3-dimensional (3D) urban mapping

A study of detection and reconstruction of building's facade through Structure-from-Motion (SfM) and Convolutional Neural Network (CNN)

MSc. Rodolfo Georjute Lotte

Doctoral thesis defense – Remote Sensing
Dr. Luiz E. O. C. Aragão and **Dr. Yosio E. Shimabukuro**
São José dos Campos, SP
August 24th, 2018

3D Mapping

- The **maps** in our lives
  - The **systematic monitoring** of forests
  - The many **urban influences** over our lives
  - The many kinds of **grown patterns**
  - **Populational** distribution
  - **Violence rates** by sectors
  - The **local** or **global** influences of deforestations

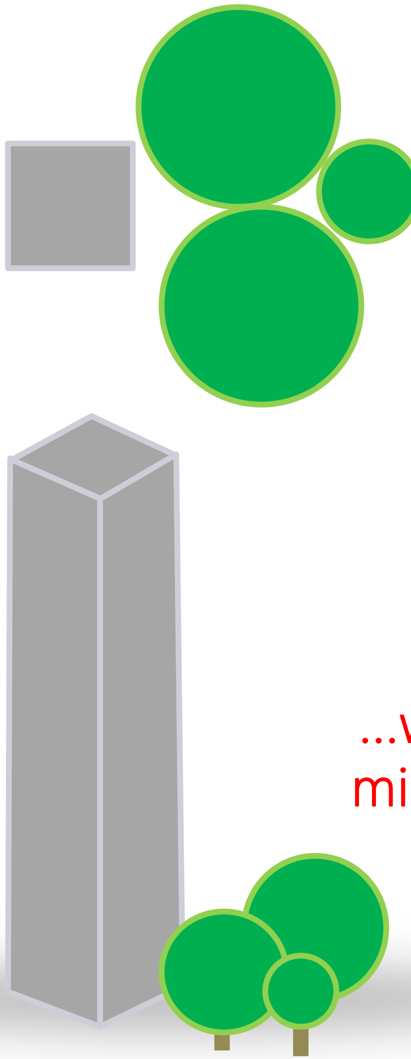
3D Mapping

- The era of **new domains**, **crownsourcing**, **social media**, **powerful computers**
- **Easy access** to sofisticated **plataforms** and **sensors**
- **High resolutions** and **processing** are now feasible
- **New scenario**, of many **alternatives** of research

The amount of trees is higher than concrete?

2D

3D

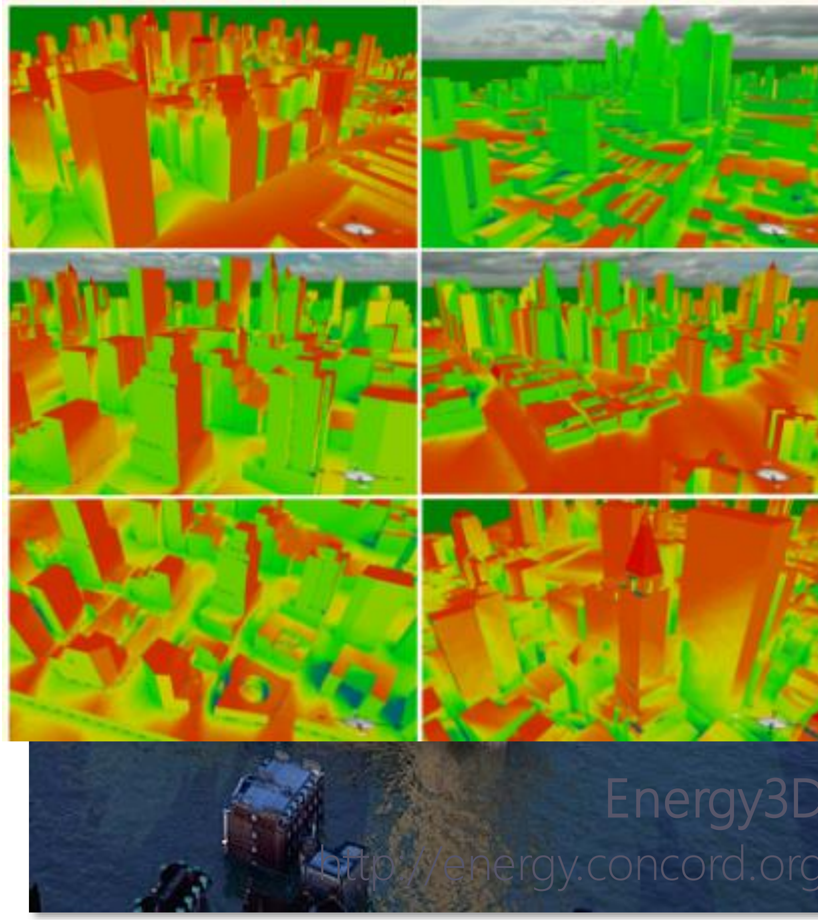


...wait a minute!

 Thesis : Pages: 1 to 3

3D Mapping

- What is the **volumetry** role on urban monitoring?
  - **Installation of solar panels**
  - **Best spots** for surveillance cameras
  - **Places** that will **never** receive sun light
  - Streets with **poor lighting**
  - **Evacuation** plans in case of flood
  - **More accurate** simulations, taking in account the volume



Vertex
Vertex Modelling: 3D Model of London. Vertex Modelling. 2015

Biljecki, Filip, et al. "**Applications of 3D city models: State of the art review**." *ISPRS International Journal of Geo-Information* 4.4 (2015): 2842-2889.

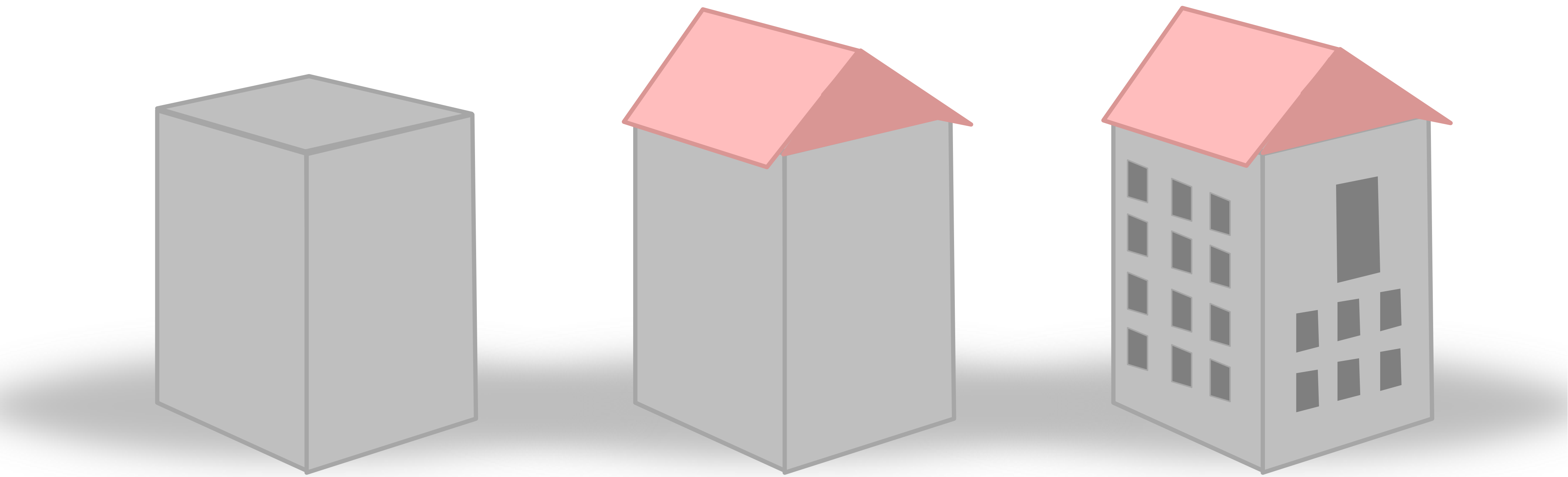
3D Mapping

- Generation of **semantic maps**
- **Once semantic**, the relation between objects can be performed
- Simple **queries**, enormous results

From roof polygons to **simple** 3D building **extrusion** in CityGML (TU Delft)

https://github.com/tudelft3d/3dfier

3D Mapping



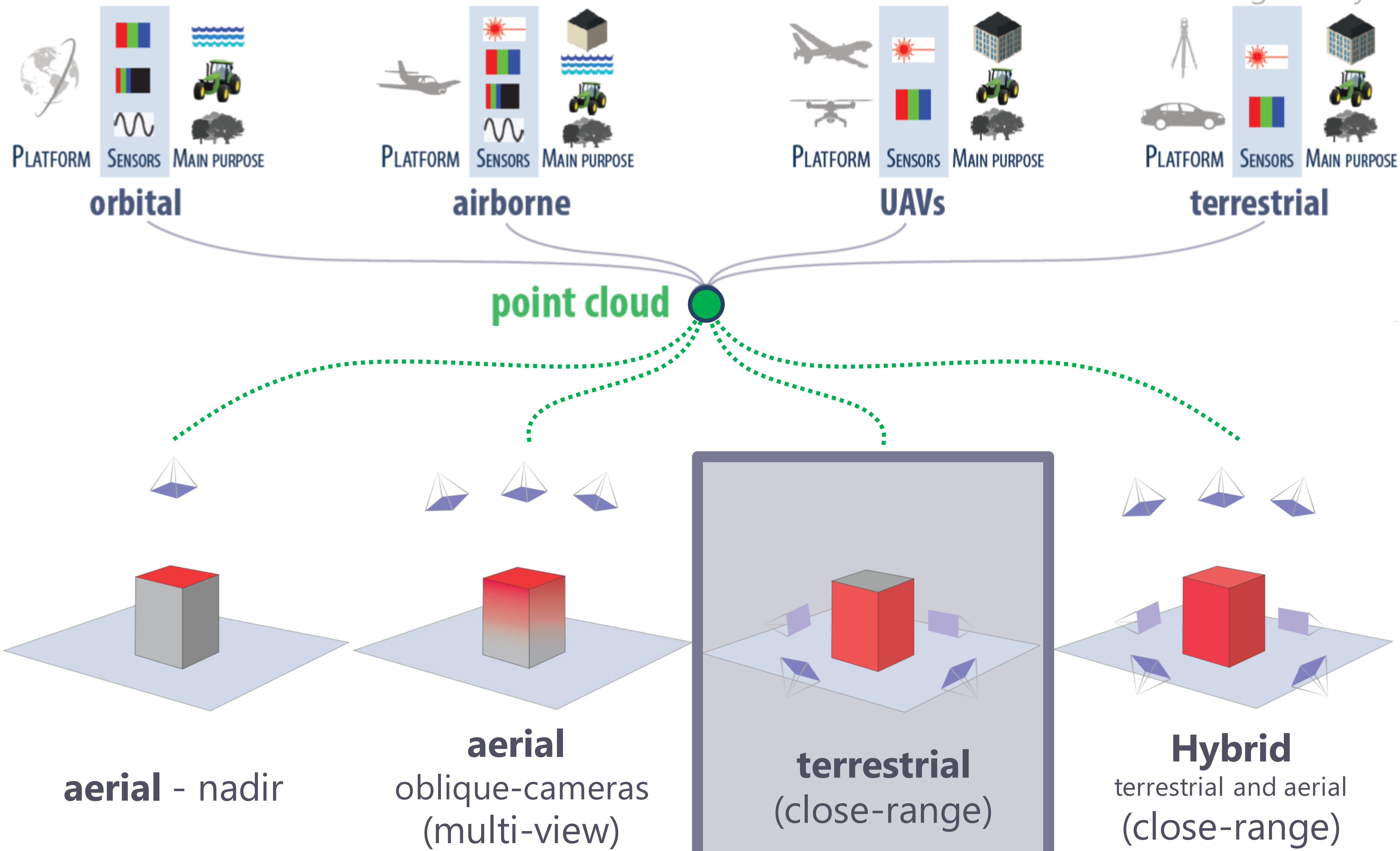
2000
airborne LiDAR

2010
airborne
LiDAR + Image-based

2015~
airborne
LiDAR + Image-based

Structural data

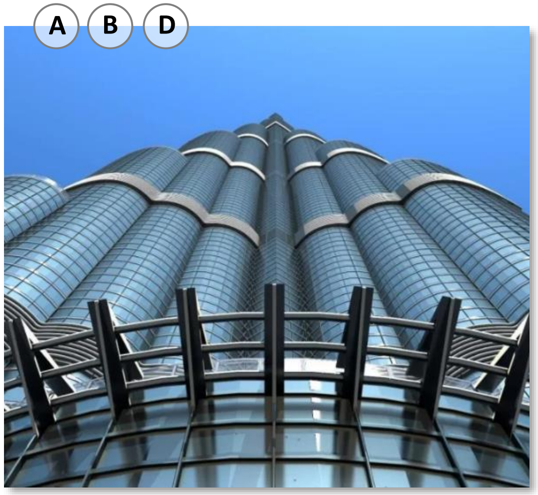
3D mapping
Geometry extraction
Deep-Learning in image analysis

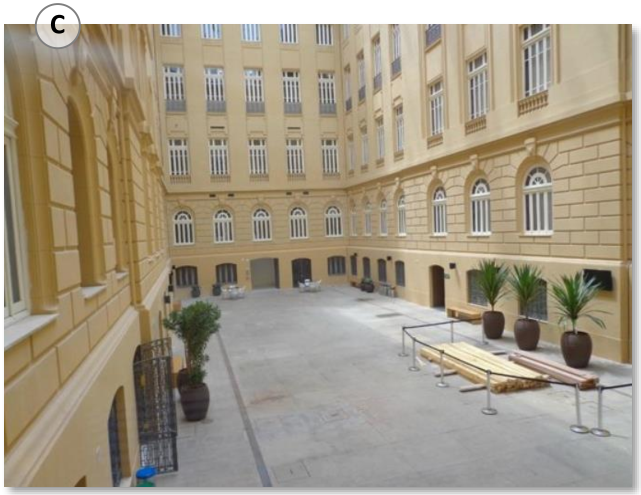


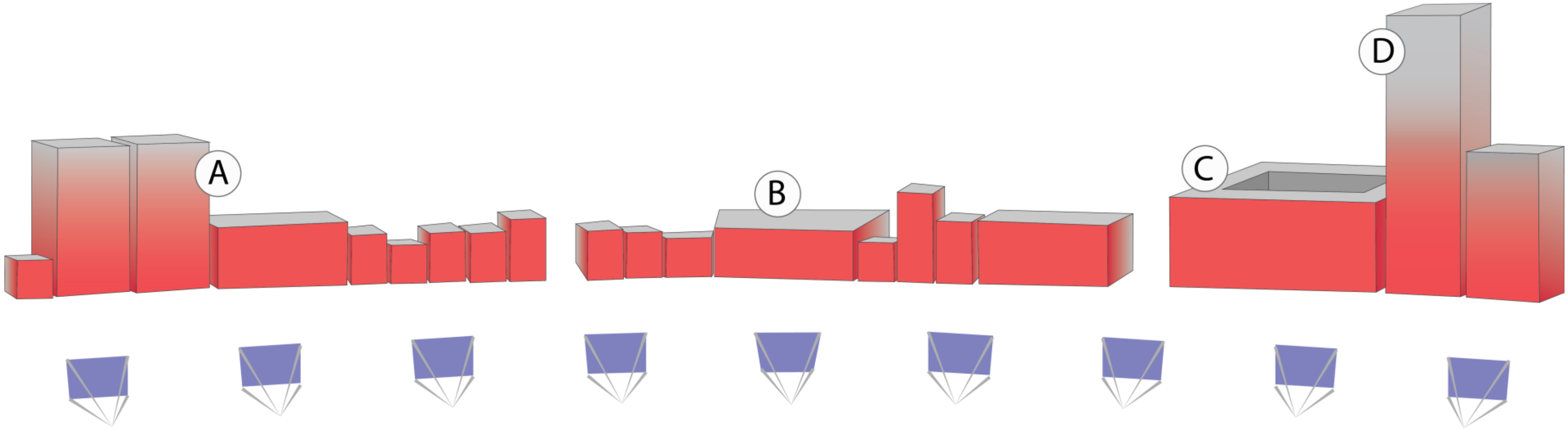
Challenges in urban scenarios

Terrestrial campaigns

- (A) Only front-view
- (B) Rooftops
- (C) Courtyards
- (D) High structures





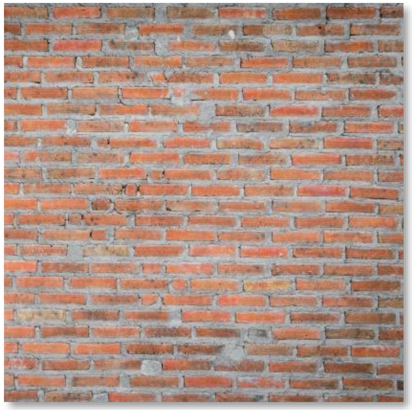


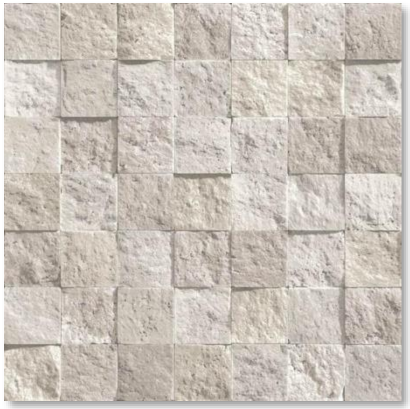
Thesis : Pages: 11 to 12

Challenges in urban scenarios

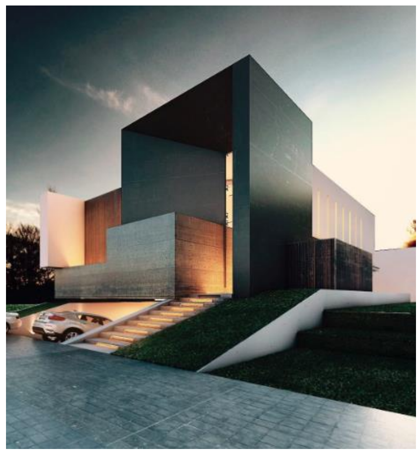
What are we dealing with?

