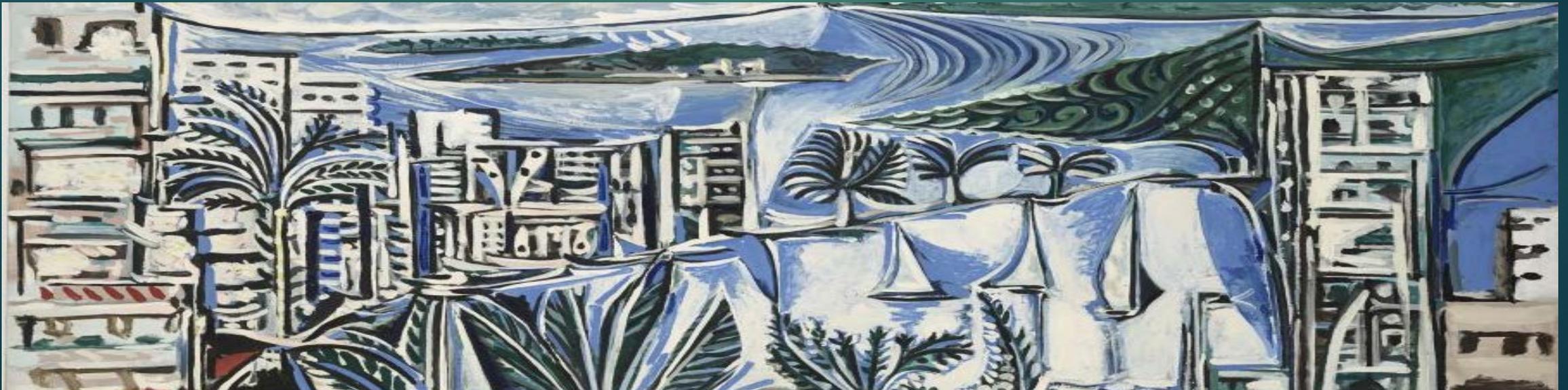
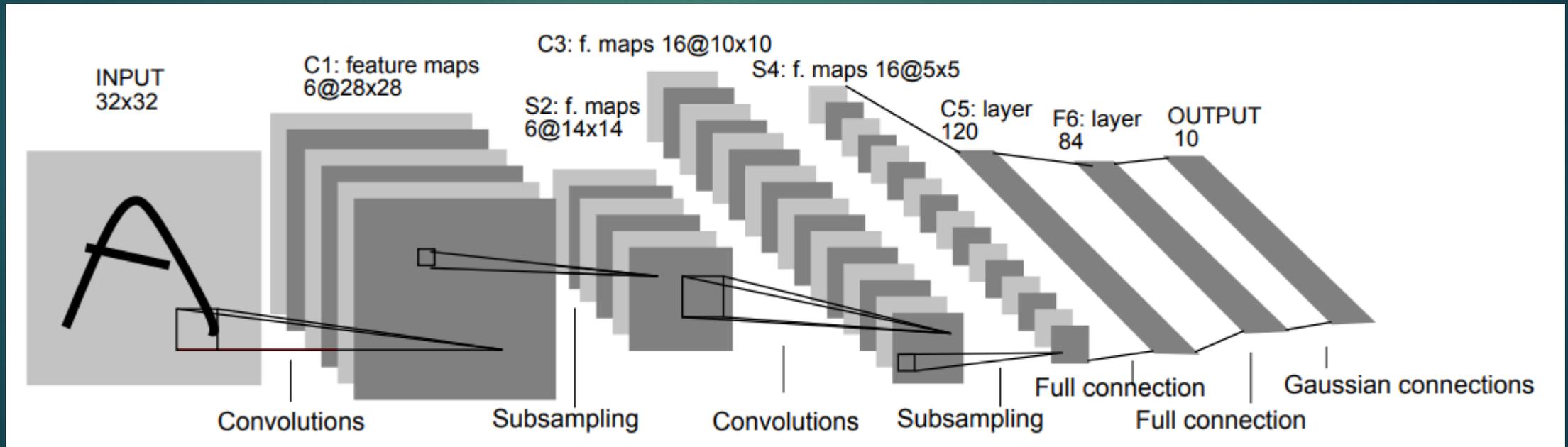


# Méthodes modernes



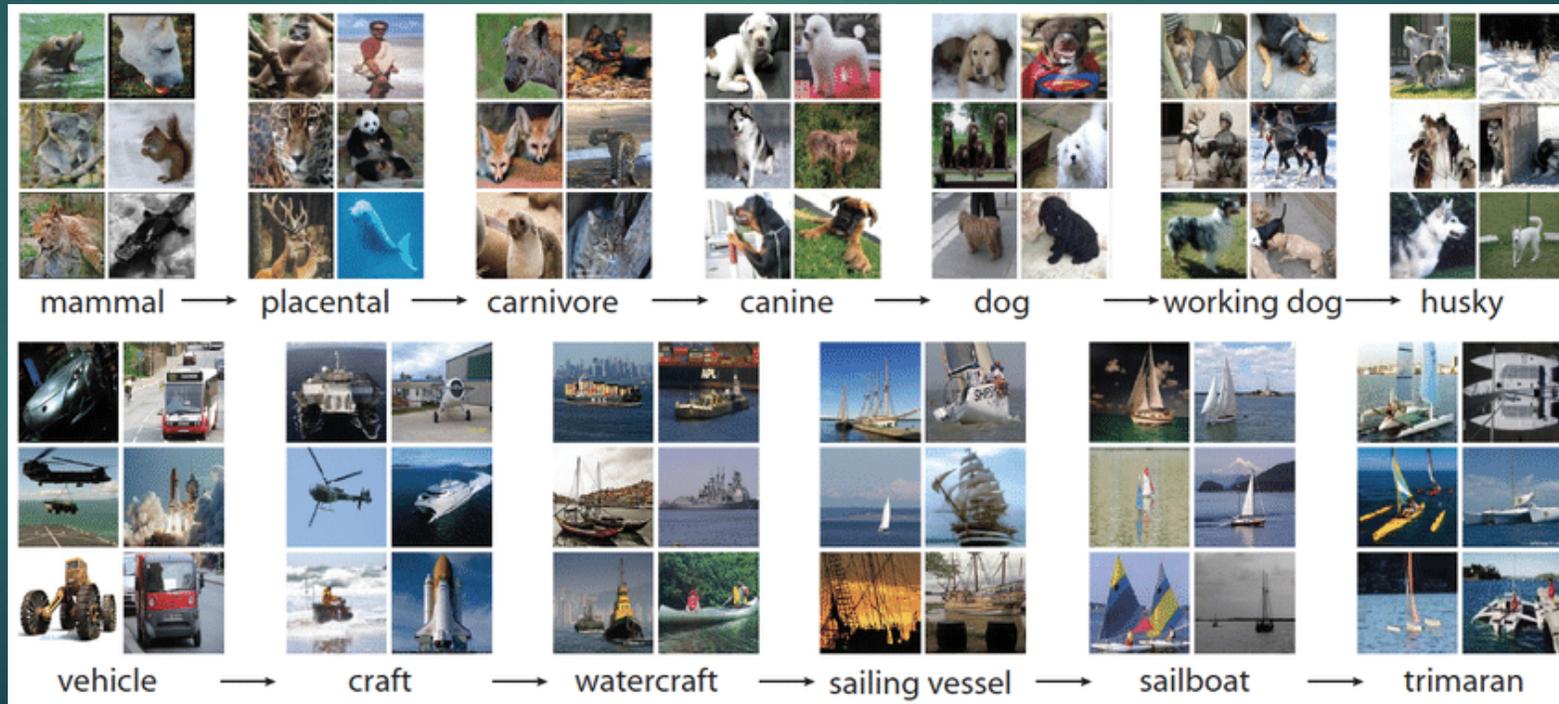
# Naissance des réseaux de neurones convolutifs (CNN)

- ▶ *Gradient-based learning applied to document recognition*, Yann LeCun, Léon Bottou, Yoshua Bengio et Patrick Haffner Proceedings of the IEEE, vol. 86, no 11, 1998



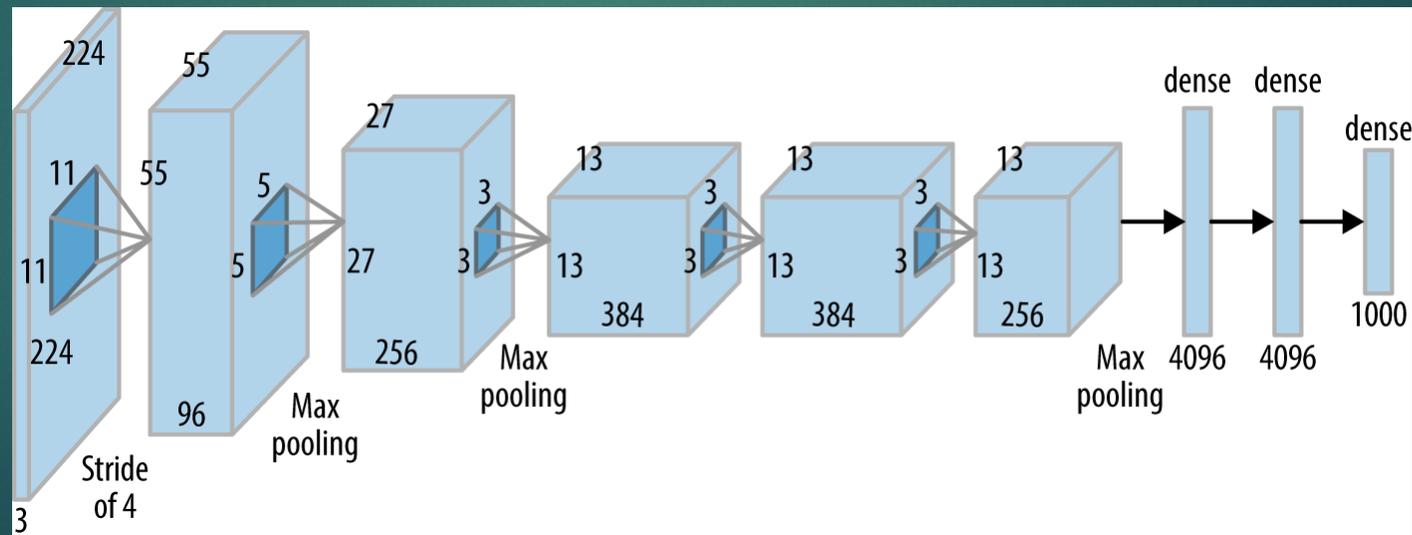
# Naissance des réseaux de neurones convolutifs (CNN)

- ▶ ImageNet : dataset conçu par des académiques pour faire avancer la recherche en vision par ordinateur
- ▶ 14 million d'images organisées et labélisées hiérarchiquement



# Naissance des réseaux de neurones convolutifs (CNN)

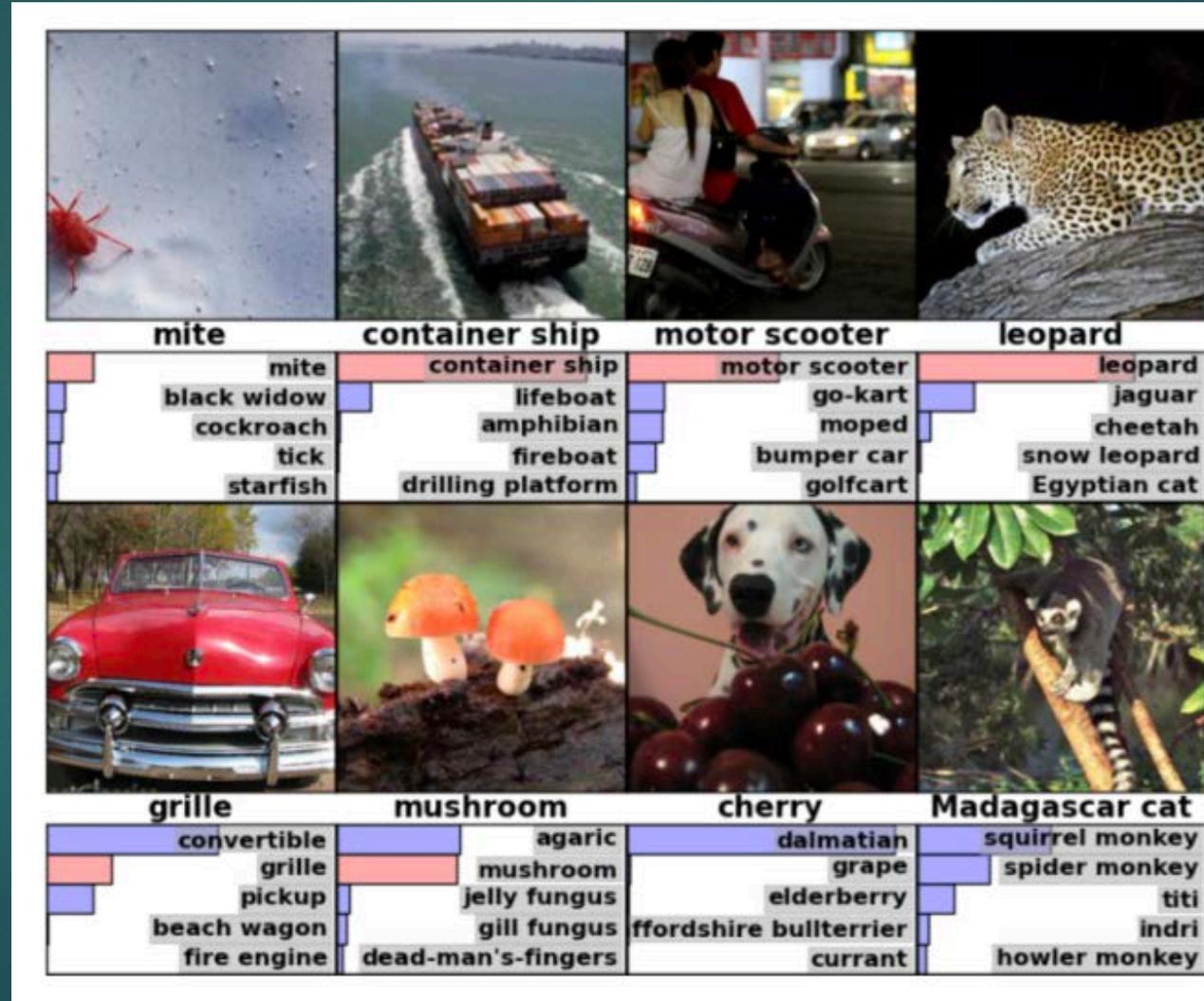
- ▶ 30 septembre 2012 : AlexNet remporte le "ImageNet Large Scale Visual Recognition Challenge" avec une erreur sur le top-5 de 15,3%, soit 10,8 points de moins que le dernier vainqueur du challenge
- ▶ Les auteurs ont montré que la profondeur du réseau est essentielle dans les performances, ce qui rend les calculs coûteux mais possibles grâce aux GPUs



Alexnet Block Diagram (source:oreilly.com)

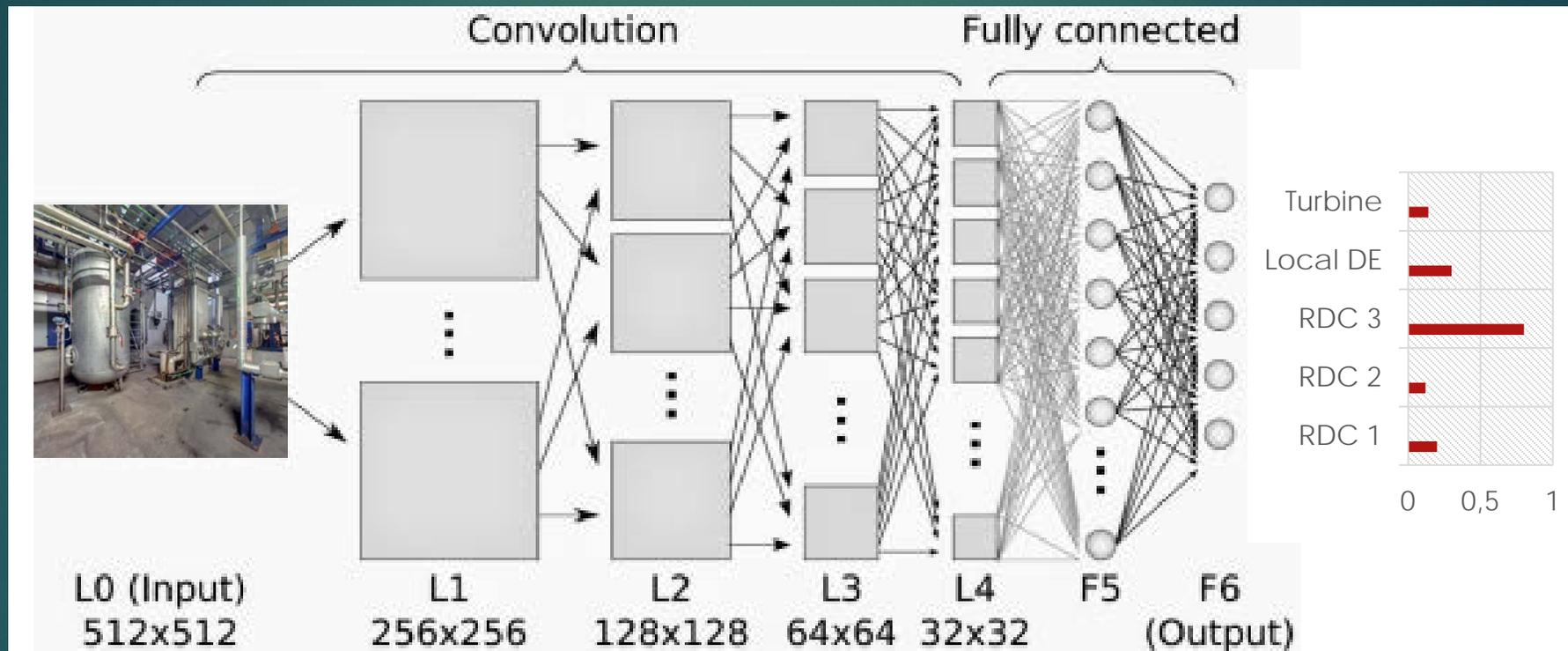
# Méthodes non génériques

# Classification par CNN



# Classification par CNN : application à la reconnaissance de scène

7

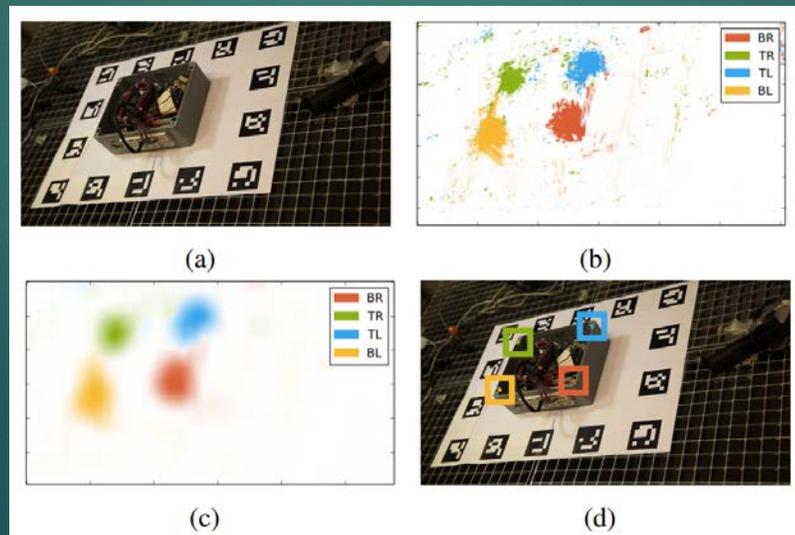
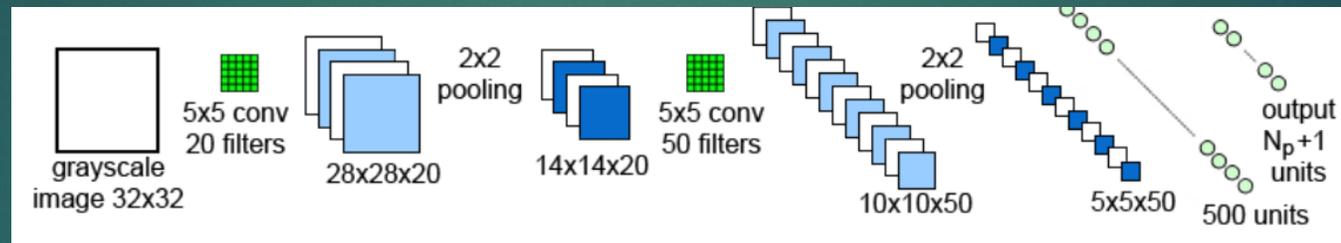


