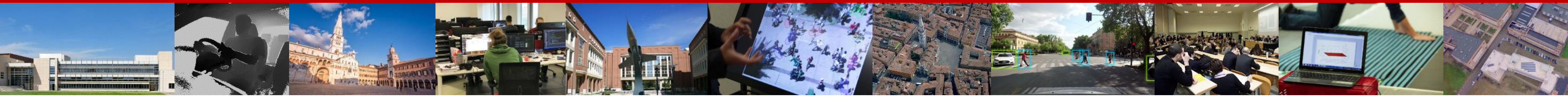
Face Verification from Depth using Privileged Information



Guido Borghi, Stefano Pini, Filippo Grazioli, Roberto Vezzani, Rita Cucchiara

{name.surname}@unimore.it

University of Modena and Reggio Emilia, Italy

BMVC 2018
Newcastle upon Tyne

UNIMORE
UNIVERSITÀ DEGLI STUDI DI MODENA E REGGIO EMILIA

AImage Lab

Centro Interdipartimentale di Ricerca
Softech: ICT per le Imprese

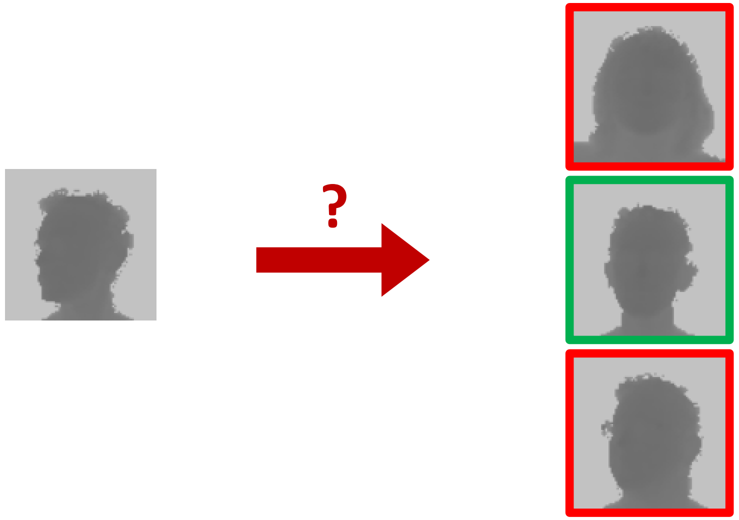
Outline

- Introduction
- Goal & Contributions
- Proposed Model
- Datasets
- Results
- Conclusion

Face Recognition

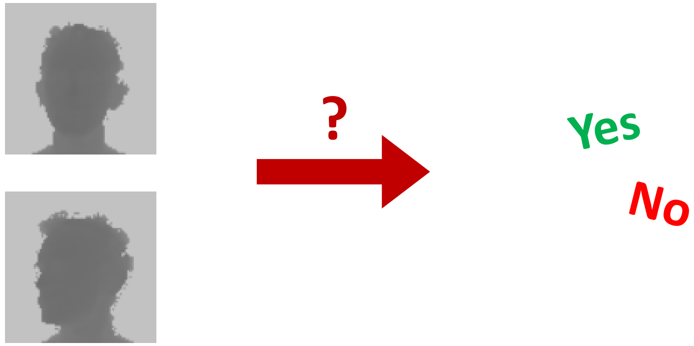
Face identification (*one-to-many*)

Comparing an unknown subject's face with a set of known faces



Face Verification (*one-to-one*)

Comparing two faces in order to determine whether they belong to the same person or not