- Texture-compositions
- Height
- Urban sector
- Shape
- Caotic-dynamic






Challenges in urban scenarios





Photogrammetry and close-range acquisition

- 3D mapping involves **Computer Science**, **Photogrammetry** and **Cartography**
- 3D cities **before** and **now**

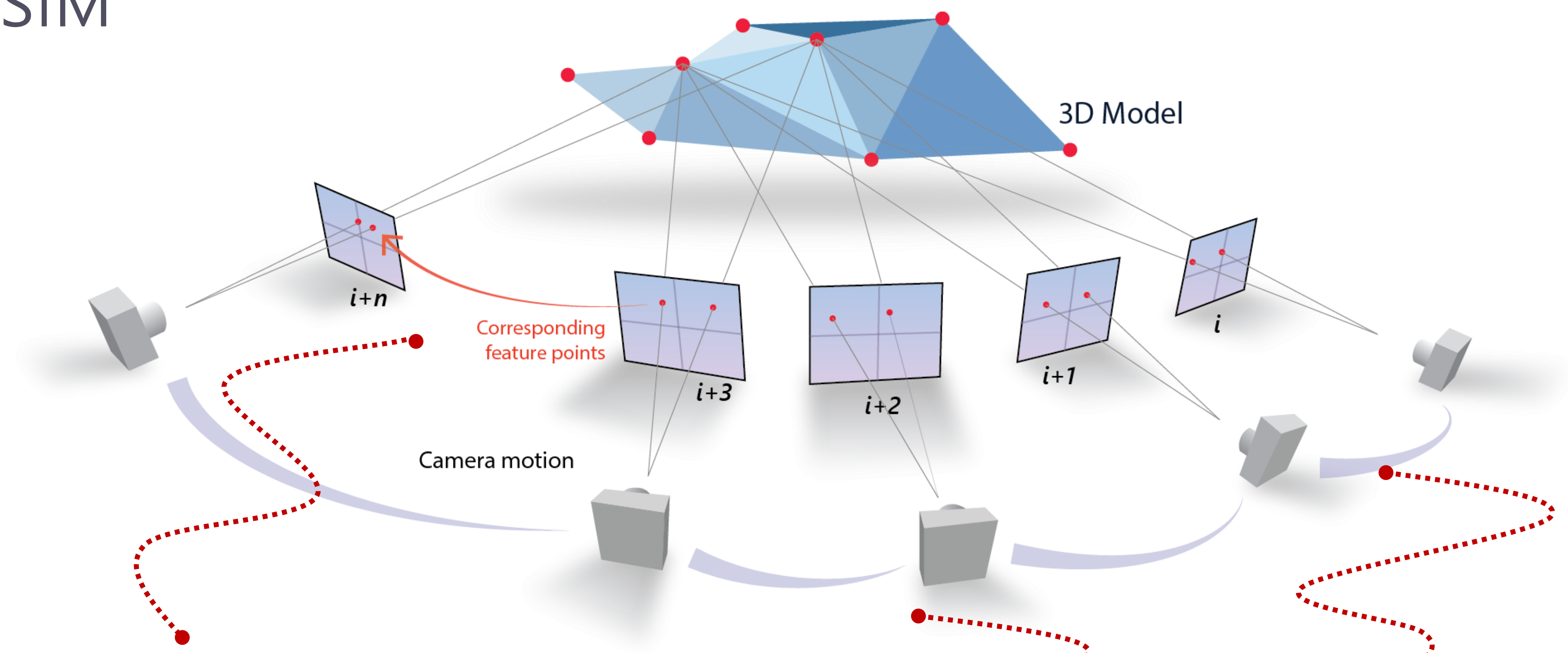
Then, how to improve the level of urban analysis?

How accurately extract the **geometry of urban elements**?

How **to detect** interest features on such scenes?

Advances in 3D reconstruction

SfM

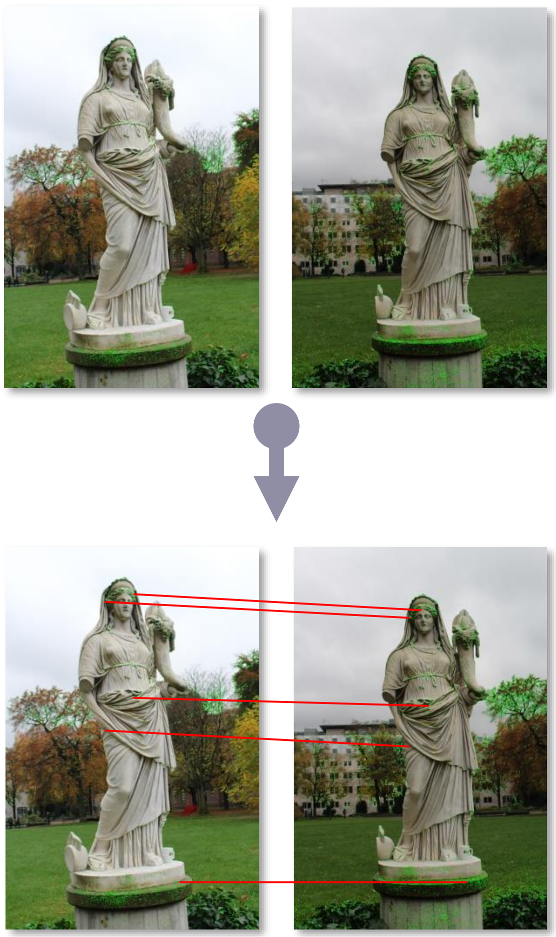


Corresponding points are expressed by the **fundamental matrix**, which is estimated with at least seven corresponding points between pairs of images

Cameras shooting **perpendicular** to the target

Small baselines to allow the overlaping among the images

Advances in 3D reconstruction



Advances in 3D reconstruction

Sparse PC

Dense-based geometry

Texturized geometry

Advances in 3D reconstruction

The Bundler

A SfM solution to reconstruct objects from unordered set of images (from internet)

Rome, Italy

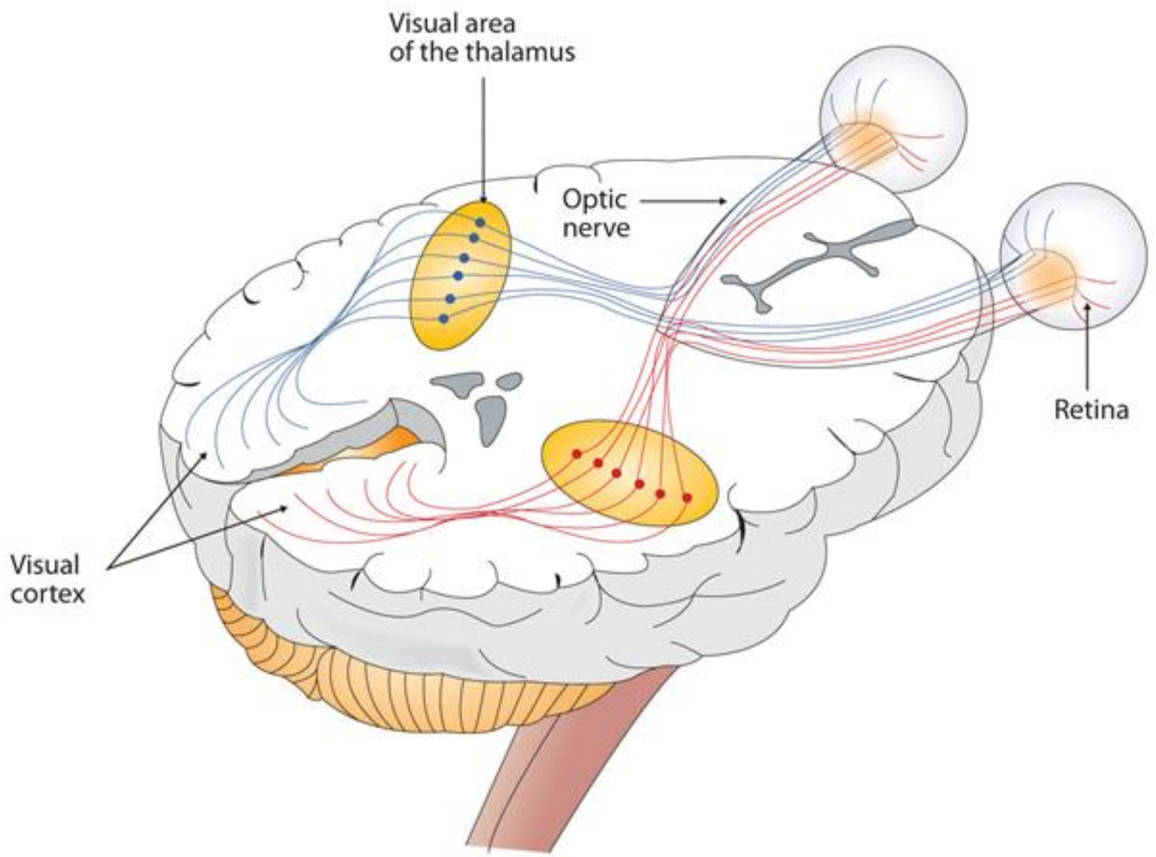
*Snavely, N. Bundler: **Structure from motion (SFM) for unordered image collections**. 2010*

Advances in 3D reconstruction

- What is **Deep-Learning**? Why this name?
- How **does it work** and actually **learn**?
- In which **natural behaviour** the concept was **based** on?
- How the **Pattern Recognition** has been **reformulated** and what is the easy-to-work **frameworks**?
- Is there a **limit** for Deep-Learning on image analysis?

Neural Network

Hierarchical learning of representations



Hubel, David H., and Torsten N. Wiesel. "**Receptive fields and functional architecture of monkey striate cortex**." The Journal of physiology 195.1 (1968): 215-243.

Neural Network

 Thesis : Pages: 27 to 29

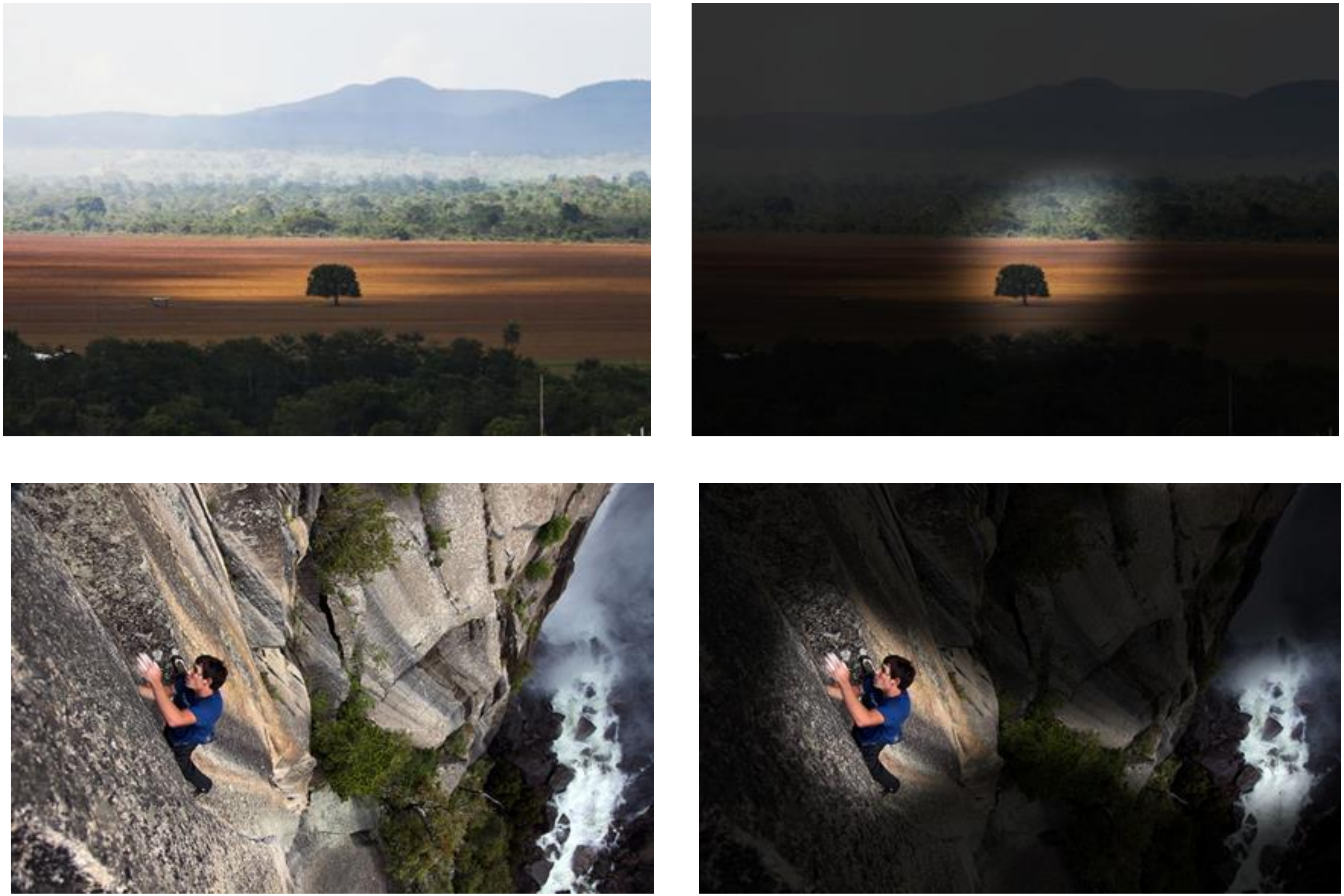
Neural Network

Neural Network

Neural Network

 Thesis : Pages: 27 to 29

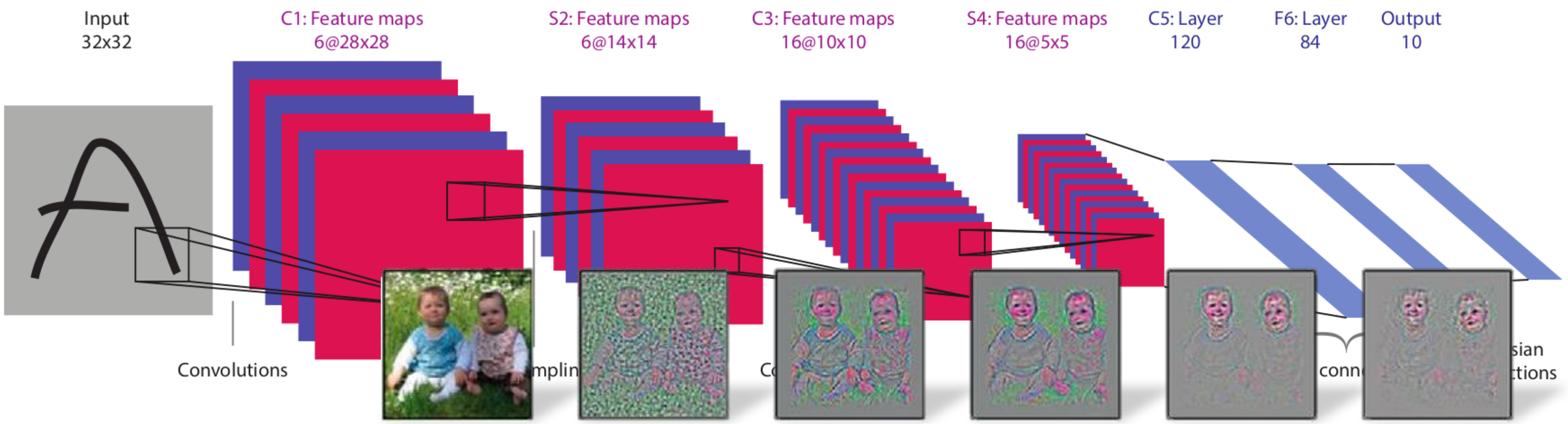
Neural Network



Neural Network

Convolutional Nets

- **Deep** neural architecture = number of internal layers
- Multidimensional inputs (**images**)
- **Learning** occurs by multiple convolutional operations and scale (**downsampling** – but same kernel size)
- **Kernel-filters** represent the **weights**
- Different **set of filters**, by different layers



LeCun, Yann, et al. "**Handwritten digit recognition with a back-propagation network**." Advances in neural information processing systems. 1990.

