



Analyse automatique d'images marines

Atelier technique ODATIS

Thibault Napoléon

*ISEN Yncréa Ouest, Laboratoire L@bISEN
Équipe Vision-AD*



Sommaire I



- 1 Contexte
- 2 Exemples
- 3 L'apprentissage profond
- 4 Les architectures neuronales
- 5 Les annotations
- 6 La puissance de calcul
- 7 Conclusion

Sommaire II



7 Conclusion

Contexte

Les données sous-marines

Contexte

Aujourd'hui les moyens d'acquisition de données optiques (c.-à-d. les appareils photo et les caméras) sont peu chers et proposent une bonne résolution spatiale. Ces qualités rendent l'acquisition vidéo particulièrement attrayante pour équiper les petits AUV pour des missions d'identification et d'observation précise.

⇒ Ceci amène à une explosion du volume de données disponibles après les missions d'observations.

Contexte

Les données sous-marines

Contexte

Aujourd'hui les moyens d'acquisition de données optiques (c.-à-d. les appareils photo et les caméras) sont peu chers et proposent une bonne résolution spatiale. Ces qualités rendent l'acquisition vidéo particulièrement attrayante pour équiper les petits AUV pour des missions d'identification et d'observation précise.

⇒ Ceci amène à une explosion du volume de données disponibles après les missions d'observations.

Notion de grandeur :

Combien de temps pour analyser 1h de vidéo à raison d'une image par seconde et de 30 secondes passées pour chaque image ?

Contexte

Les données sous-marines

Contexte

Aujourd'hui les moyens d'acquisition de données optiques (c.-à-d. les appareils photo et les caméras) sont peu chers et proposent une bonne résolution spatiale. Ces qualités rendent l'acquisition vidéo particulièrement attrayante pour équiper les petits AUV pour des missions d'identification et d'observation précise.

⇒ Ceci amène à une explosion du volume de données disponibles après les missions d'observations.

Notion de grandeur :

Combien de temps pour analyser 1h de vidéo à raison d'une image par seconde et de 30 secondes passées pour chaque image ? **30 heures !!**

Contexte

Les données sous-marines

Contexte

Aujourd'hui les moyens d'acquisition de données optiques (c.-à-d. les appareils photo et les caméras) sont peu chers et proposent une bonne résolution spatiale. Ces qualités rendent l'acquisition vidéo particulièrement attrayante pour équiper les petits AUV pour des missions d'identification et d'observation précise.

⇒ Ceci amène à une explosion du volume de données disponibles après les missions d'observations.

Notion de grandeur :

Combien de temps pour analyser 1h de vidéo à raison d'une image par seconde et de 30 secondes passées pour chaque image ? **30 heures !!**

Solution : utiliser des outils d'apprentissage automatique

Contexte

L'apprentissage automatique

C'est quoi ?

C'est un champ de l'intelligence artificielle qui concerne la conception, l'analyse, le développement et l'implémentation de méthodes permettant à une machine d'évoluer.

Cette évolution se fait par un processus systématique lui permettant de répondre à des tâches difficiles ou impossibles à remplir par des moyens algorithmiques plus classiques.

Un savant mélange entre l'informatique et les statistiques :

- Comment construire des machines qui résolvent des problèmes ?
- Quels problèmes sont intrinsèquement faisables/infaisables ?



Contexte

Apprentissage Vs Système de règles

Système de règles :

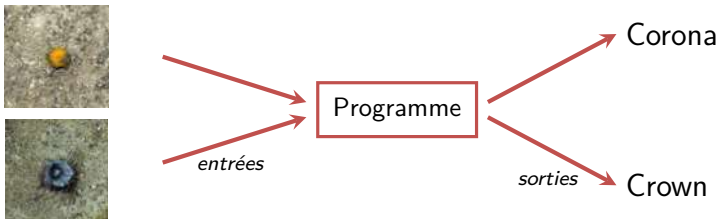


FIGURE 1 – *Algorithme de classification à base d'un système de règles.*

Contexte

Apprentissage Vs Système de règles

Système de règles :

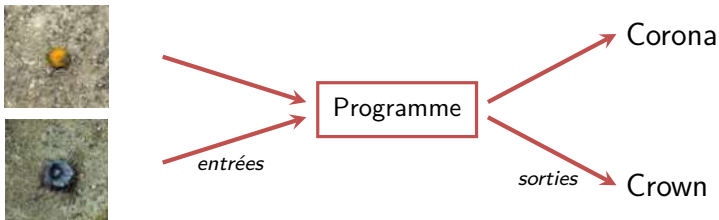


FIGURE 1 – Algorithme de classification à base d'un système de règles.

