



Université de la Manouba  
Ecole Nationale des Sciences de l'Informatique



Projet de Fin d'Etudes

# Segmentation d'une Cartographie 3D d'un Environnement Intérieur par l'Aspect Visuel des Matériaux



Superviseur Académique :  
Dr. Nesrine BEN YAHIA

Présenté par :  
Ghassene TANABENE

Superviseurs Professionnels :  
Mr. Ali AMAMOU  
Mr. Ahmed DHAHRI

**I.**

**PRÉSENTATION GENERALE**

**II.**

**RECHERCHE**

**III.**

**DÉVELOPPEMENT**

**IV.**

**EXPÉRIMENTATIONS ET RÉSULTATS**

**V.**

**CONCLUSION ET PERSPECTIVES**

# PLAN

---





# I. PRÉSENTATION GÉNÉRALE

# I. PRÉSENTATION GÉNÉRALE (1/6)

## Présentation de l'organisme d'accueil



Institut de Recherche sur l'Hydrogène  
- IRH -



Université du Québec à  
Trois-Rivières (UQTR)



1994



50-200 chercheurs



- Hydrogène
- Réseaux Intelligents
- Mobilité
- Matériaux innovants
- Installations

# I. PRÉSENTATION GÉNÉRALE (2/6)

## Problématique



- Les robots de désinfection à lampes UV aident à lutter contre la pandémie du covid-19.
- L'IRH développe en 2021 la 3<sup>ème</sup> version de son robot de désinfection.
- Ce robot a besoin de comprendre la nature des matériaux qui existent dans son environnement.

**Comment permettre au robot de voir et de comprendre les types de matériaux ?**

# I. PRÉSENTATION GÉNÉRALE (4/6)

---

## Objectif du stage

- Développer la perception du robot de désinfection à travers :

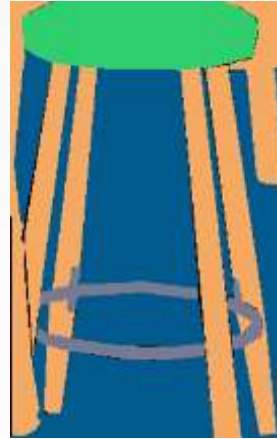
La **segmentation** de son environnement intérieur en se basant sur l'**aspect visuel des matières constitutives**.

# I. PRÉSENTATION GÉNÉRALE (5/6)

## Défis



Image originale



Résultat souhaité

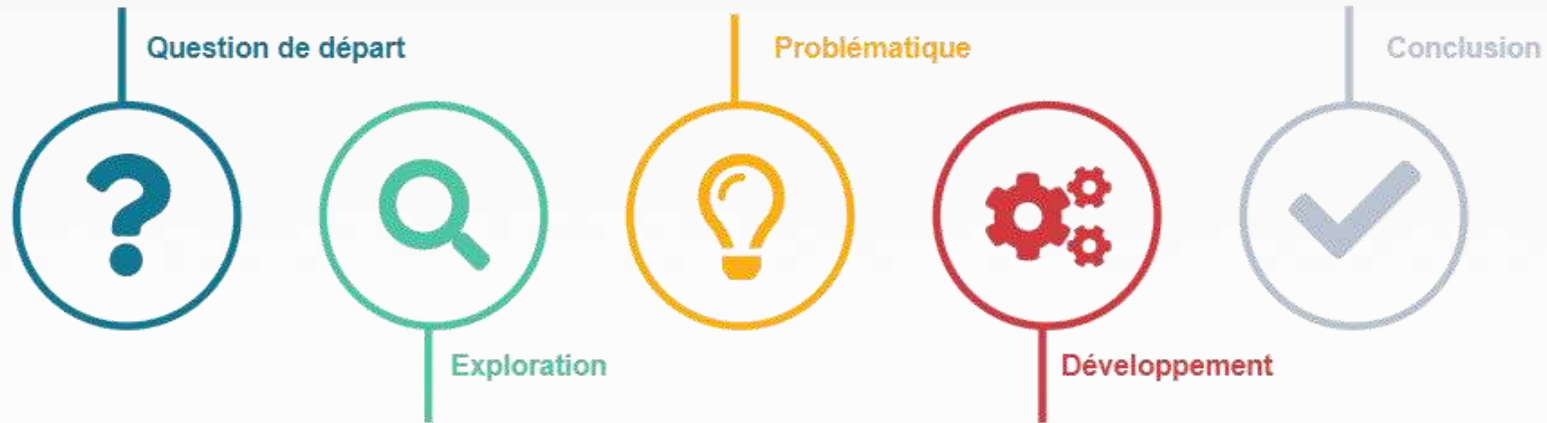
Cuir, Bois, Métal, Pierre

- La ressemblance des matériaux
- La peinture des matériaux
- L'insuffisance dans quelques cas de l'information visuelle pour déterminer la nature du matériau
- Des objets sont partiellement cachés par d'autres



# I. PRÉSENTATION GÉNÉRALE (6/6)

## Méthodologie de travail



Méthodologie de recherche et développement



## II. RECHERCHE

## II. RECHERCHE (1/7)

### Segmentation d'image

- Partitionnement d'une image numérique en plusieurs segments.
- Attribution des étiquettes à chaque groupement de pixels qui partagent certaines caractéristiques.
- **Applications :** Voiture autonome, Médecine, Perception Robotique ...

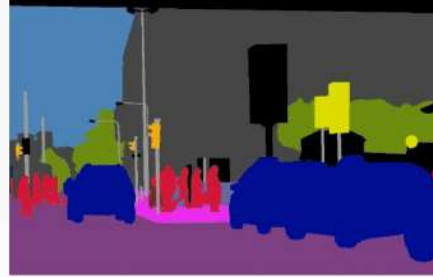


## II. RECHERCHE (2/7)

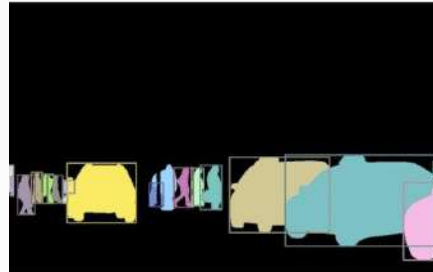
Types de segmentation d'image



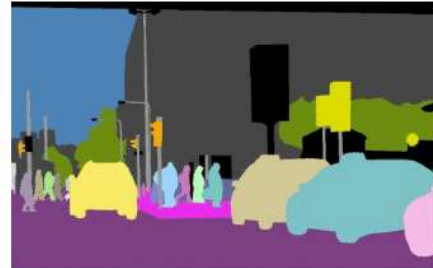
Image



Sémantique



Des instances

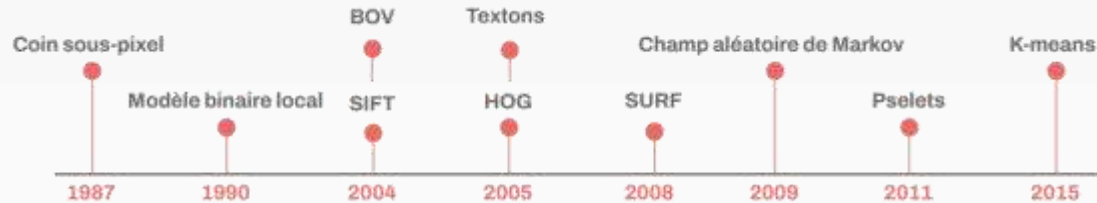


Panoptique

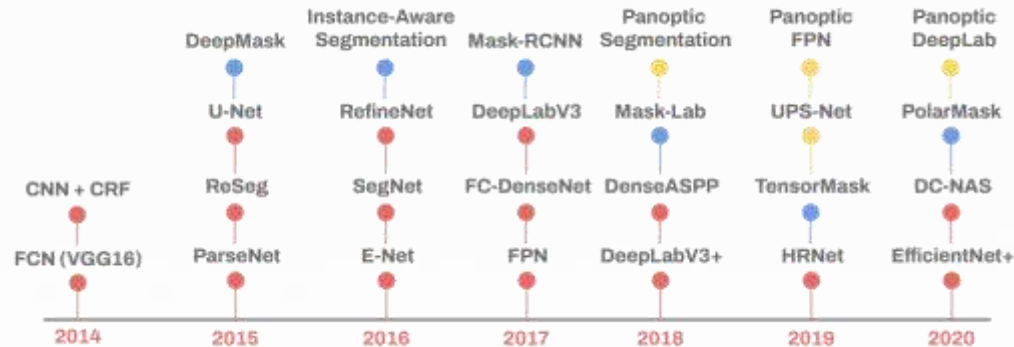
## II. RECHERCHE (3/7)