Neural Network

- **Artifical intelligence** software will able to detect images of child abuse.
- The system **still can't tell** the difference between a **desert** and a **naked body**

Hypotheses

- The **volumetry of buildings**, as well as their **facade features**, can be accurately extracted through optical images and SfM/MVS technique

- Facade features can be **automatically detected** even under complex scenarios with no preprocessing need

- The **geometric quality** of the 3D model, as well as the **quality** of the **3D labeling**, is a direct function of the **point cloud density**

- Point cloud quality by SfM/MVS depends on the camera parameter estimation, image spectral and spatial characteristics. Therefore, the **targets geometry** and **texture** are **fundamental** in the process of **reconstruction** and classification

Objectives

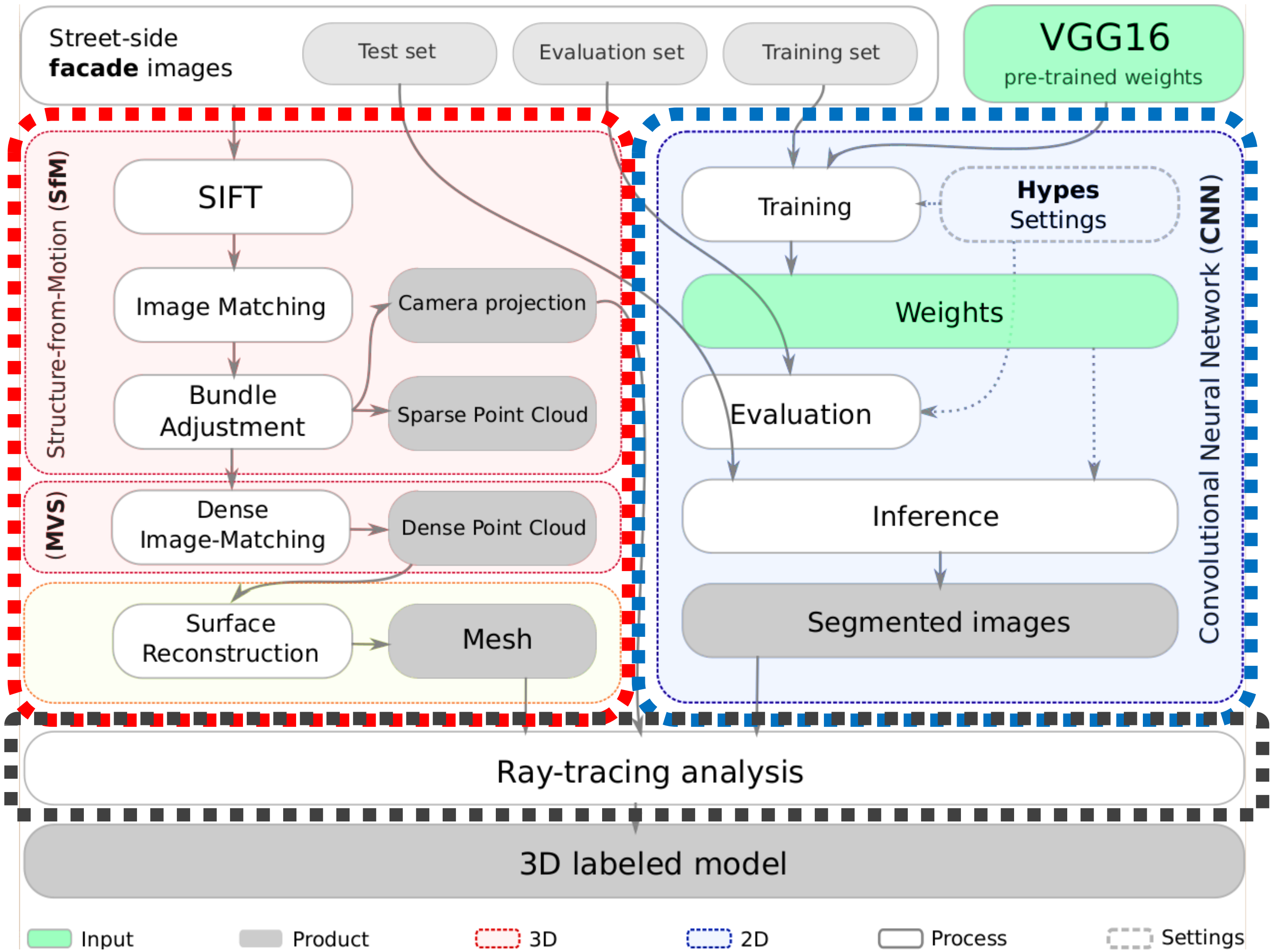
*From **structural data**, explore the extraction of **building geometric**, simultaneously, **detect** their facade features. Then, **associate both** information in one single **3D labeled model***

- Develop a routine to classify facade elements in 2D, using a **CNN** neural architecture
- Using the same images, obtain the facade geometry using **SfM/MVS**
- **Evaluate the performance** of the neural model for different urban scenarios and architectural styles
- **Evaluate** a **case study** with a real application in **Brazil**, whose architecture differs from the datasets used during the neural model training
- Classify the **3D model** of the extracted facade using the images previously segmented in the 2D domain by **Ray-Tracing**

Methodology

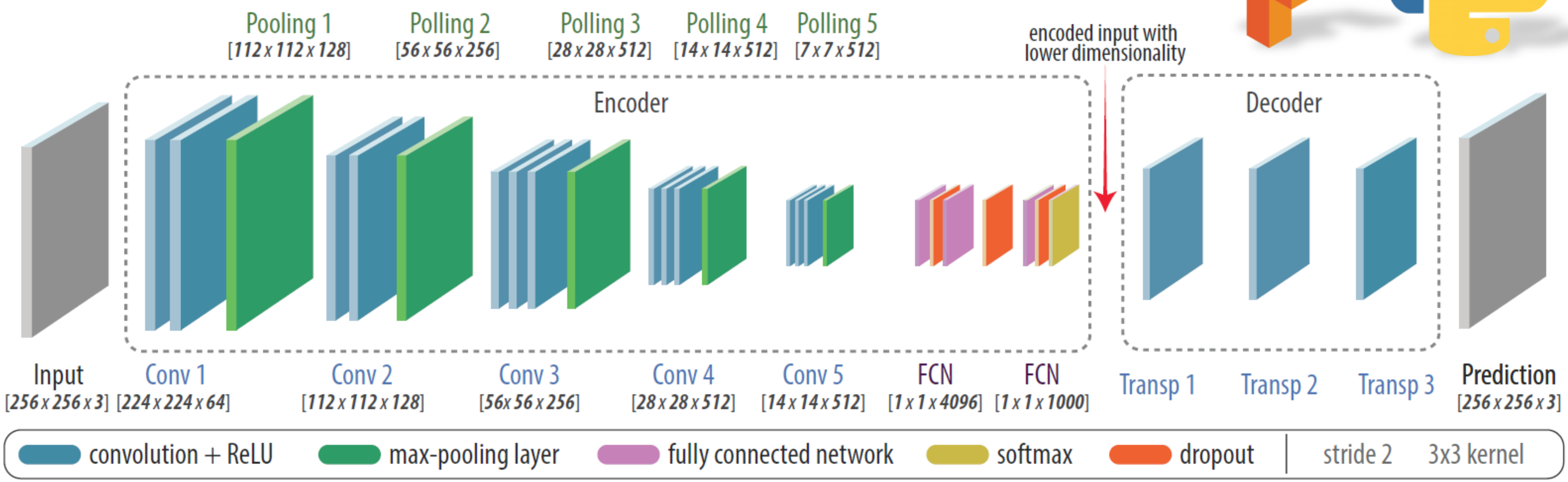
#	Name	Architecture	Image/Label	Rectified	PC generation
1	RueMonge2014	Hausmaniann	428/219	○	●
2	CMP	Multiple	378/378	●	○
3	eTRIMS	Multiple	60/60	○	○
4	ENPC	Hausmaniann	79/79	●	○
5	ECP	Hausmaniann	104/104	●	○
6	Graz	Multiple	50/50	●	○
7	SJC	Undefined	175/10	○	●

Methodology



Methodology

The neural model

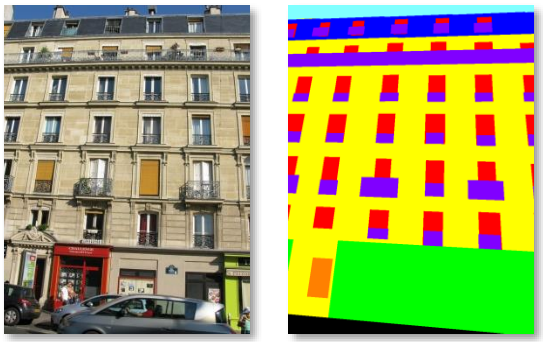


VGG16

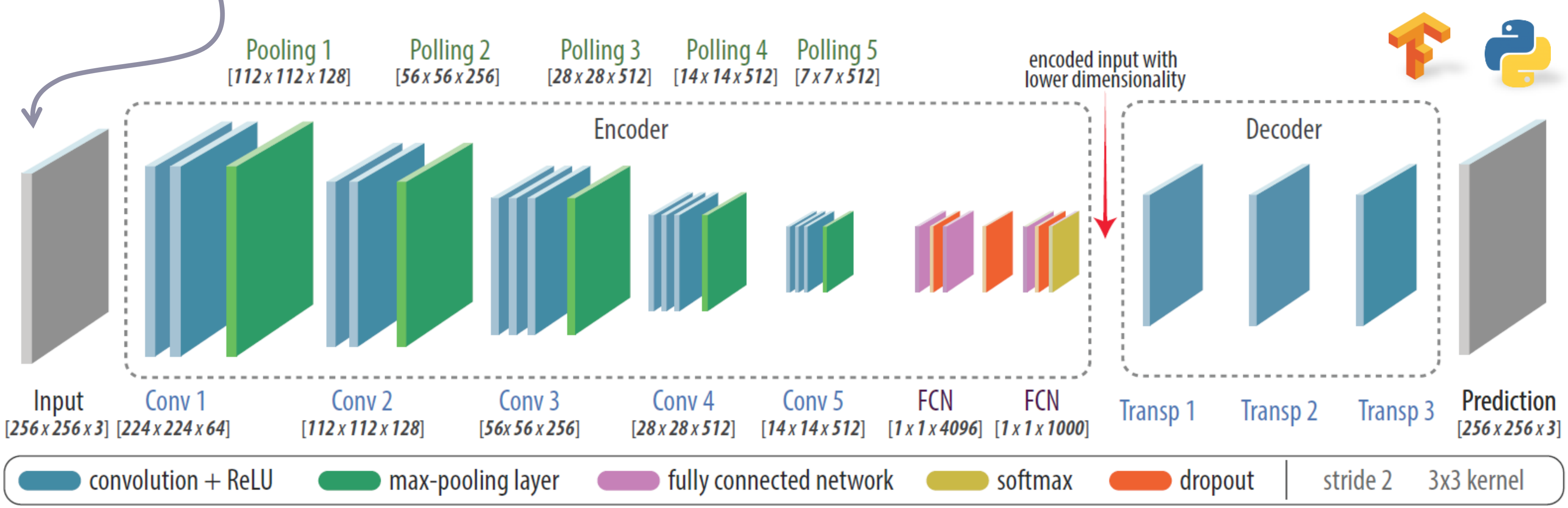
deconvolution

Simonyan, Karen, and Andrew Zisserman. "**Very deep convolutional networks for large-scale image recognition.**" *arXiv preprint arXiv:1409.1556* (2014).
Teichmann, Marvin, et al. "**Multinet: Real-time joint semantic reasoning for autonomous driving.**" *arXiv preprint arXiv:1612.07695* (2016).

Methodology



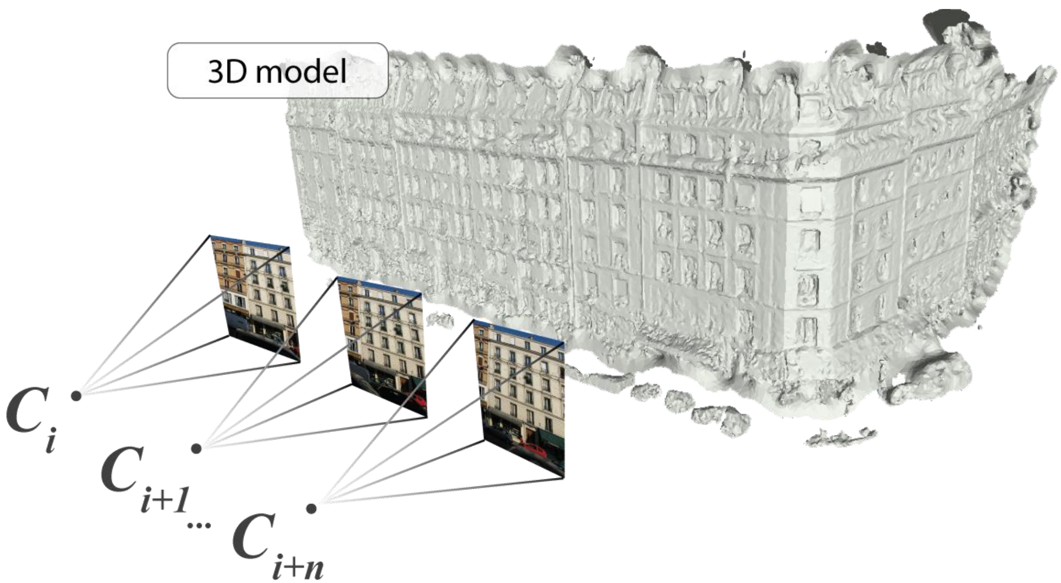
- The pair: original + annotation, is **mandatory** for the CNN neural model used (**supervised**)
- Only **80%** from the total of images are selected to be **training samples**. The other **20%**, are left for **validation**
- **The predictions** are function of the quality of annotations
- **Pre-trained** weights (**ImageNet**)

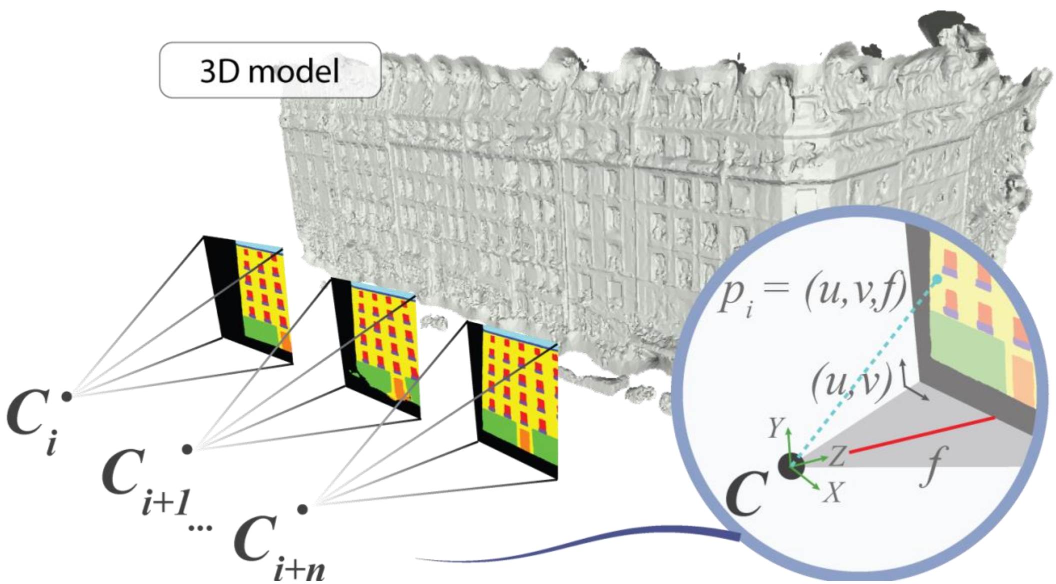


Methodology

- Due to the range of pictures, and geometry of acquisition (overlaping), only **RueMonge2014** and **SJC** datasets were able to fit in the SfM/MVS procedure
- **Agisoft Photoscan** was adopted to the SfM/MVS pipeline (sparse, dense, mesh, camera param estimation)
- For **visualization** and **point cloud processing**, the software **MeshLab** and **CloudCompare** were used
- **CGAL** is a open-source C++ library, which allow the implementation of new routines for geometrical processing. Used in this study as a **preprocessing** alternative