Problème : lorsqu'une nouvelle classe est introduite, le CNN doit être réappris

# Classification par CNN : application à la détection de « points de contrôle »



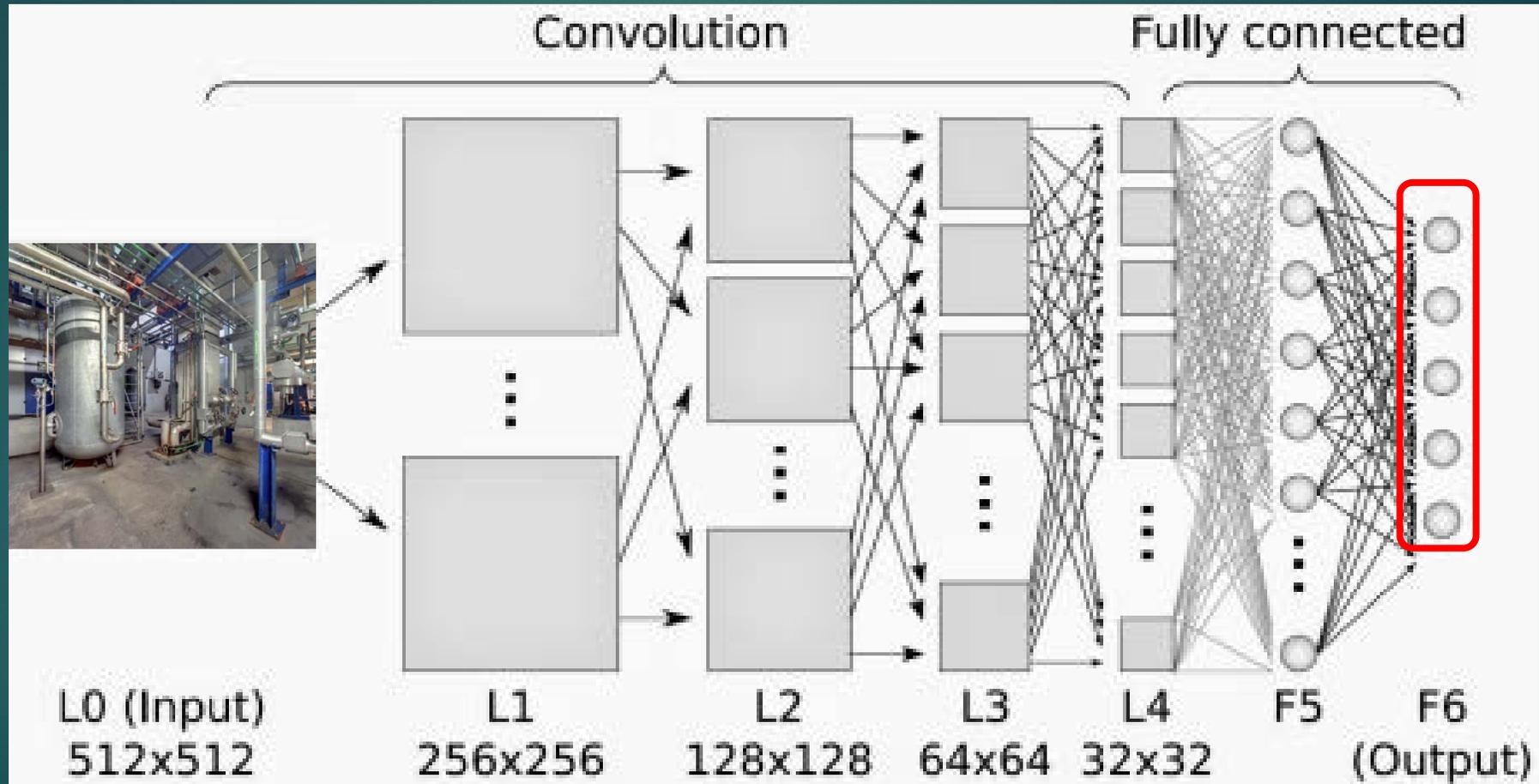
- ▶ *A Novel Representation of Parts for Accurate 3D Object Detection and Tracking in Monocular Images*, Crivellaro et al., ICCV 2015



La pose peut alors être calculée par PnP

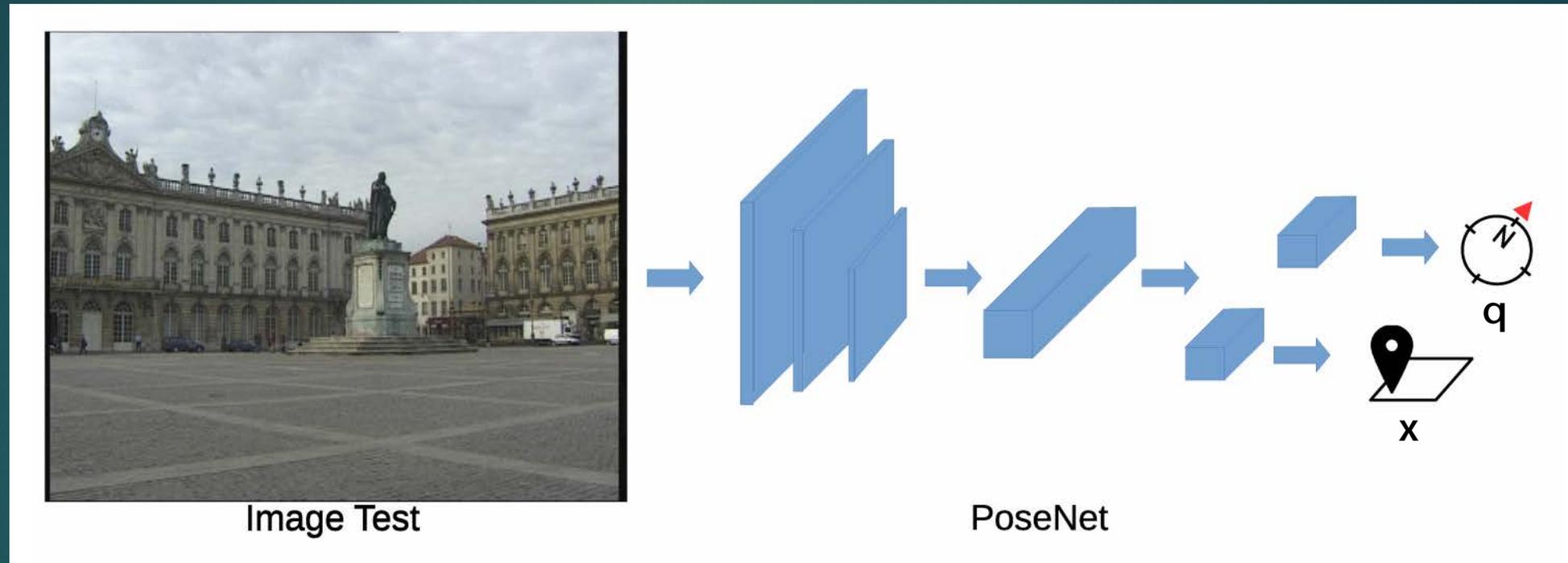
Problème : ce réseau doit être réappris pour tout nouvel objet

# Régression par CNN



# Régression par CNN : PoseNet

- ▶ PoseNet: A Convolutional Network for Real-Time 6-DOF Camera Relocalization. Kendall et al., ICCV'2015

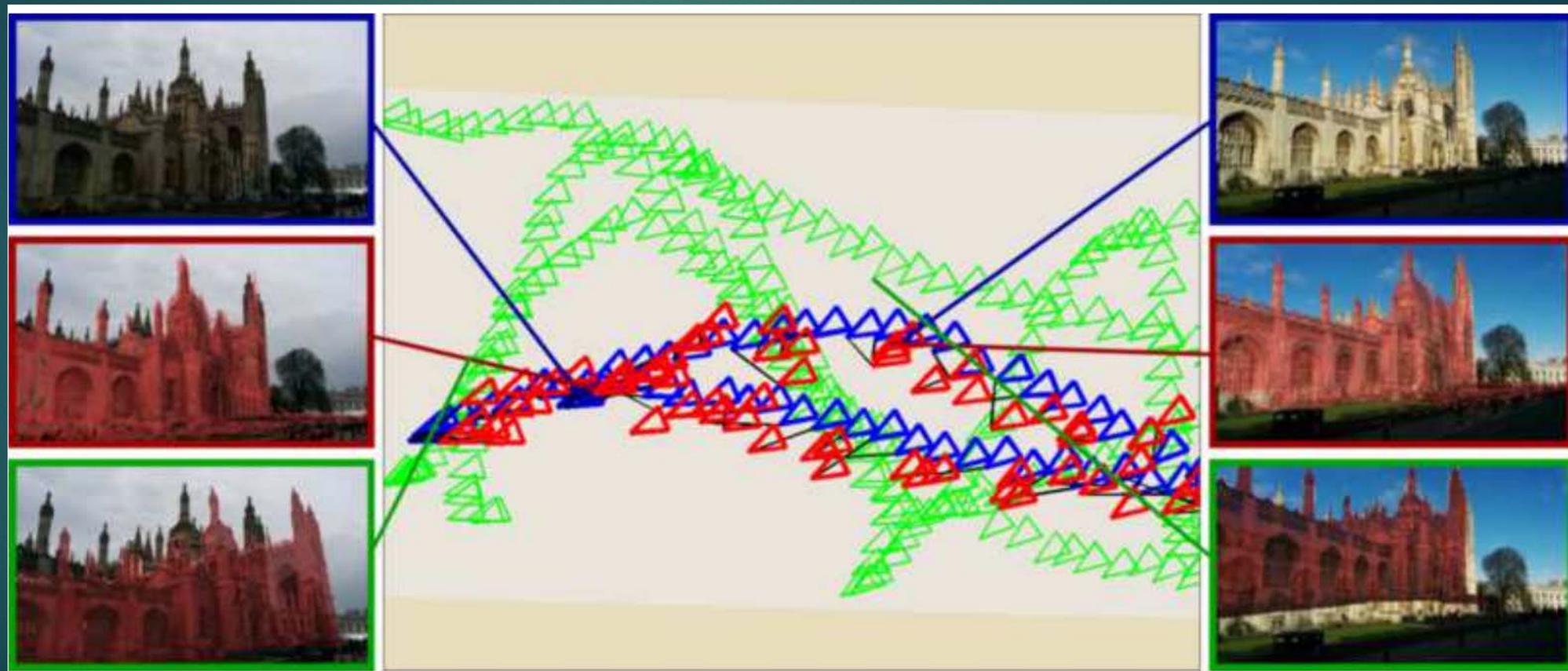


$$loss(I) = \|\hat{\mathbf{x}} - \mathbf{x}\|_2 + \beta \left\| \hat{\mathbf{q}} - \frac{\mathbf{q}}{\|\mathbf{q}\|} \right\|_2$$

# Régression par CNN : PoseNet

11

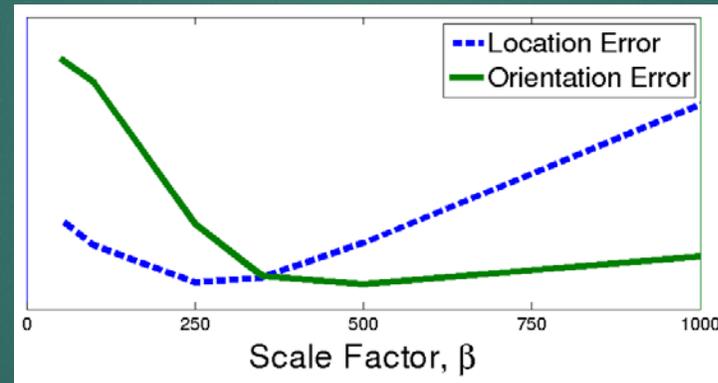
- ▶ PoseNet: A Convolutional Network for Real-Time 6-DOF Camera Relocalization. Kendall et al., ICCV'2015



# Régression par CNN : PoseNet

12

- ▶ PoseNet: *A Convolutional Network for Real-Time 6-DOF Camera Relocalization*. Kendall et al., ICCV'2015
- ▶ Problèmes :
  - ▶ Problème du choix de  $\beta$

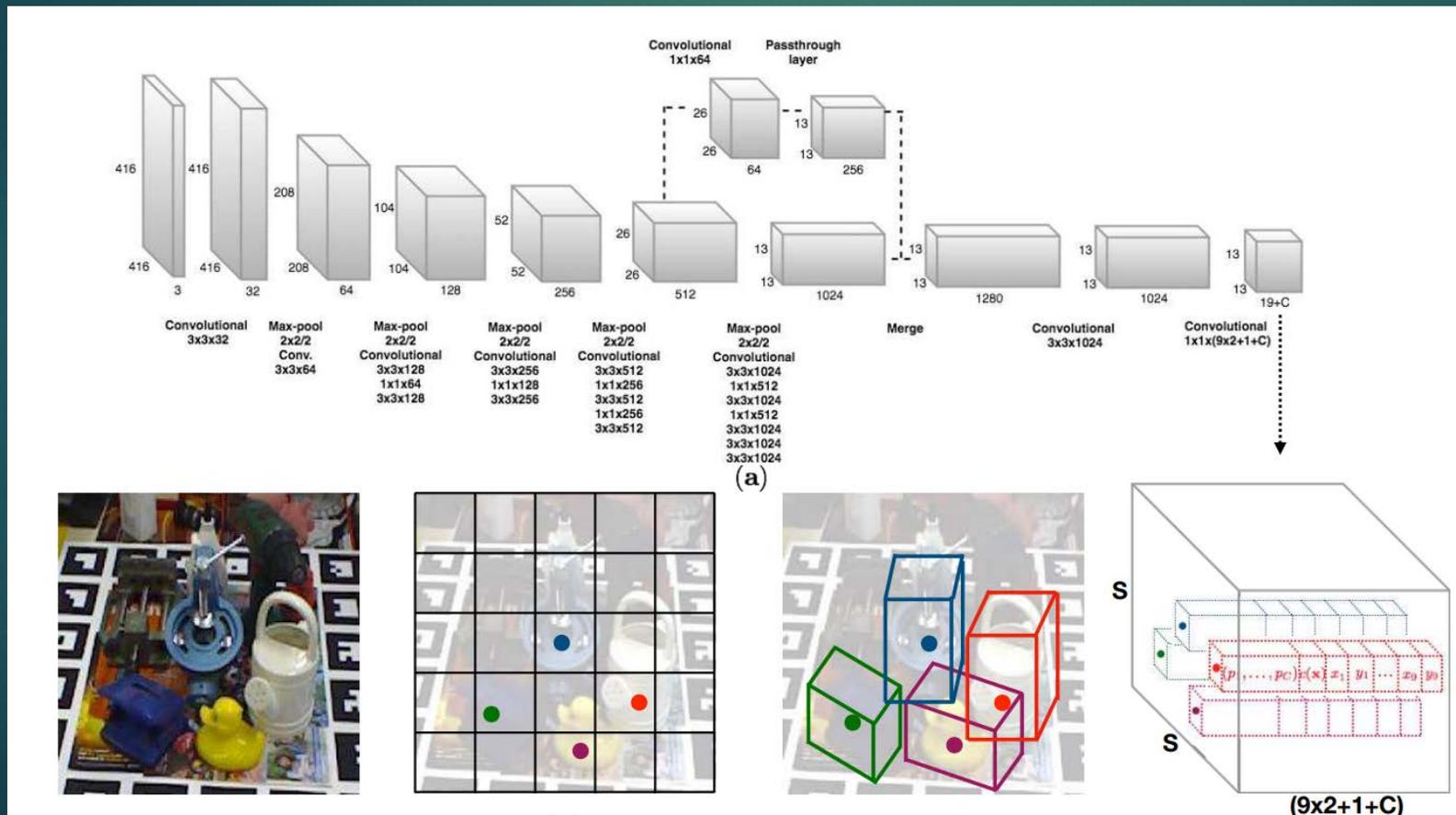


- ▶ Généralement moins précis qu'un calcul PnP
- ▶ Problème de la représentativité des données (équivalent à une recherche par plus proche voisin, cf. [Sattler 2019])
- ▶ Doit être réappris pour chaque nouvelle scène

# Régression par CNN : détection des sommets de boîtes englobantes

13

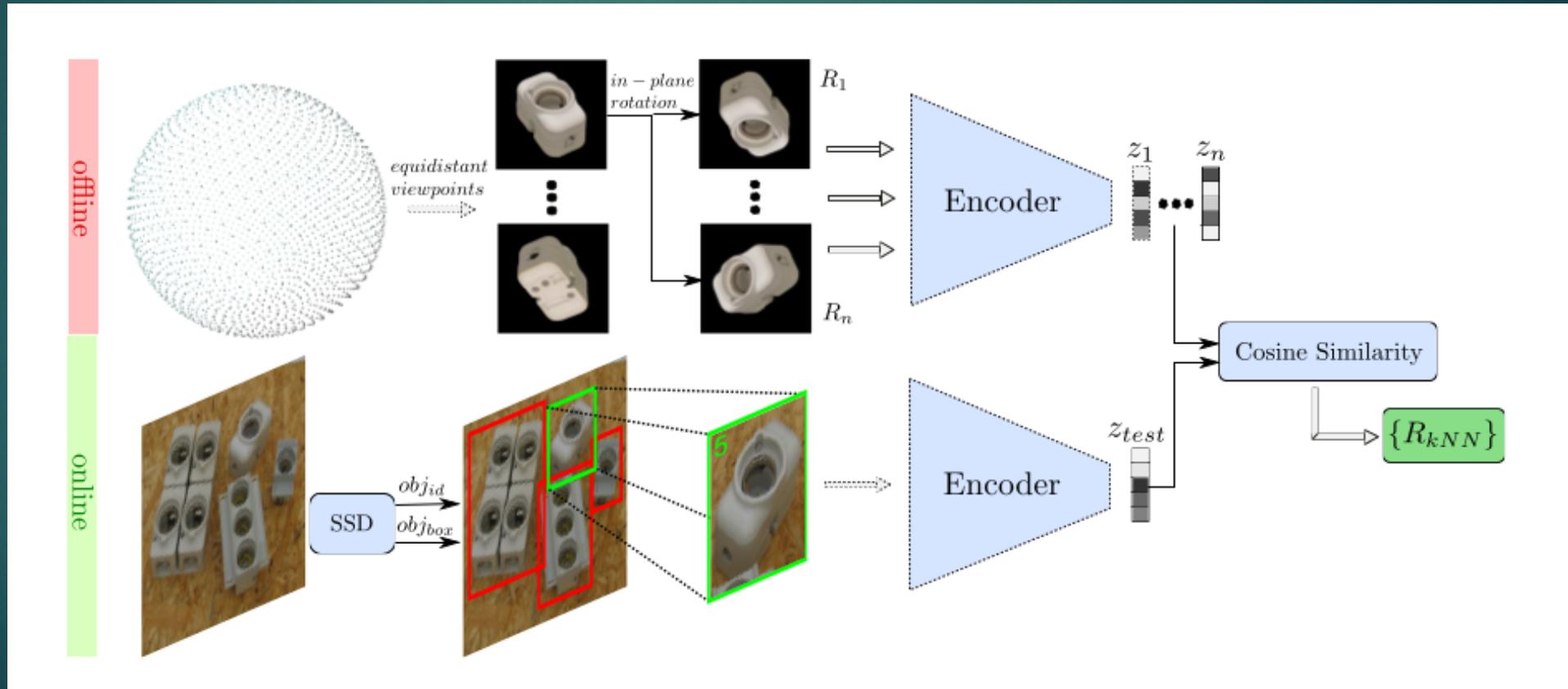
- Real-Time Seamless Single Shot 6D Object Pose Prediction [Tekin et al. CVPR 2018]



# Architectures Encoder-Decoder : tracking par synthèse revisité

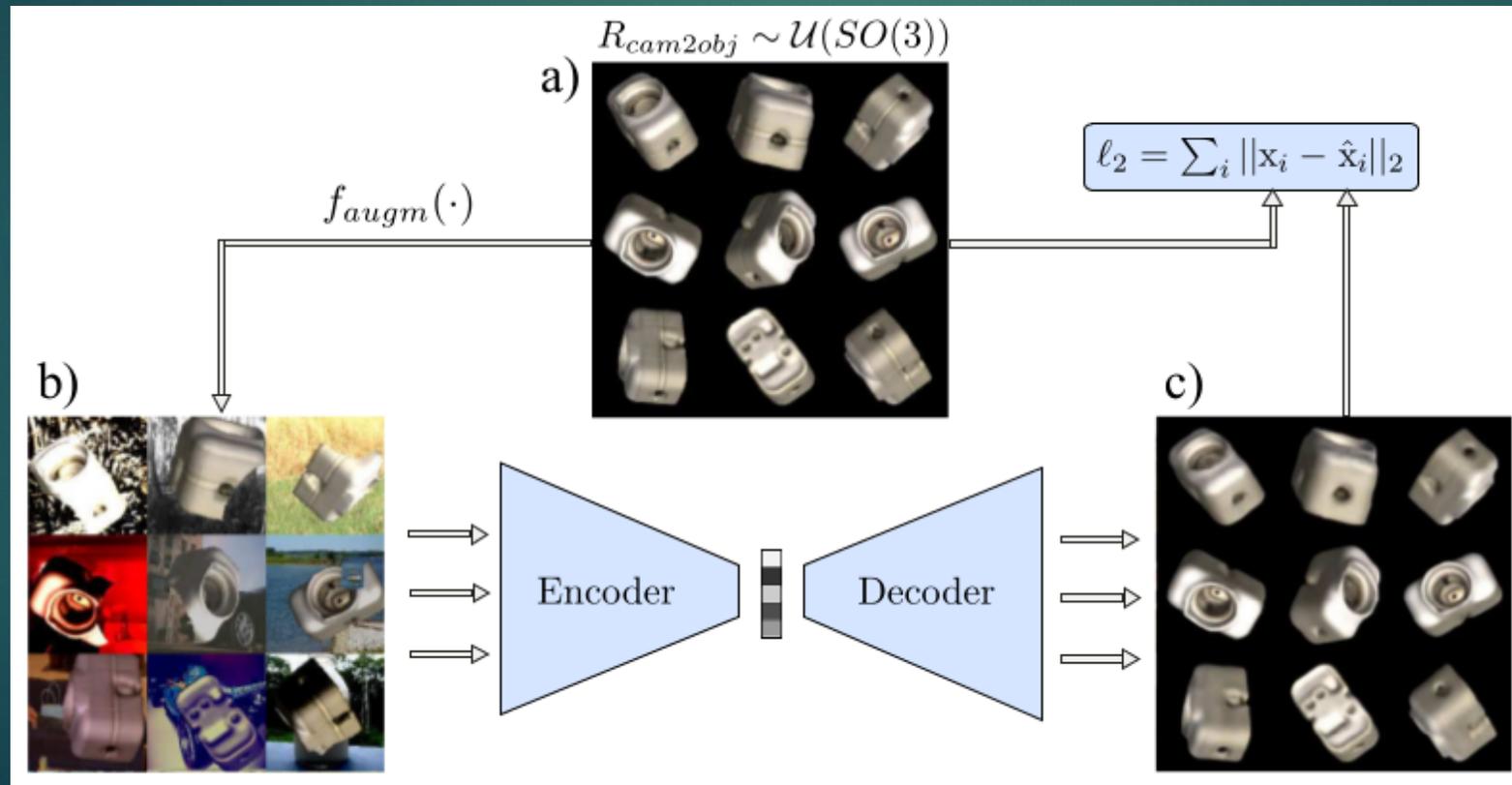
14

- ▶ *Implicit 3D Orientation Learning for 6D Object Detection from RGB Images*, Sundermeyer et al., ECCV 2018



# Architectures Encoder-Decoder : tracking par synthèse revisit 

- Utilisation d'une architecture E/D pour rendre la m thode robuste aux conditions d'acquisition (occultations, illumination, arri re-plan, ...)