### Méthodes de segmentation d'image

#### A- Méthodes traditionnelles



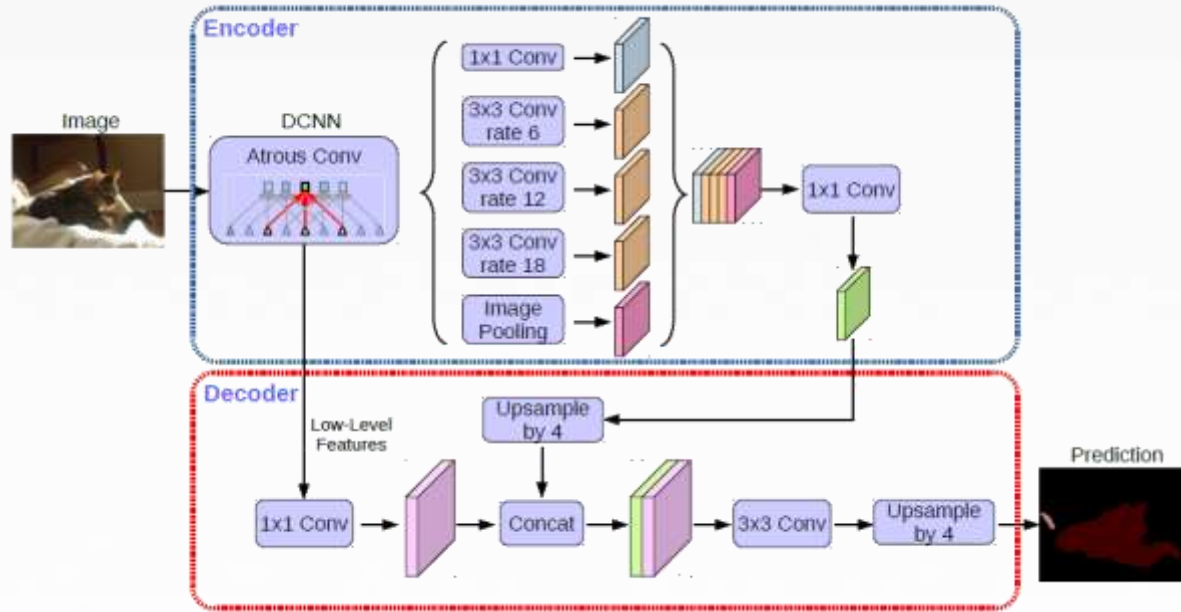
#### B- Méthodes modernes basées sur l'apprentissage profond



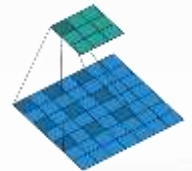
## II. RECHERCHE (4/7)

### Modèles de segmentation sémantique

#### 1- DEEPLABV3+



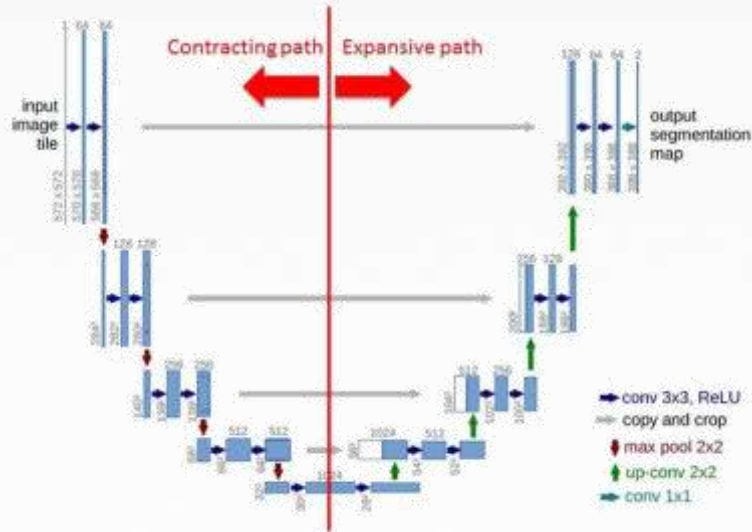
- Année : 2018
- 4<sup>ème</sup> version de la famille DeepLab
- 89 % précision sur « Pascal voc »
- Structure encodeur – décodeur
- ASPP
- Convolution à trous



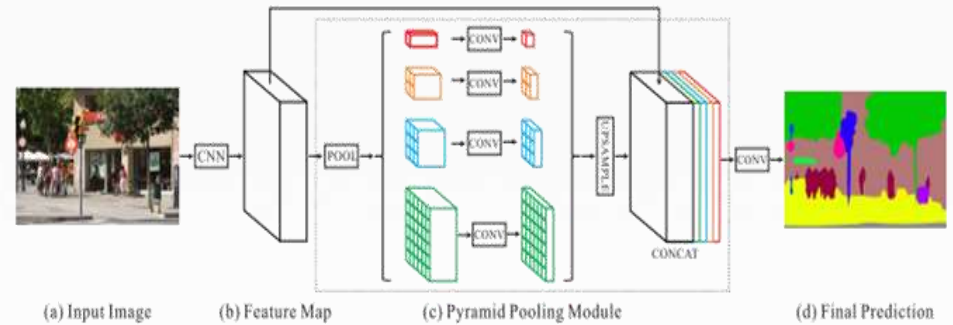
## II. RECHERCHE (5/7)

### Autres modèles

#### U-NET

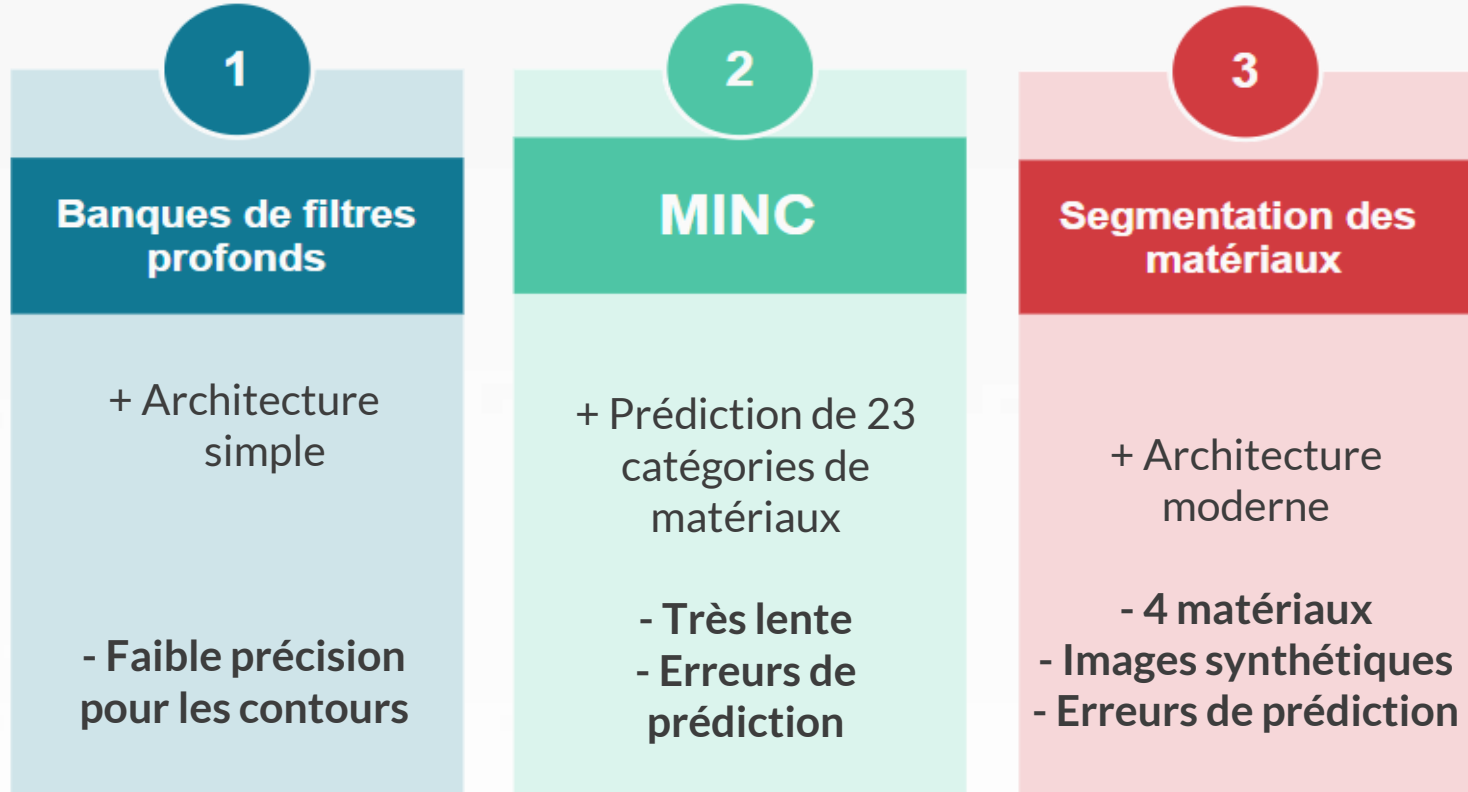


#### PSPNET



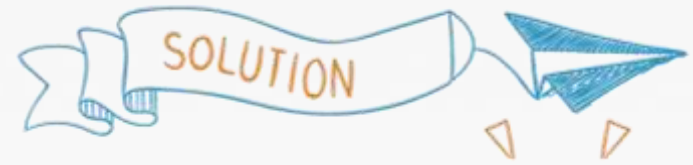
## II. RECHERCHE (6/7)

Etude de l'existant





## II. RECHERCHE (7/7)



Solution proposée

Type de segmentation	Modèles	Extracteur des caractéristiques (Backbones)	Données
Segmentation Sémantique	<ul style="list-style-type: none"><li>• <b>DeepLabV3+</b></li><li>+</li><li>• PSPNet et U-Net</li></ul>	<ul style="list-style-type: none"><li>• ResNet</li><li>• Xception</li></ul>	Nouvel ensemble de données spécifique à nos besoins.



