based architectures such as EfficientNet

Roof type



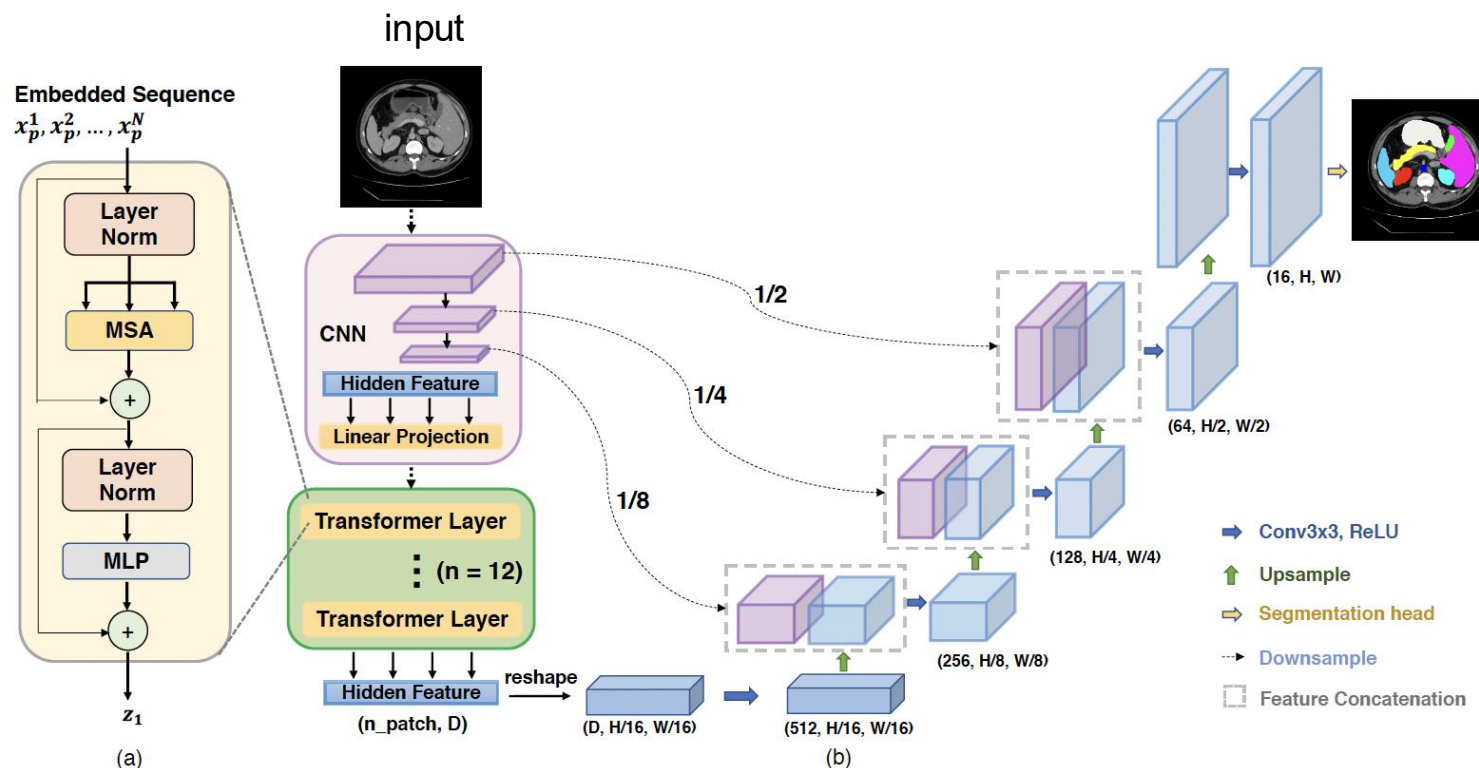
Roof material



Semantic segmentation network

TransUNet is a hybrid model combining:

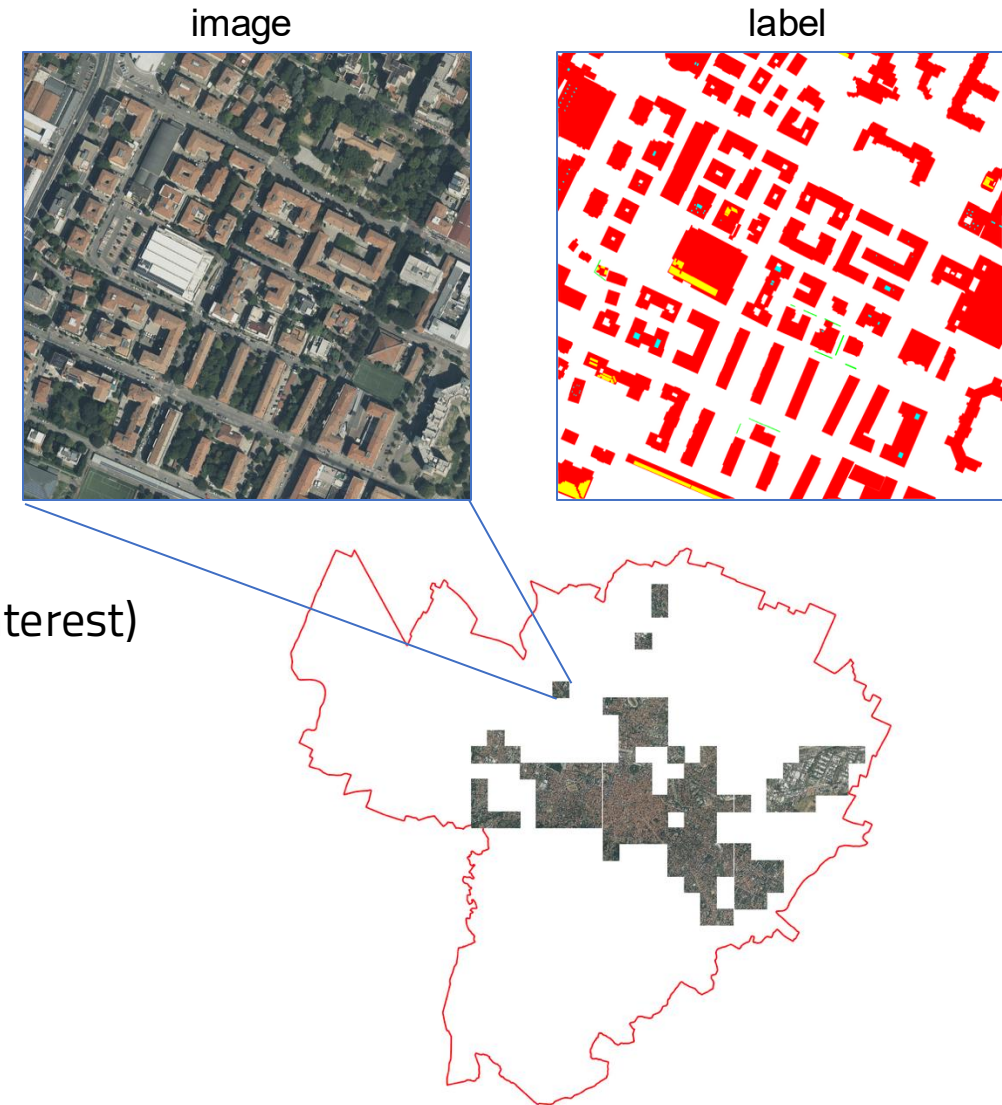
- A **Vision Transformer (ViT)** encoder for capturing global context
- A **U-Net decoder** for precise, pixel-level prediction



Chen, Jieneng, et al. "Transunet: Transformers make strong encoders for medical image segmentation." *arXiv preprint arXiv:2102.04306* (2021).