Probl me : ce r seau doit  tre r appri  pour tout nouvel objet

# Méthodes génériques

# Méthodes génériques

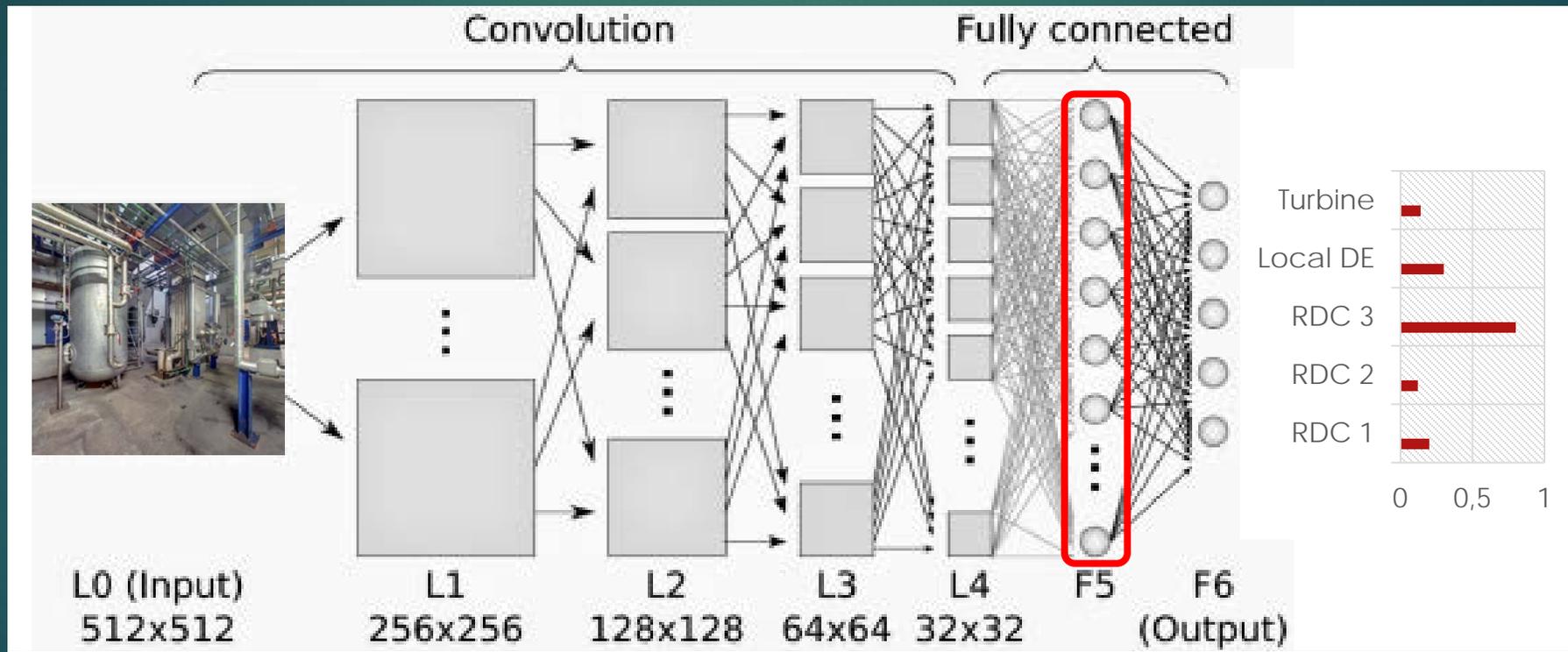
17

- ▶ Toutes les méthodes qui précèdent requièrent un entraînement spécifique à chaque scène / objet
- ▶ Deux manières d'éviter ce problème :
  - ▶ Utiliser des descripteurs CNN (ConvNet features)
  - ▶ Considérer une sémantique de classe plutôt que d'instance
    - ▶ Cartes de sémantique
    - ▶ Détection de classes d'objets

# Méthodes génériques

DESCRIPTEURS CNN

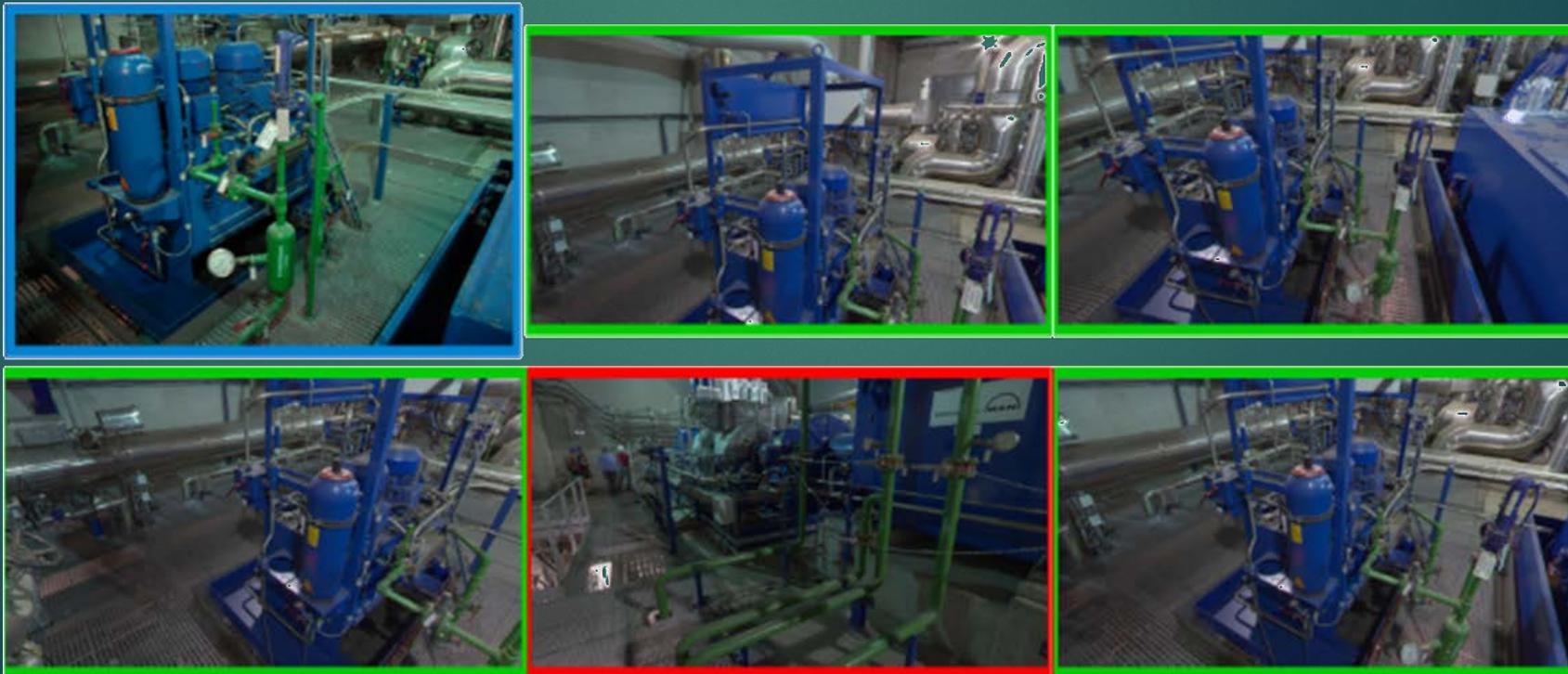
# Descripteurs CNN



# Descripteurs CNN : application à la reconnaissance de lieu

20

- ▶ Mêmes méthodes qu'avec les BoW, en remplaçant les histogrammes de mots visuels par des descripteurs CNN calculés sur toute l'image (descripteurs globaux)
- ▶ Exemple [Gaudillière et al. 2018]



# Descripteurs CNN : application à la reconnaissance de lieu

21

- ▶ Mêmes méthodes qu'avec les BoW, en remplaçant les histogrammes de mots visuels par des descripteurs CNN calculés sur toute l'image (descripteurs globaux)
- ▶ Exemple [Gaudillière et al. 2018]



# Descripteurs CNN : application à la reconnaissance de lieu

22

- ▶ Mêmes méthodes qu'avec les BoW, en remplaçant les histogrammes de mots visuels par des descripteurs CNN calculés sur toute l'image (descripteurs globaux)
- ▶ Exemple [Gaudillière et al. 2018]



# Region proposal

- ▶ Méthodes de proposition de boites (box proposal ou region proposal) susceptibles de contenir des objets
- ▶ Exemples :
  - ▶ Selective Search [Uijlings et al. 2012]
  - ▶ Edge Boxes [Zitnick et al. 2014]



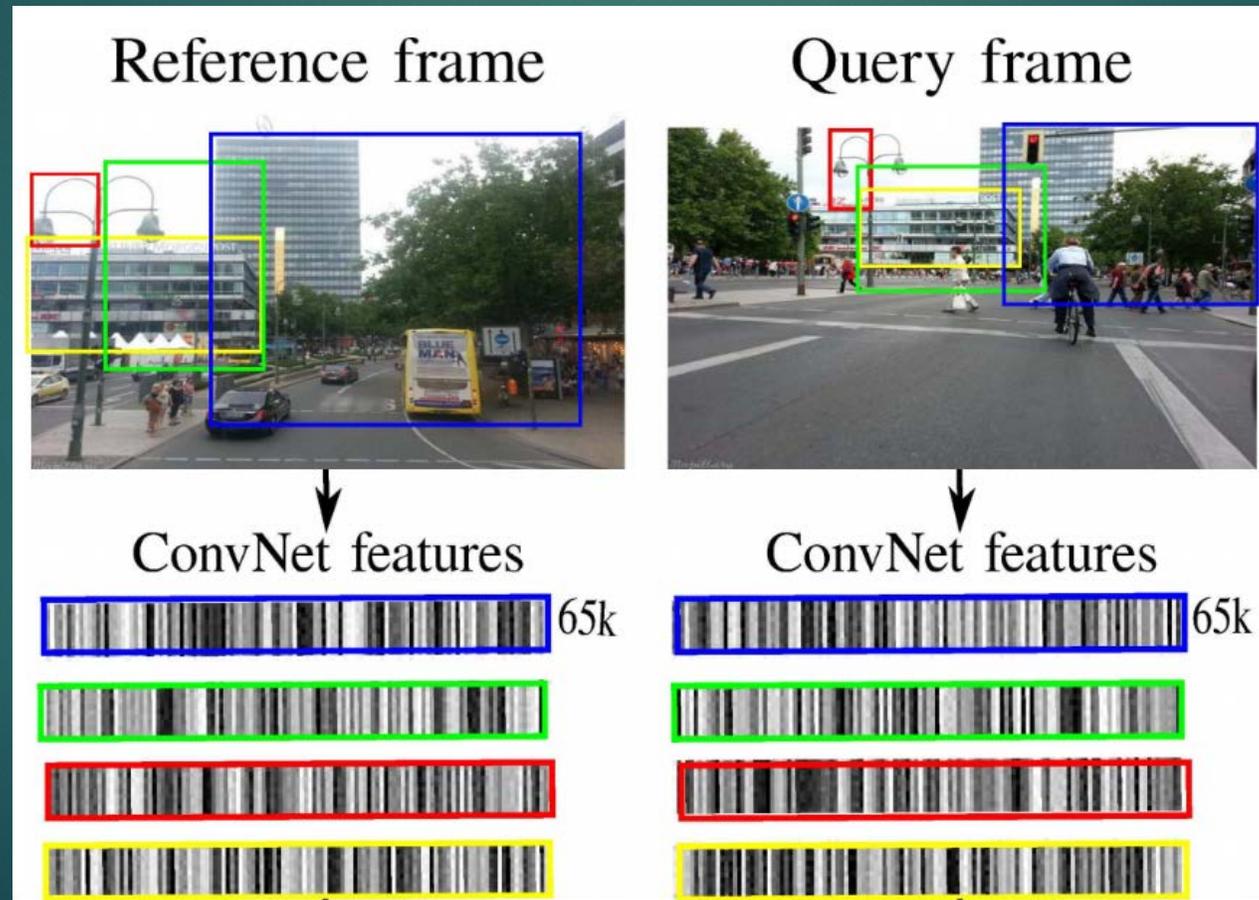
Selective Search



Edge Boxes

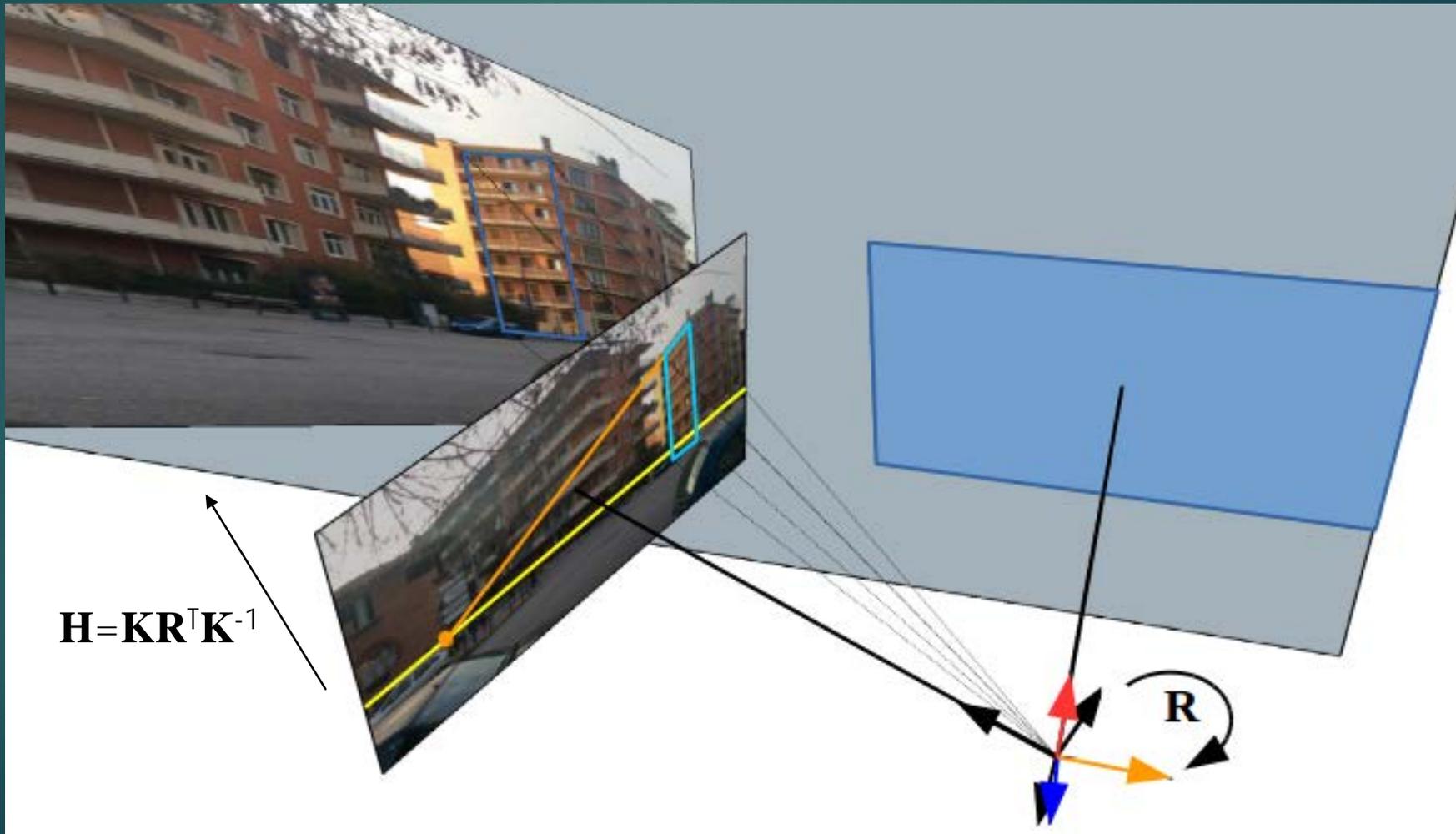
# Region proposal + descripteurs CNN : reconnaissance de lieux

- ▶ Place Recognition with ConvNet Landmarks [Sünderhauf, 2015]



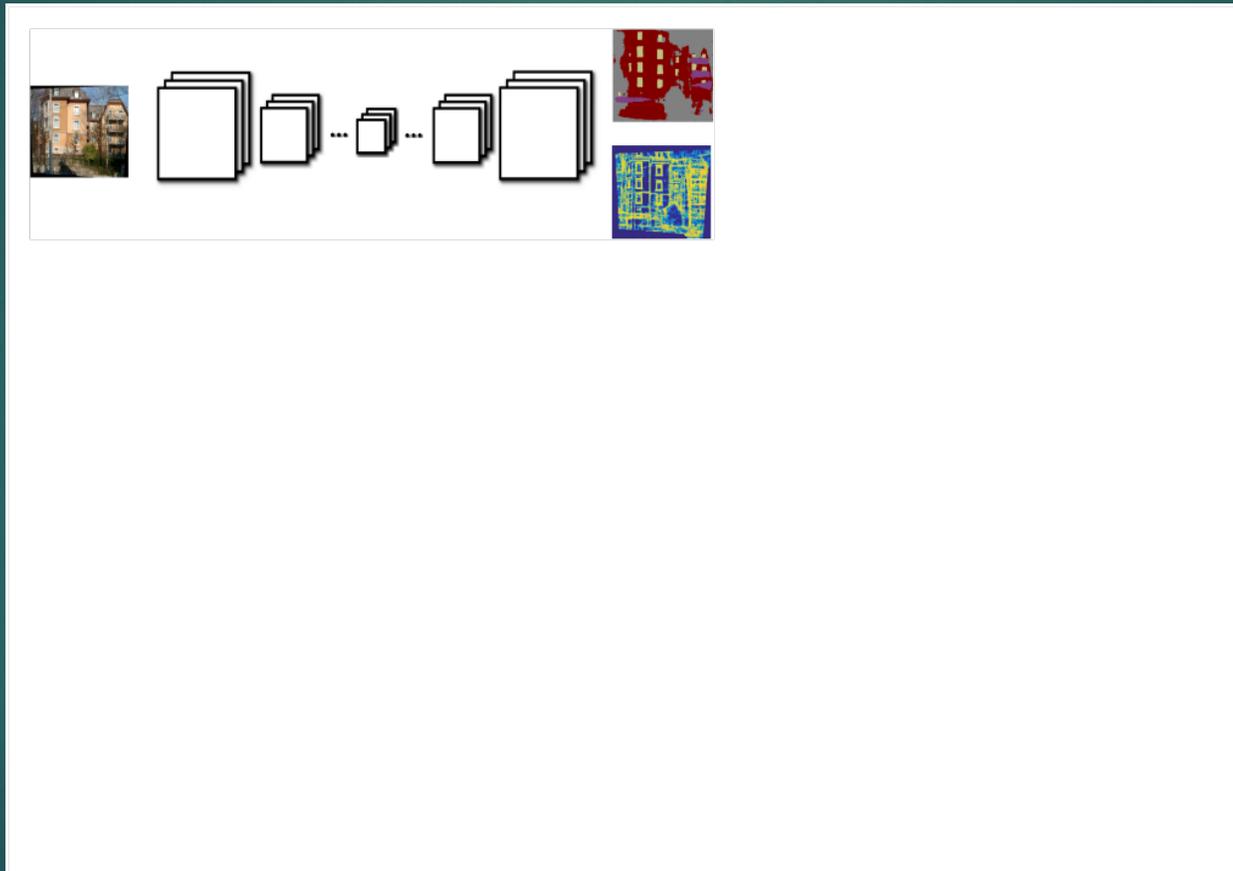
# Region proposal + descripteurs CNN : cas particulier des façades

25



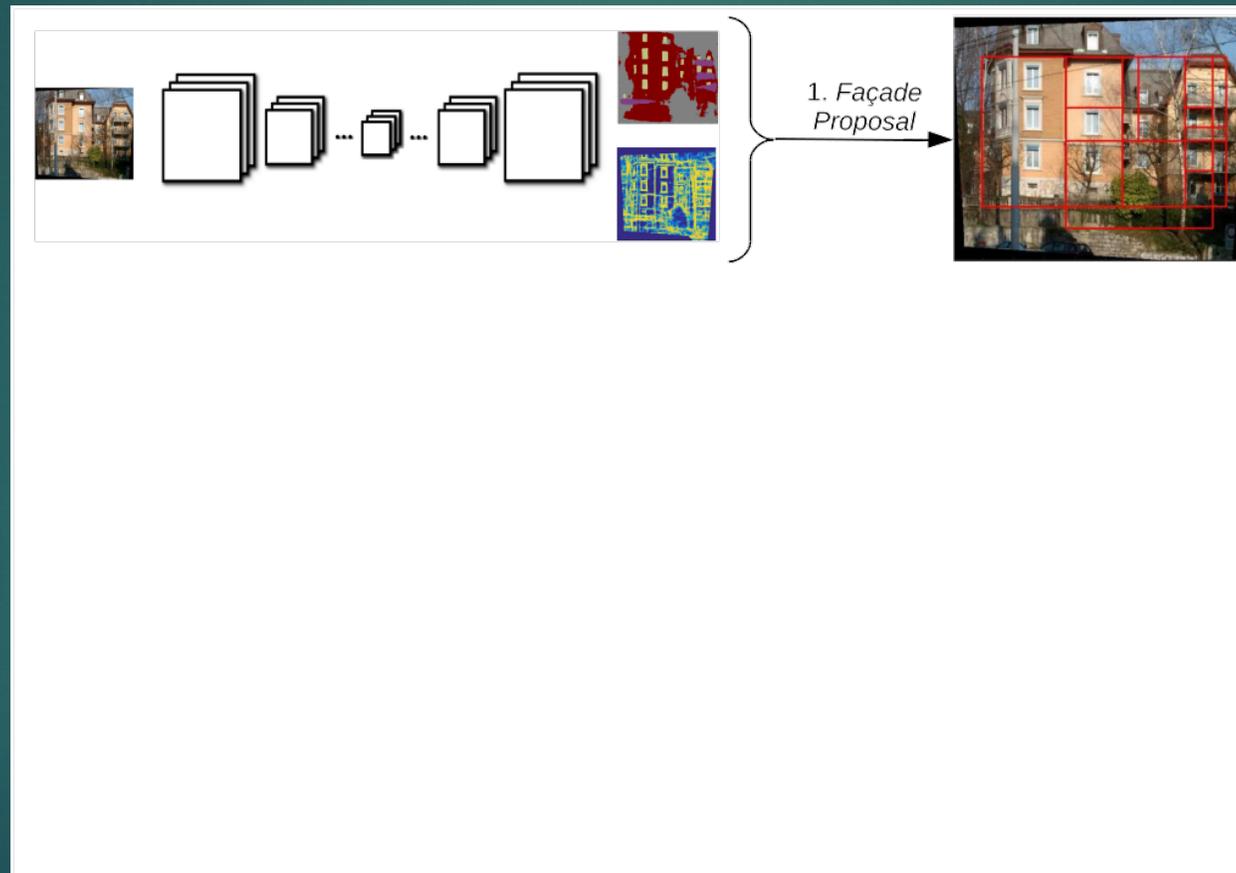
# Region proposal + descripteurs CNN : cas particulier des façades