Difficultés :

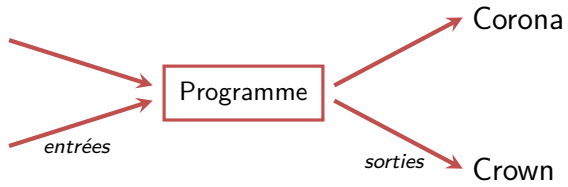
- Impossible de créer des règles sur les pixels, car trop de variété.
- Difficulté pour trouver des règles pour chaque espèce et entre chaque espèce.

Contexte



Apprentissage Vs Système de règles

Apprentissage automatique :



————— Inférence

FIGURE 2 – Algorithme de classification à base d'apprentissage automatique.

Contexte

Apprentissage Vs Système de règles

Apprentissage automatique :

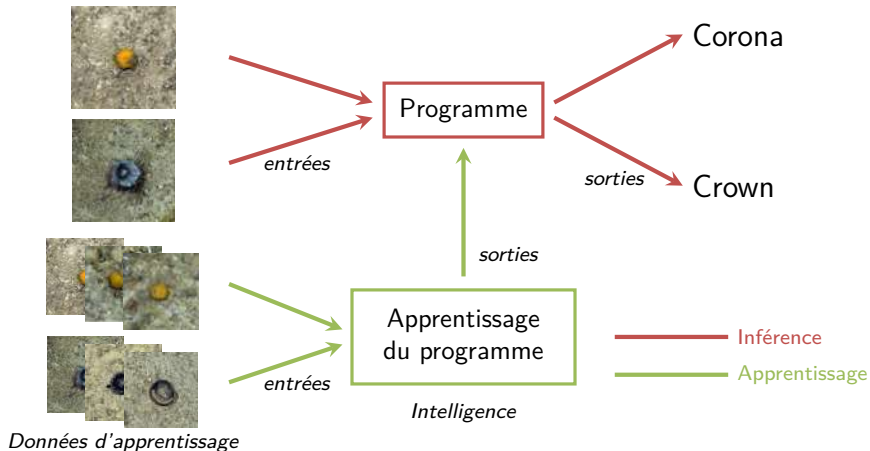


FIGURE 2 – Algorithme de classification à base d'apprentissage automatique.

Contexte



Remarques

De quel type d'apprentissage automatique allons nous parler ?

- Non supervisé ? Semi-supervisé ? Supervisé ? ⇒ Supervisé
- À base de noyau (SVM...) ? À base d'arbre (Forêt aléatoire...) ? Apprentissage Profond ? ⇒ Apprentissage Profond

Contexte

Remarques

De quel type d'apprentissage automatique allons nous parler ?

- Non supervisé ? Semi-supervisé ? Supervisé ? ⇒ **Supervisé**
- À base de noyau (SVM...) ? À base d'arbre (Forêt aléatoire...) ? Apprentissage Profond ? ⇒ **Apprentissage Profond**

Pourquoi ce choix ?

L'apprentissage profond (c.-à-d. Deep Learning) est une méthode d'apprentissage automatique fortement supervisé très performante nécessitant :

- Pas/Peu de prétraitement des données.
- Pas/Peu de connaissances a priori sur le fonctionnement des algorithmes.

Attention : quels que soient les choix effectués, un outil automatique ne fera jamais mieux qu'un expert humain. Par contre il le fera infiniment plus vite.

Sommaire I



- 1 Contexte
- 2 Exemples
 - Classification automatique d'éponges marines
 - Étude de l'habitat dans le canal du Mozambique
 - Segmentation sémantique en rade de Brest
- 3 L'apprentissage profond
- 4 Les architectures neuronales
- 5 Les annotations
- 6 La puissance de calcul

Sommaire II



7 Conclusion

Exemples

Classification automatique d'éponges marines

Le projet

Reconnaissance et classification automatique de morphotypes d'éponges marines présentes sur des images de fonds marins en collaboration avec Ifremer.



FIGURE 3 – Extraction des 6 classes d'éponges traitées dans ce projet.

Problématique :

Comment détecter et classifier les éponges sur plus de 20000 images ?