Methodology

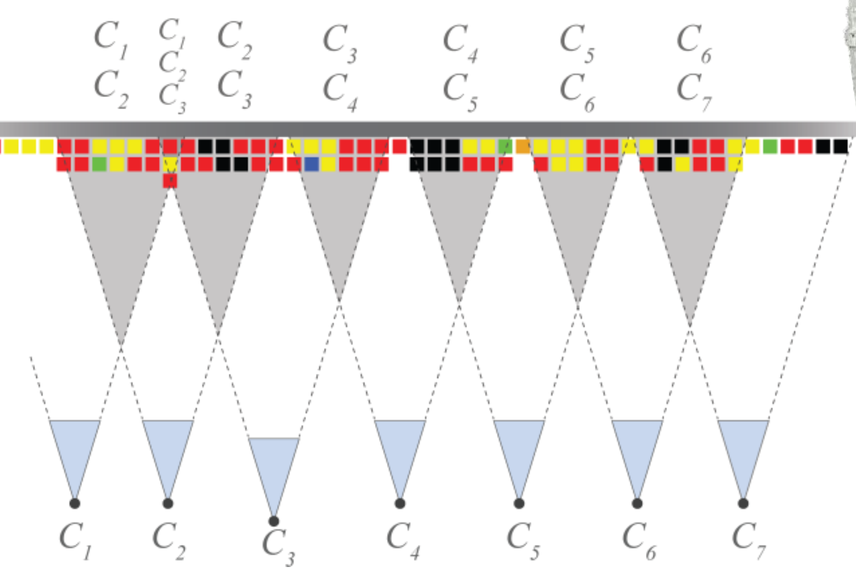




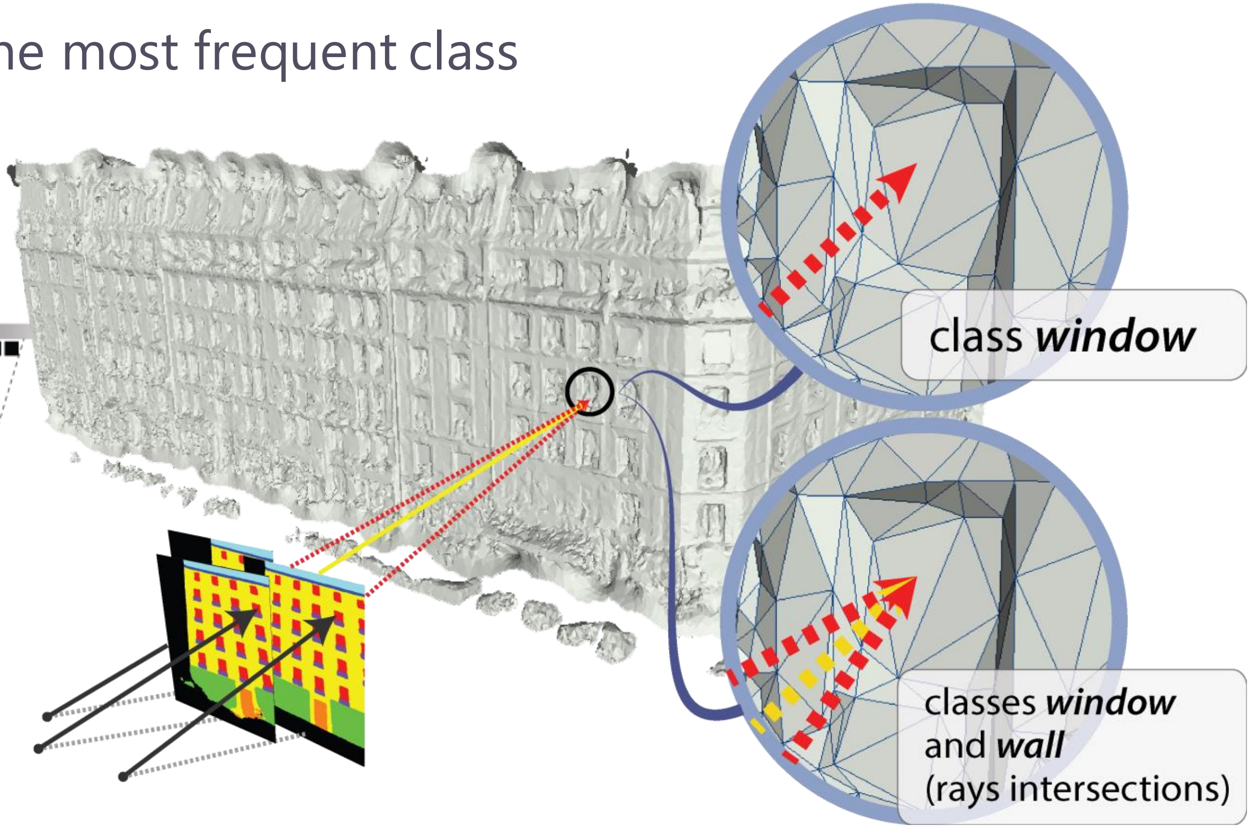
- Camera **intrinsic** and **extrinsic** parameter estimated from SfM/MVS are **preserved**
- The images from **CNN predictions** are then **replaced**
- Projection using the **Camera Frame** model transformation
  - A **"reverted" ray-tracing** is applied from the center of projection from each picture **toward** onto **mesh**
  - The routines were mainly made on **C++**
- Then, the **3D labeling algorithm** only demands the **camera estimated parameters**, the **mesh**, and the **segmented images**

Methodology

- Due to the overlapped images, the predictions might **intersect** with one or more pictures
- The **intersectioned classes**, only one can be choose **to assign each triangle mesh**
- The **mode** is applied to get the most frequent class



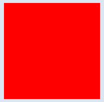
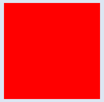
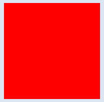
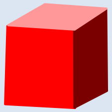
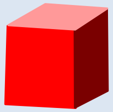
Top view



Front view

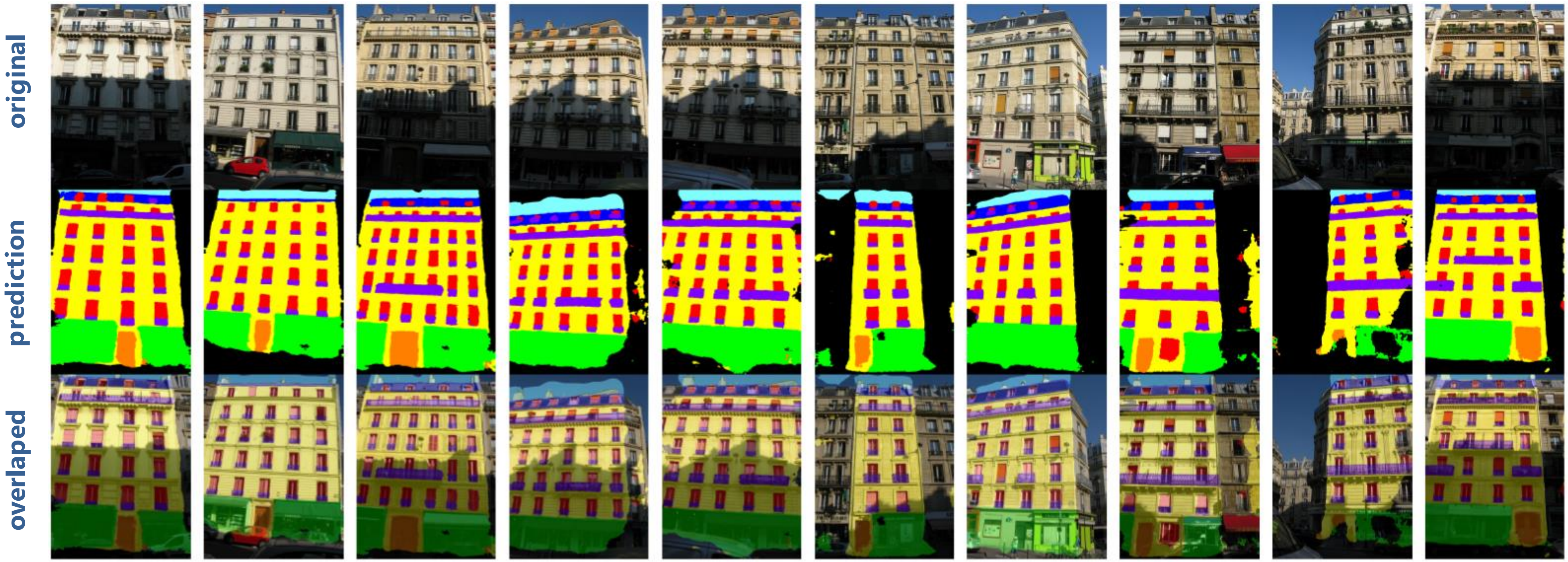
Methodology

Strategy of analysis

#	Domain	Dataset	Goal
1		Online datasets **(individually)**	Evaluate the performance of the neural model **according to each online dataset**
2		Online datasets **(individually)**	Evaluate the performance of the neural model **according to SJC**, **individually**
3		Online datasets **(all-together)**	Evaluate the performance of the neural model **according to SJC**, all online datasets **together**
4		RueMonge2014	Evaluate how accurate the 3D labeling is according to the point cloud density, under a known dataset
5		SJC	Evaluate how accurate the 3D labeling is according to the point cloud density, under an unknown dataset

Results

RueMonge2014 CMP eTRIMS ENPC ECP Graz SJC



#		Classes	Predicted 1	2	3	4	5	6	7	8
Ground-Truth 1	●	Background	0.842	0.004	0.008	0.079	0.014	0.015		0.035
2	●	Roof	0.056	0.826	0.058	0.021	0.015	0.025		
3	●	Sky	0.061	0.013	0.924	0.002				
4	●	Wall	0.042			0.919	0.010	0.015	0.005	0.007
5	●	Balcony	0.044	0.011		0.039	0.892	0.013		
6	●	Window	0.031	0.009		0.072	0.023	0.863		
7	●	Door	0.047			0.073			0.875	0.005
8	●	Shop	0.111			0.014				0.873

Scale: Success (1.0) – Error (0.0)

Rates:	Evaluation	
	Accuracy:	0.9563
	F1-Score:	0.8624

- At **50k iterations**, the CNN reached **95%** of accuracy and **86%** of F1-score regarding the reference
- When **global context** is lost, the facade **is not** segmented
- Only the **main facades** are considered by the net
- The segmentation occurs even under **complex scenarios** with no preprocessing need

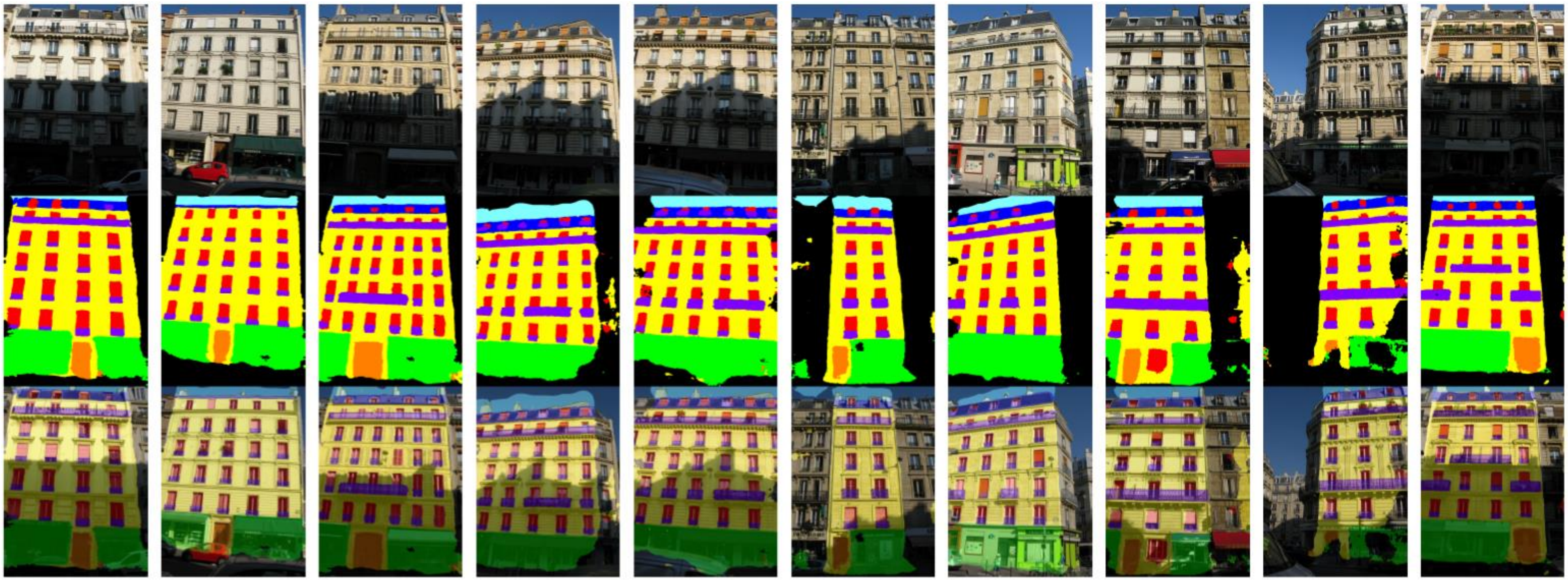
Results

RueMonge2014 CMP eTRIMS ENPC ECP Graz SJC

original

prediction

overlaped



Ground-Truth	Classes	Predicted 1	2	3	4	5	6	7	8
1	Background	0.842	0.004	0.008	0.079	0.014	0.015		0.035
2	Roof	0.056	0.826	0.058	0.021	0.015	0.025		
3	Sky	0.061	0.013	0.924	0.002				
4	Wall	0.042			0.919	0.010	0.015	0.005	0.007
5	Balcony	0.044	0.011		0.039	0.892	0.013		
6	Window	0.031	0.009		0.072	0.023	0.863		
7	Door	0.047			0.073			0.875	0.005
8	Shop	0.111			0.014				0.873

Scale: Success (1.0) – Error (0.0)

Rates:	Evaluation	
	Accuracy:	0.9563
	F1-Score:	0.8624

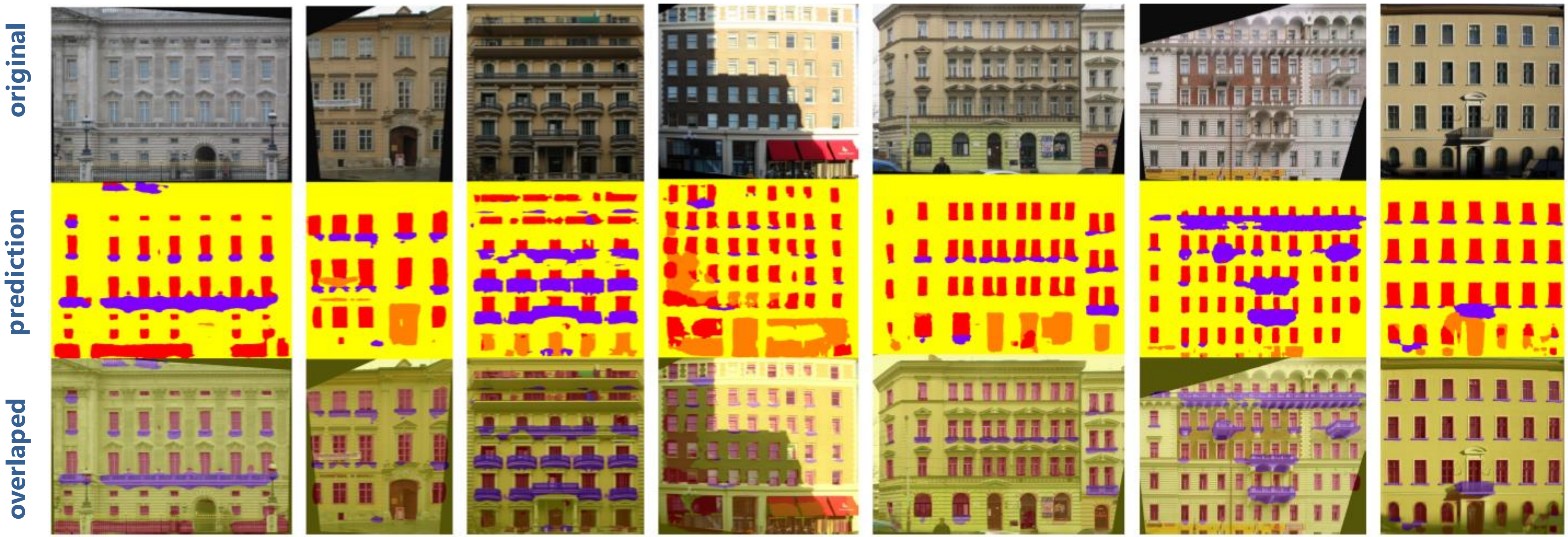
- The confusion matrix shows **good predictions** over all classes
- Tiny confusion between **bg** and **wall** against the others
- The respective classes **share boundaries** with all the others
- The **boundaries** are acctually the reason of **low accuracies**

 Thesis : Pages: 53 to 54

Results

- The target are always **well identify** but still **not delineated** properly
- **Obstructing objects** are often negligenciated by the CNN (good generalization – **enough internal layers**)
- Facades that does not cover the entire image **is not segmented**

Results



Ground-Truth		Classes	Predicted 1	2	3	4	5	6	7	8	Scale	Evaluation
	1	Background									Success (1.0)	
	2	Roof										
	3	Sky										
	4	Wall				0.937	0.022	0.022	0.019			
	5	Balcony				0.390	0.508	0.085	0.017			
	6	Window				0.109	0.013	0.825	0.052			
	7	Door				0.268		0.037	0.694			
	8	Shop									Error (0.0)	
		Rates:									Accuracy:	0.9357
											F1-Score:	0.7418

- The CMP got only **4 facade classes**
- The **rectification** of the images were badly made, as well as the **annotations**
- At 50k was **not enough** to get **good delineation** of the features (**74%** of F1-score)
- The classes **balcony** and **door** got **confused with wall**