26



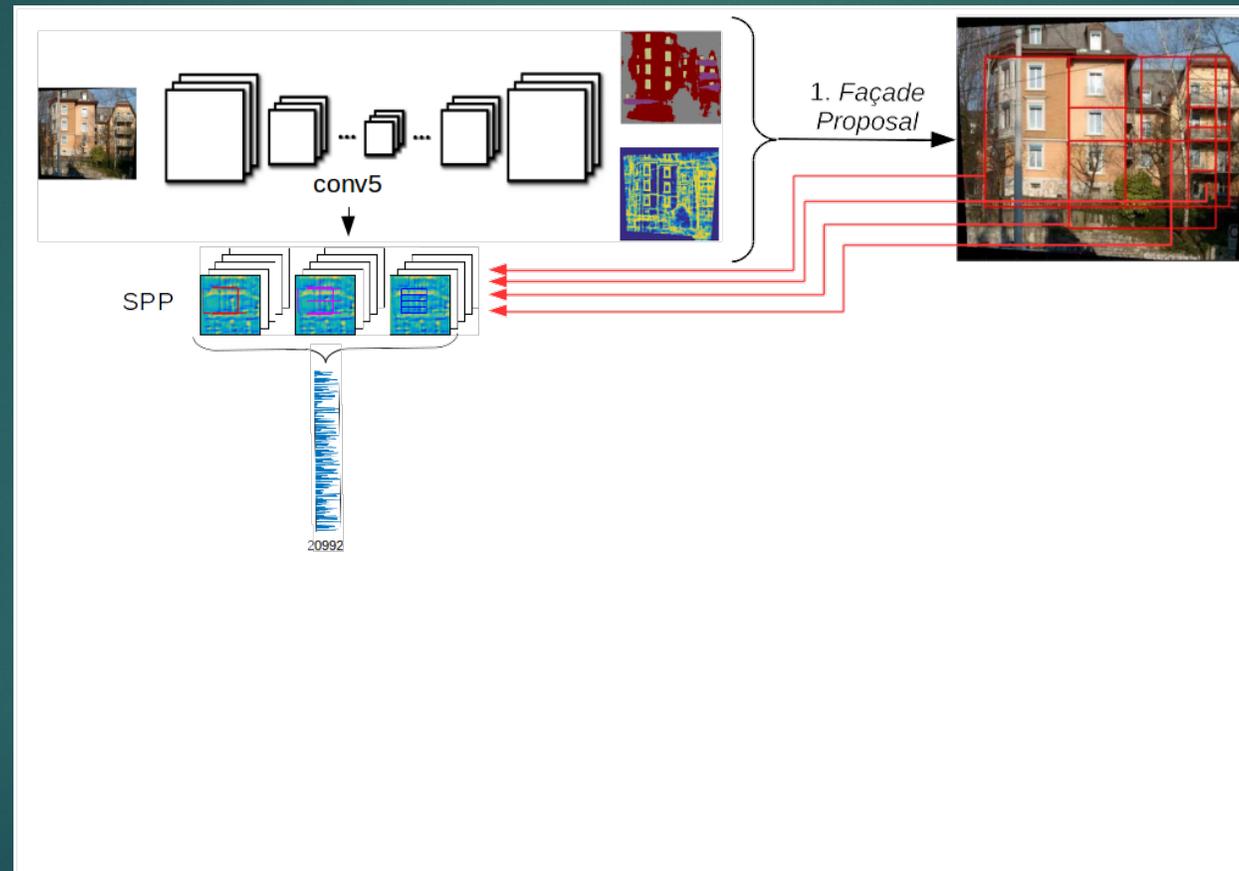
# Region proposal + descripteurs CNN : cas particulier des façades

27



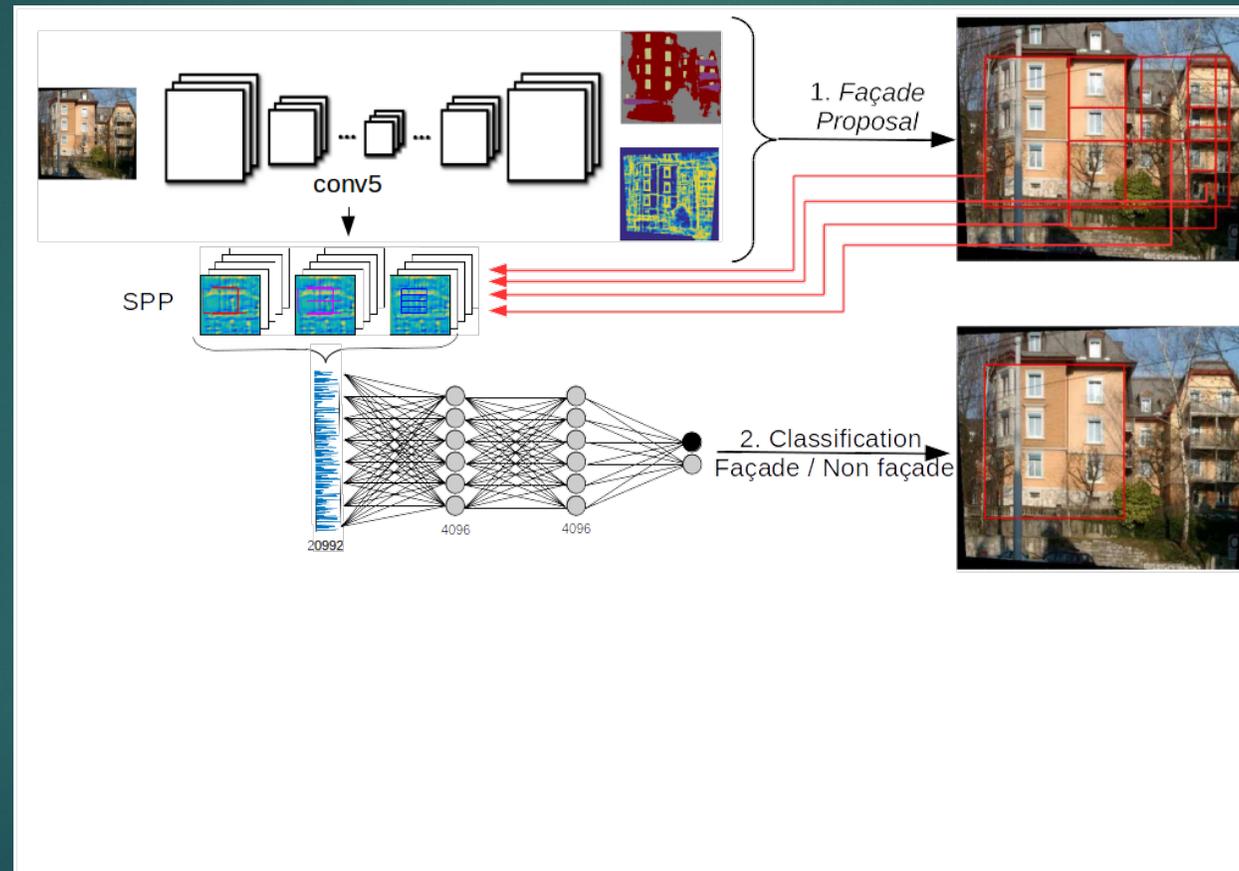
# Region proposal + descripteurs CNN : cas particulier des façades

28



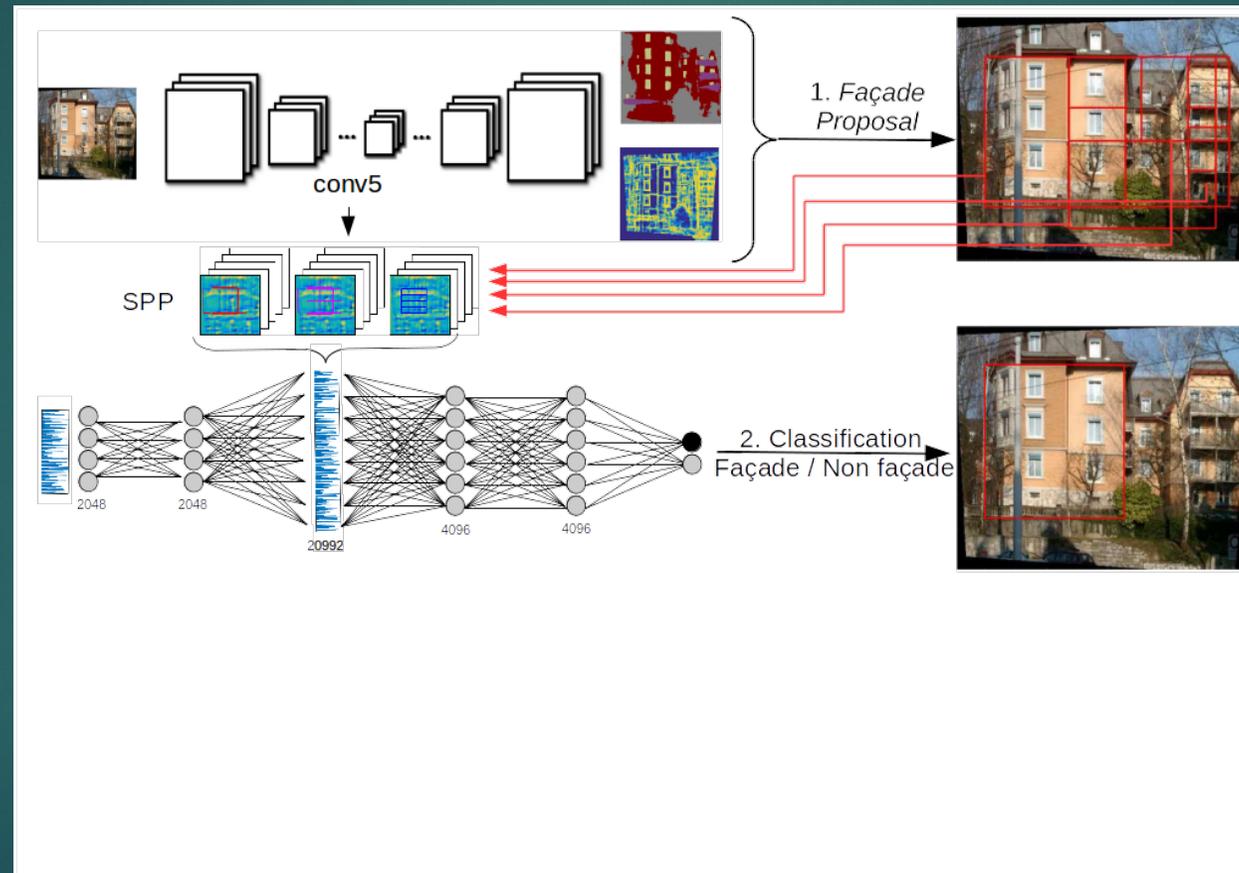
# Region proposal + descripteurs CNN : cas particulier des façades

29



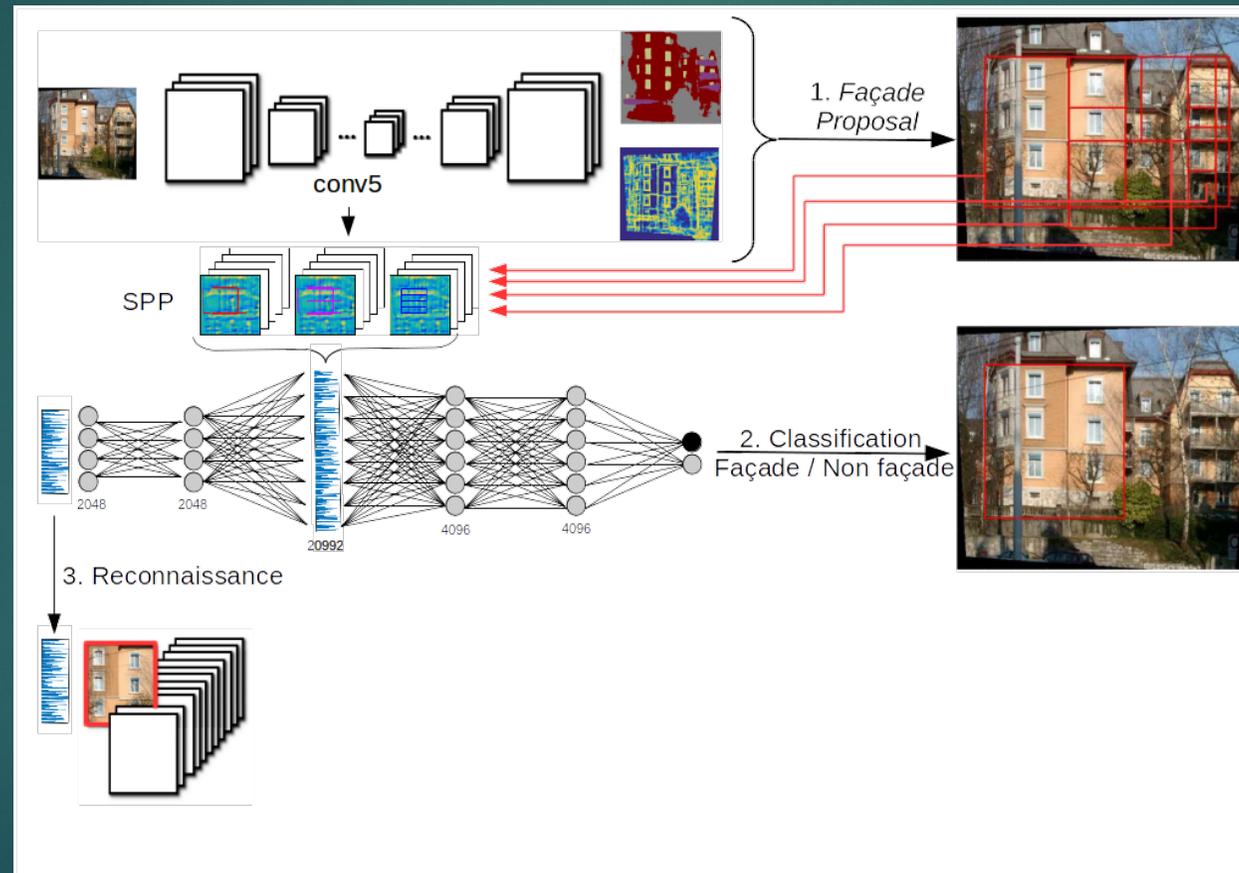
# Region proposal + descripteurs CNN : cas particulier des façades

30



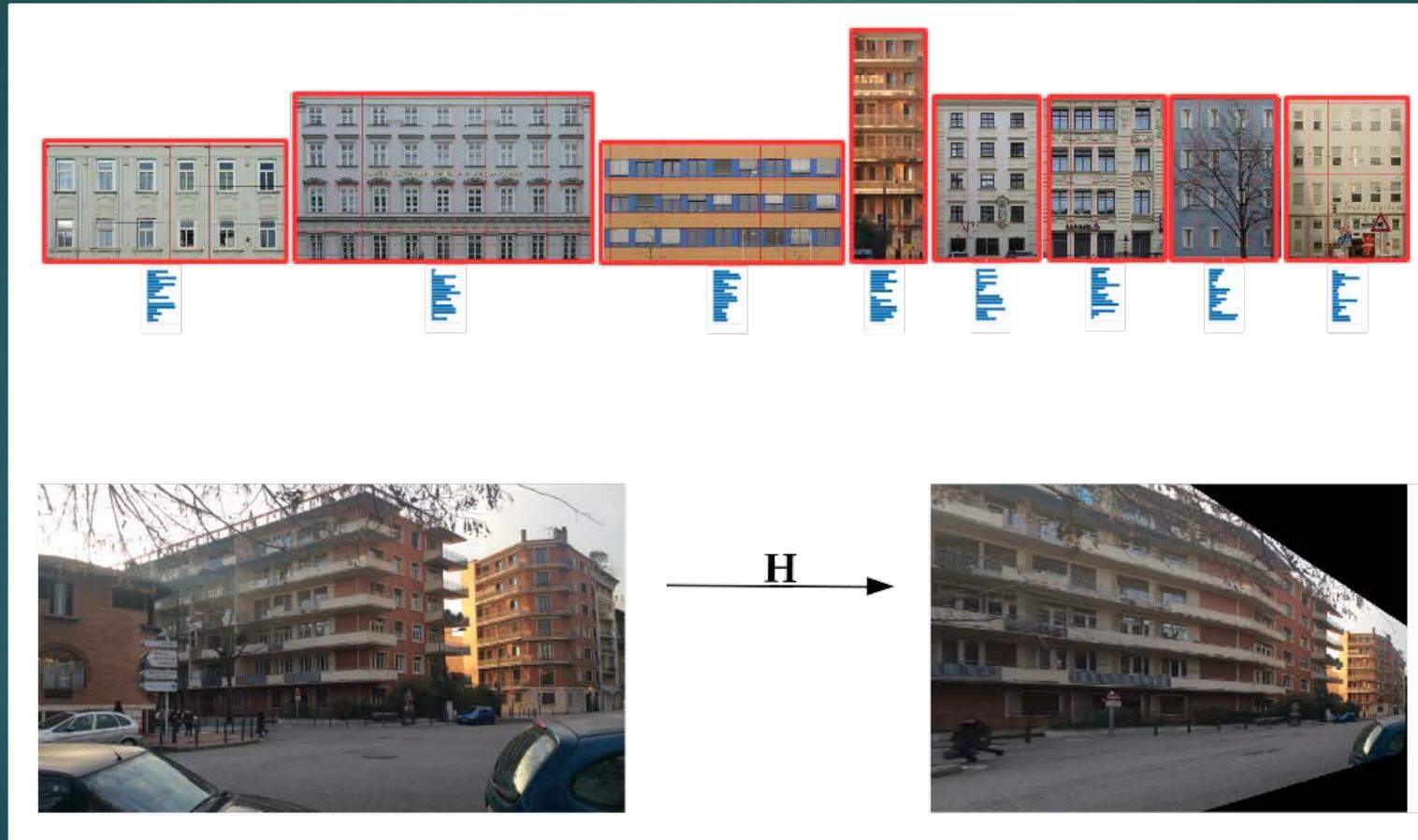
# Region proposal + descripteurs CNN : cas particulier des façades