III.

# DÉVELOPPEMENT

# III. DÉVELOPPEMENT (1/9)

---

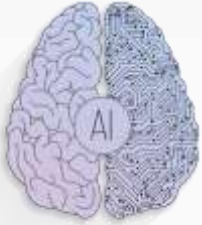
## Besoins fonctionnels :

- Segmentation sémantique des matériaux
- Interfaçage ROS de simulation

## Besoins non-fonctionnels :

- Efficacité
- Rapidité
- Maintenabilité

# III. DÉVELOPPEMENT (2/9)



Intelligence  
Artificielle

1



Cloud  
Computing

2

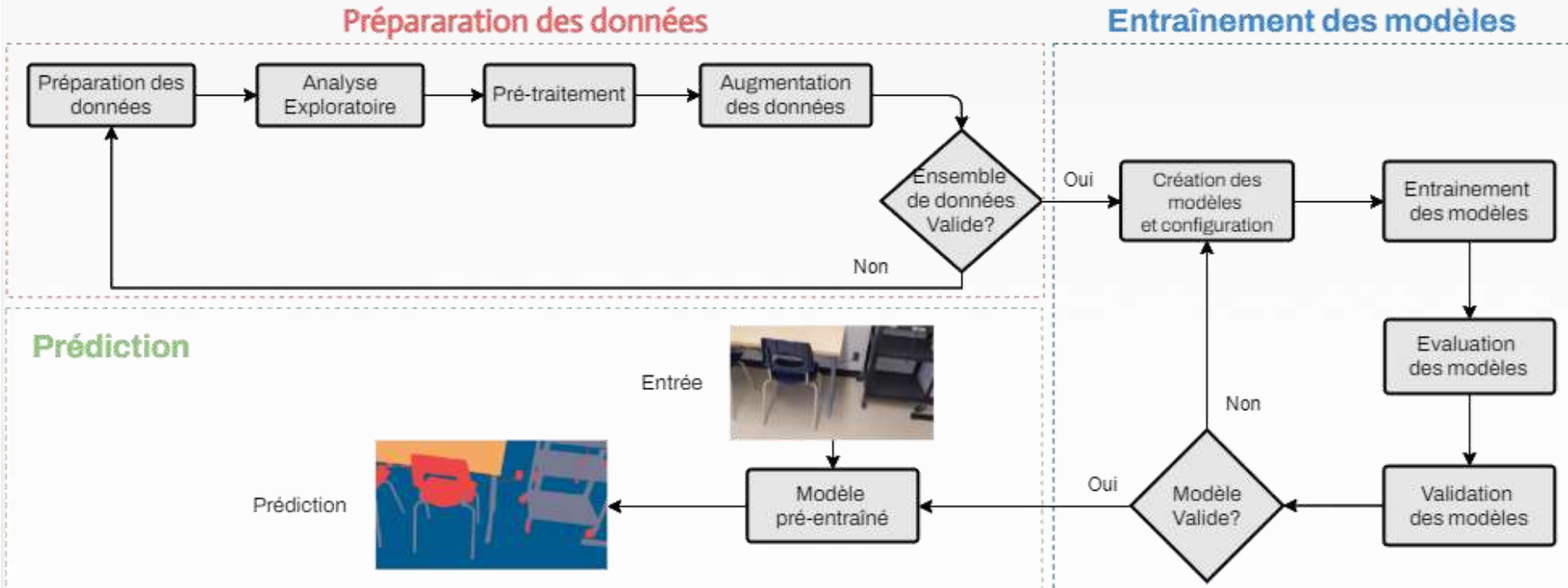


Interfaçage ROS  
(Robot Operating  
System)

3

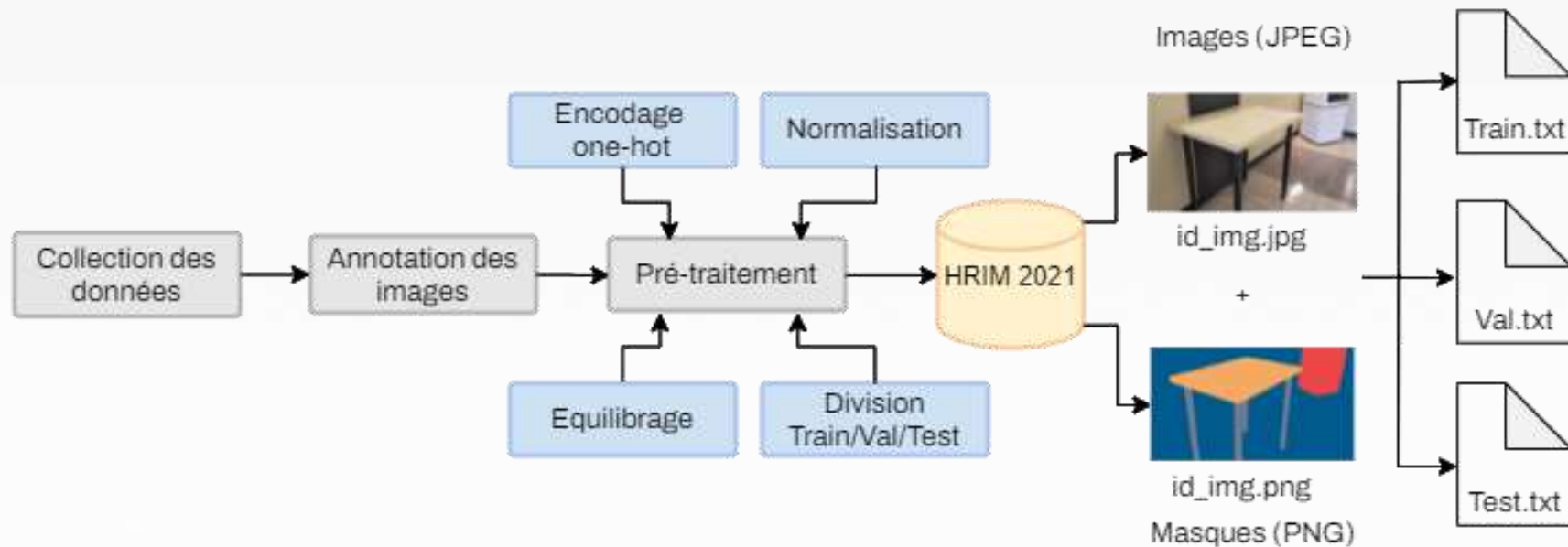
# III. DÉVELOPPEMENT (3/9)

## Processus de développement de la solution



### III. DÉVELOPPEMENT (4/9)

#### Processus de préparation de données



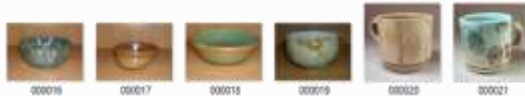
# III. DÉVELOPPEMENT (5/9)

## 1 COLLECTION DES DONNÉES

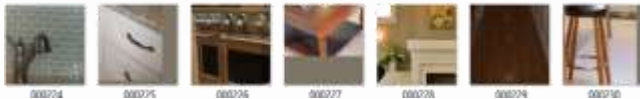
- Environnement intérieur de l'IRH



- Flickr.com

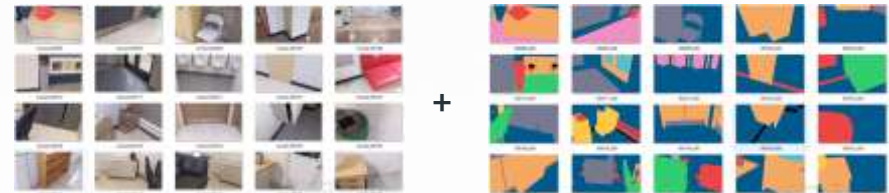
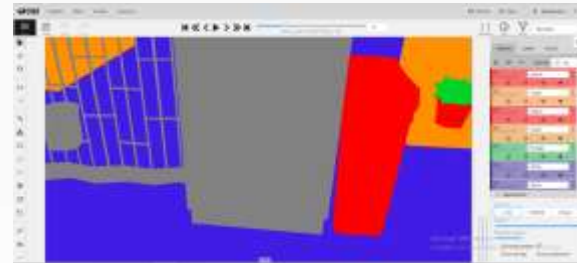