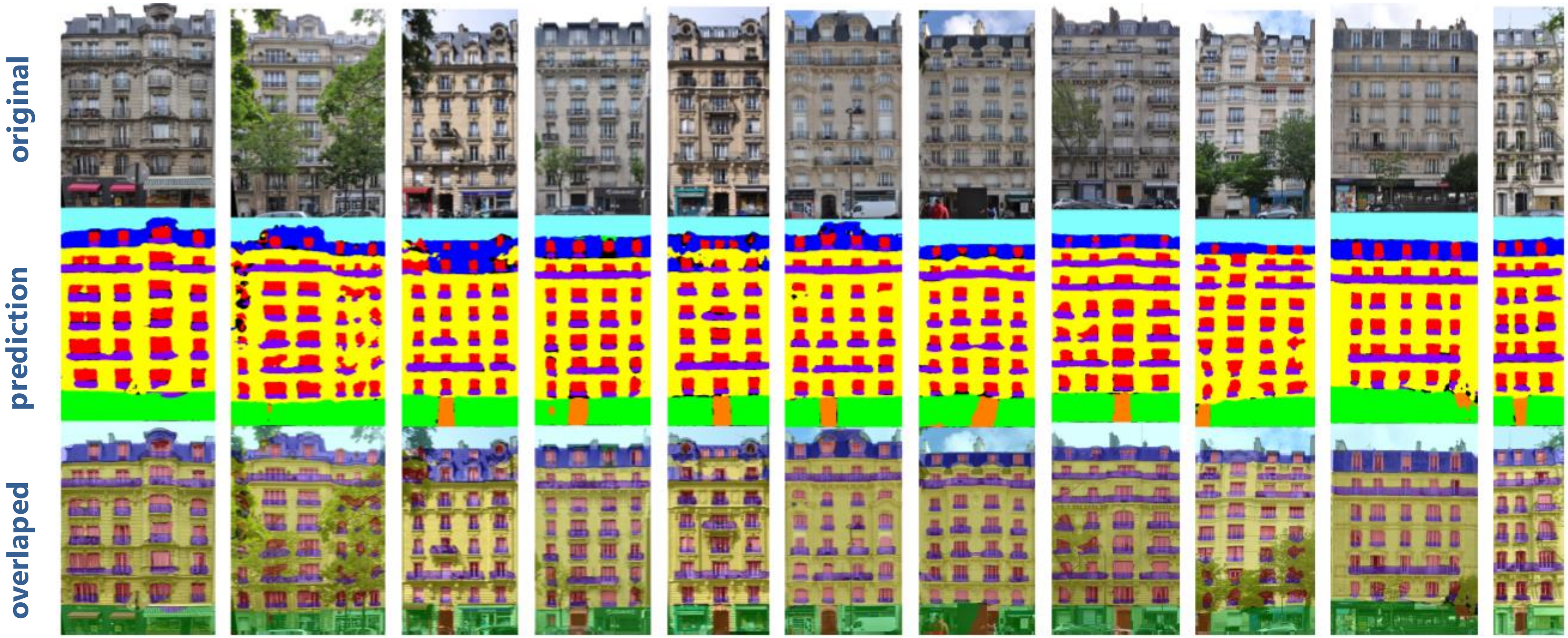
Results

	Classes	Predicted								Scale	Evaluation
		1	2	3	4	5	6	7	8		
1	Background	0.956			0.037		0.006			Success (1.0)	
2	Roof										
3	Sky										
4	Wall	0.024			0.947		0.020	0.009	0.007		
5	Balcony										
6	Window				0.206		0.773	0.017			
7	Door	0.049			0.204		0.078	0.669			
8	Shop									Error (0.0)	
	Rates:									**Accuracy:**	**0.9632**
										F1-Score:	**0.8291**

(Rows: Ground-Truth)

- Only **4 classes useful** for this dataset. The quality of annotations is better
- The reduction of classes, contributed **positively** during the training
- **Roof** is treat as **wall** in eTRIMS, due to the incorrect annotation
- As in RueMonge, eTRIMS reached **good predictions** and **delineation**
- **Window** and **door** confused with wall
- **eTRIMS** facades are **most similar** to SJC dataset

Results

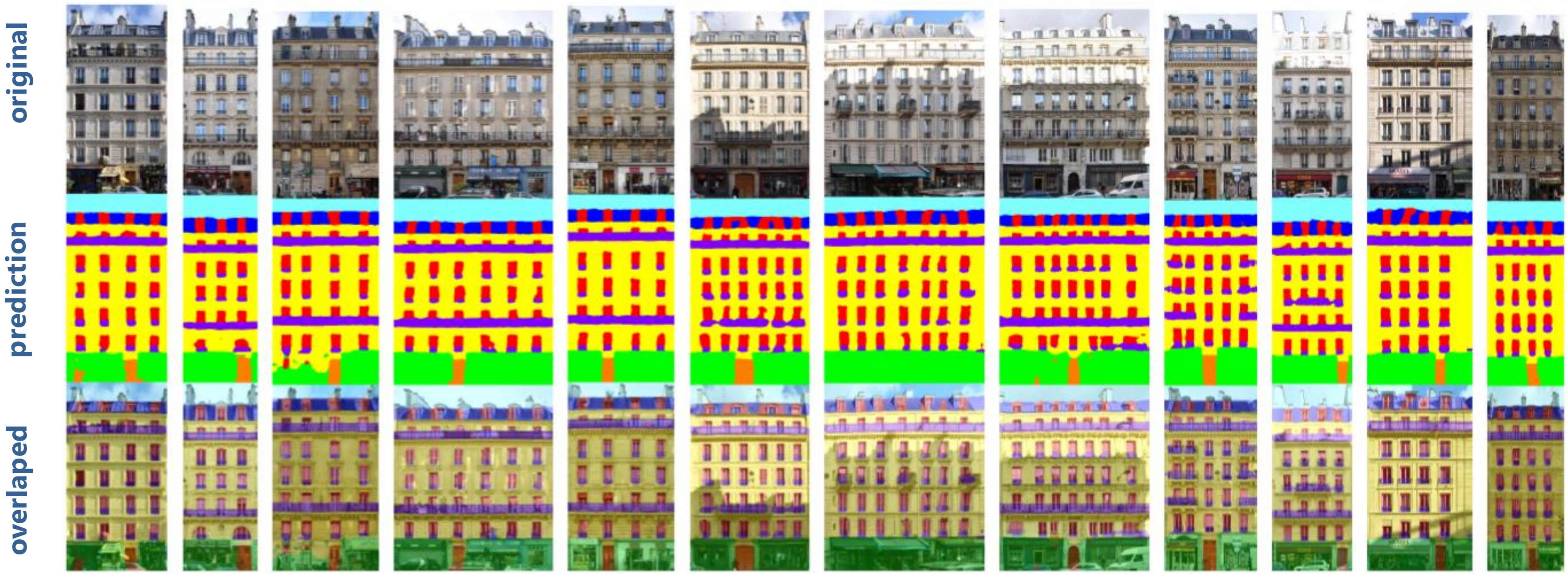


	Classes		Predicted 1	2	3	4	5	6	7	8	Scale
Ground-Truth	1	Background	0.105	0.060	0.033	0.315	0.235	0.190	0.013	0.049	Success (1.0)
	2	Roof	0.041	0.783	0.055	0.072	0.015	0.034			
	3	Sky	0.006	0.015	0.977						
	4	Wall	0.011	0.007		0.916	0.028	0.034		0.004	
	5	Balcony	0.034	0.004		0.125	0.794	0.042			
	6	Window	0.020	0.023		0.113	0.023	0.817			
	7	Door	0.022			0.033			0718	0.227	
	8	Shop	0.007			0.038	0.004		0.017	0.935	Error (0.0)

Rates:	Evaluation
Accuracy:	0.9636
F1-Score:	0.7655

- The **obstructing objects** are annotated as being part of the facades
- Consequently, the training **treats** it as being an **urban element** (second figure – left-to-right)
- **Poor annotations** regarding the roof-parts
- The accuracy was **high** (**96%**) and **poor** delineation (F1-score **76%**)

Results



		Classes	Predicted 1	2	3	4	5	6	7	8	Scale
Ground-Truth	1	Background									
	2	Roof		0.844	0.043	0.030		0.083			Success (1.0)
	3	Sky		0.009	0.986						
	4	Wall			0.390	0.938	0.030	0.025			
	5	Balcony				0.109	0.854	0.032		0.004	
	6	Window		0.030		0.081	0.041	0.841			
	7	Door				0.014			0.873	0.113	
	8	Shop			0.006	0.023			0.013	0.955	Error (0.0)

Rates:	Evaluation	
	Accuracy:	0.9762
	F1-Score:	0.8946

- The most significant scores, accuracy of **97%** and F1-score of **89%**
- **Poor annotations** regarding the roof-parts
- **Poor annotations** lies on **poor delineations**

Results

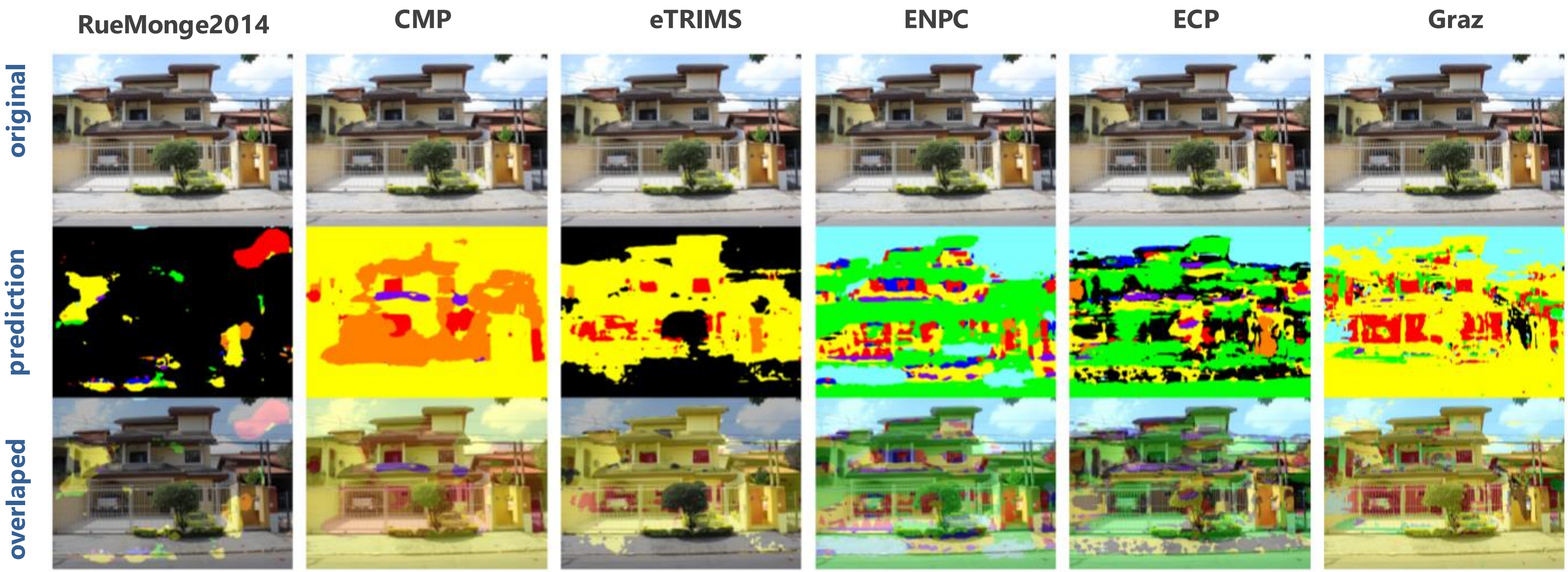
RueMonge2014 CMP eTRIMS ENPC ECP **Graz** SJC

original
prediction
overlaped

Ground-Truth	Classes		1	2	3	4	5	6	7	8	Scale	Evaluation
			Predicted									
1	●	Background				0.846		0.013	0.029	0.112	Success (1.0)	
2	●	Roof		0.039	0.171	0.733		0.056				
3	●	Sky	0.059	0.004	0.757	0.171		0.005				
4	●	Wall				0.941		0.040		0.012		
5	●	Balcony				0.836	0.051	0.102		0.008		
6	●	Window				0.169		0.816		0.009		
7	●	Door				0.242		0.045	0.661	0.052		
8	●	Shop				0.624		0.068	0.056	0.245	Error (0.0)	
	Rates:										Accuracy:	0.9368
											F1-Score:	0.7698

- **Smallest** set of images (50)
- **A lot** of **confusion** involving the classes wall
- **Similar architectural style** than CMP, ECP, and ENPC

Results



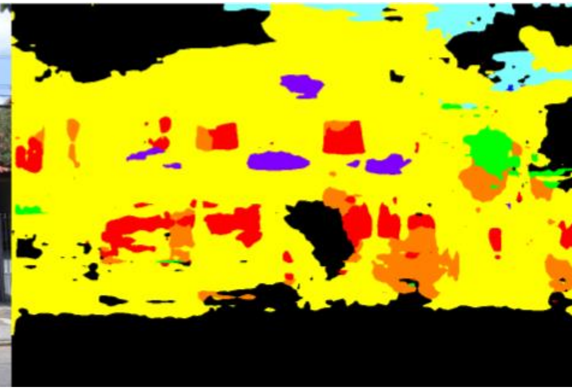
- Using the “knowledge” acquired from the **different trainings**, the predictions under **SJC** got mostly incorrect results: **accuracy 85%**, **F1-score 47%**
- The classes **sky**, **shop** and **door**, does not fit to the architectural style
- Consequently, **eTRIMS** that has only 3 classes, got better predictions

Results

RueMonge2014 CMP eTRIMS ENPC ECP Graz **SJC**

original

prediction



overlaped

Ground-Truth	Classes	Predicted 1	2	3	4	5	6	7	8
1	● Background	0.811			0.158		0.006	0.020	0.004
2	● Roof	0.035	0.002		0.870	0.042	0.014	0.009	0.028
3	● Sky	0.297	0.001	0.211	0.487				0.003
4	● Wall	0.113			0.833	0.003	0.011	0.020	0.020
5	● Balcony	0.029			0.514	0.213	0.051	0.193	
6	● Window	0.027			0.176	0.009	0.683	0.106	
7	● Door	0.075			0.543	0.002	0.116	0.244	0.020
8	○ Shop								

Scale: Success (1.0) – Error (0.0)

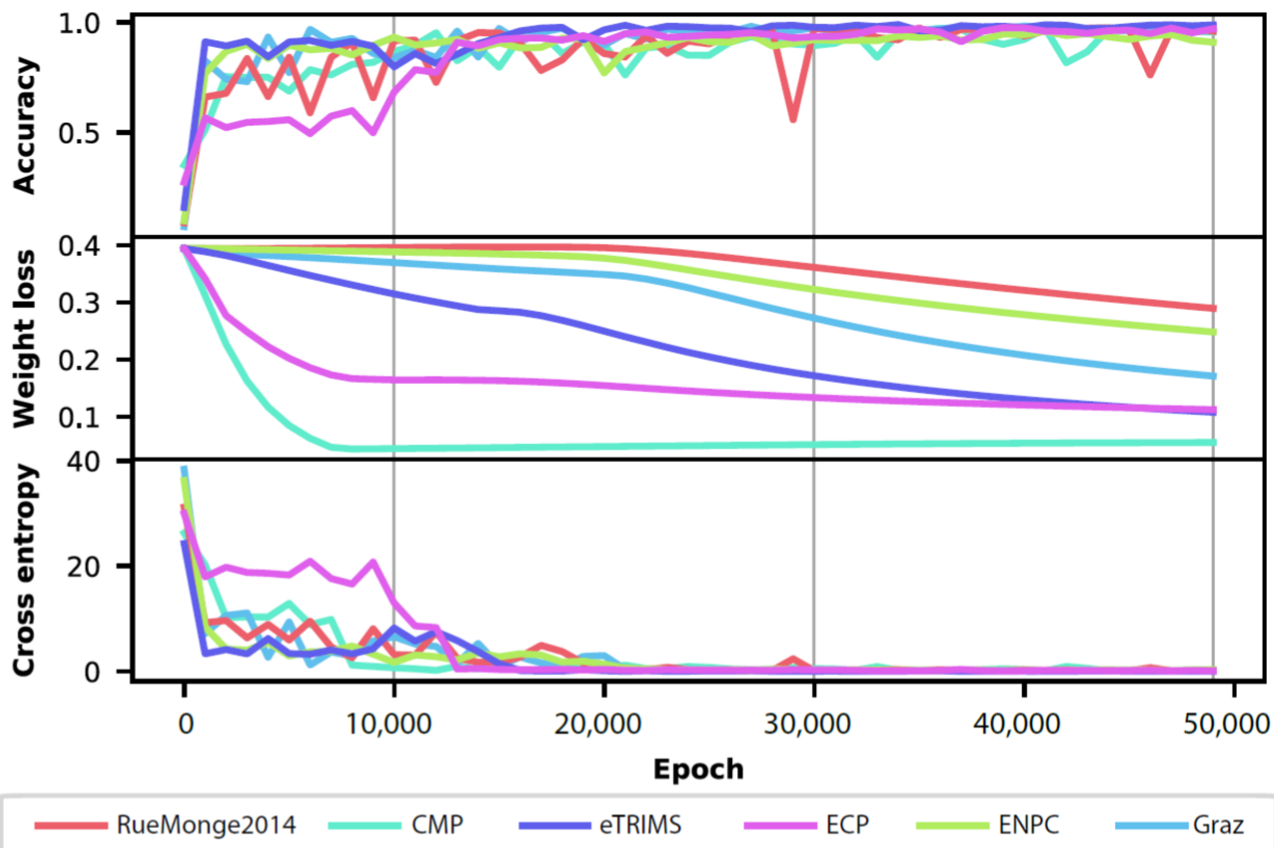
Rates:

Evaluation	
Accuracy:	**0.8591**
F1-Score:	**0.4706**

 Thesis : Pages: 63 to 65

Results

CNN training and performance



- All online datasets got **similar performance** during training
- Stabilizing **around 12k** iterations
- The objective function (**cross-entropy**) is already **minimal around 15k** for all them
- The **accuracy** was **high** for most of them
- With no GPU, the training reached **7 days** of processing

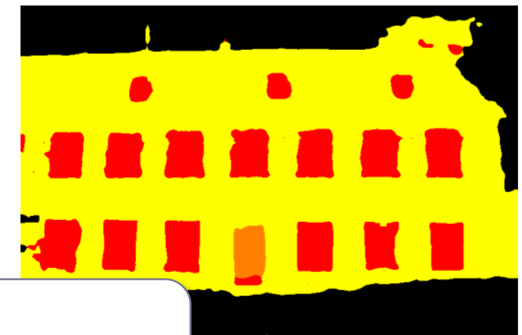
 Thesis : Pages: 51 to 53

Results

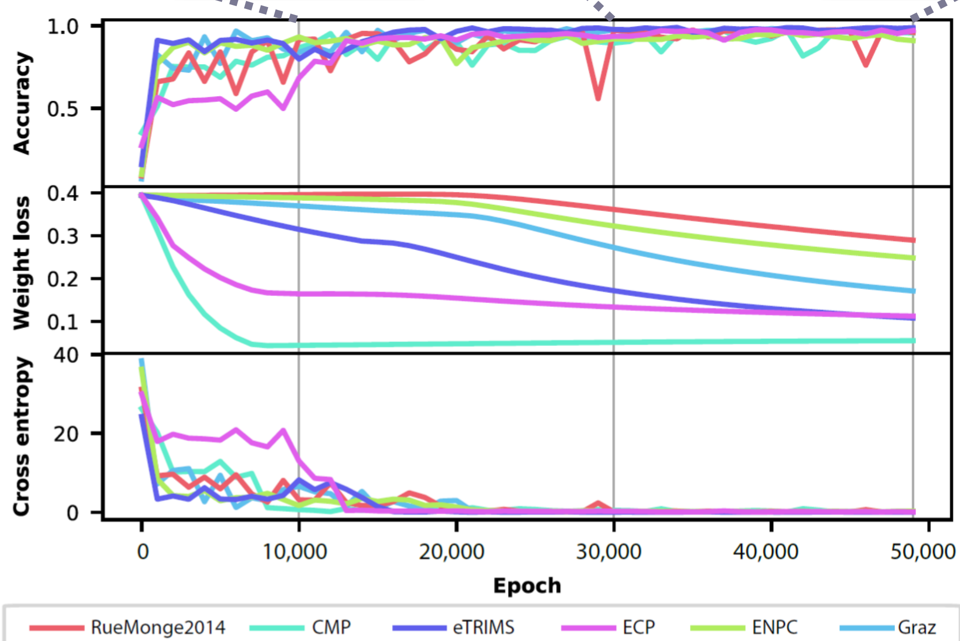
CNN training and performance

10k
Accuracy: **0.79**
Weight loss: **0.32**
Cross-entropy: **9.2**

30k
Accuracy: **0.92**
Weight loss: **0.18**
Cross-entropy: **0.213**

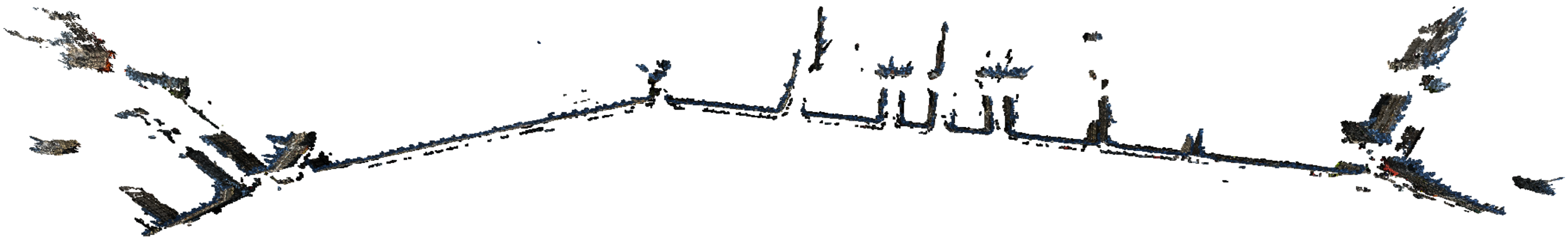


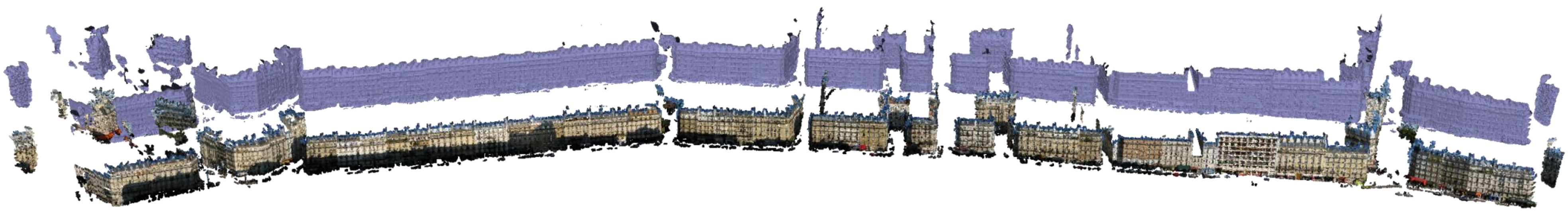
50k
Accuracy: **0.96**
Weight loss: **0.11**
Cross-entropy: **0.011**

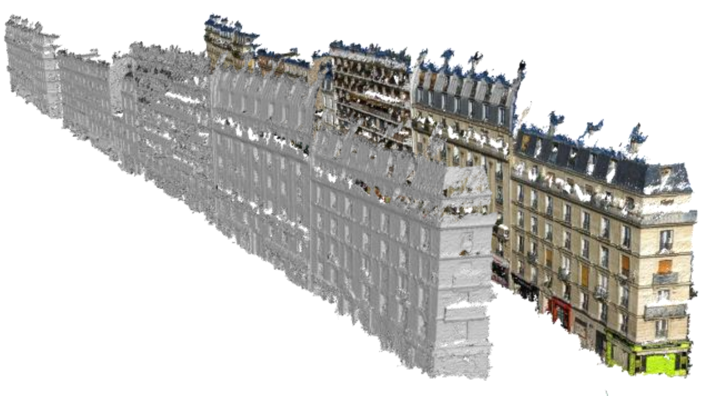


 Thesis : Pages: 51 to 53

Results

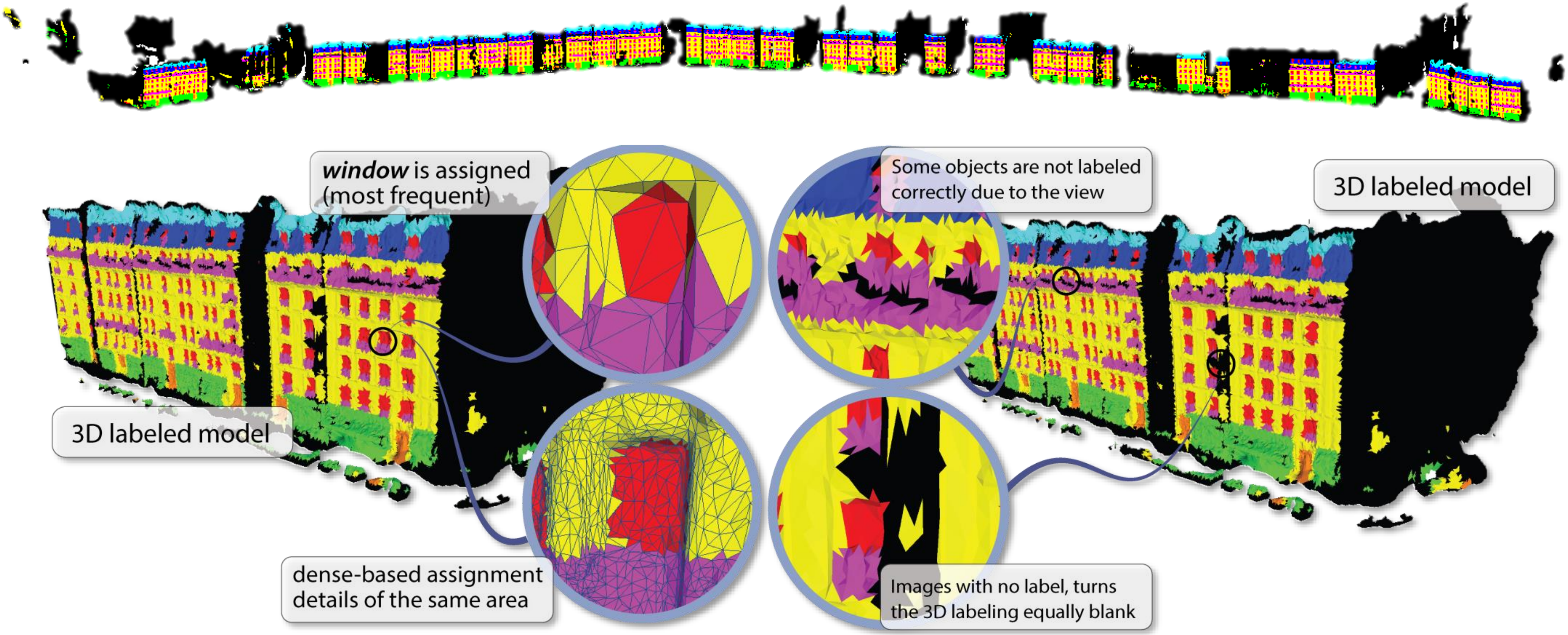






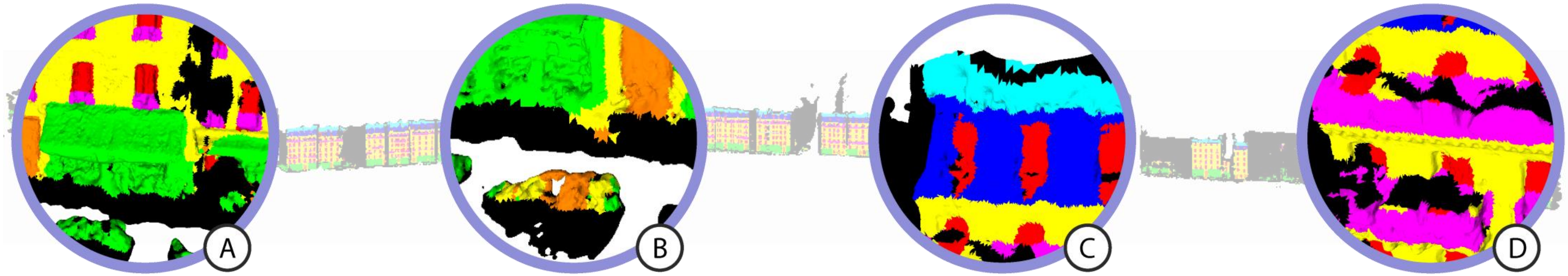
- Effects by occlusion **were minimal**
- The **facade texture** from RueMonge2014 **benefit** the reconstruction
- Once the acquirement is **close-range**, the dense point cloud was more than expected (**well defined**)

Results



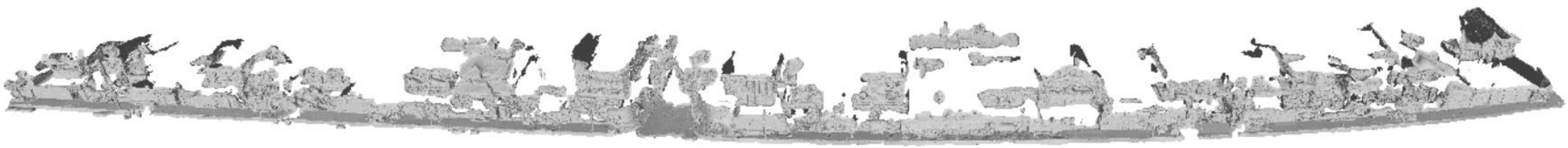
Dataset	Point Cloud density	Num. faces (triangles)	3D reconstruction - SfM/MVS (min)	Ray-tracing (sec)
RueMonge2014	Sparse	1,072,646	21.4	12.42
RueMonge2014	Dense	9,653,679	46.4	27.13

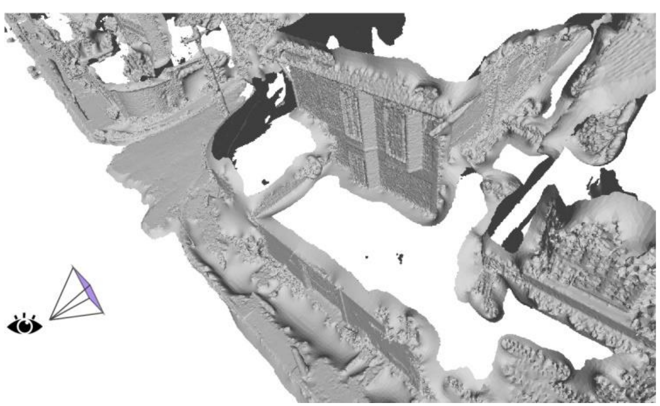
Results



- In **“A”** and **“B”**, despite the presence of **vehicles**, the classification of **stores** is not much impaired
- **“C”** and **“D”,** the classification of features at the top of the building, is impaired due to the **perspective**
- **Roof** and **balconies** can not be entirely assigned

Results

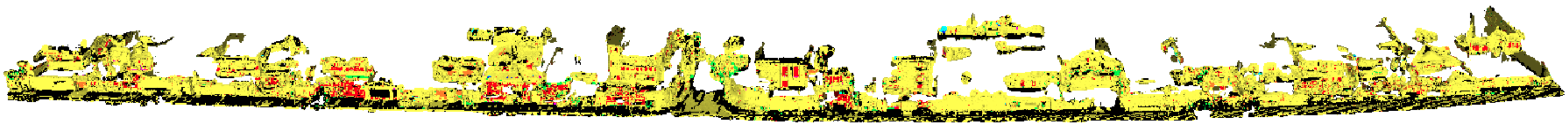


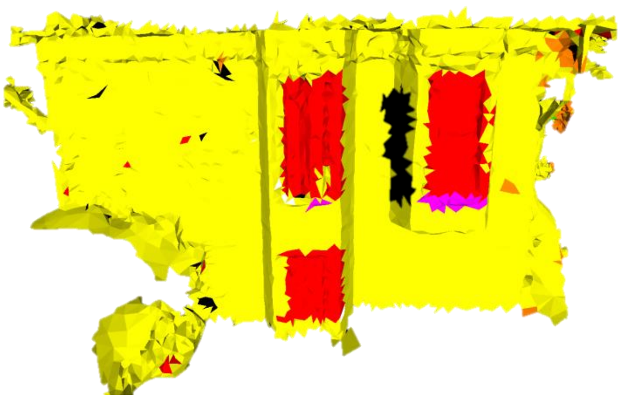


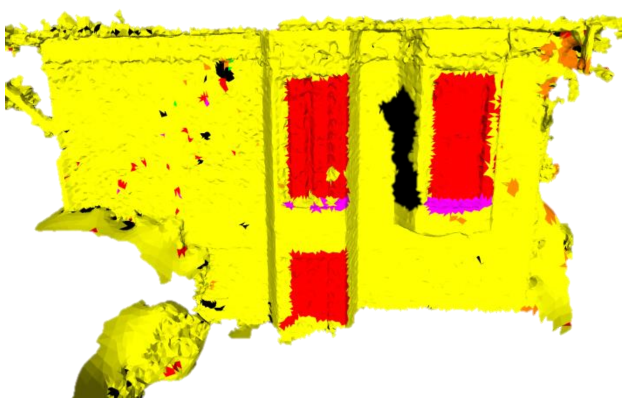
- **In Brazil**, the houses are typical erected first with a **wall** together **with a gate**, then, the house itself
- With this geometry, the **terrestrial** campaign is **not sufficient**
- A **lot of trees** and **cars** in front of the houses

Dataset	Point Cloud density	Num. faces (triangles)	3D reconstruction - SfM/MVS (min)	Ray-tracing (sec)
SJC	Sparse	800,000	13.6	20.89
SJC	Dense	3,058,329	35.5	41.12