Segmentation analysis on Bologna dataset

- 115 airborne images (4116×4116 pixels) $\simeq 44 \text{ km}^2$
- spatial resolution of 15 cm.
- RGB imagery
- ground-truth comprising 5 classes:
 - - 70,0 % "other" (areas whose segmentation is not of interest)
 - - 0.17 % "skylight"
 - - 0.34 % "solar panel"
 - - 29.3 % "footprint"
 - - 0.22 % "fence"



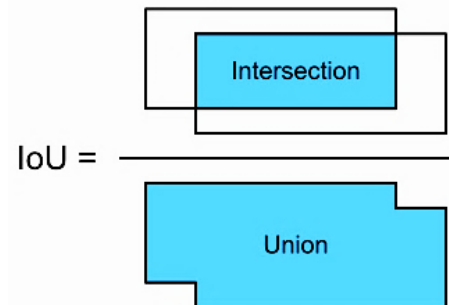
Segmentation: performance metrics

	other	solar panel	skylight	footprint	fence
IoU	0.933 ± 0.003	0.608 ± 0.041	0.142 ± 0.046	0.848 ± 0.008	0.093 ± 0.019
Precision	0.965 ± 0.002	0.756 ± 0.074	0.338 ± 0.066	0.915 ± 0.005	0.303 ± 0.094
Recall	0.965 ± 0.002	0.763 ± 0.062	0.206 ± 0.075	0.921 ± 0.005	0.124 ± 0.04

Overall Results:
Accuracy: 0.948 ± 0.002

$$\text{Precision} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Positives}}$$

$$\text{Recall} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Negatives}}$$



Pretraining on ImageNet21K

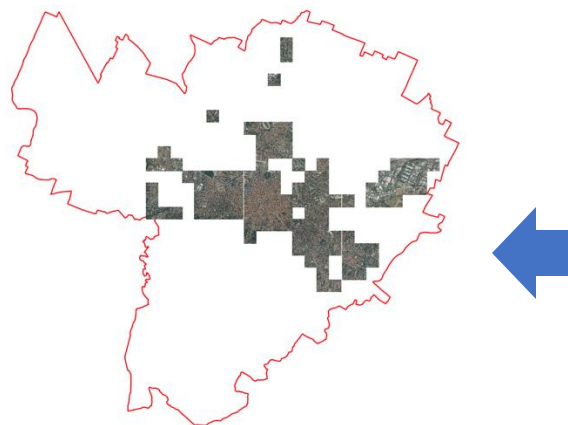
Training: Epochs 60;
Batch size: 24; Tile size: 512

Training dataset: 92 images,
5888 no overlapped tiles (84%
training; 16% validation)

Test Set: 23 cropped images
5175 tiles with a 50% overlap,
metrics calculated on the
merged images

Segmentation: performance metrics

	other	solar panel	skylight	footprint	fence
IoU	0.933 ± 0.003	0.608 ± 0.041	0.142 ± 0.046	0.848 ± 0.008	0.093 ± 0.019
Precision	0.965 ± 0.002	0.756 ± 0.074	0.338 ± 0.066	0.915 ± 0.005	0.303 ± 0.094
Recall	0.965 ± 0.002	0.763 ± 0.062	0.206 ± 0.075	0.921 ± 0.005	0.124 ± 0.04