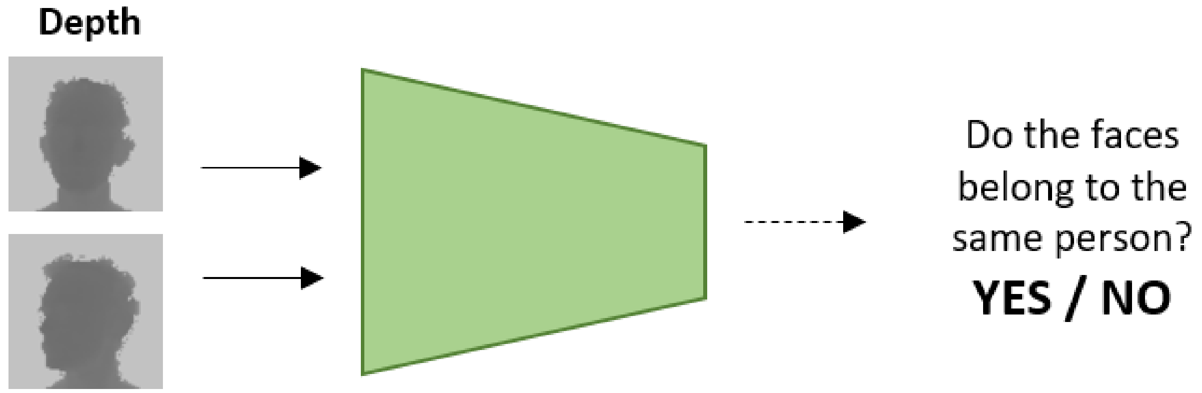


Privileged Information

- Also called **side information**, only few papers have investigated the use of this learning paradigm introduced in 2009 by Vapnik *et al.*[1].
- **Additional (privileged) knowledge** about the training examples is provided **only during the training** phase to **improve the performance** of the system at testing time
- The **privileged knowledge is not available at testing time**, but the system can leverage on the information learned at training time

1. *Vapnik et al., "A new learning paradigm: Learning using privileged information", Neural Networks, 2009.*
2. *Hoffman et al., "Learning with Side Information through Modality Hallucination", IEEE CVPR, 2016.*

Goal & Contributions



- We tackle the *face verification* task analyzing **raw depth images** at testing time
- We use **shallow Siamese networks** to address the **limited size of existing depth-based datasets**
- We exploit the ***learning using privileged information*** paradigm
- We **directly detect the identity of a person** without using strong a-priori hypotheses, like facial landmarks or nose tip localization

JanusNet Architecture

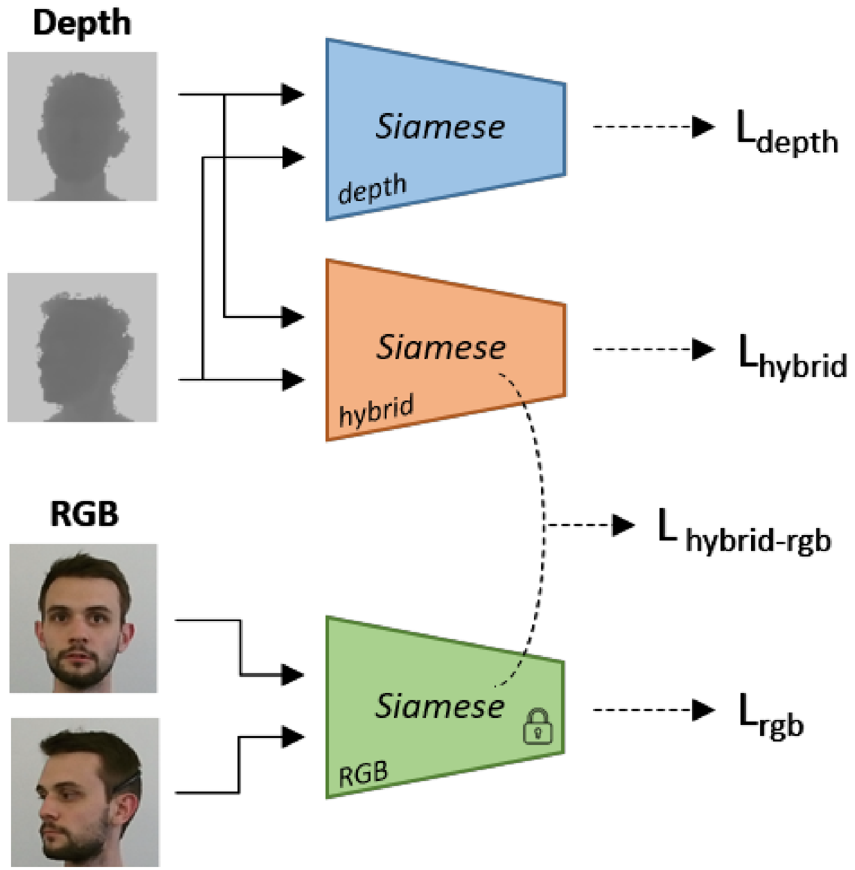
Training phase:

- The model is composed of three modules: a *depth*, an *hybrid* and an *RGB* Siamese network
- Each *Siamese* network predicts the similarity between an input pair of images
- Privileged Information Loss:

$$L_{hybrid-rgb_{1,2}} = \frac{1}{N}\sum_n^N (y_n^{hybrid} - y_n^{rgb})^2$$

- Final Loss:

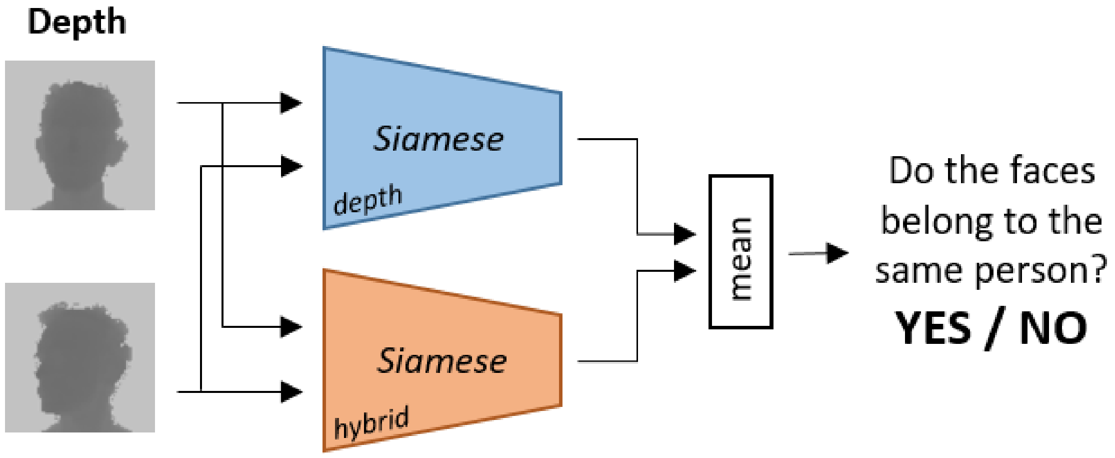
$$L = \alpha\left(L_{hybrid-rgb_1} + L_{hybrid-rgb_2}\right) + \beta\left(L_{depth} + L_{hybrid} + L_{rgb}\right)$$



JanusNet Architecture

Testing phase:

- Only the depth and the hybrid network are employed for the face verification task