Exemples

Classification automatique d'éponges marines

Le projet

Reconnaissance et classification automatique de morphotypes d'éponges marines présentes sur des images de fonds marins en collaboration avec Ifremer.



FIGURE 3 – Extraction des 6 classes d'éponges traitées dans ce projet.

Problématique :

Comment détecter et classifier les éponges sur plus de 20000 images ?
⇒ **S'appuyer sur un outil de classification automatique utilisant l'apprentissage automatique.**

Exemples

Classification automatique d'éponges marines

Nombre d'annotations pour chaque classes :

Le nombre d'annotations n'est pas homogène entre tous les morphotypes. Certains, comme « Vase » sont sur-représentés avec plus de deux fois plus d'annotations que « Grey_white ». Au total, nous disposons d'environ 1600 annotations.

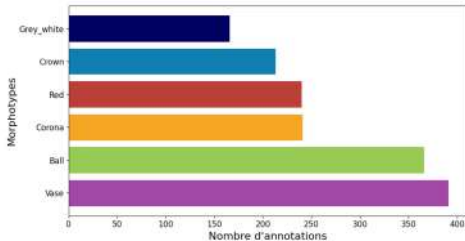


FIGURE 4 – Répartition des annotations dans les 6 classes utilisées lors du projet.

Exemples

Classification automatique d'éponges marines



FIGURE 5 – Exemple de résultats de détection et de classification des éponges marines.

Les performances obtenus sont de $\approx 70\%$

Exemples

Classification automatique d'éponges marines



FIGURE 6 – Exemple de résultats de détection et de classification des éponges marines.

Exemples

Étude de l'habitat dans le canal du Mozambique

Le projet

Étude de l'habitat des espèces benthiques dans le canal du Mozambique en collaboration avec Ifremer.

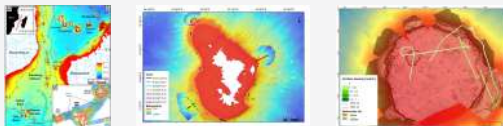


FIGURE 7 – Canal du Mozambique (à gauche) avec les deux lieux de plongées : les pentes insulaires de Mayotte (au milieu), le mont sous-marin « bank hall » (à droite)

Problématique :

Comment étudier la répartition des fonds marins (sédiments, roche volcanique...) sur plus de 9000 images ?

Exemples

Étude de l'habitat dans le canal du Mozambique

Le projet

Étude de l'habitat des espèces benthiques dans le canal du Mozambique en collaboration avec Ifremer.

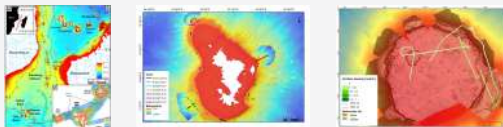


FIGURE 7 – Canal du Mozambique (à gauche) avec les deux lieux de plongées : les pentes insulaires de Mayotte (au milieu), le mont sous-marin « bank hall » (à droite)

Problématique :

Comment étudier la répartition des fonds marins (sédiments, roche volcanique...) sur plus de 9000 images ?

⇒ **S'appuyer sur un outil de classification automatique utilisant l'apprentissage automatique.**

Exemples

Étude de l'habitat dans le canal du Mozambique

Objectifs et contraintes :

- Annotation et quantification automatique du fond marin selon quatre classes :
 - ⇒ sédiment, graviers, roche volcanique, roche calcaire
- Peu d'annotation (< 1000), pas d'annotation sémantique
- Image fortement texturée, sans structure ⇒ difficile d'extraire des caractéristiques

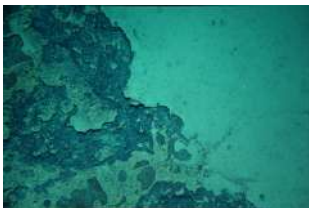


FIGURE 8 – Image originale (à gauche) et image automatiquement segmentée (c.-à-d. annotée) (à droite)

Exemples

Étude de l'habitat dans le canal du Mozambique

Le réseau de neurones proposé :

Extraction des caractéristiques du fond marin par la combinaison d'un réseau VGG19 et de couches d'extraction d'ondelettes (Wavelet CNN) pour extraire des informations sur les textures.