- MINC-2500



## 2 ANNOTATION MANUELLE DES IMAGES

Computer Vision Annotation Tool



# III. DÉVELOPPEMENT (6/9)

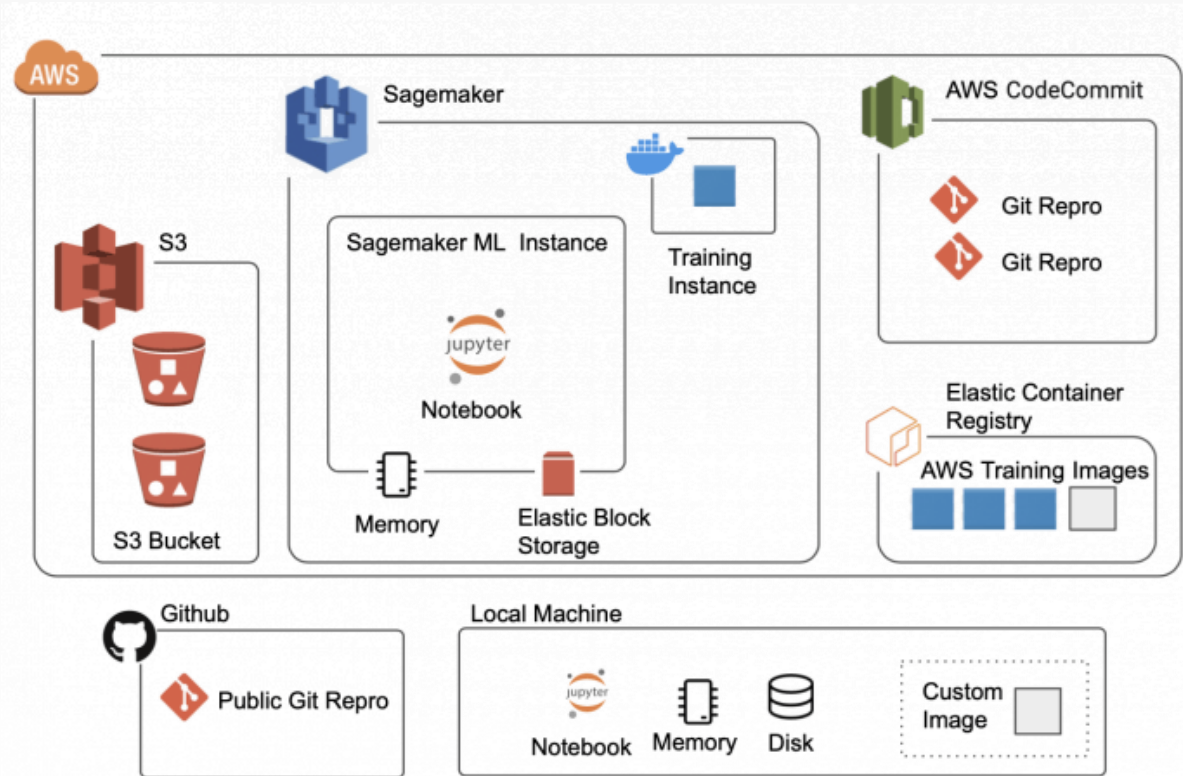
## Environnement matériel

### AMAZON SAGEMAKER

Instance : ml.g4dn.xlarge

- ✓ GPU : 1
- ✓ vCPU : 4
- ✓ Mémoire (Gio) : 16
- ✓ Mémoire GPU (Gio) : 16

### STOCKAGE : SERVICES S3 ET EBS





# III. DÉVELOPPEMENT (7/9)

---

## Outils



---

## Parmi les bibliothèques utilisées



# III. DÉVELOPPEMENT (8/9)

## Modèles d'extraction des caractéristiques des images (backbones)

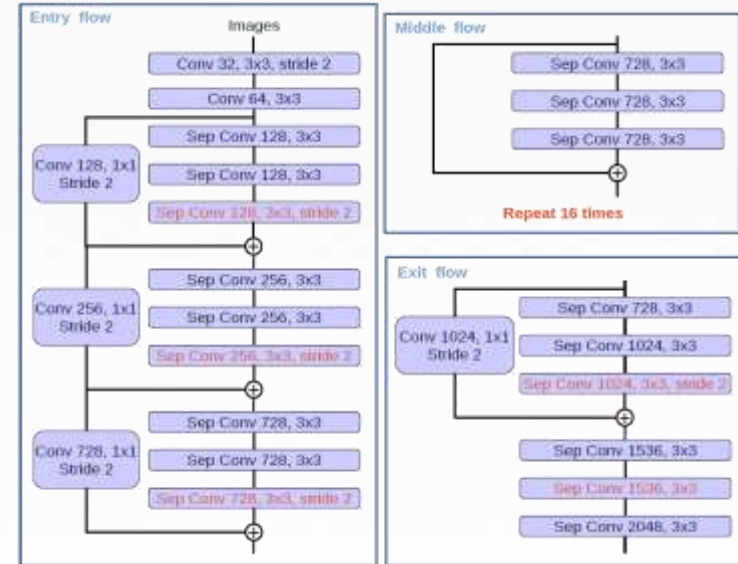
### MODÈLE RESNET



5 Variants selon la profondeur du réseau de neurones :

- ResNet 18 couches.
- ResNet 34 couches.
- ResNet 50 couches.
- ResNet 101 couches.
- ResNet 152 couches.