31



# Region proposal + descripteurs CNN :

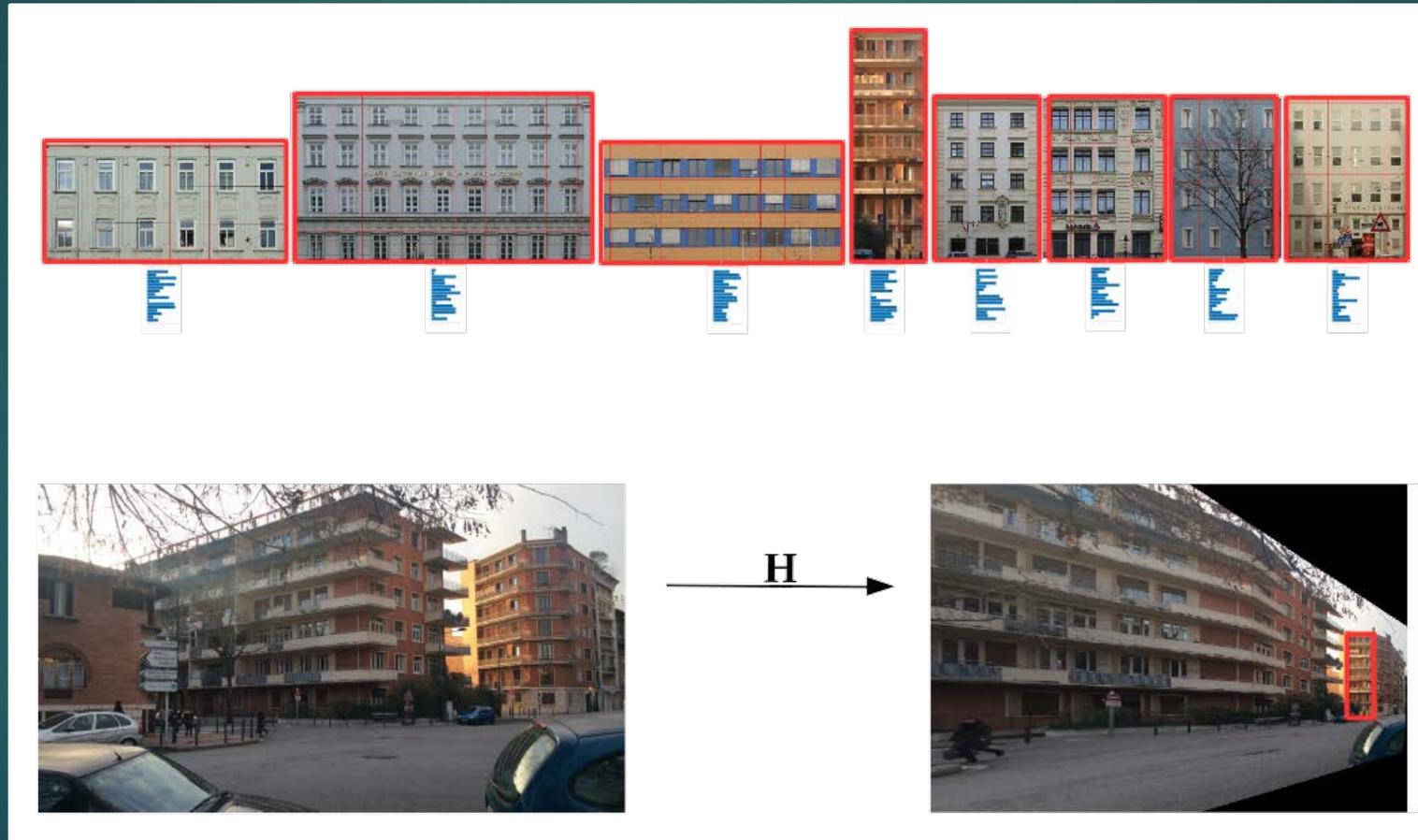
32



# Region proposal + descripteurs CNN :

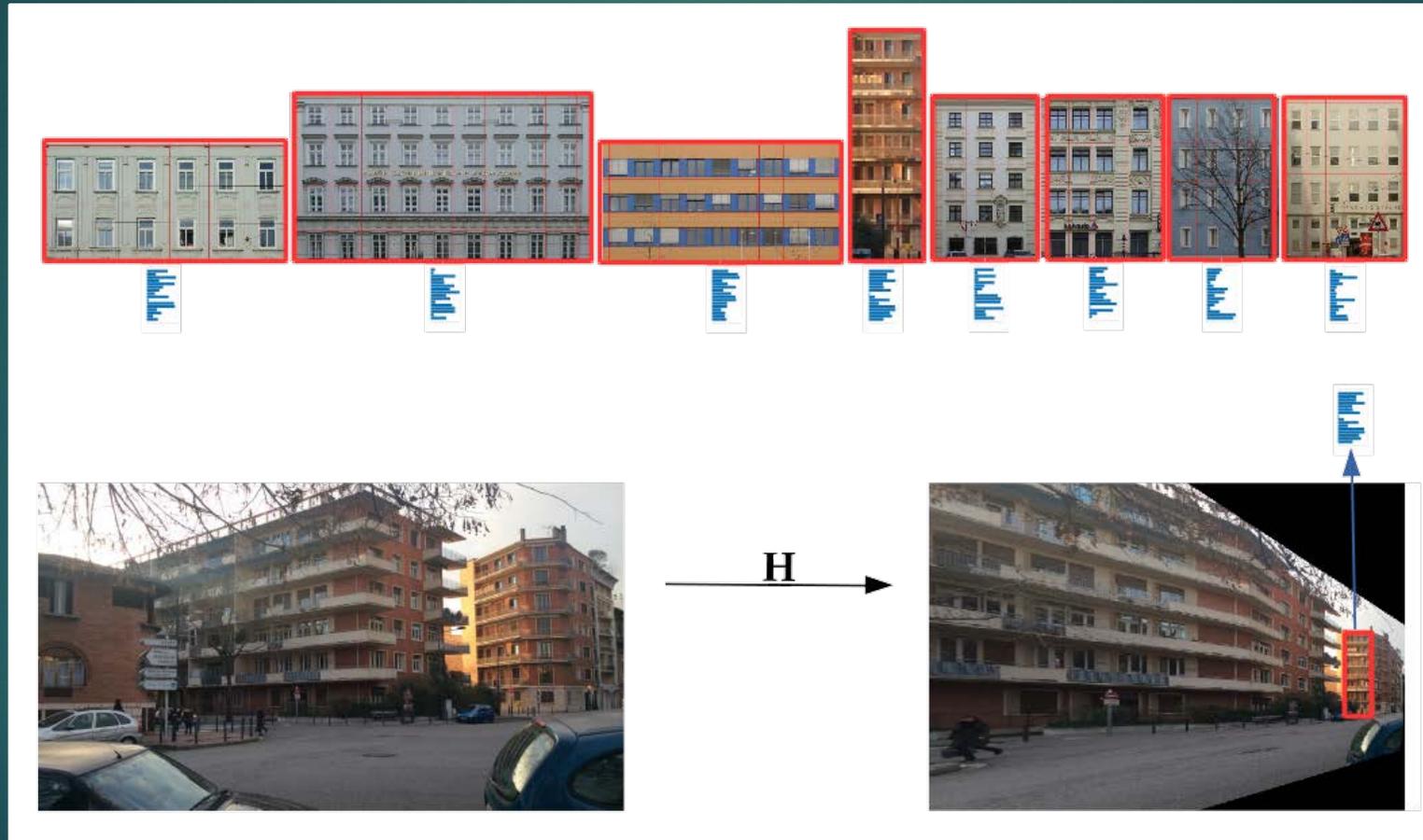
cas particulier des façades

33



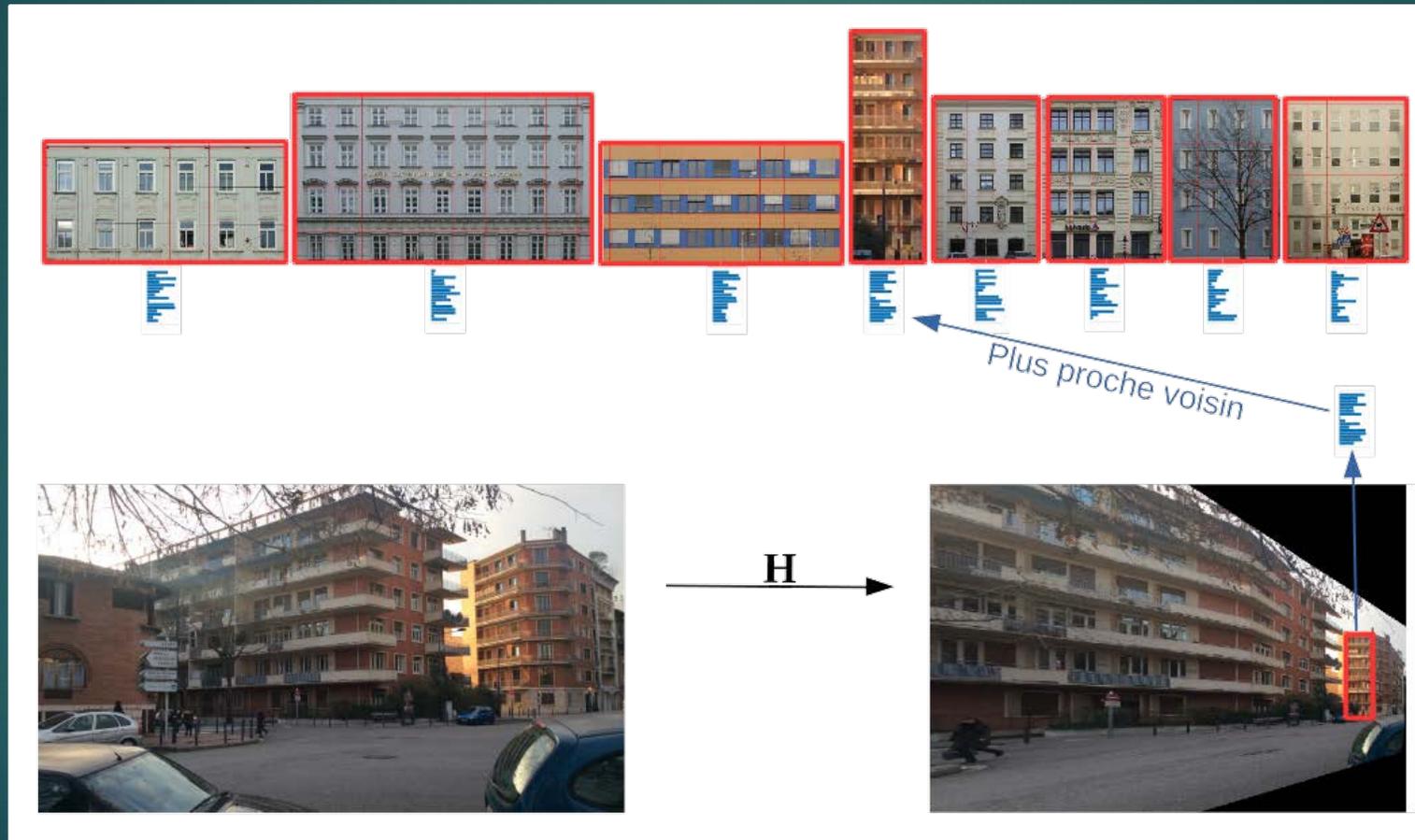
# Region proposal + descripteurs CNN :

34



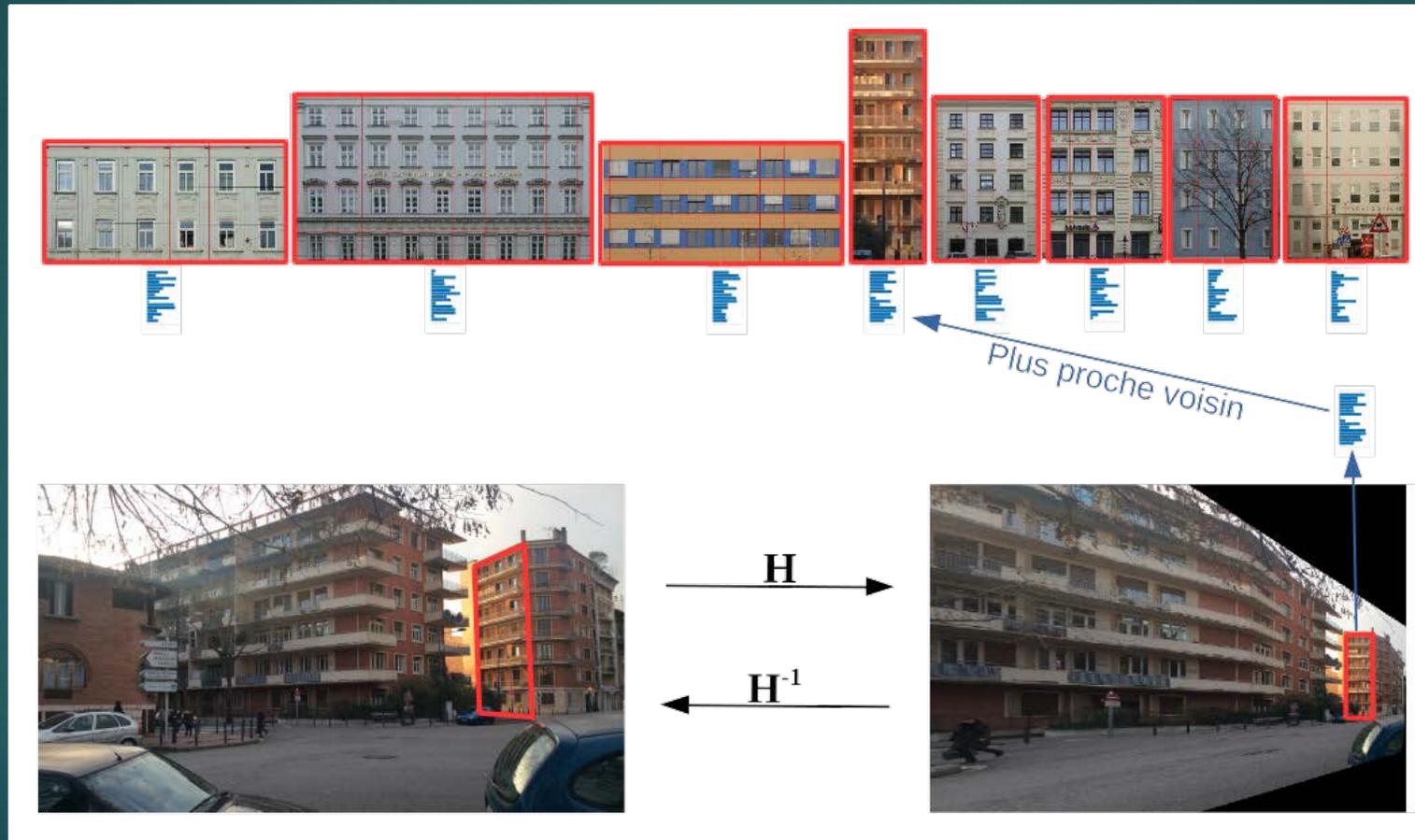
# Region proposal + descripteurs CNN :

35



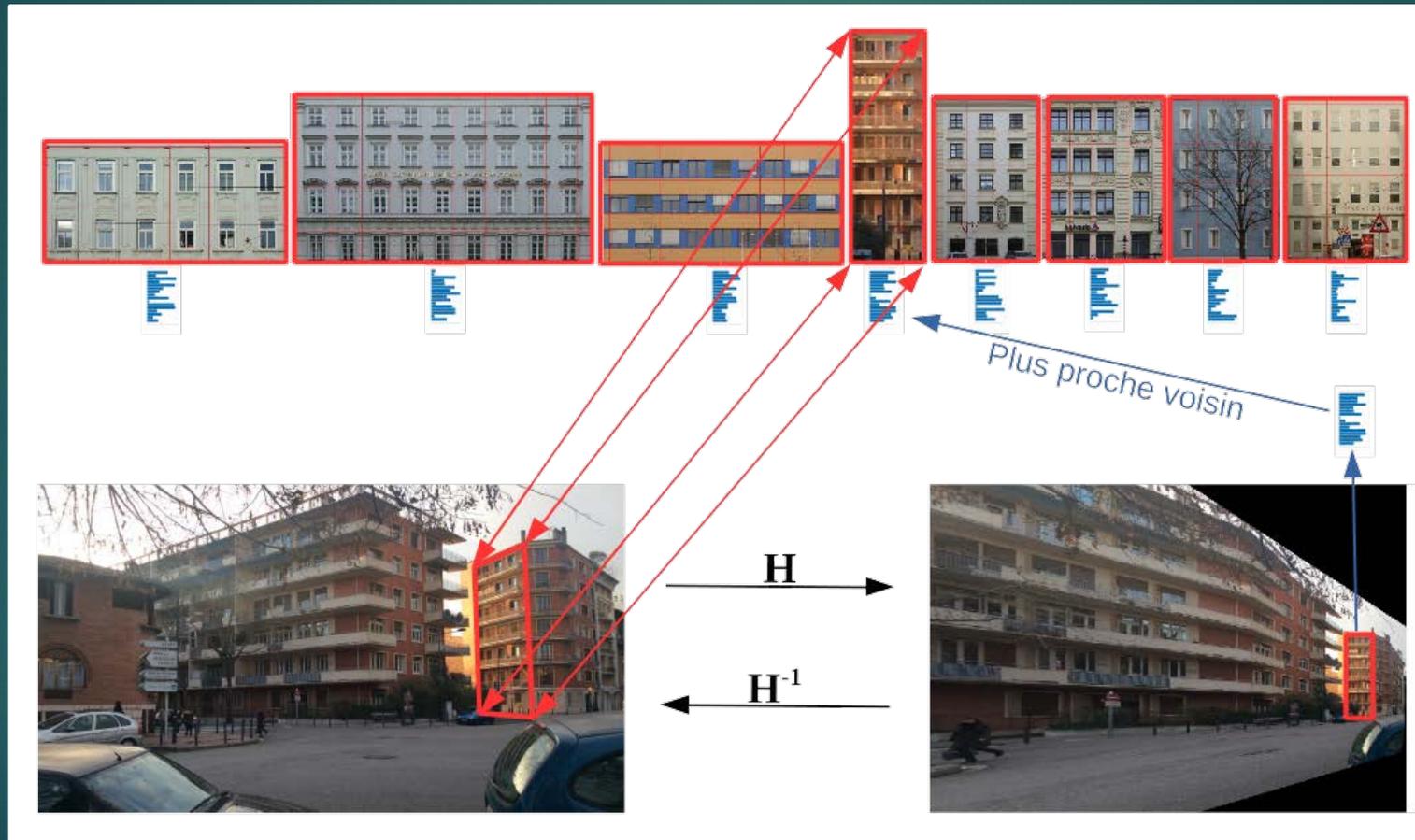
# Region proposal + descripteurs CNN :

36



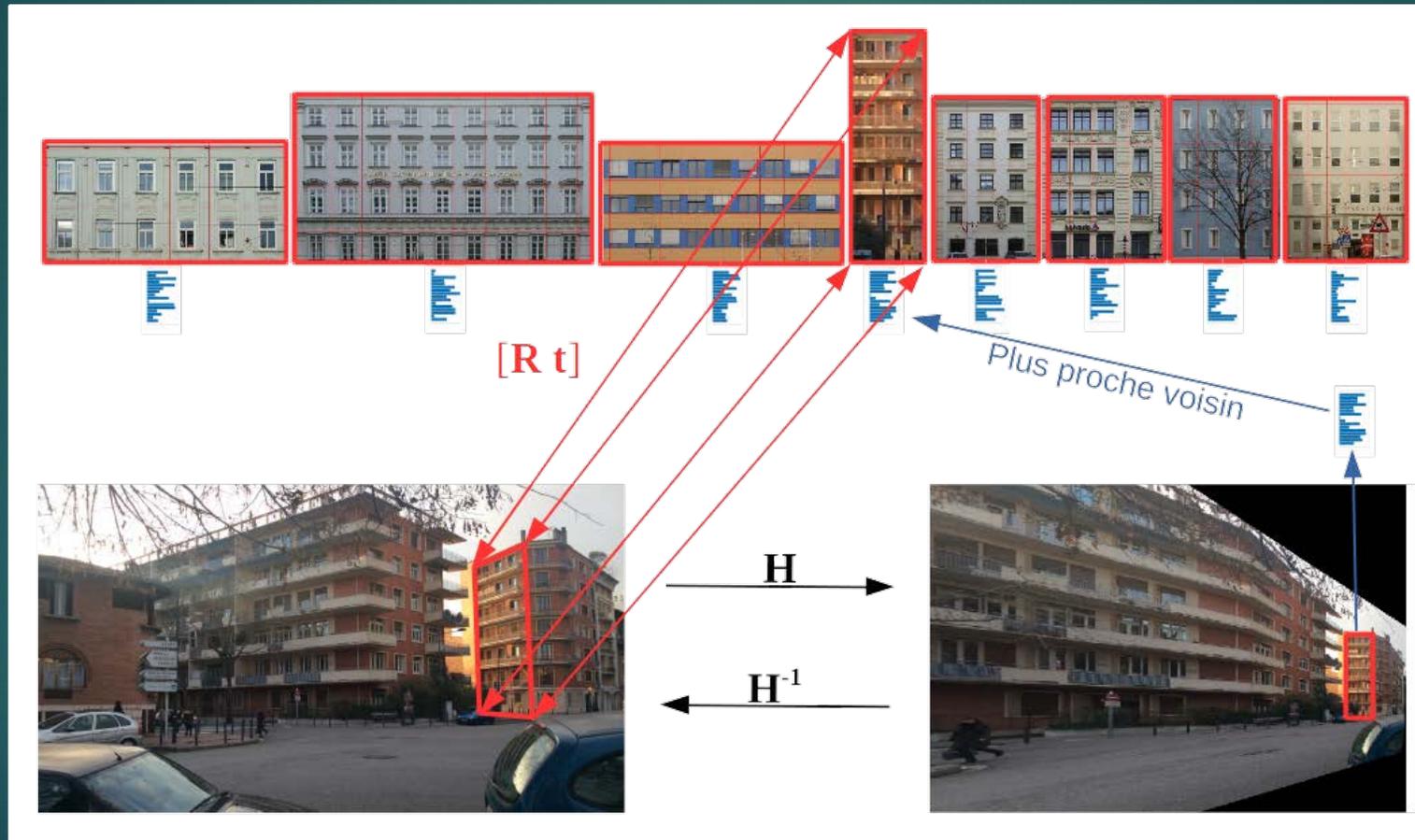
# Region proposal + descripteurs CNN : cas particulier des façades

37



# Region proposal + descripteurs CNN : cas particulier des façades

38



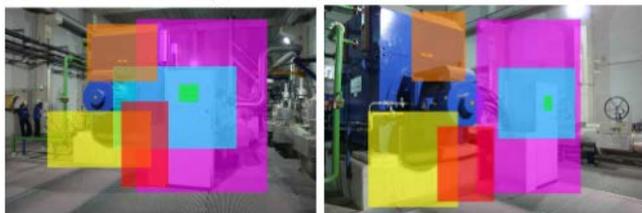
# Region proposal + descripteurs CNN : cas général

39

- ▶ Region-based epipolar and planar geometry estimation in low-textured environment [Gaudillière et al., 2018]



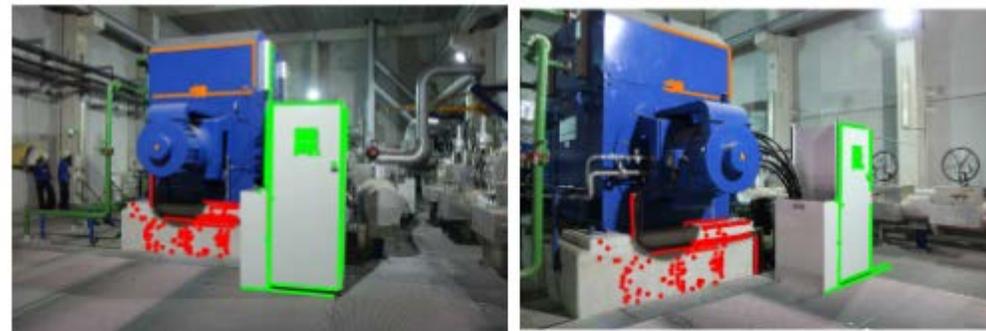
(a) Deux vues de l'UEM.