Pretraining on ImageNet21K

Training: Epochs 60;
Batch size: 24; Tile size: 512

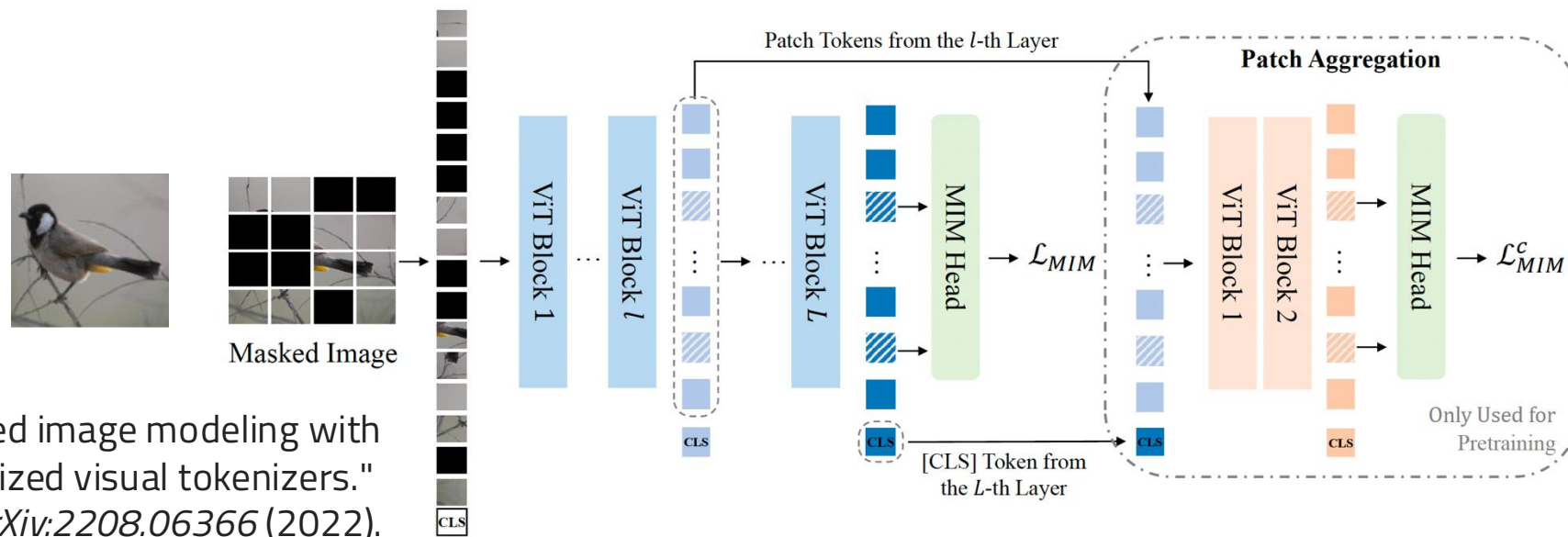
Training dataset: 92 images,
5888 no overlapped tiles (84%
training; 16% validation)

Test Set: 23 cropped images
5175 tiles with a 50% overlap,
metrics calculated on the
merged images

Classification network

BEiT v2 (Bidirectional Encoder representation from Image Transformers) is a **vision transformer (ViT)** model developed by Microsoft Research:

- BEiT v2 is trained using a **masked image modeling (MIM)** task
- The model learns **without any class labels**
- After pretraining, it is **fine-tuned on labeled data** for downstream tasks like classification, using **supervised learning**



Peng, Zhiliang, et al. "Beit v2: Masked image modeling with vector-quantized visual tokenizers."
arXiv preprint arXiv:2208.06366 (2022).

Classification: preliminary performance metrics

- Total buildings in all dataset: 14864, 8 classes
- Model: beit_base_patch16_224
 - ❑ Patch size 16x16 pixel
 - ❑ Input resolution 224 x224
 - ❑ Number of layers (transformer blocks)

Class	No. Instances	% of total
Tetto a due falde	1679	11.30%
Tetto a mezzo padiglione	854	5.75%
Tetto a padiglione	3966	26.68%
Tetto complesso	2519	16.95%
Tetto curvo	136	0.91%
Tetto obliquo	1299	8.74%
Tetto piatto	3602	24.23%
Tetto piramidale	809	5.44%



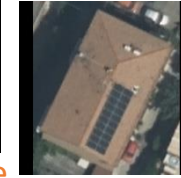
complex roof



curved
roof



simple gable



half hip



Skillion



simple hip



flat



pyramidal

Roof shape

	Precision	Recall
Tetto_complesso	0.573	0.505
Tetto_curvo	0.611	0.407
Tetto_due_falde	0.741	0.649
Tetto_mezzo_padiglione	0.424	0.339
Tetto_obliquo	0.677	0.411
Tetto_padiglione	0.687	0.845
Tetto_piatto	0.789	0.897
Tetto_piramidale	0.723	0.543