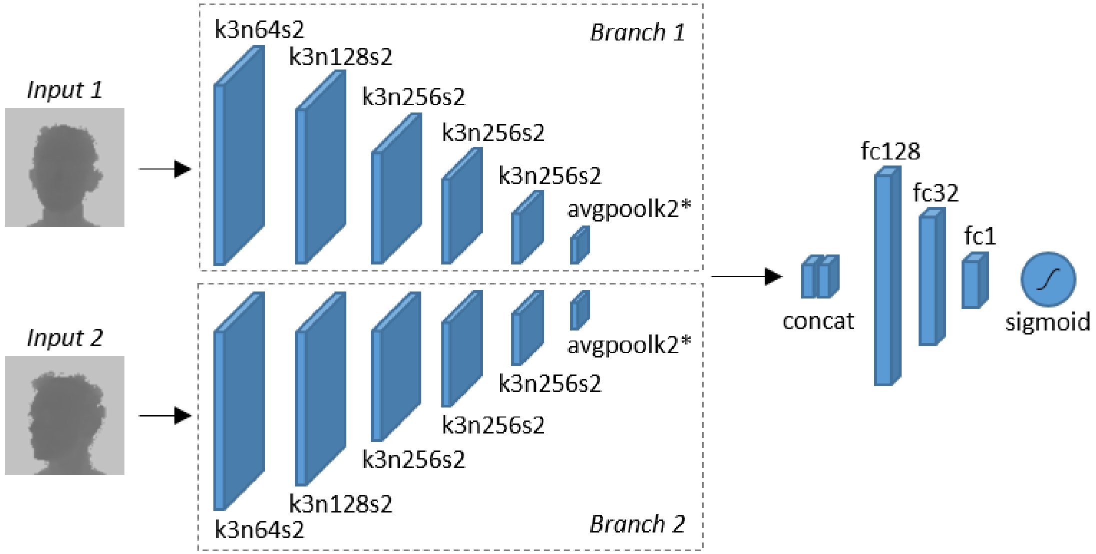
Siamese Architecture

Input: a couple of facial depth maps or RGB images.

Dynamic crop: $$w, h_H = \frac{f_{x,y} \cdot R_{x,y}}{D}$$

Output: a similarity score in the $[0,1]$ range

Network architecture: k, n, s, fc correspond to kernel size, no. of feature maps, stride and units

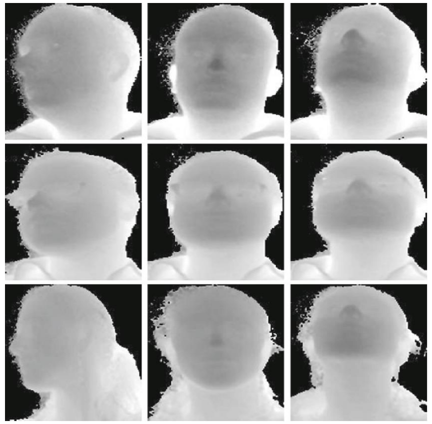


Pandora Dataset[1]

- **250k frames** from 110 sequences
- **22 subjects** (10 males and 12 females)
- Both **depth and RGB frames** and skeleton annotations frame by frame
- Acquired with **Microsoft Kinect One v2**
- Designed for Head and Shoulder Pose estimation

HRRFaceDatabase[2]

- 20k frames
- 18 subjects
- Depth frames only
- Acquired with **Microsoft Kinect One v2**
- Designed for **Face Recognition**



1. G. Borghi et al. *"Poseidon: Face-from-depth for driver pose estimation", CVPR, 2017.*
2. T. Mantecon et al. *"Depth-based face recognition using local quantized patterns adapted for range data", ICIP, 2014.*

Ablation Study

- We evaluated the privileged information framework, verifying that the JanusNet architecture outperforms the single Siamese models.
- Besides, we verified that the face verification accuracy of the proposed model was comparable to a well-known deep architecture pre-trained on RGB images.

Model	Data type	Accuracy
FaceNet[1]	RGB	0.8232
Hybrid network	Depth	0.7553
RGB network	RGB	0.7631
Depth network	Depth	0.7950
JanusNet	**Depth**	**0.8142**

Accuracy for the face verification task on the Pandora test set.

1. *Schroff et al., "FaceNet: A Unified Embedding for Face Recognition and Clustering", CVPR, 2015.*

Comparison on HRRFaceDataset

- We compare our model with [1, 2] for the face identification task.

	Pegasos SVM				JanusNet		
	LBP	SIFT	DLQP	Bag-D3P	*max*	*avg*	*voting*
Accuracy	0.5917	0.7194	0.7347	0.9430	0.9756	**0.9877**	0.9804
Improvement	-	+12.7	+14.3	+35.1	+38.4	**+39.6**	+38.9

Accuracy for the face identification task on the HRRFaceD dataset.