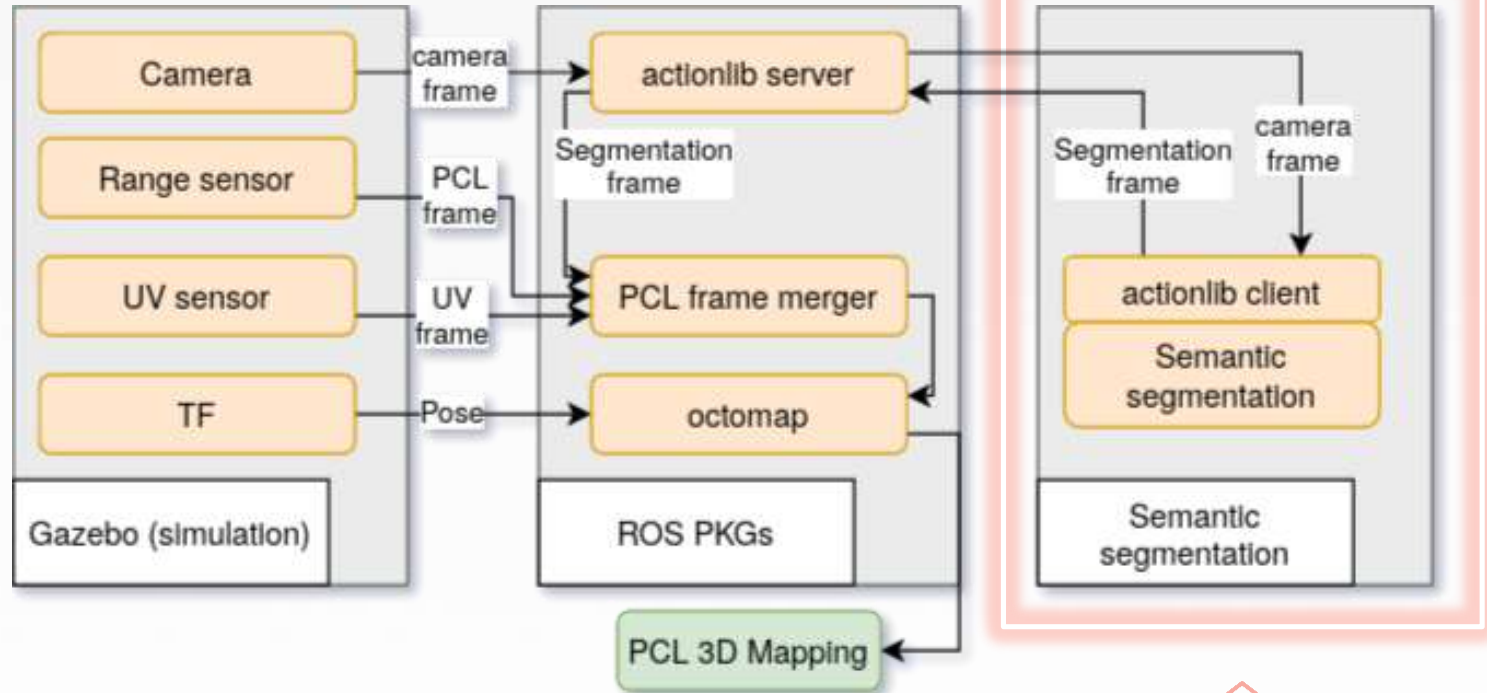
### MODÈLE XCEPTION

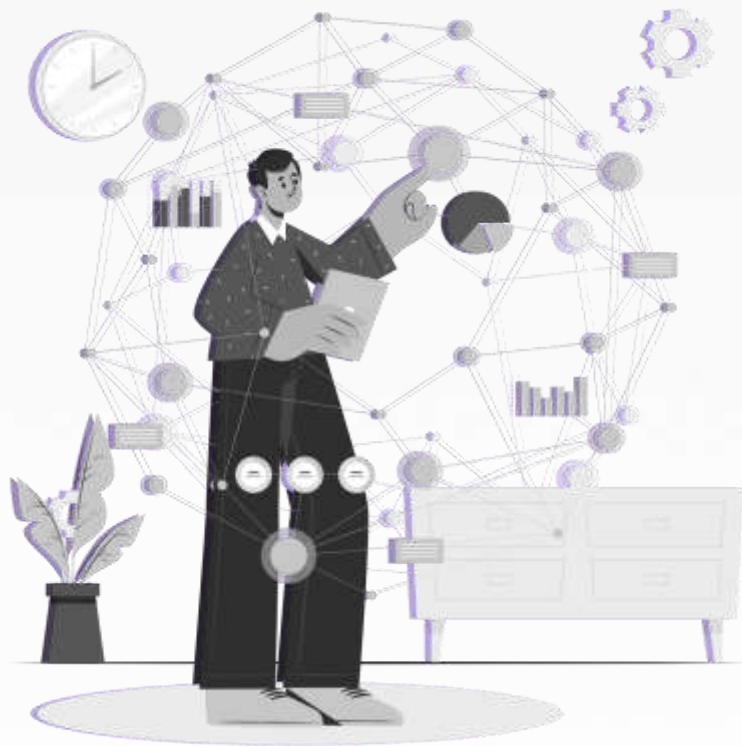


# III. DÉVELOPPEMENT (9/9)

Intégration de la solution dans l'interface ROS

Module de segmentation

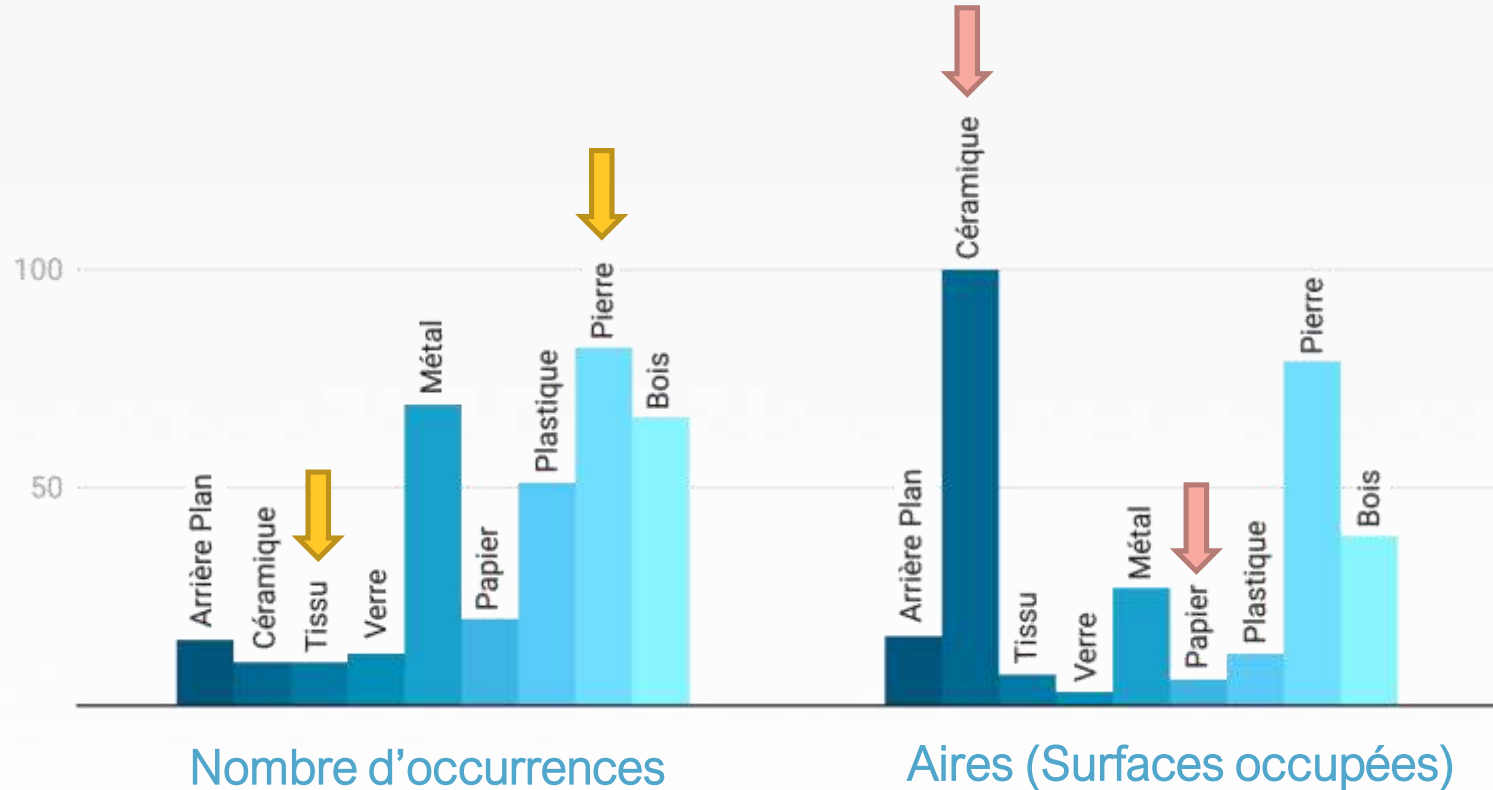




# IV. EXPÉRIMENTATIONS ET RÉSULTATS

# IV. EXPÉRIMENTATIONS ET RÉSULTATS (1/11)

Analyse de notre jeu de donnée : « HRIM 2021 »



# IV. EXPÉRIMENTATIONS ET RÉSULTATS (2/11)

## Stratégie d'entraînement des modèles

### Expérimentation 1

- \* Petite base de données

### Expérimentation 3

- \* Grande base de données
- \* Techniques d'augmentation
- \* Enrichissement de données

Data scraping

Création d'un  
algorithme  
d'annotation  
automatique


### Expérimentation 2

- \* Base de données plus grande
- \* Techniques d'augmentation

# IV. EXPÉRIMENTATIONS ET RÉSULTATS (3/11)

## Métriques d'évaluation

### Intersection over Union (IoU)


$$\text{IoU} = \frac{\text{Area of Overlap}}{\text{Area of Union}}$$

### Pixel Accuracy

$$\text{accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

# IV. EXPÉRIMENTATIONS ET RÉSULTATS (4/11)

## Expérimentation 1 (1/2)

- Petite base de données : 112 images (de l'IRH seulement)

	Modèle	Backbone	Moyenne IoU	Précision Moy. des Pixels
1 <sup>er</sup>	DeepLabV3+	ResNet101	30 %	66.4 %
	DeepLabV3+	ResNet18	29.2 %	67.8%
	DeepLabV3+	ResNet50	26.2 %	62 %
	DeepLabV3+	ResNet34	26 %	57.9 %
2 <sup>ème</sup>	DeepLabV3+	ResNet152	24.4 %	60.7 %
	PSPNet	ResNet50	23.5 %	60.3 %
	DeepLabV3+	Xception	13.6 %	47 %
3 <sup>ème</sup>	UNet	ResNet50	12.8 %	49.4 %






# IV. EXPÉRIMENTATIONS ET RÉSULTATS (5/11)

## Expérimentation 1 (2/2)

- Petite base de données : 279 images (IRH + Flickr + MINC-2500)



Modèle	Backbone	Moyenne IoU	Précision Moy. des Pixels
PSPNet	ResNet101	58.7 %	77.2 %
DeepLabV3+	ResNet101	53 %	73.6 %
UNet	ResNet101	19.1 %	53.3 %



Modèle	Backbone	Moyenne IoU	Précision Moy. des Pixels
DeepLabV3+	ResNet-18	55.1 %	74.9 %
DeepLabV3+	ResNet-34	55 %	75.2 %
DeepLabV3+	ResNet-50	53.7 %	74.2 %
DeepLabV3+	ResNet-101	53 %	73.6 %
DeepLabV3+	ResNet-152	52 %	72.9 %
DeepLabV3+	Xception	18.6 %	49.3 %