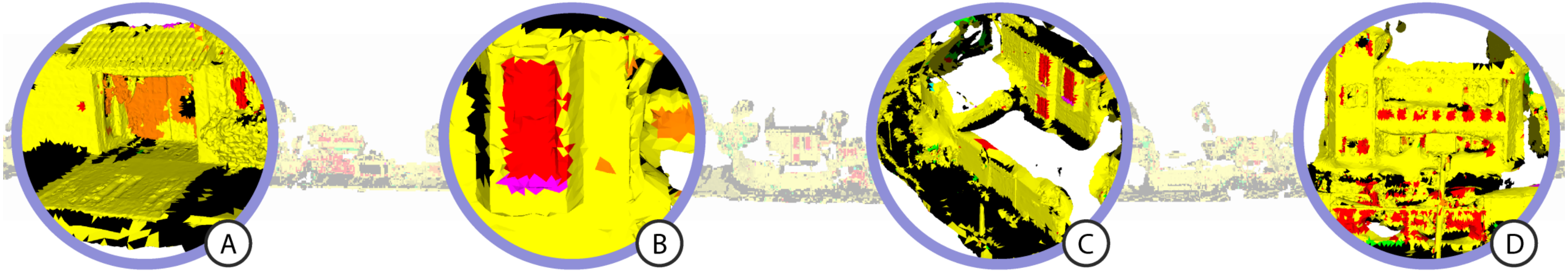
Results

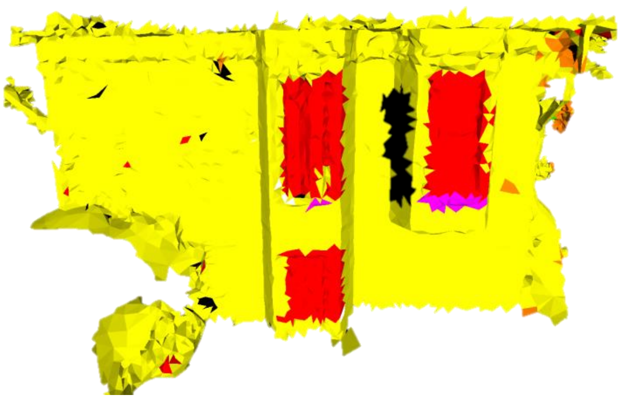


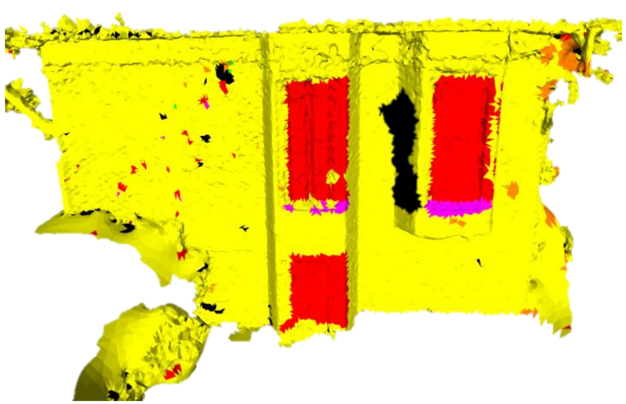




Results







- The **painting** and **façade layout**, contribute negatively on SfM/MVS processing (too homogenous)
- **In Brazil**, the houses are typical erected first with a **wall** together **with a gate**, then, the house itself
- In **"A",** the **roof part** could be segmented as such, but was segmented as **wall**
- **Features** similar to the online datasets, could be identify in SJC

Conclusions

- Present a methodology for **classifying facade geometry** under different scenarios
- Used **state-of-art** methods, which in summary can be easily **reproduced** and **extended**
- Although no preprocessing has been applied, the **accuracy** on CNN predictions for **individual datasets** were **above 93%**
- For SJC, the **predictions failed** to delineate the object, but still the **accuracy** was **85%** using all-together “knownledge”
- The **objects delineation** is a “on going” topic in CNN approaches
- The use of SfM/MVS **was adequated** for structural data acquisition, and tends to improve along the next years

Conclusions

- The **dependency** of **training datasets** is a **negative aspect** when the goal is to reach a generic segmenter

- The increasing of **iterations** (epochs) and **datasets** would affect **positively** since the neural network gets more time and example in *"how to learn"*

- The **Ray-tracing** technique was **fundamental** and adequaded to the **geometric classification (3D labeling)**

- The **scene geometry** is impaired according to the point cloud density. Too sparse point clouds might generalize much the interest features

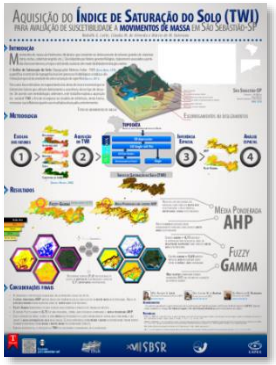
Conclusions

- **Brazilian architecture** differs much from the data used for training. However, it was able to identify **windows**, **doors** and **wall**
- By that, it is clear that a **bigger datasets** could get better predictions over an **unknown dataset**
- **Architectures** tends to evolve, the extraction of feature has to consider current layouts, but also follow this **eminent evolution**
- All the **source-code** produced were **shared** in a public plataform
- The routines were mainly made by **open-source** libraries and softwares

Conclusions

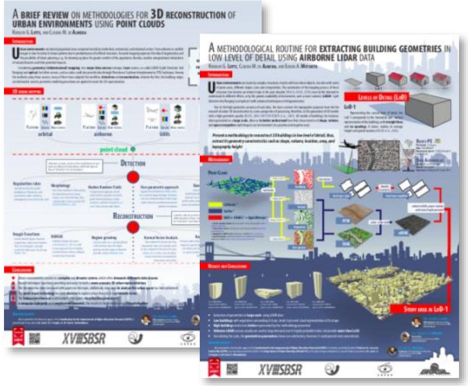
- **Pre-classifications** of: architectural style, area (suburbs, center, residential, so on)
- **Stardardization**: figure how to “translate” a classified model to a well-known language, such as **CityGML (mesh2cityGML)**
- Use **different neural models** for **different scenarios**, mitigate which of these could fit more to the urban context
- Use **unsupervised** approaches to avoid training samples issues
- Take areas with **typical irregularities**, apply the methodology, find those irregularities automatically

- Aquisição do índice de saturação do solo (TWI) para avaliação de suscetibilidade a movimentos de massa em São Sebastião-SP
  SBSR – 2015 - Santos, SP
  Rodolfo G. Lotte, Cláudia M. de Almeida, Márcio de M. Valeriano

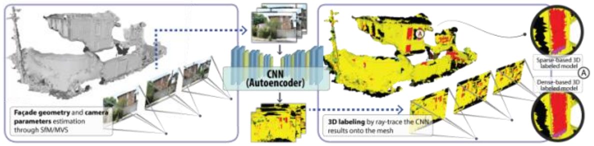


- Object-based image analysis for urban land cover classification in the city of Campinas - SP, Brazil
  JURSE – 2015 - Lausanne, Suíça
  David G. M. França, **Rodolfo G. Lotte**, Cláudia M. de Almeida, Sacha M. O. Siani, Thales S. Körting, Leila G. M. Fonseca, Luiz T. da Silva

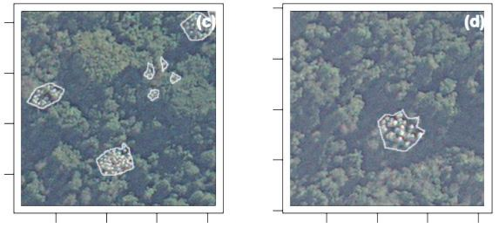
- A methodological routine for extracting building geometries in low level of detail using airborne lidar data
  SBSR – 2017 - Santos, SP
  Rodolfo G. Lotte, Cláudia M. de Almeida, Edson A. Mitishita
- A brief review on methodologies for 3D reconstruction of urban environments using point clouds
  SBSR – 2017 - Santos, SP
  Rodolfo G. Lotte, Cláudia M. de Almeida



- A2 3D Façade Labeling over Complex Scenarios: A Case Study Using Convolutional Neural Network and Structure-From-Motion
  Remote-Sensing Journal – 2018 - Special Issue in Deep Learning
  Rodolfo G. Lotte, Norbert Haala, Mateusz Karpina, Luiz E. O. C. Aragão, Yosio E. Shimabukuro



- A2 Using the U-net convolutional network to map forest types and disturbance in the Atlantic rainforest with very high resolution images
  Remote-Sensing Journal – 2018 - Special Issue in Deep Learning
  Fabien H. Wagner, Alber Sanchez, Yuliya Tarabalka, **Rodolfo G. Lotte**, Matheus P. Ferreira, Marcos P. M. Aguiar, Emanuel Gloor, Oliver L. Phillips, Luiz E. O. C. Aragão



- A1 Ionospheric TEC Climatology over Latin America – Daytime
  JGR – 2018 - Journal of Geophysical Research
  Esmeralda Romero-Hernandez, Clezio Denardini, Hisao Takahashi, J. Americo Gonzalez-Esparza, Paulo Nogueira, Marcelo de Padua, **Rodolfo G. Lotte**, Patricia Negreti, Olusegun Jonah, Laysa Resende, Mario Rodriguez-Martinez, Maria Sergeeva, Paulo Barbosa Neto, Victor De la Luz, João Galera Monico, E. Aguilar-Rodriguez

Acknowledgment

INPE

CAPES

CNPq

ifp

Universität Stuttgart

drones

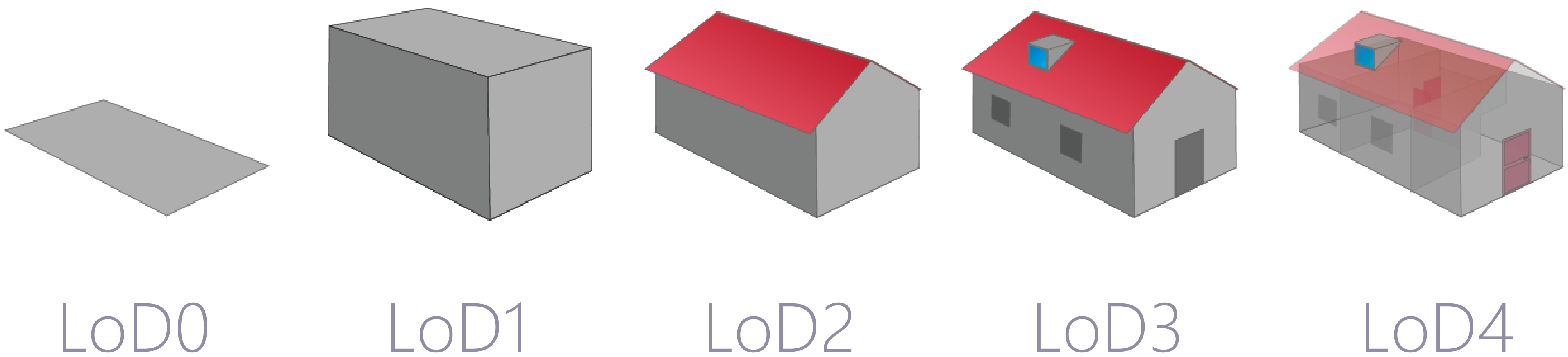
EXÉRCITO BRASILEIRO
SERVIÇO GEOGRÁFICO

"No one will motivate you more than your own determination and courage"

R.G. Lotte

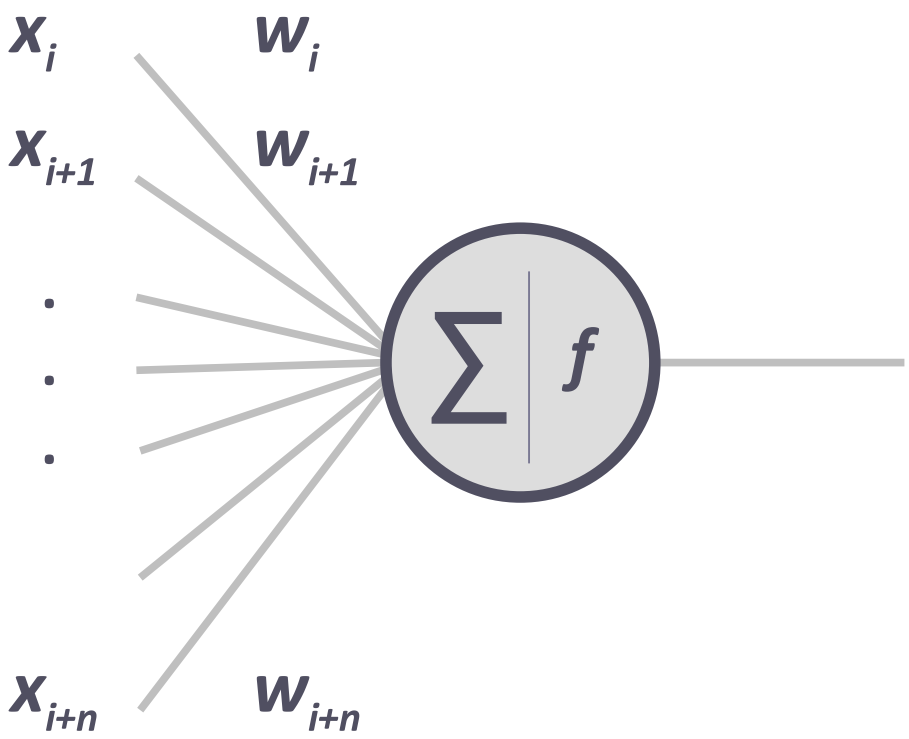
Level of Detail (LoD)

OGC – CityGML 3.0



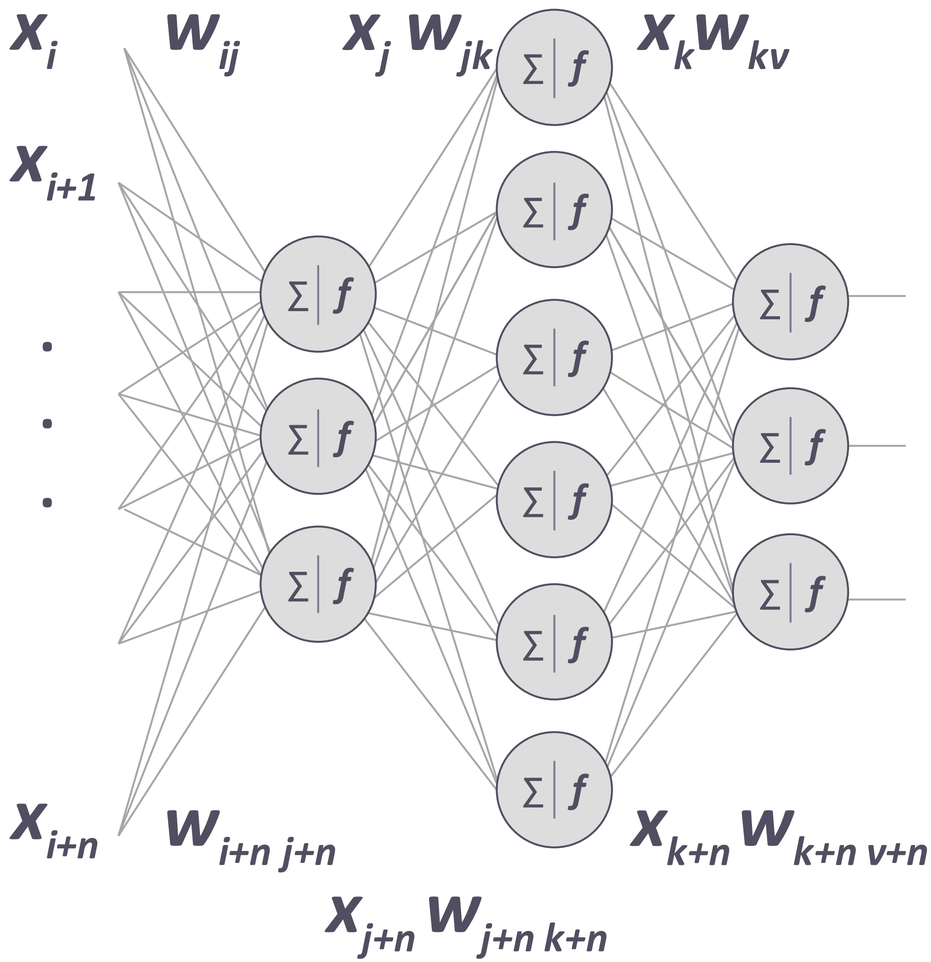
Neural Network

The XOR problem (linear)



McCulloch, Warren S., and Walter Pitts. "**A logical calculus of the ideas immanent in nervous activity**." The bulletin of mathematical biophysics 5.4 (1943): 115-133.