- To deal with the one-to-many comparison, JanusNet is used to obtain a similarity score between every possible pair of images contained in the dataset. Then, to determine the final identity, we combine the results with the following functions.

$$y = \operatorname*{argmax}_i J(s, s'),\ \forall s' \in S_i \qquad y = \operatorname*{argmax}_i \operatorname*{avg}_{s' \in S_i} J(s, s') \qquad y = \operatorname*{argmax}_i \#\{S_i \mid J(s, s') > t\},\ \forall s' \in S_i$$

Competitor results are taken from:

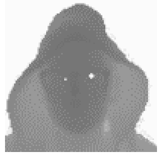
1. *T. Mantecon et al. "Depth-based face recognition using local quantized patterns adapted for range data", ICIP, 2014.*
2. *T. Mantecon et al. " Visual face recognition using bag of dense derivative depth patterns", IEEE SPL, 2016.*

Results

Pandora Subsets

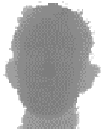
Sequence-based splits:

- $\{S_1, S_2, S_3\}$: actions performed with constrained movements
- $\{S_4\ , S_5\}$: complex movements, occlusions and camouflage



Angle-based splits:

- $\{A_1\} = \{s_{\varrho\theta\sigma} \mid \forall\gamma \in \{\varrho, \theta, \sigma\}\colon\ -10° \leq \gamma \leq 10°\}$ frontal faces
- $\{A_2\} = \{s_{\varrho\theta\sigma} \mid \exists\gamma \in \{\varrho, \theta, \sigma\}\colon\ \gamma < -10° \vee \gamma > 10°\}$ non-frontal faces
- $\{A_3\} = \{s_{\varrho\theta\sigma} \mid \forall\gamma \in \{\varrho, \theta, \sigma\}\colon\ \gamma < -10° \vee \gamma > 10°\}$ extremely-rotated faces



where ϱ, θ, σ are roll, pitch and yaw, respectively.

Pandora Subsets

Verification accuracy on different sequence subsets

Train \ Test	$\{S_1, S_2, S_3\}$	$\{S_4, S_5\}$	$\{S_1, S_2, S_3, S_4, S_5\}$
$\{S_1, S_2, S_3\}$	0.8442	0.7464	0.7734
$\{S_4, S_5\}$	0.7921	0.7127	0.7426
$\{S_1, S_2, S_3, S_4, S_5\}$	0.8049	0.7323	0.7620

Verification accuracy on different angle subsets

Train \ Test	A_1	A_2	A_3	$\{A_1, A_2\}$
A_1	0.8016	0.6603	0.6179	0.6888
A_2	0.8337	0.7859	0.7664	0.7950
A_3	0.5054	0.5028	0.5044	0.5002
$\{A_1, A_2\}$	0.7984	0.7505	0.7273	0.7620

Conclusion

- We proposed a framework, namely ***JanusNet***, that tackles the face verification task using only **depth maps** at testing time.
- During the training procedure, the model can leverage on **RGB images**, provided as **privileged information**, to improve its performance at testing time, when only depth maps are available.
- Existing depth datasets provide depth maps of a limited number of subjects. Bigger datasets would be useful to further analyze the proposed model and broadly evaluate its generalization capabilities.