(b) Exemple de régions mises en correspondance.



(c) Correspondances pertinentes (points et segments) des homographies estimées entre régions en correspondance (une couleur par homographie).



(d) Correspondances pertinentes après fusion des homographies.



(e) Trois premiers plans obtenus avec un algorithme multi-RANSAC 4 points.

# Region proposal + descripteurs CNN : cas général

40

- ▶ Region-based epipolar and planar geometry estimation in low-textured environment [Gaudillière et al., 2018]



Correspondences inliers obtenues avec SIFT  
[Lowe04]

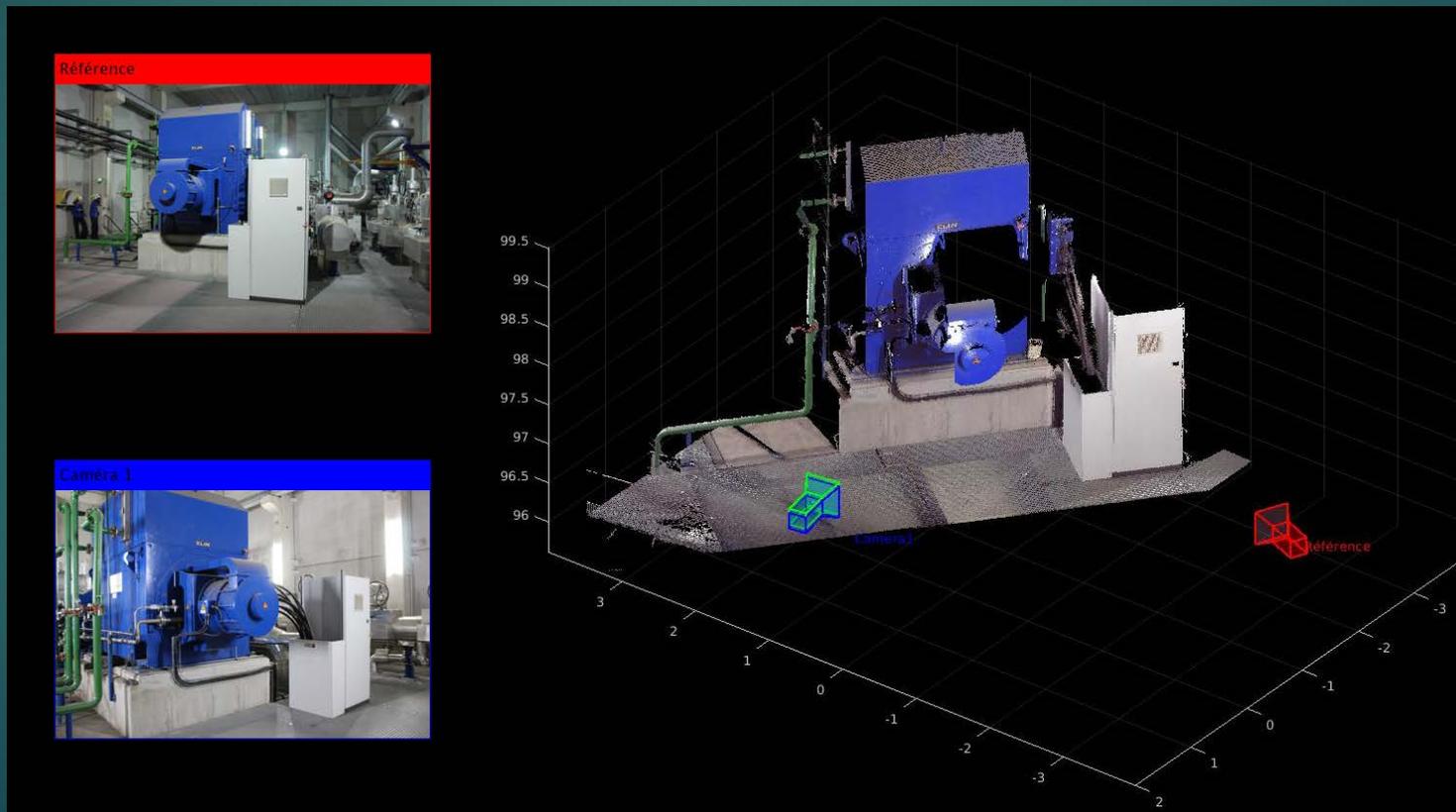


Correspondences inliers obtenues avec la méthode  
de [Gaudillière et al., 2018]

# Region proposal + descripteurs CNN : cas général

41

- ▶ Region-based epipolar and planar geometry estimation in low-textured environment [Gaudillière et al., 2018]



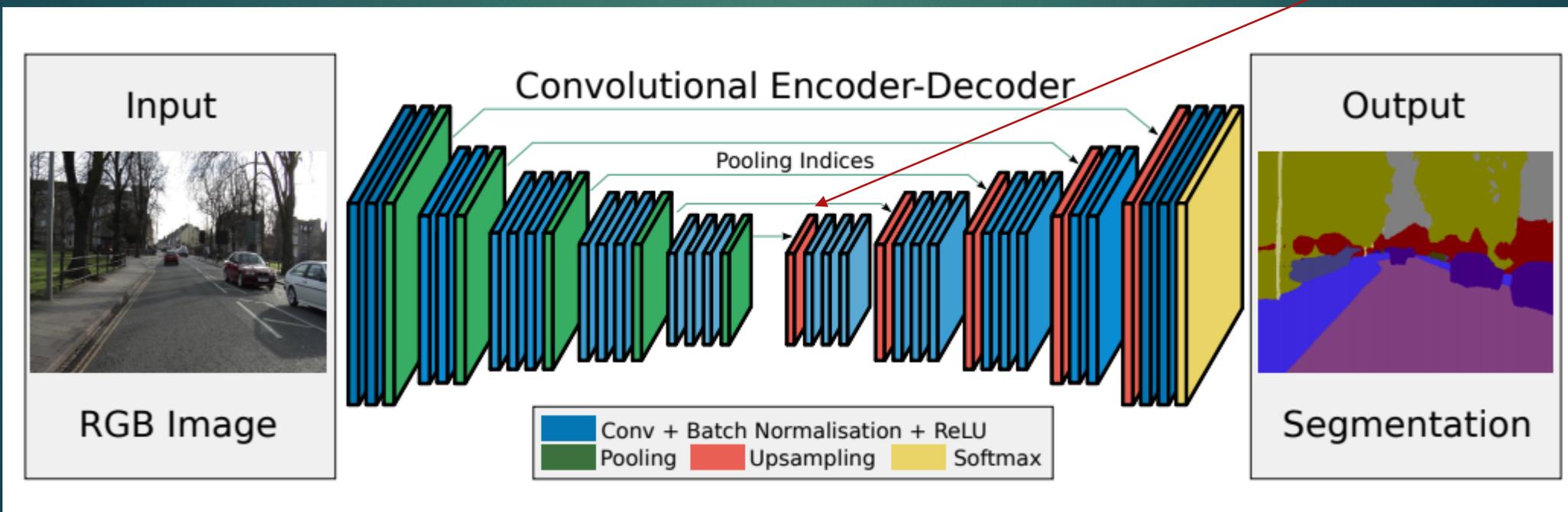
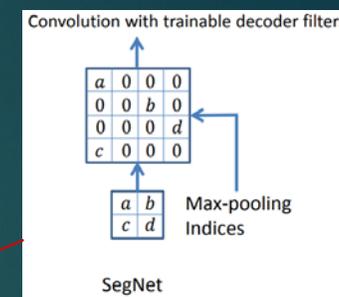
# Méthodes génériques

SÉMANTIQUE DE CLASSE

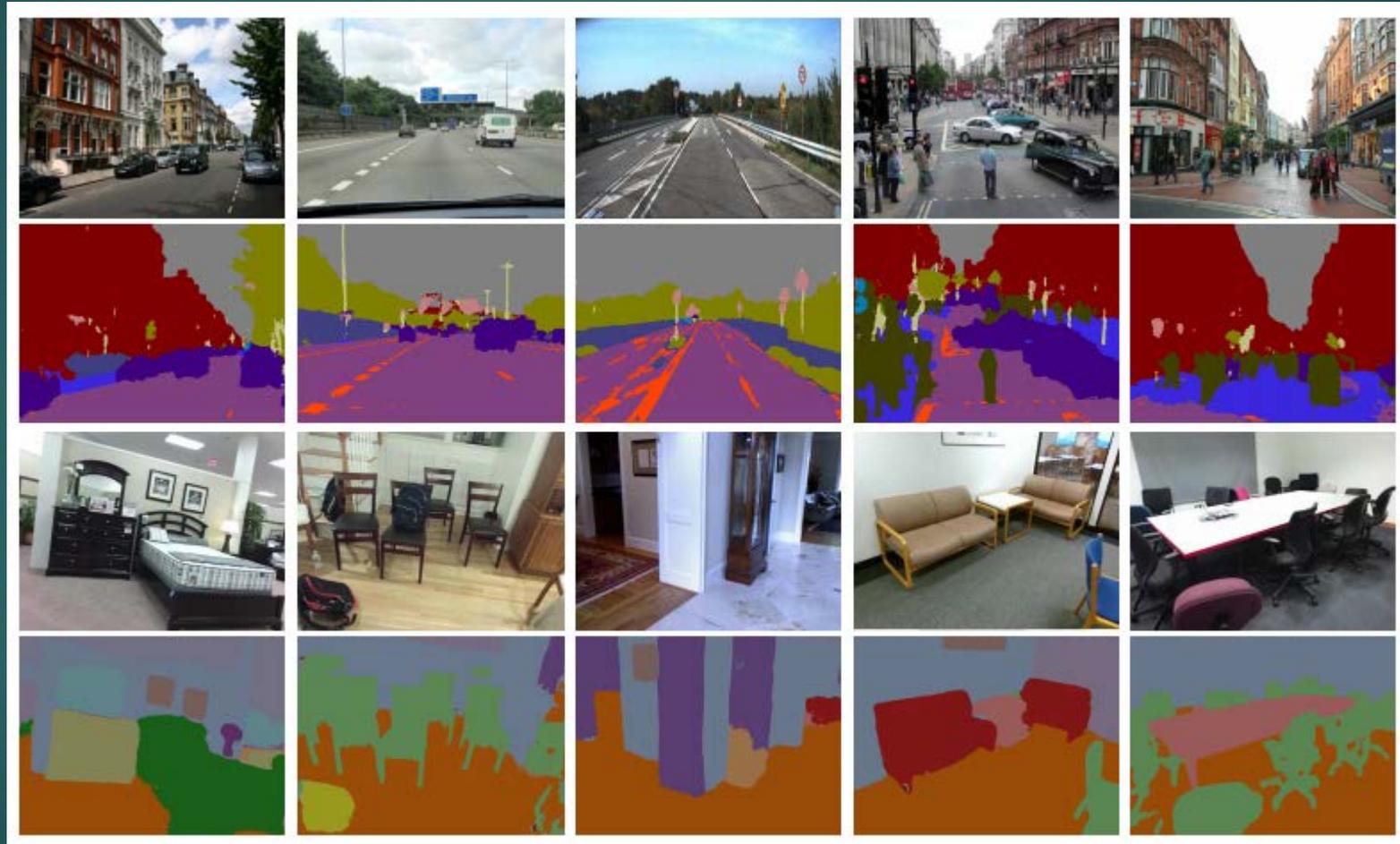
# Cartes sémantiques

43

- ▶ SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation, Badrinarayanan et al., CoRR 2015



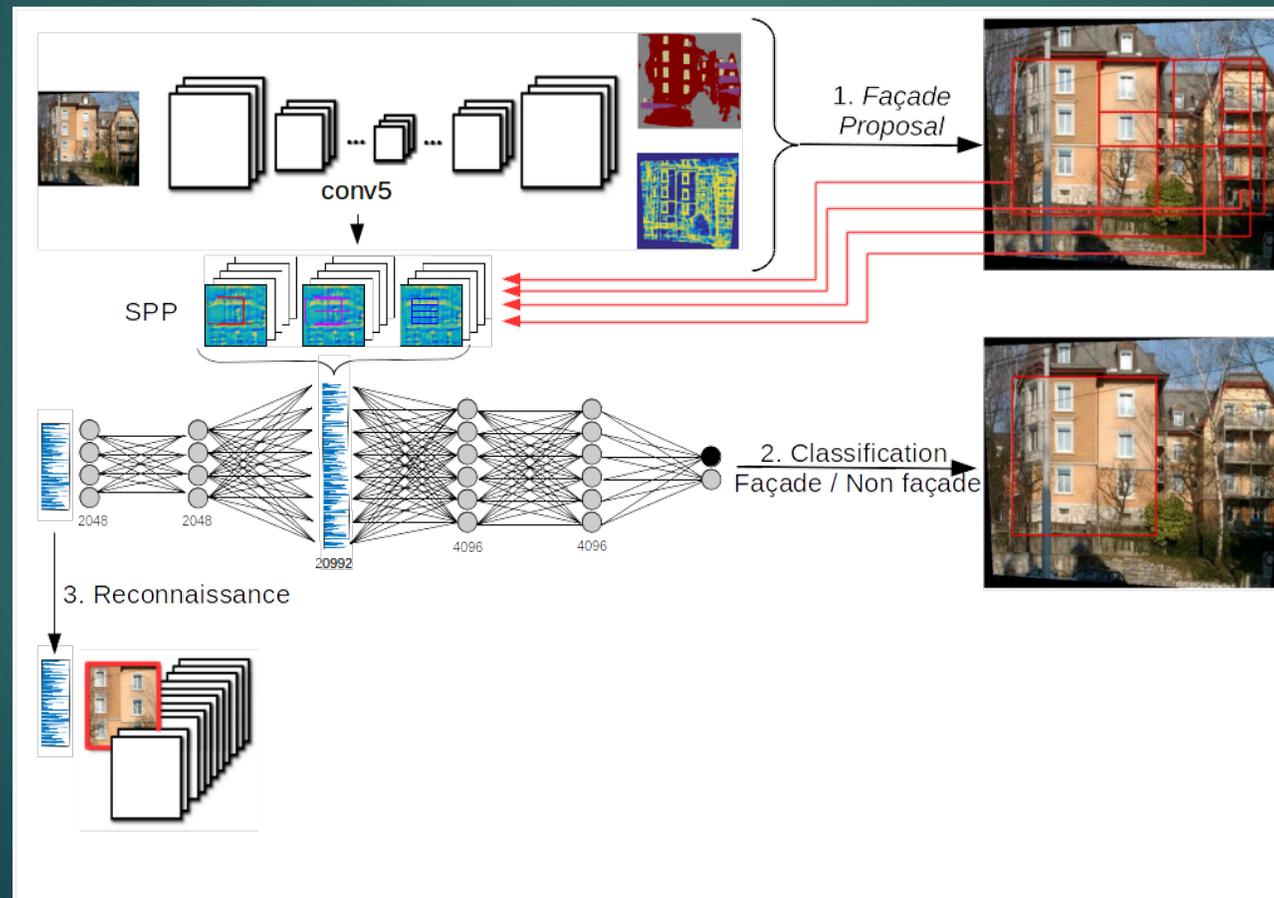
# Cartes sémantiques



# Cartes sémantiques : application au recalage rigide

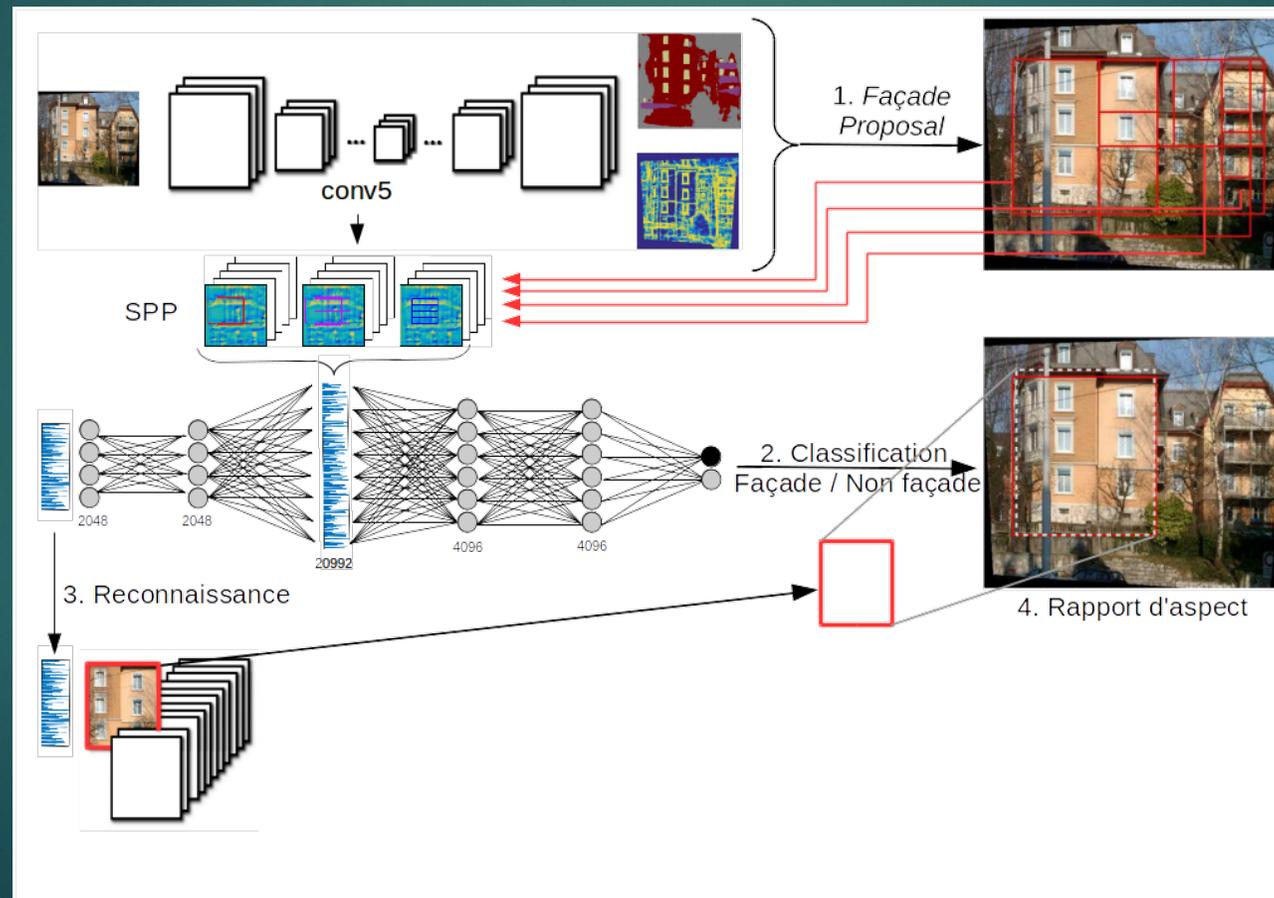
45

- ▶ Article en cours de soumission...



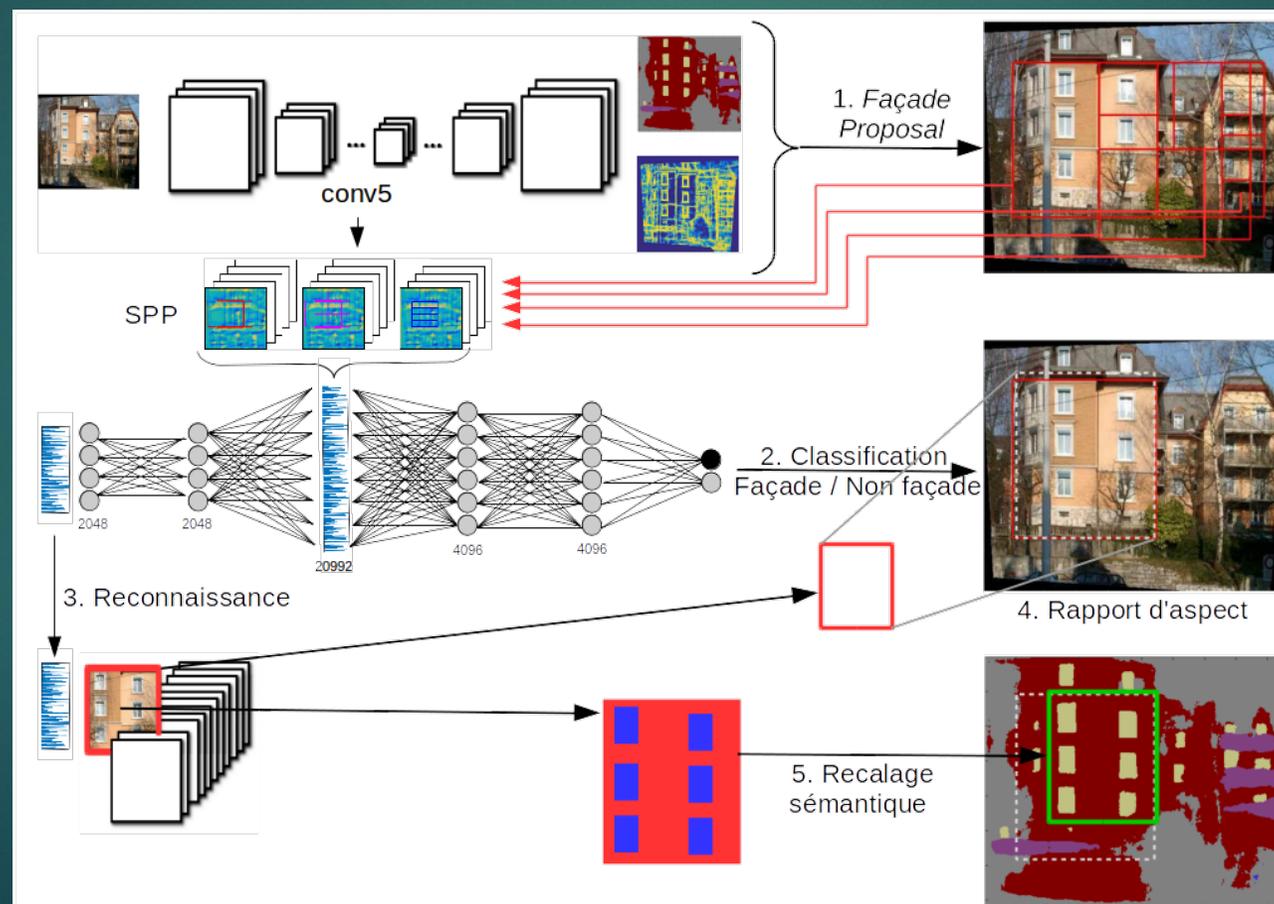
# Cartes sémantiques : application au recalage rigide

► Article en cours de soumission...