# IV. EXPÉRIMENTATIONS ET RÉSULTATS (6/11)

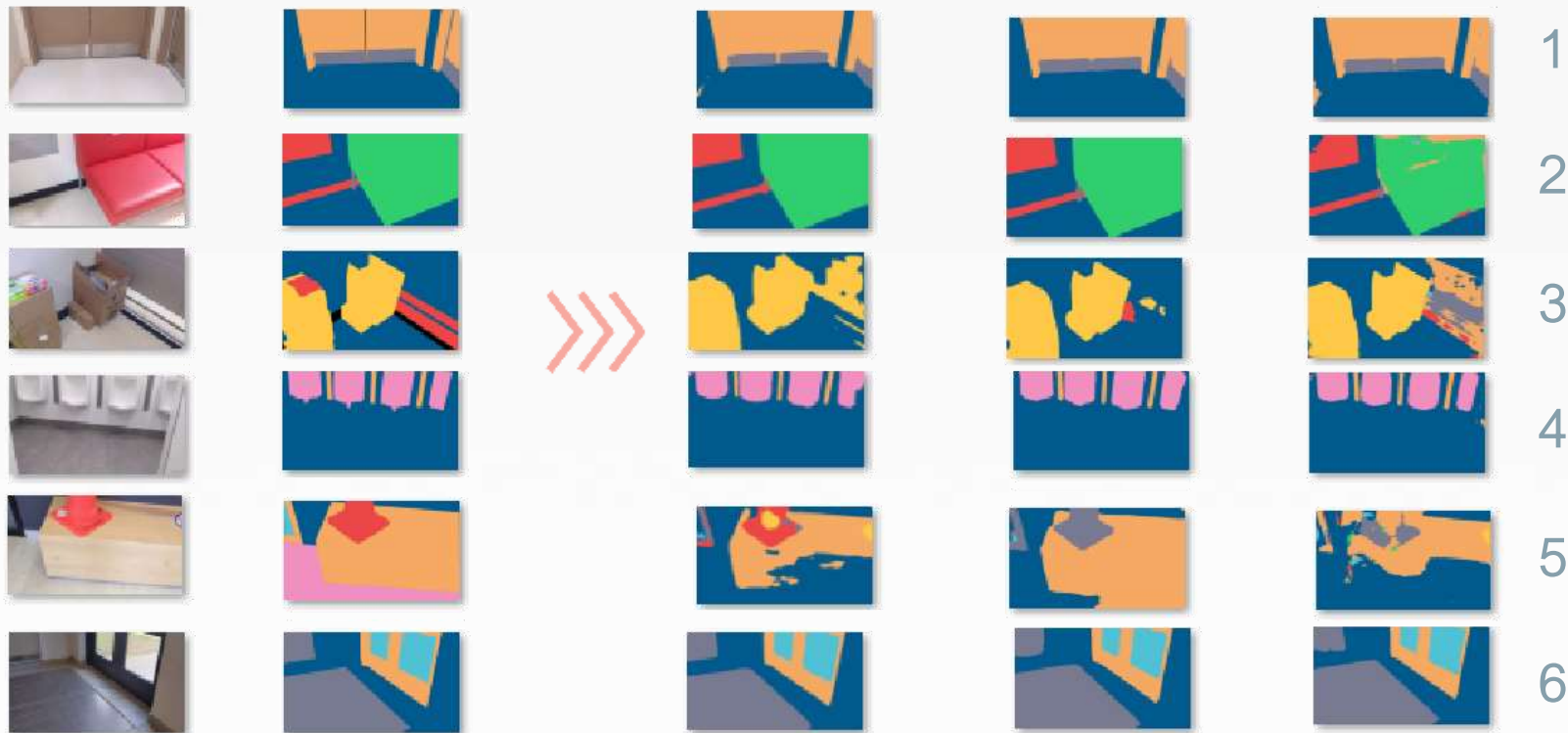
## Expérimentation 2

- Augmentation de données : Transformations spatiales et au niveau des pixels
- Plus grande base de données : (279 images x 27 ) = 7533 images

	Modèle	Backbone	Moyenne IoU	Précision Moy. des Pixels
1 <sup>er</sup>	DeepLabV3+	ResNet101	76.6 %	88.4 %
	DeepLabV3+	ResNet18	76.3 %	88.7 %
2 <sup>ème</sup>	PSPNet	ResNet101	75.2 %	88.1 %
3 <sup>ème</sup>	UNet	ResNet101	56.5 %	76.8 %
	DeepLabV3+	Xception	48.1 %	70.6 %



# IV. EXPÉRIMENTATIONS ET RÉSULTATS (7/11)



Images

Vérité Terrain

DeepLabV3+

PSPNet

U-Net

Arrière-Plan

Pierre

Céramique

Tissu

Verre

Métal

Papier

Plastique

Bois

# IV. EXPÉRIMENTATIONS ET RÉSULTATS (8/11)

## Expérimentation 3

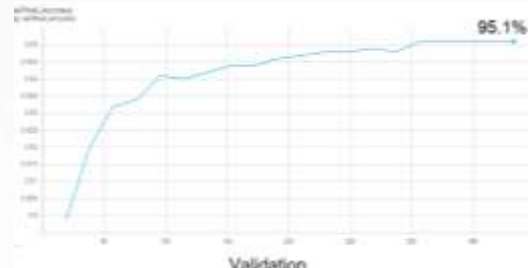
➤ Plus large base de données : 15 315 images

Modèle	Backbone	Moyenne IoU	Précision Moy. des Pixels
DeepLabV3+	ResNet101	83.2 %	95.1 %

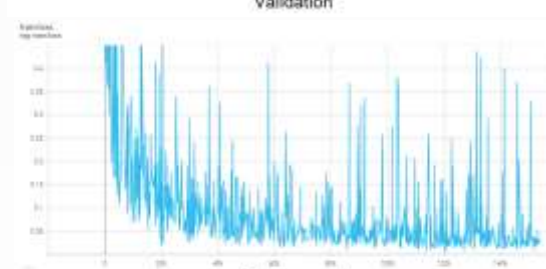
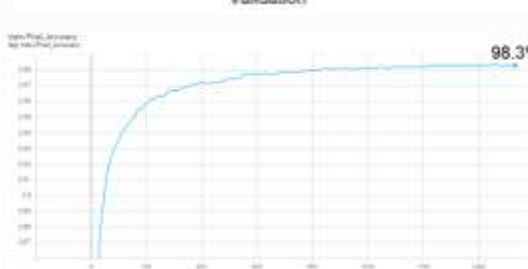
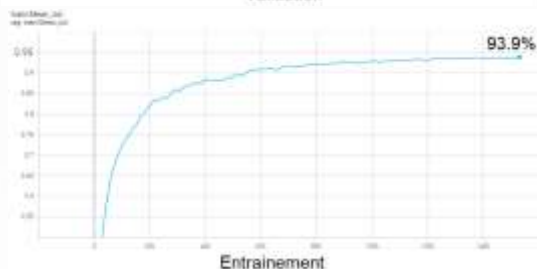
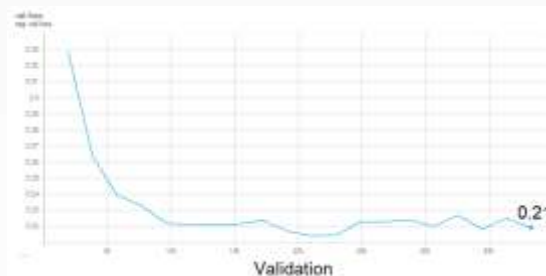
Moyenne IoU