True Positives

False Positives

False Negatives

Thank you for your attention

Face Verification from Depth using Privileged Information

Guido Borghi, Stefano Pini, Filippo Grazioli, Roberto Vezzani, Rita Cucchiara

{name.surname}@unimore.it

University of Modena and Reggio Emilia, Italy

Thank you for your attention

Face Verification from Depth using Privileged Information

Guido Borghi, Stefano Pini, Filippo Grazioli, Roberto Vezzani, Rita Cucchiara

{name.surname}@unimore.it

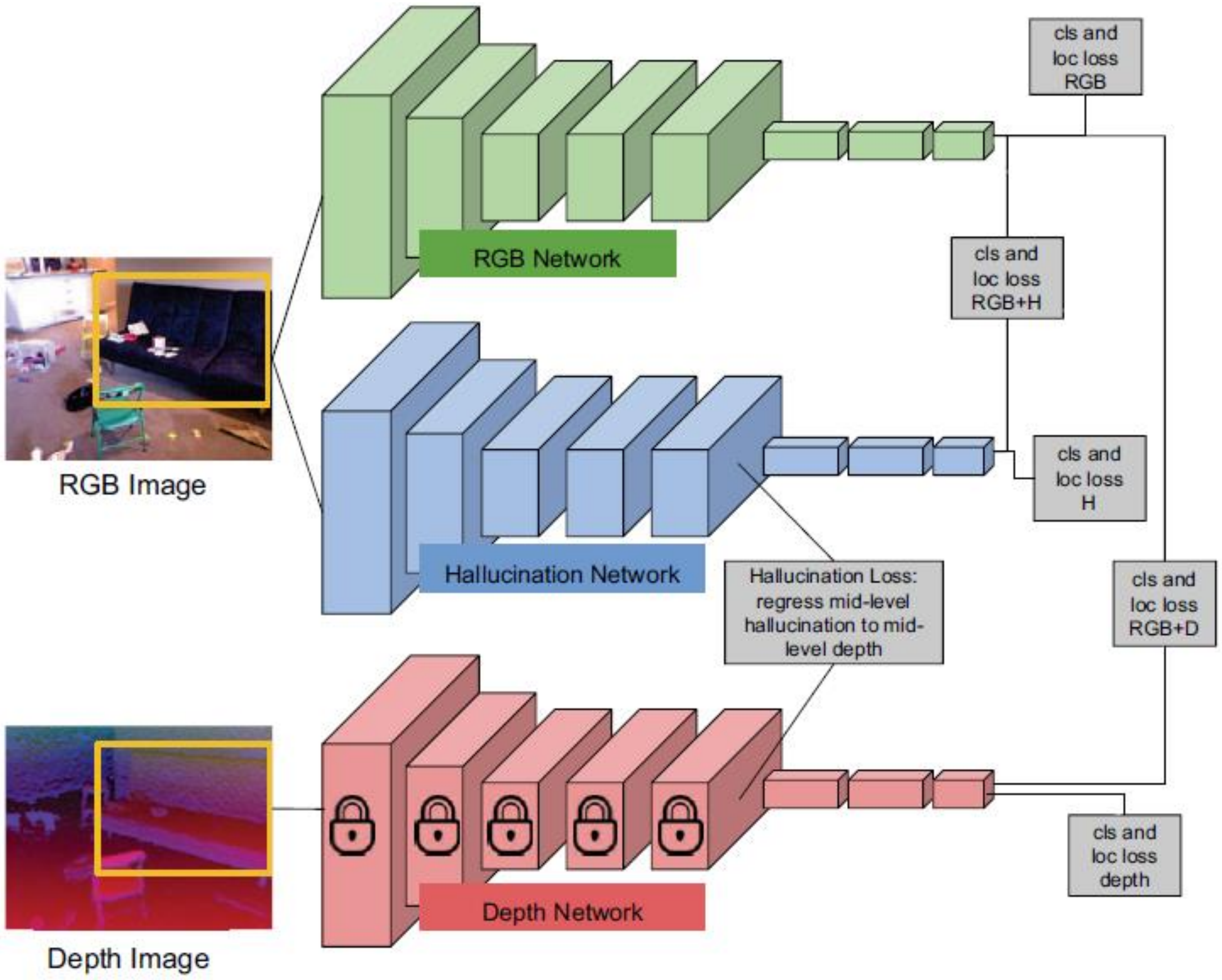
University of Modena and Reggio Emilia, Italy

Pandora Sequences

Frames from sequences $\{S_1, S_2, S_3\}$

Frames from sequences $\{S_4, S_5\}$

Learning with Side Information through Modality Hallucination



1. Hoffman et al., *"Learning with Side Information through Modality Hallucination", IEEE CVPR, 2016.*

Thank you for your attention

Face Verification from Depth using Privileged Information

Guido Borghi, Stefano Pini, Filippo Grazioli, Roberto Vezzani, Rita Cucchiara

{name.surname}@unimore.it

University of Modena and Reggio Emilia, Italy