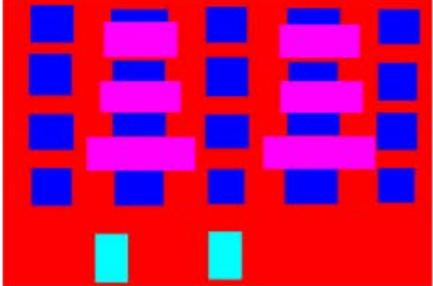
# Cartes sémantiques : application au recalage rigide

► Article en cours de soumission...



# Cartes sémantiques : application au recalage rigide

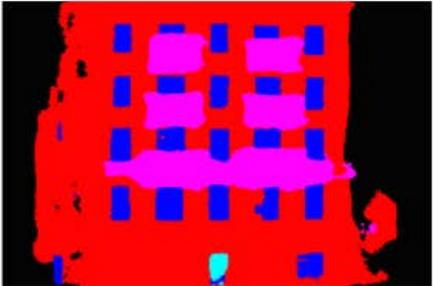
- ▶ Article en cours de soumission...



$I_{ref}$



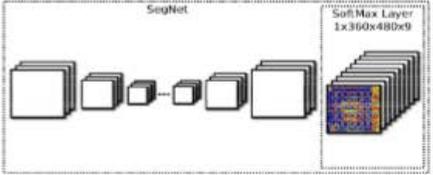
$$\mathbf{T} = \begin{bmatrix} s & 0 & t_x \\ 0 & s & t_y \\ 0 & 0 & 1 \end{bmatrix}$$



$I$

Espérance-Maximisation

$$\Theta = (\mathbf{T}, \pi_{k_j}, \alpha)$$



$$P(X_i|I_{ref}) = \sum_{j=1}^K \sum_{k_j=1}^{m_j} \pi_{k_j} \mathcal{N}_p(\mu_{k_j}, \Sigma_{k_j})$$

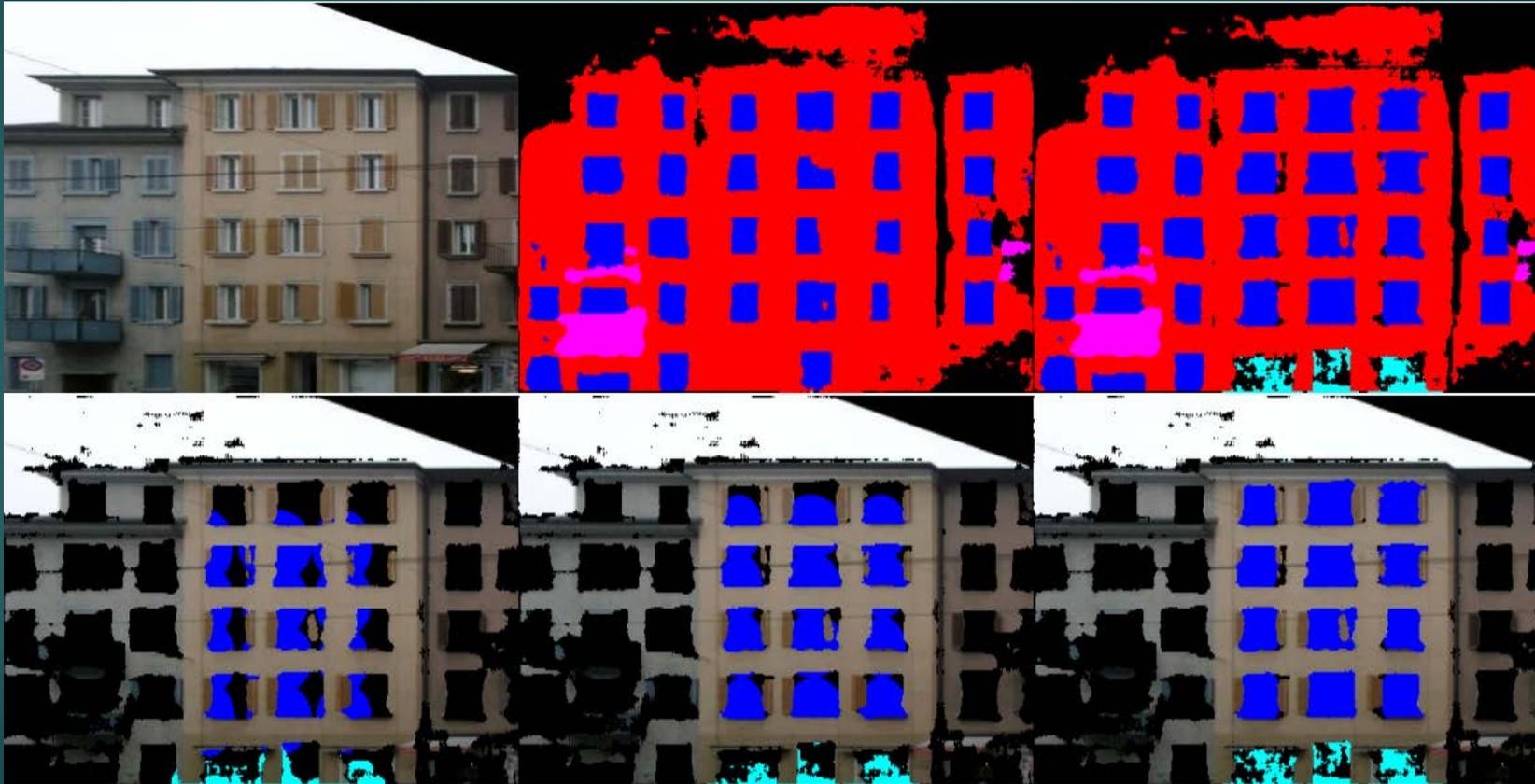
$$\{P(I_j|i, I)\}_{1 \leq j \leq K}$$

$$\{P(o|i, I) = \alpha\}_{1 \leq j \leq K}$$

# Cartes sémantiques : application au recalage rigide

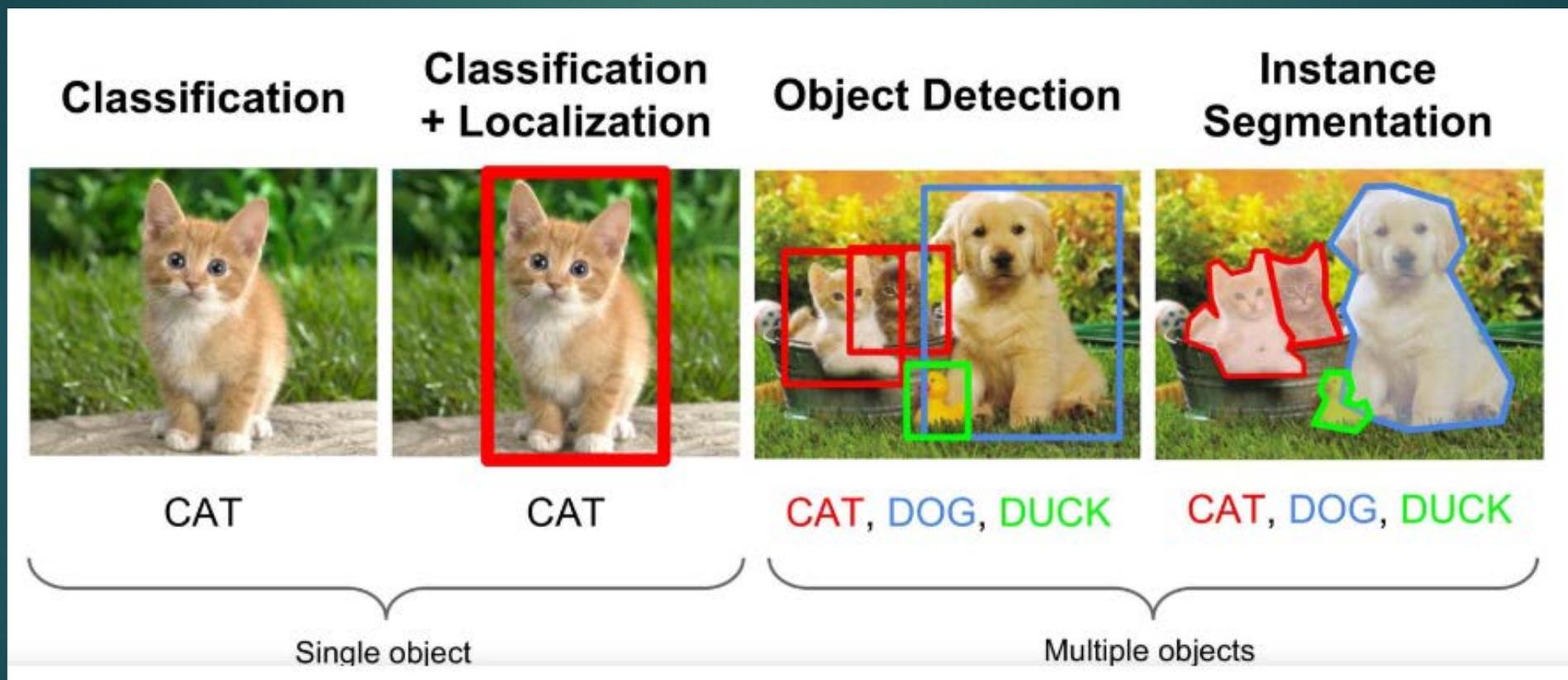
49

- ▶ Article en cours de soumission...



# Détection de classes d'objets

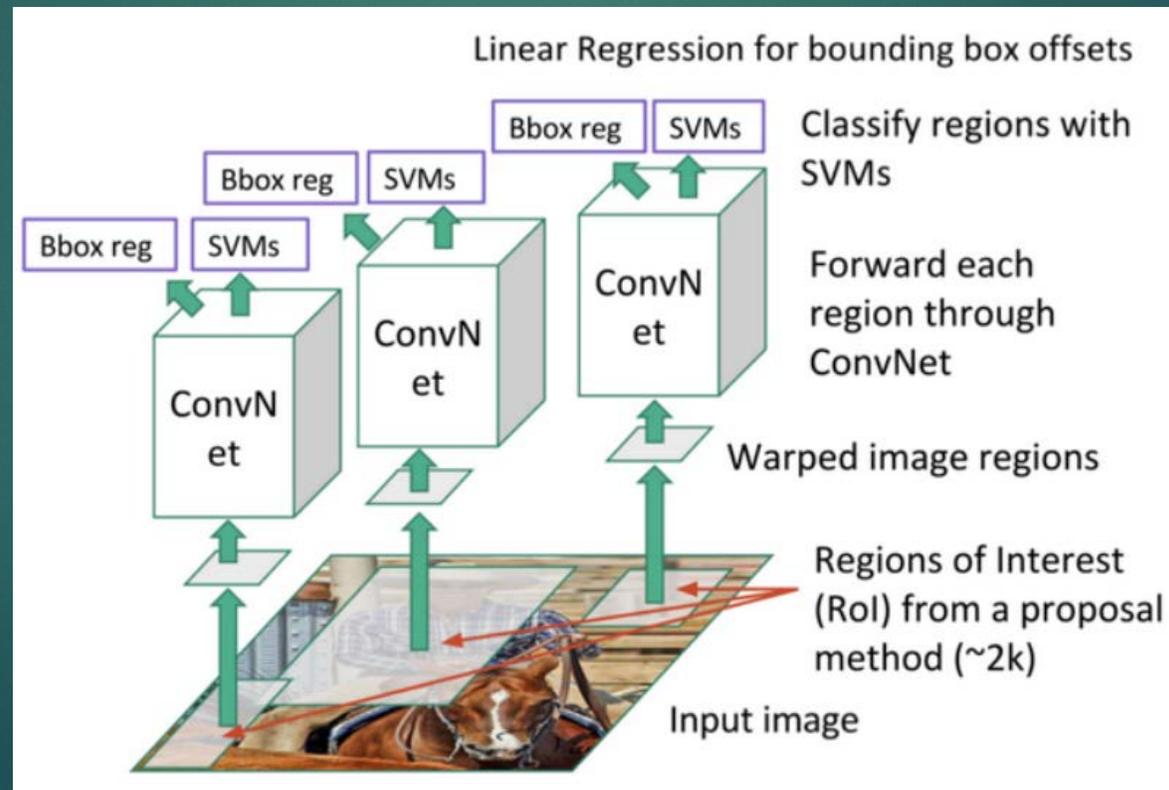
50



# Détection de classes d'objets

51

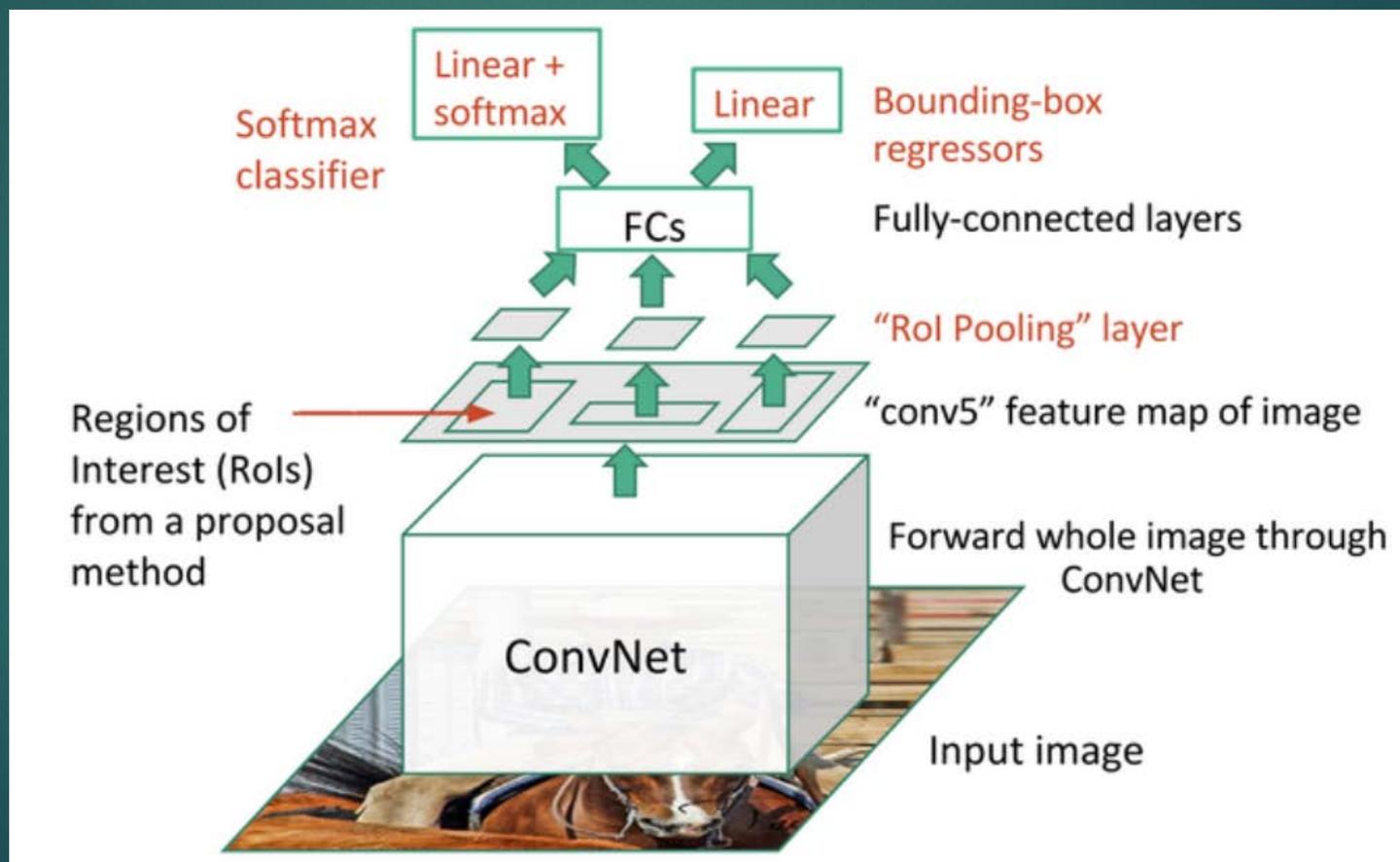
- ▶ **R-CNN** : Region-based Convolutional Network (R. Girshick et al., 2014)
  - ▶ Proposition de boîtes : selective search [Uijlings et al. 2012]
  - ▶ Application d'un CNN + SVM sur chaque boîte proposée



# Détection de classes d'objets

52

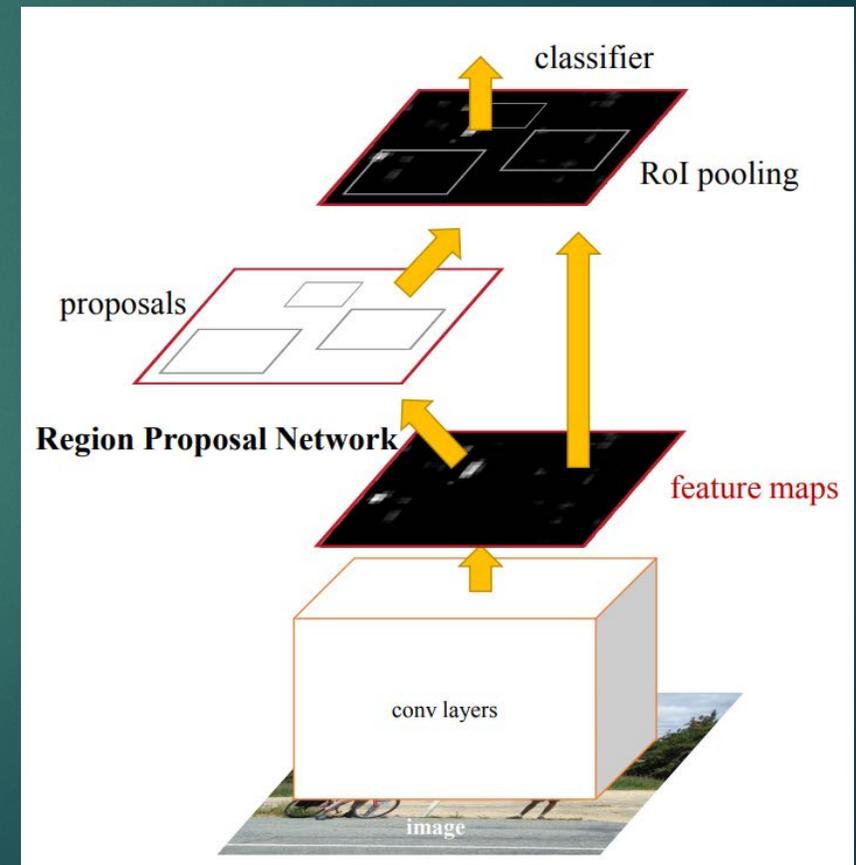
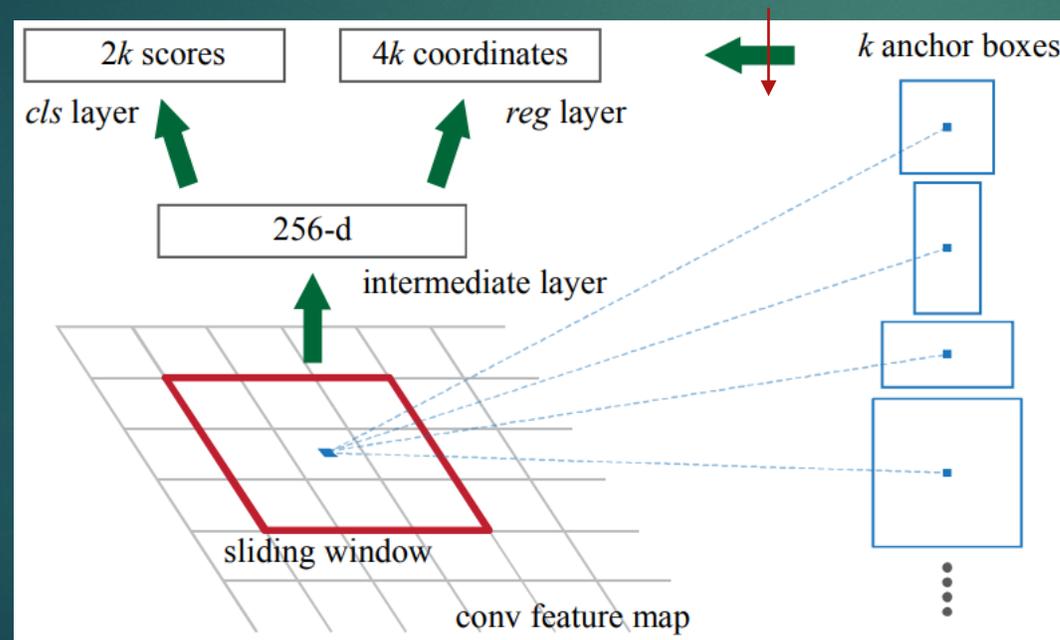
- ▶ **Fast R-CNN** : Fast Region-based Convolutional Network (Girshick 2015)