Précision Moy. des pixels

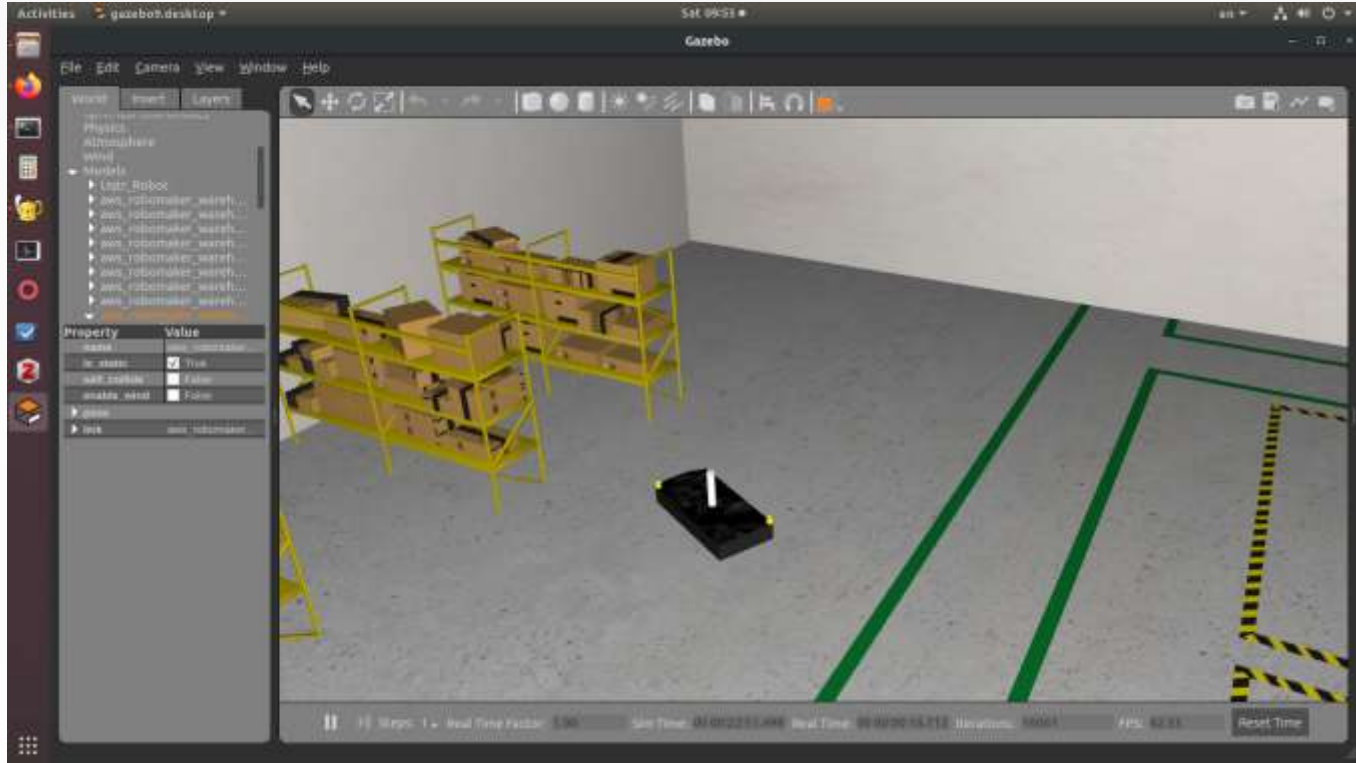


Perte



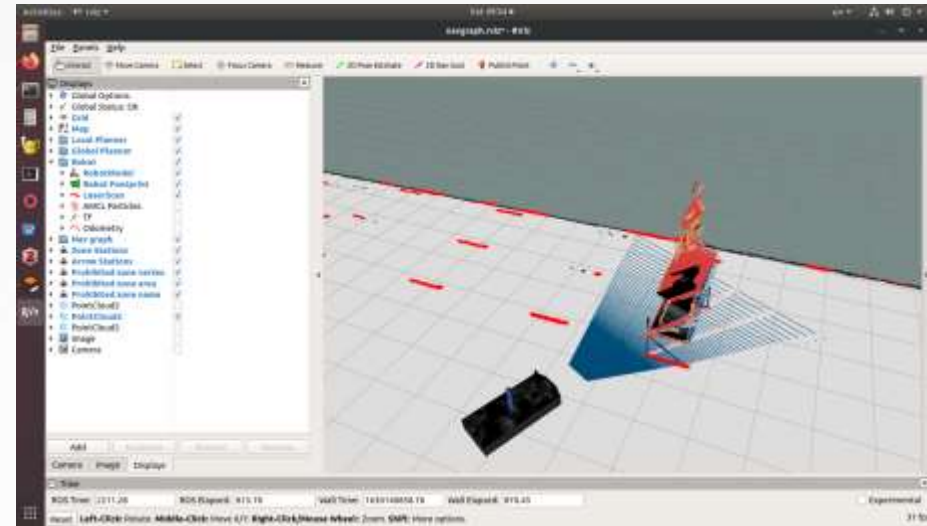
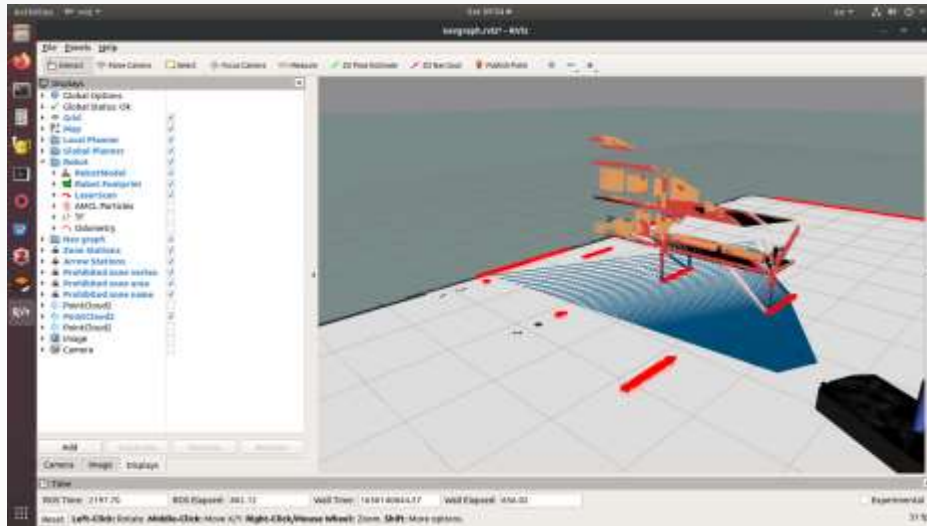
# IV. EXPÉRIMENTATIONS ET RÉSULTATS (9/11)

Intégration de la solution dans l'environnement ROS de simulation



## IV. EXPÉRIMENTATIONS ET RÉSULTATS (10/11)

# Module de segmentation sémantique de l'environnement



Le robot de désinfection navigue dans son environnement et arrive à reconnaître la nature des matériaux qui existent.

## IV. EXPÉRIMENTATIONS ET RÉSULTATS (11/11)

Segmentation d'une cartographie 3D de l'environnement intérieur par le robot

Bois, Plastique, Pierre



Entrée de la caméra



Résultat de notre module  
de segmentation sémantique

# V. CONCLUSION ET PERSPECTIVES

---

- Nous avons développé un modèle **DeepLabV3+ (Resnet101)** de segmentation sémantique des matériaux avec une précision **83.2% (Moy. IoU)** réalisant ainsi une contribution dans ce domaine.
- Dans cette présentation, nous avons commencé par une présentation générale. Ensuite, nous avons présenté les résultats des recherches. Puis, nous avons exposé les étapes de développement jusqu'à arriver aux expérimentations et résultats.
- Au cours de ce stage, nous avons rencontré de nombreux problèmes lors de la préparation des données de matériaux ainsi que la nécessité de **mémoire** et **GPU**.
- Pour les futures améliorations, nous pourrions continuer ce projet en ajoutant encore des données et en utilisant d'autres modèles d'extraction de caractéristiques.



**MERCI DE VOTRE ATTENTION**

---

## Enrichissement des données (Expérimentation 3) pour améliorer le résultat :

1- **Data Scraping** : Création d'un algorithme d'extraction des images présentant un seul type de matériau par image sur un arrière-plan blanc afin de simplifier la cette tâche d'annotation.  
(Beautiful Soup + Selenium)

2- **Annotation automatique** : Création d'un algorithme d'annotation automatique de ces images par un traitement de pixels à l'aide de la bibliothèque OpenCV.

Data Scraping

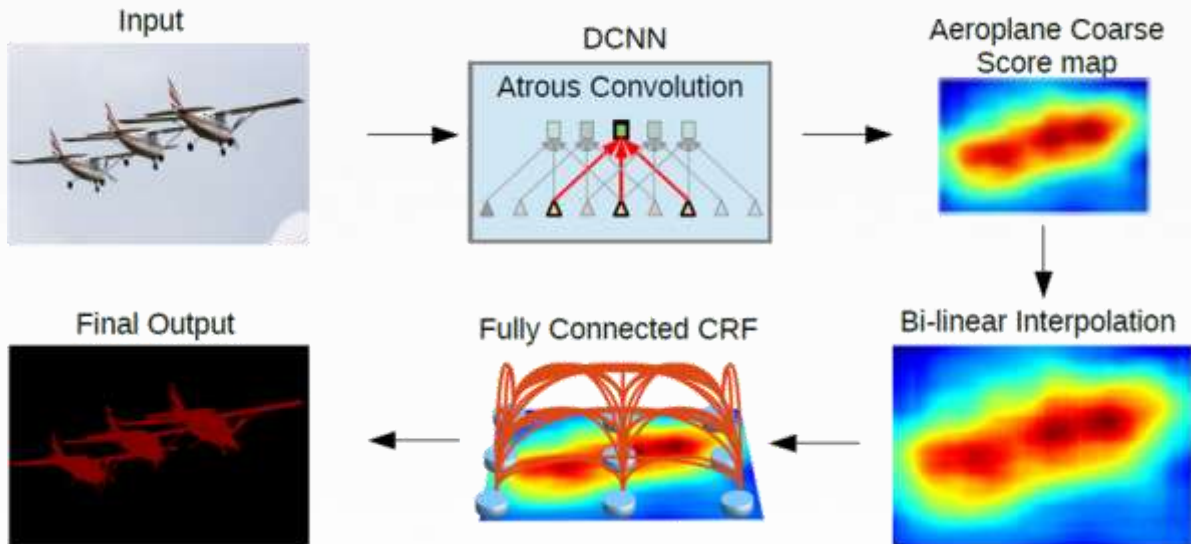


Génération  
Automatique des  
Masques



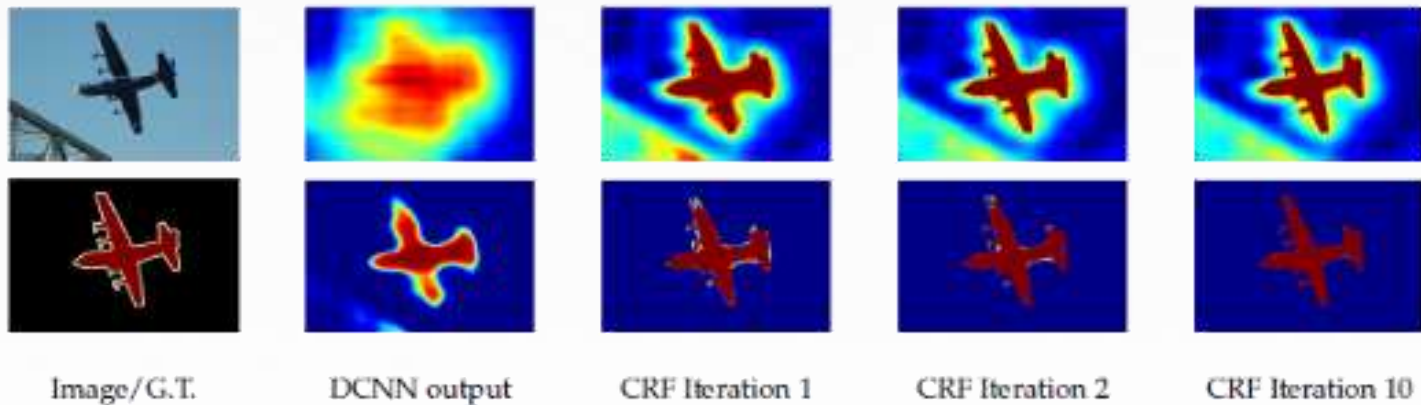
## → Principe de fonctionnement de DeepLabv1:

- **Dilated Convolution** (Algorithme À trous)
- **Fully Connected Conditional Random Field (CRF)**
- Utilisation **VGG-16** comme backbone (Feature extractor)