Global Precision: 0.653

Global Recall: 0.575

Global Accuracy: 0.692

Classification: preliminary performance metrics

- Total buildings in all dataset: 14864, 8 classes
- Model: beit_base_patch16_224
 - Patch size 16x16 pixel
 - Input resolution 224 x224
 - Number of layers (transformer blocks)

Class	No. Instances	% of total
Tetto a due falde	1679	11.30%
Tetto a mezzo padiglione	854	5.75%
Tetto a padiglione	3966	26.68%
Tetto complesso	2519	16.95%
Tetto curvo	136	0.91%
Tetto obliquo	1299	8.74%
Tetto piatto	3602	24.23%
Tetto piramidale	809	5.44%



complex roof



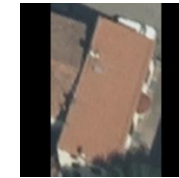
curved roof



simple gable



half hip



Skillion



simple hip



flat



pyramidal

Roof shape

	Precision	Recall
Tetto_complesso	0.573	0.505
Tetto_curvo	0.611	0.407
Tetto_due_falde	0.741	0.649
Tetto_mezzo_padiglione	0.424	0.339
Tetto_obliquo	0.677	0.411
Tetto_padiglione	0.687	0.845
Tetto_piatto	0.789	0.897
Tetto_piramidale	0.723	0.543

Global Precision: 0.653

Global Recall: 0.575

Global Accuracy: 0.692

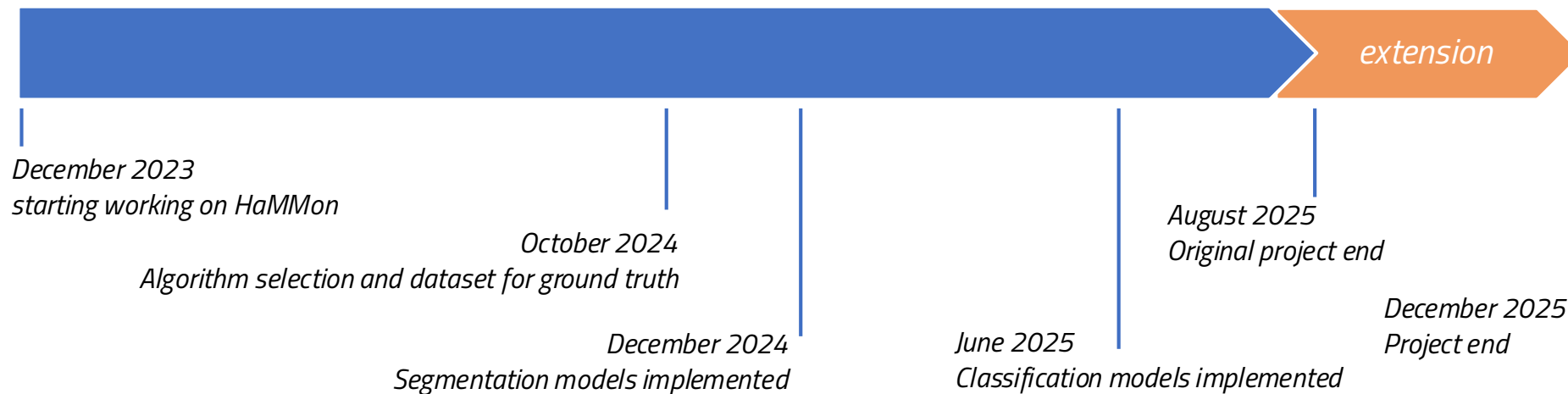
Conclusions

Main achievements:

- Segmentation network implemented on the Bologna Dataset
- Initiated roof shape classification using images from the Bologna dataset

Next steps:

- Develop and implement neural networks for roof material classification
- Extend the analysis to the Milan dataset, which features higher-resolution imagery





Finanziato
dall'Unione europea
NextGenerationEU



Ministero
dell'Università
e della Ricerca



Italiadomani
PIANO NAZIONALE
DI RIPRESA E RESILIENZA



Thanks for your attention!