# Détection de classes d'objets

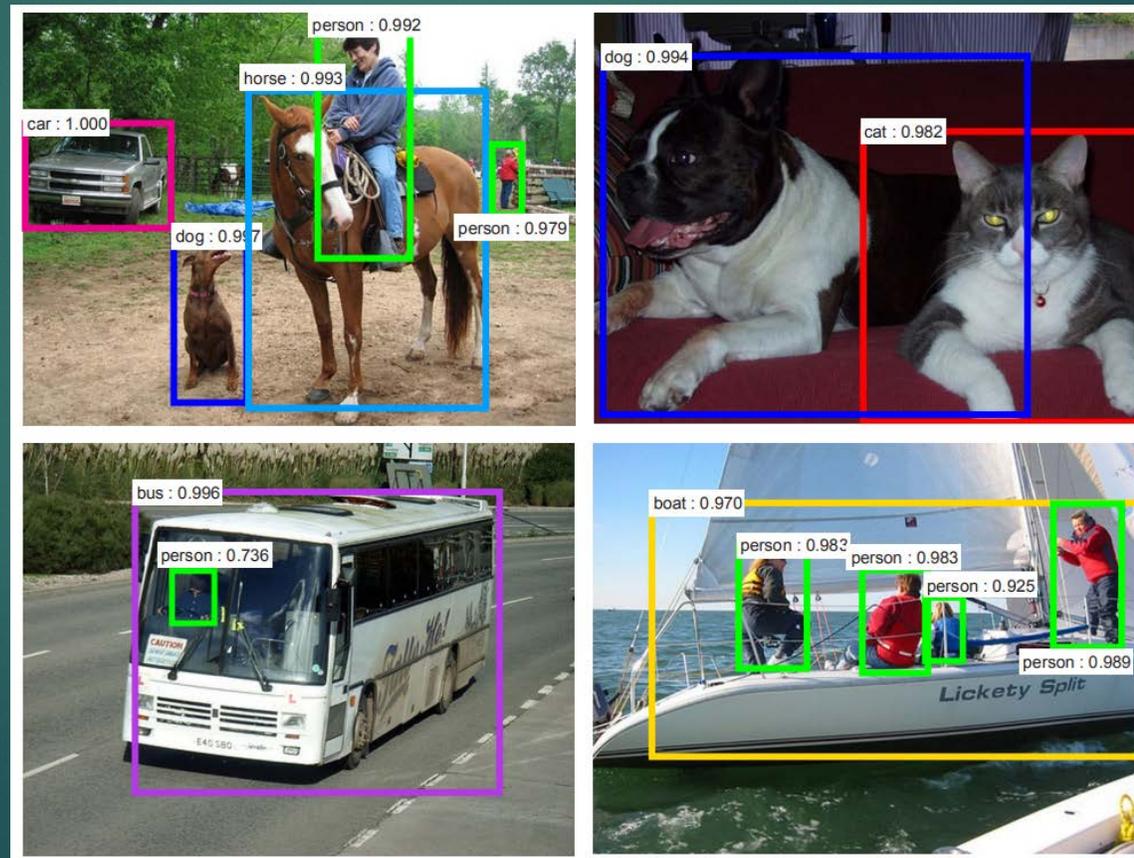
53

- ▶ **Faster R-CNN** : Faster Region-based Convolutional Network (S. Ren and al. 2016) : Region Proposal Network (RPN) + Fast R-CNN



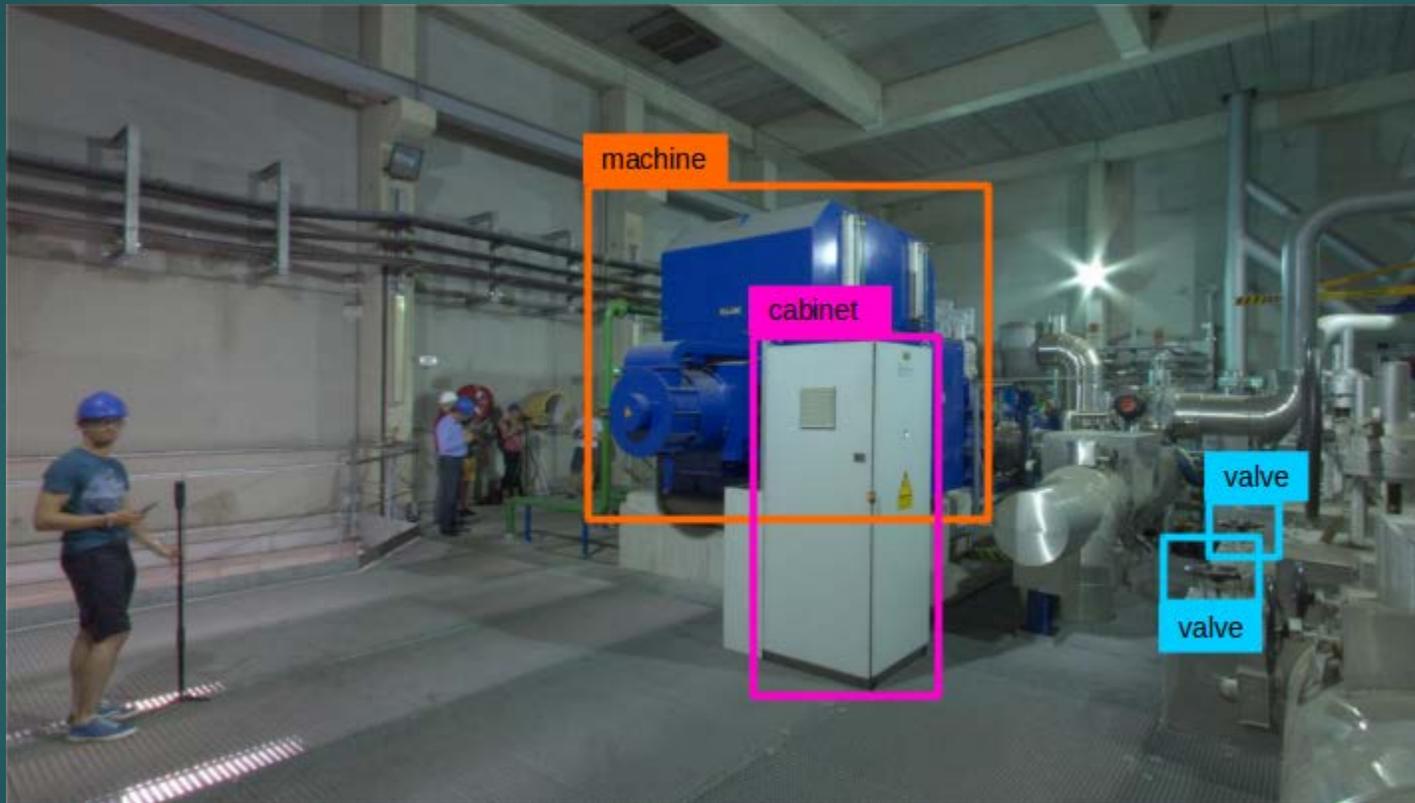
# Détection de classes d'objets

- ▶ **Faster R-CNN** : Faster Region-based Convolutional Network (S. Ren and al. 2016) : Region Proposal Network (RPN) + Fast R-CNN



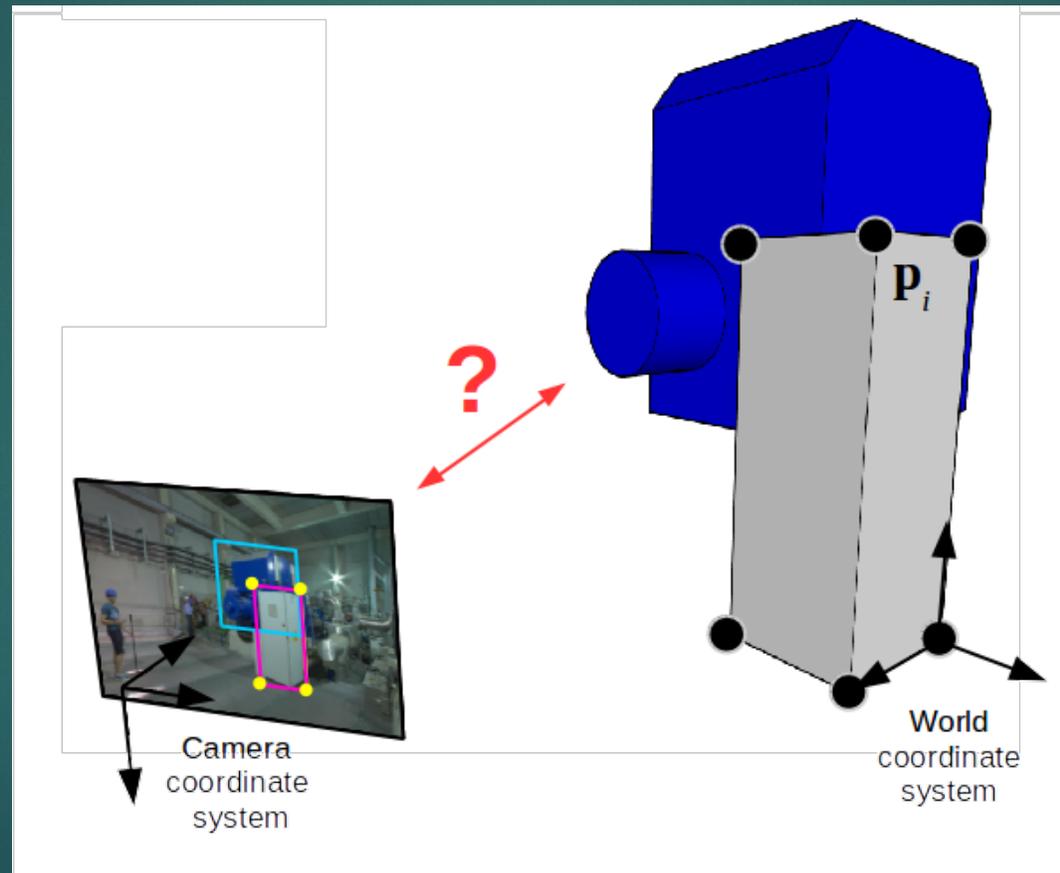
# Peut-on calculer la pose à partir d'une détection d'objets ?

55



# Le positionnement peut-il bénéficier d'une détection d'objets ?

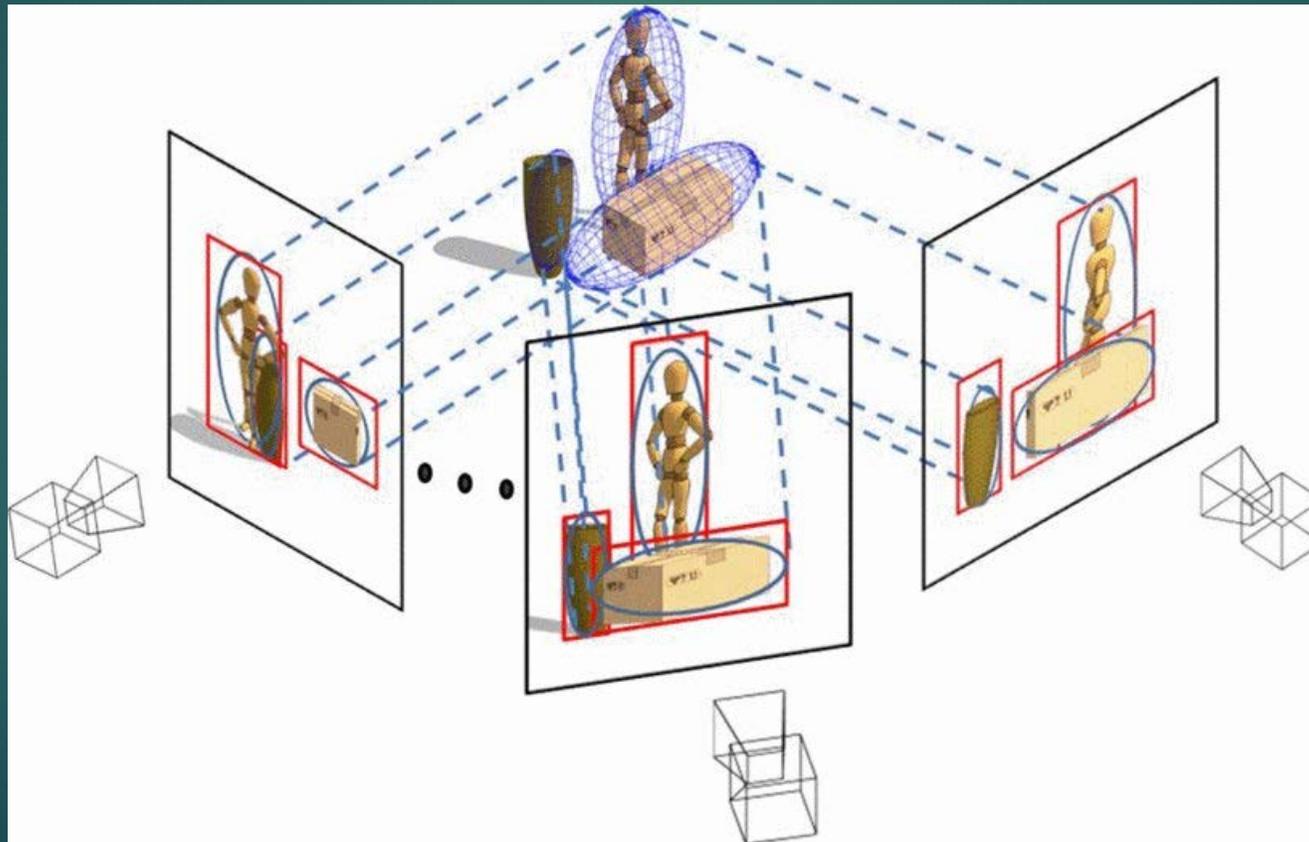
56



# Le positionnement peut-il bénéficier d'une détection d'objets ?

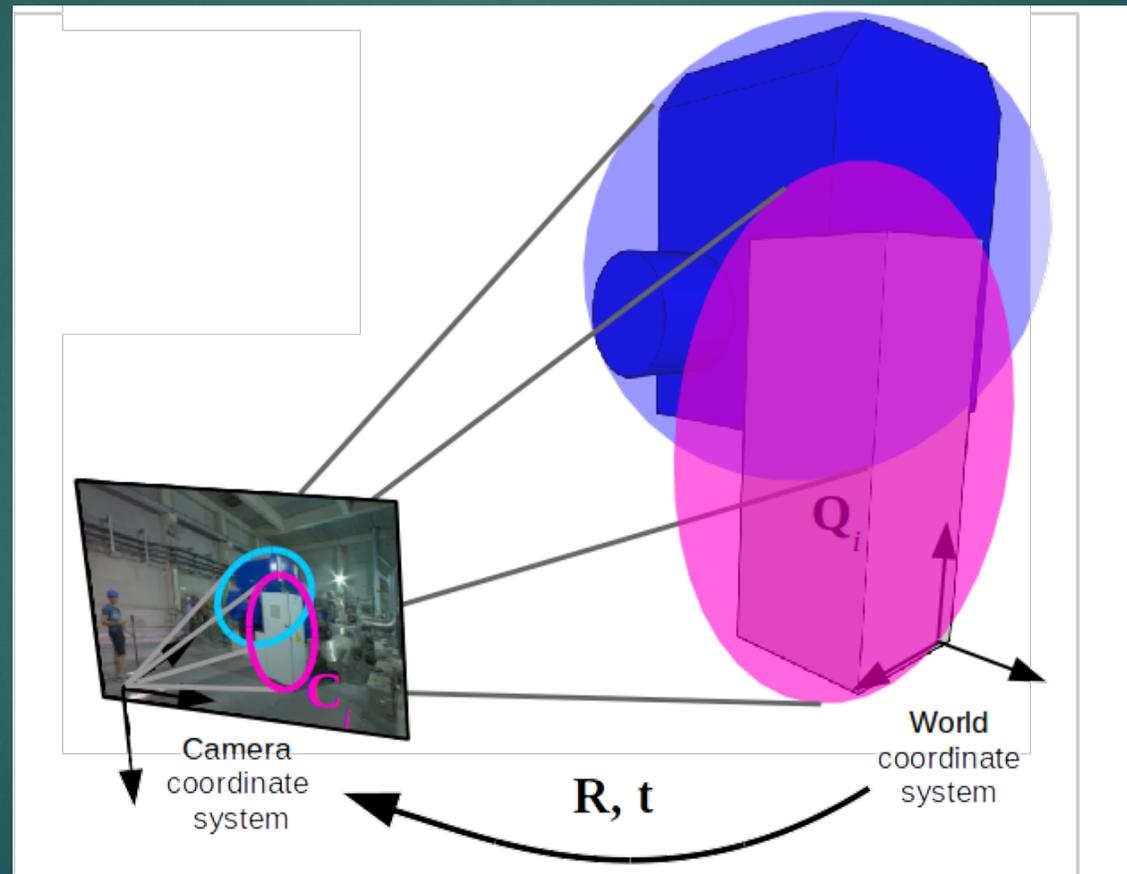
57

- ▶ [Rubino et al., 2018]



# Le positionnement peut-il bénéficier d'une détection d'objets ?

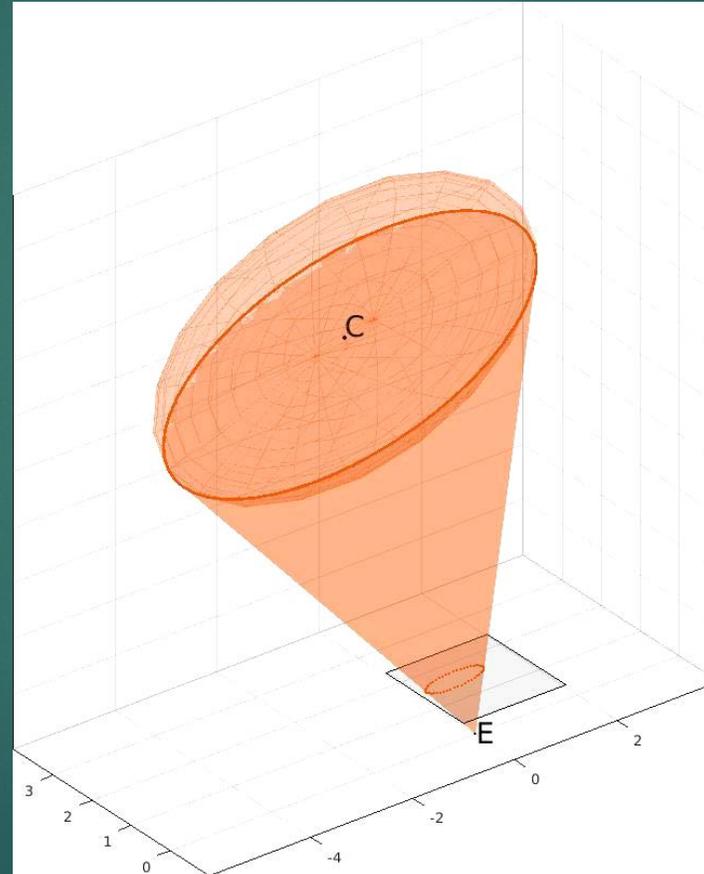
58



[Gaudillère et al., IROS 2019]

# Le positionnement peut-il bénéficier d'une détection d'objets ?

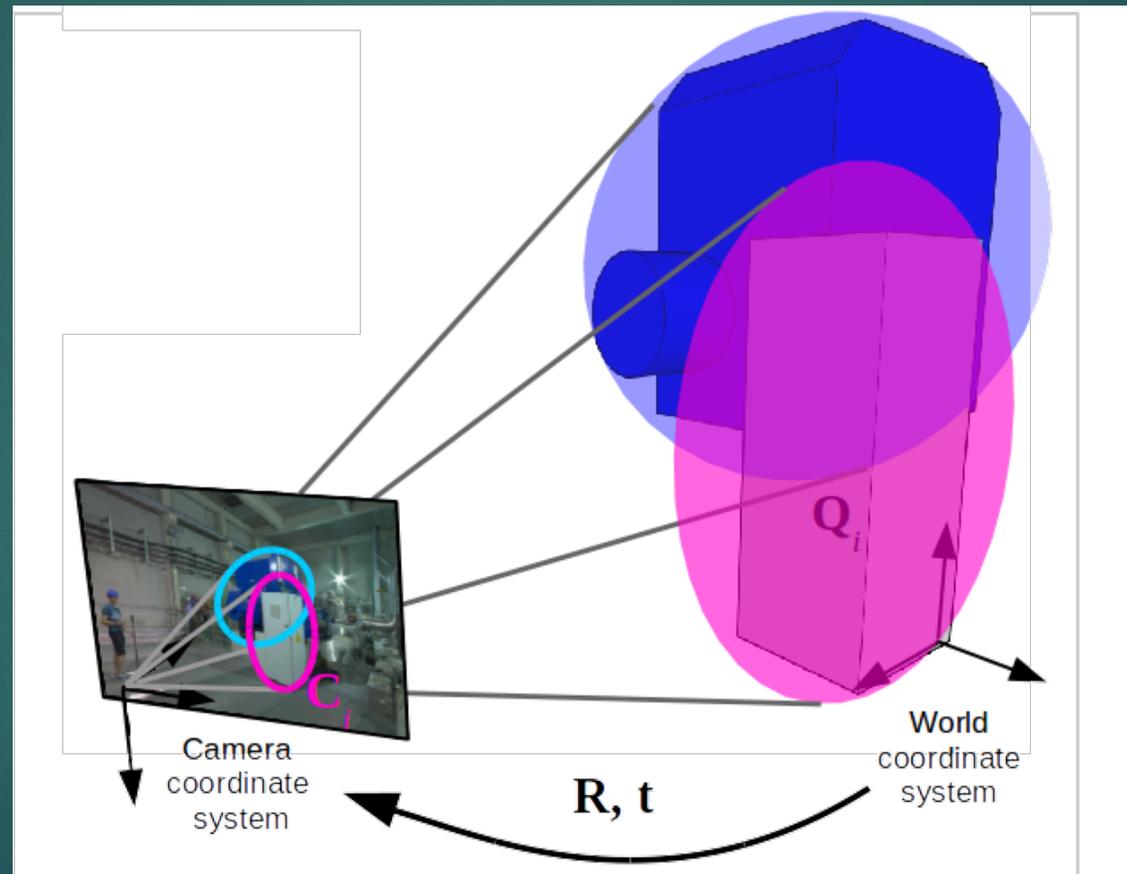
59



[Gaudillère et al., IROS 2019]

# Le positionnement peut-il bénéficier d'une détection d'objets ?

60



# Le positionnement peut-il bénéficier d'une détection d'objets ?

61