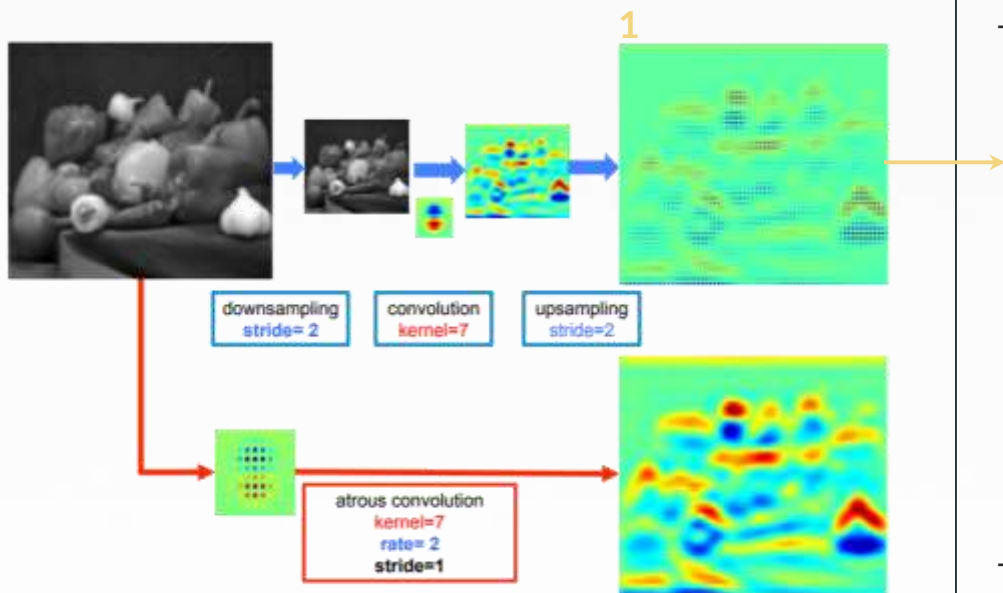
- Fully Connected Conditional Random Field (CRF)

Pour résoudre le problème de faible précision aux limites des objets et améliorer la capacité du modèle à capturer des détails fins.



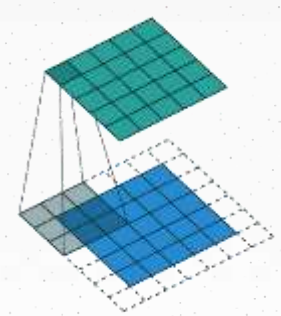
- Dilated Convolution (Atrous Conv)

Résout le problème de réduction de résolution des cartes de caractéristiques (feature map).



- Causes :

- Convolution



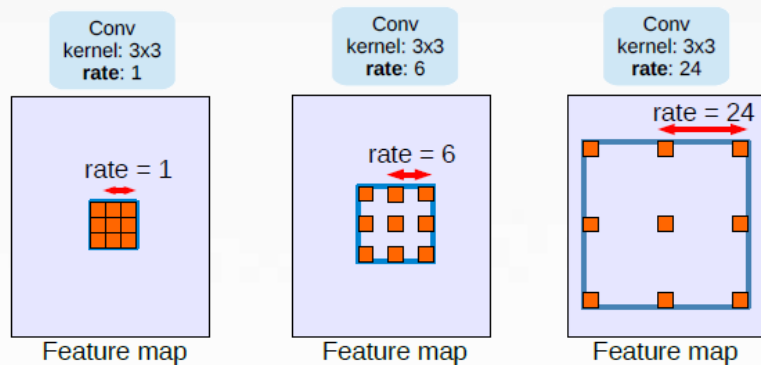
12	20	30	0
8	12	2	0
34	70	37	4
112	100	25	12

$2 \times 2$  Max-Pool

20	30
112	37

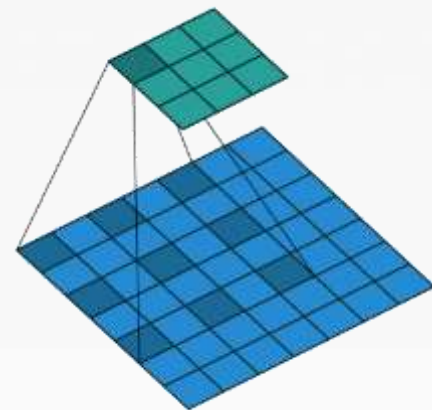
- Max pooling

- Convolution à trous

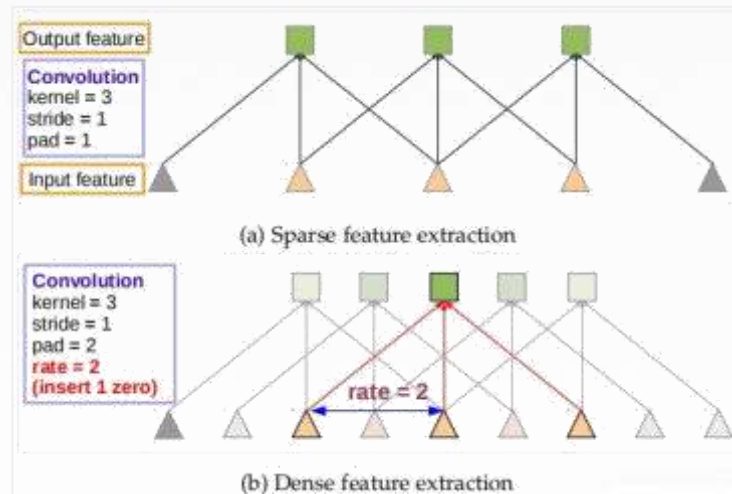


Atrous Convolution with Different Rates  $r$

$$y[i] = \sum_k x[i + r \cdot k] w[k]$$



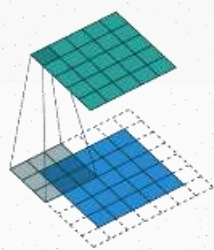
Convolution à trous



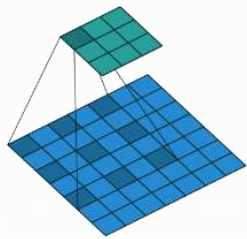
Problème de classe multi-échelles  
Solution : ASPP



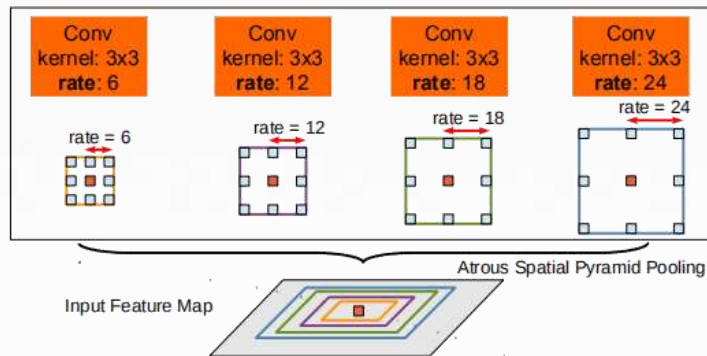
## Les techniques de DeepLabV3+



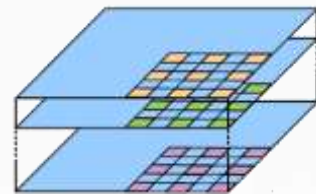
Convolution standard



Convolution à trous



ASPP : Mise en commun de  
pyramide spatiale dilatée



Convolution à  
trous séparable  
et en profondeur



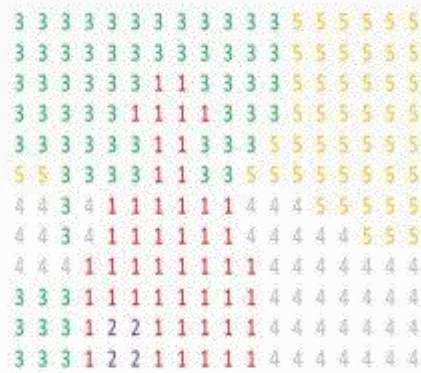
## ➤ ONE HOT ENCODING



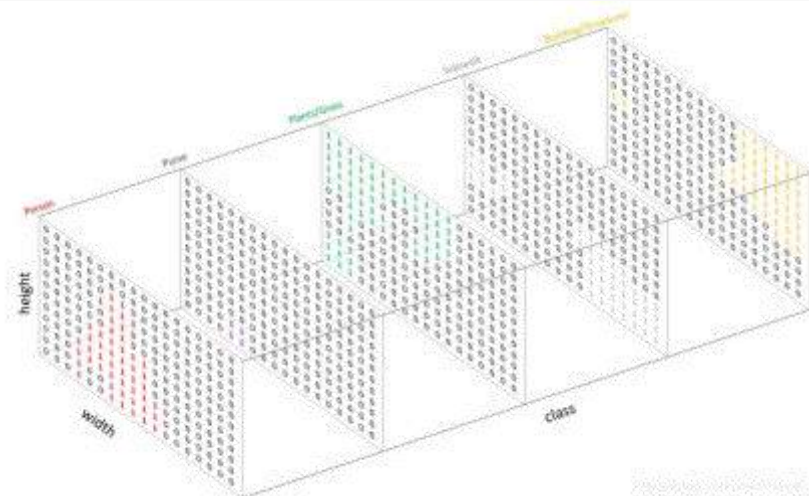
Input



- 1: Person
- 2: Purse
- 3: Plants/Green
- 4: Sidewalk
- 5: Building/Structures



Semantic Labels



## Les techniques d'augmentation