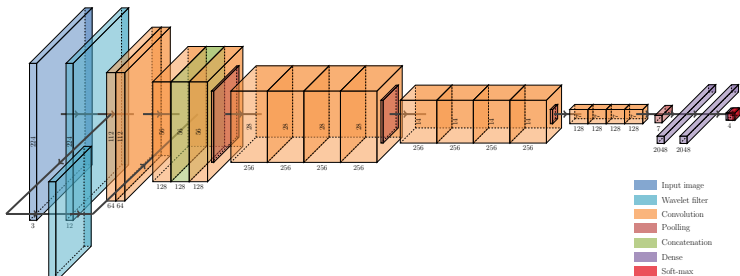


FIGURE 9 – Réseau de neurones utilisé pour la segmentation des images.

Les performances obtenus sont de $\approx 83\%$

Exemples

Étude de l'habitat dans le canal du Mozambique

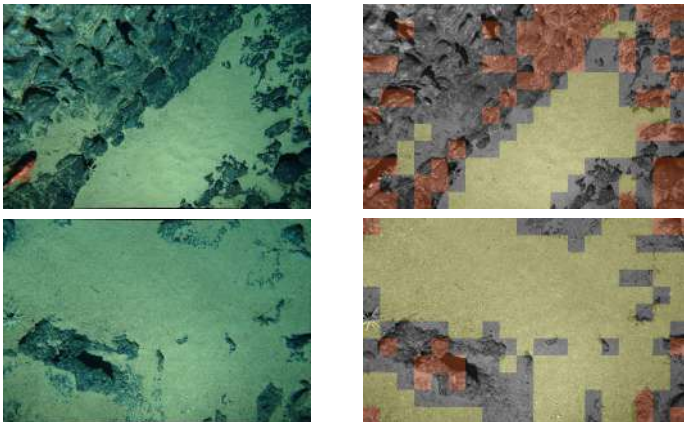


FIGURE 10 – Résultats qualitatifs sur l'approche proposée Les images originales (à gauche), les images annotées (à droite)

Exemples

Segmentation sémantique en rade de Brest

Segmentation sémantique
La segmentation sémantique vise à attribuer une classe à chaque pixel de l'image.

Annotation de chaque pixel avec 4 classes :

- Eau (400 annotations)
- Fond lithique (400 annotations)
- Fond sableux (400 annotations)
- Algues/herbiers (400 annotations)



FIGURE 11 – Image issue de la rade de Brest

Exemples

Segmentation sémantique en rade de Brest

Réseaux de neurones utilisés :

Pour segmenter sémantiquement nos images, nous utilisons deux réseaux de neurones :

- SqueezeNet : pour la classification par patch (et la génération des annotations sémantiques).
- U-Net : pour la segmentation sémantique.



FIGURE 12 – Schéma du processus de segmentation sémantique

Exemples

Segmentation sémantique en rade de Brest



FIGURE 13 – Résultat qualitatif de la segmentation sémantique (extraction d'une image traitée en temps réel sur un GPU Titan V)

Exemples

Segmentation sémantique en rade de Brest



FIGURE 14 – Résultat qualitatif de la segmentation sémantique (extraction d'une image traitée en temps réel sur un GPU Titan V)

Exemples

Segmentation sémantique en rade de Brest

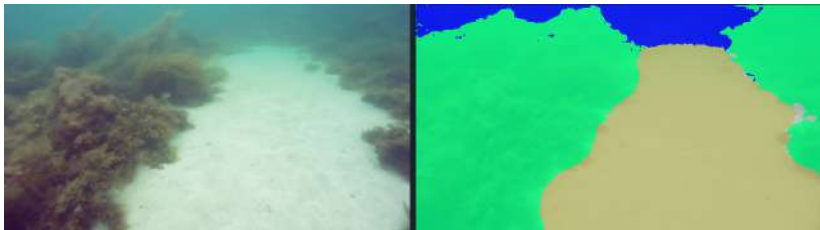


FIGURE 15 – Résultat qualitatif de la segmentation sémantique (extraction d'une image traitée en temps réel sur un GPU Titan V)

Sommaire I



- ① Contexte
- ② Exemples
- ③ L'apprentissage profond
 - C'est quoi ?
 - Pour faire quoi ?
- ④ Les architectures neuronales
- ⑤ Les annotations
- ⑥ La puissance de calcul
- ⑦ Conclusion

L'apprentissage profond

C'est quoi ?

C'est quoi ?