Neural Network

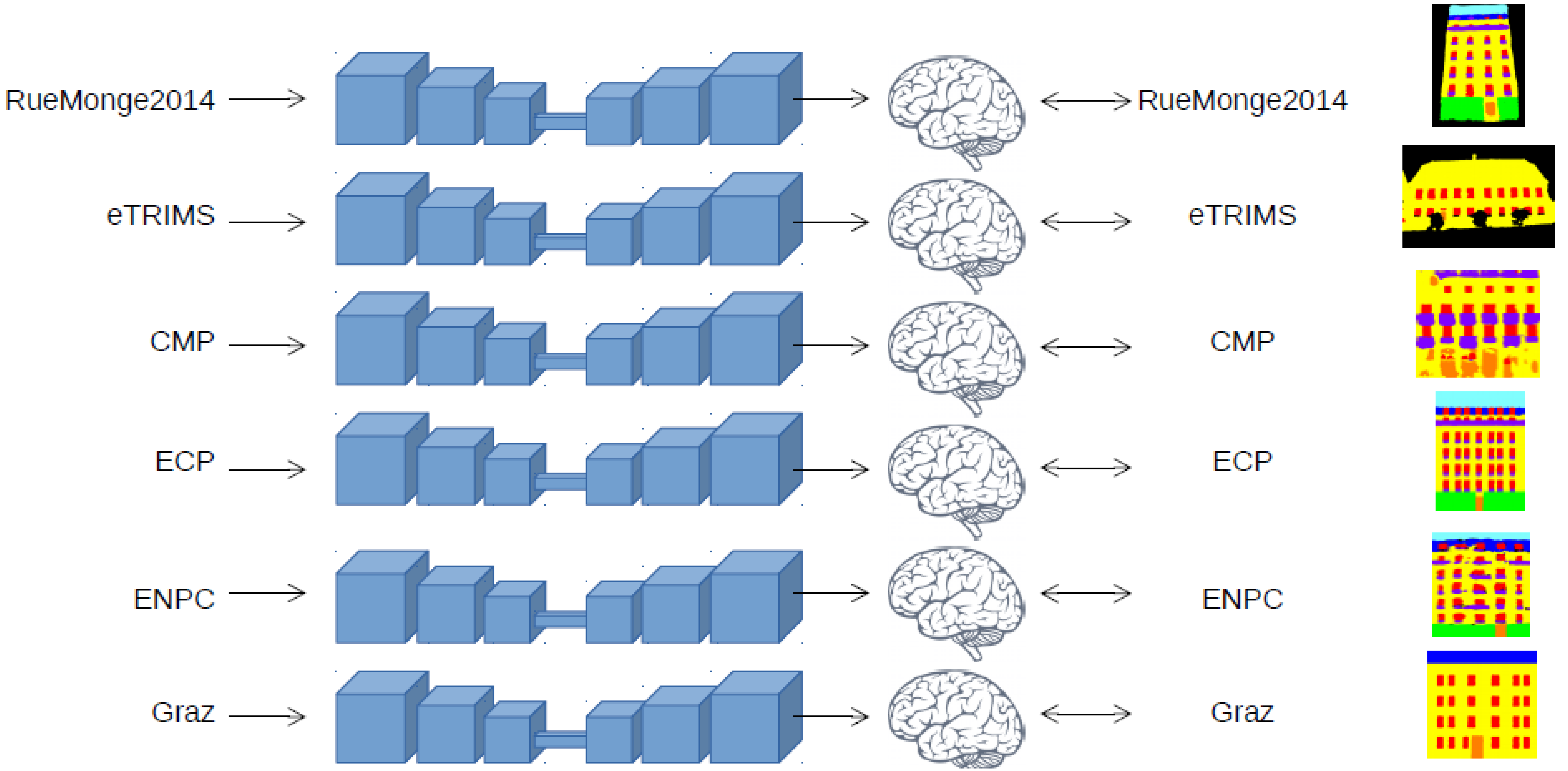


Rosenblatt, Frank. "**The perceptron: a probabilistic model for information storage and organization in the brain**." Psychological review 65.6 (1958): 386.

 Thesis : Pages: 27 to 29

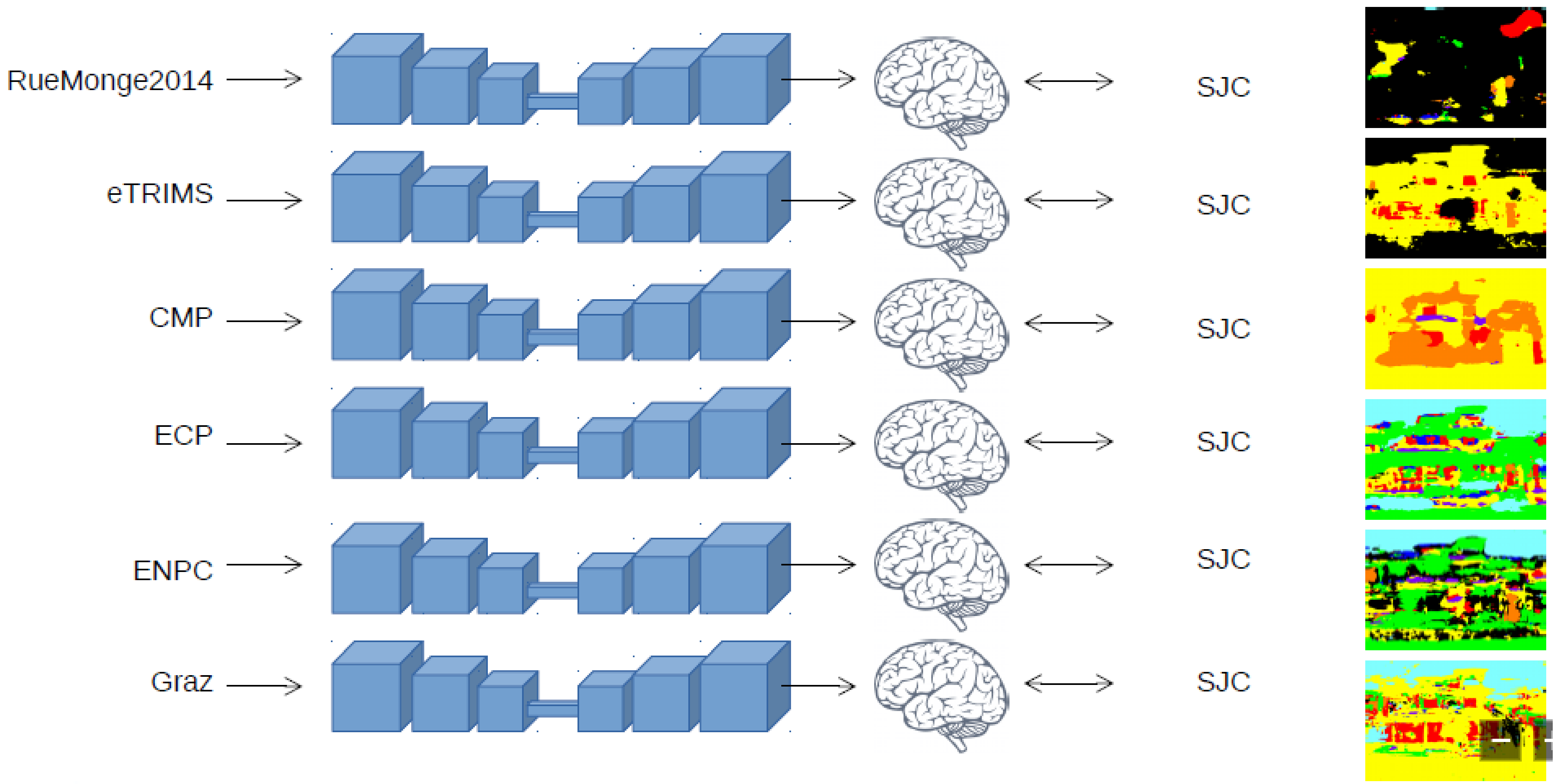
Experiments

Experiment 1



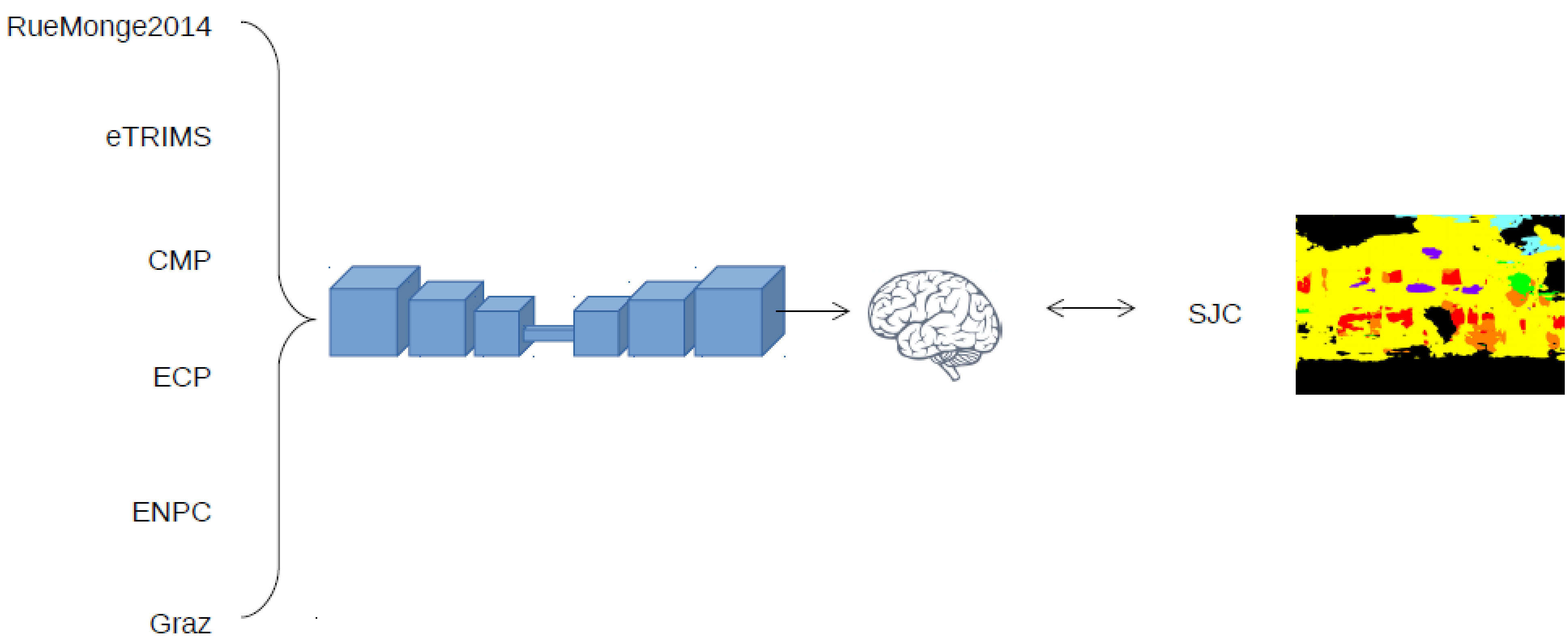
Experiments

Experiment 2

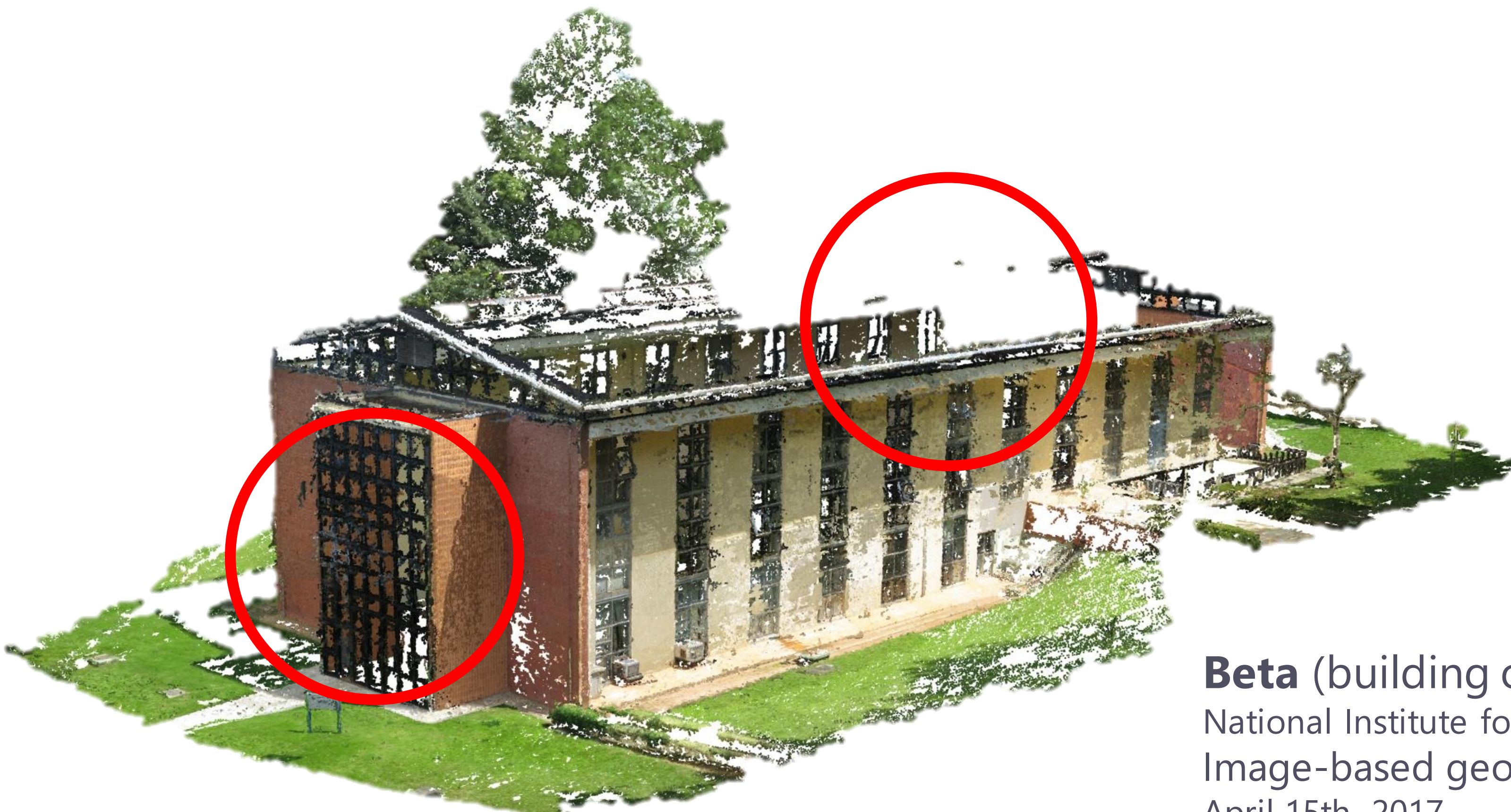


Experiments

Experiment 3



Difficulties using SfM/MVS



Beta (building on the left)
National Institute for Space Research (**INPE**)
Image-based geometry information
April 15th, 2017

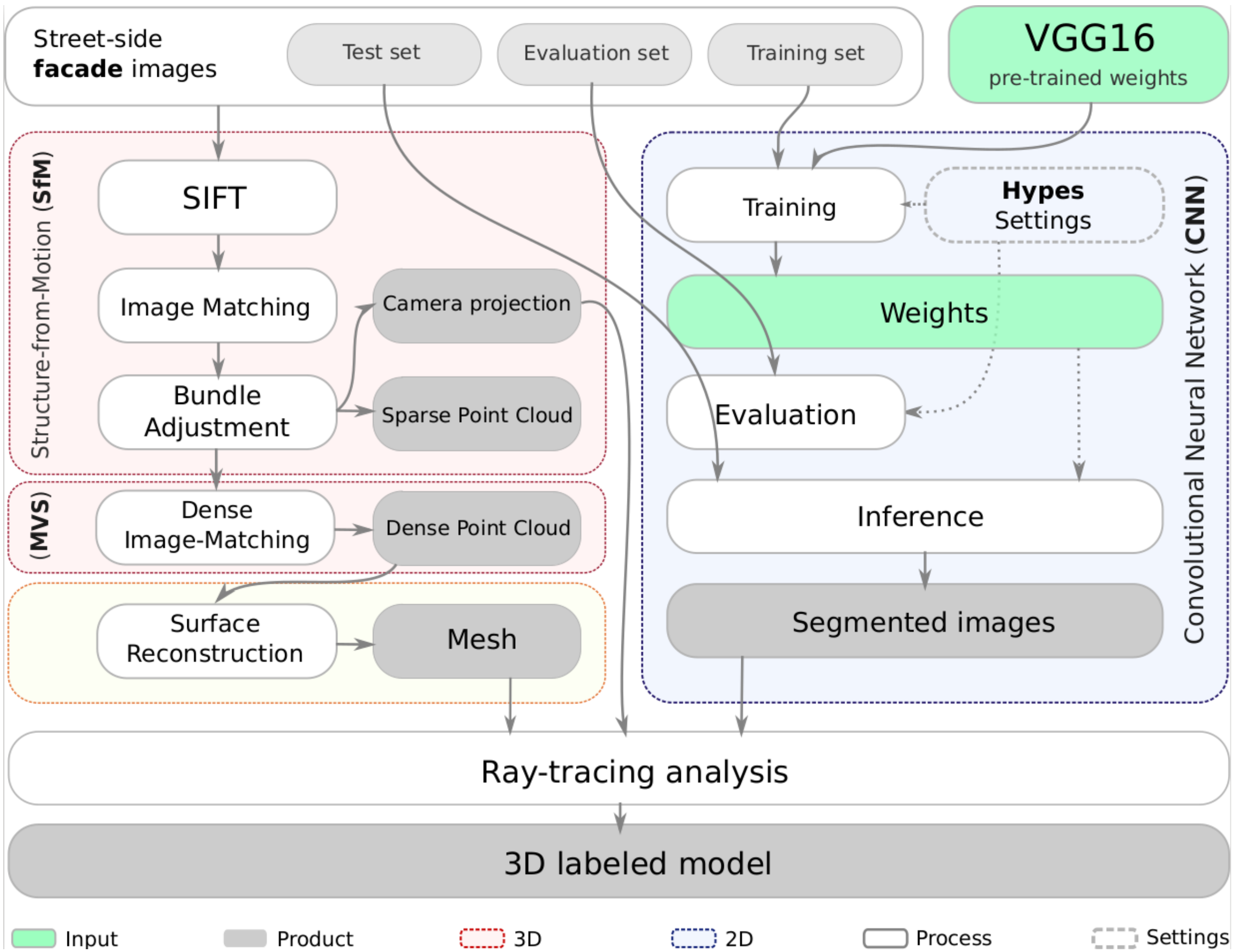
drones

Difficulties using SfM/MVS

CCST
National Institute for Space Research (**INPE**)
Image-based geometry information
April 15th, 2017

drones

Source-code



rodolfolotte/3d-reconstruction

rodolfolotte/deep-learning

rodolfolotte/image

rodolfolotte/point-cloud

rodolfolotte/documentation