C'est un champ de l'apprentissage automatique qui permet de créer des processus de décision complexes où l'intervention de l'utilisateur est minimale.

Les données en entrée sont les images brutes et les données en sorties sont souvent les probabilités d'appartenance de l'objet à chacune des classes apprises.

Inconvénients de cette approche :

- Elle nécessite généralement un très gros volume de données annotées.
- Elle demande une puissance de calcul considérable pour la phase d'apprentissage.



L'apprentissage profond

C'est quoi ?

Avant les réseaux de neurones :

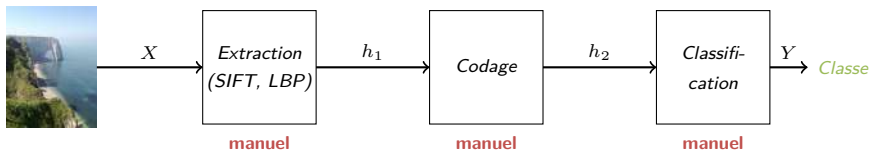


FIGURE 16 – *Extraction, codage et classification manuelle de caractéristiques.*

L'apprentissage profond

C'est quoi ?

Avant les réseaux de neurones :

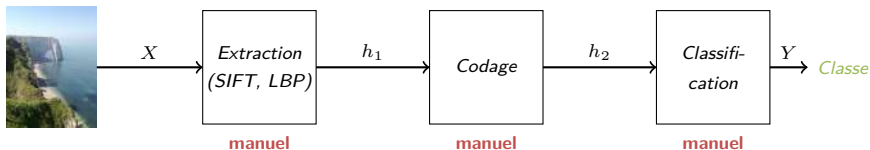


FIGURE 16 – Extraction, codage et classification *manuelle* de caractéristiques.

Après les réseaux de neurones :

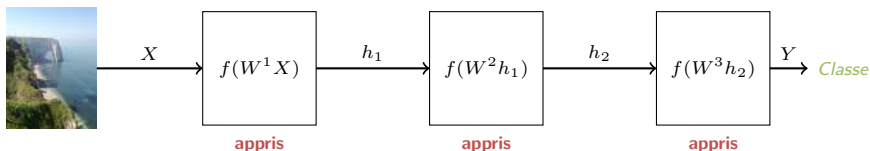


FIGURE 17 – Extraction, codage et classification *automatique* de caractéristiques.

L'apprentissage profond

Pour faire quoi ?

Quels types de problèmes peut-on traiter ?

Les approches d'apprentissage automatique permettent généralement de répondre aux problèmes classiques de la vision humaine tels que :

L'apprentissage profond

Pour faire quoi ?

Quels types de problèmes peut-on traiter ?

Les approches d'apprentissage automatique permettent généralement de répondre aux problèmes classiques de la vision humaine tels que :

- **La reconnaissance d'objets** : étant donné une image, le but est d'attribuer un label (c.-à-d. une classe), parmi celles apprises, à l'image.



Scène de circulation

L'apprentissage profond

Pour faire quoi ?

Quels types de problèmes peut-on traiter ?

Les approches d'apprentissage automatique permettent généralement de répondre aux problèmes classiques de la vision humaine tels que :

- **La reconnaissance d'objets** : étant donné une image, le but est d'attribuer un label (c.-à-d. une classe), parmi celles apprises, à l'image.



Scène de circulation

- **La détection d'objets** : en plus de la reconnaissance, le but ici est de localiser les objets dans la scène en fournissant leur boîte minimale englobante.



Voiture

L'apprentissage profond

Pour faire quoi ?

- **L'identification d'objets** : dans cette approche, on cherche à donner une « identité » à l'objet présent dans l'image. Pour cela il est nécessaire d'avoir une base de données des individus connus.



Mercedes C36

L'apprentissage profond



Pour faire quoi ?

L'identification d'objets : dans cette approche, on cherche à donner une « identité » à l'objet présent dans l'image. Pour cela il est nécessaire d'avoir une base de données des individus connus.

-



Mercedes C36

La segmentation de scènes : cette tâche consiste à découper l'image en zones cohérentes (c.-à-d. de même classe). Cette segmentation peut se faire par blocs de pixels ou directement au niveau pixel.

-



Segmentation sémantique

Sommaire I



- 1 Contexte
- 2 Exemples
- 3 L'apprentissage profond
- 4 Les architectures neuronales
Mise en place
Les architectures
- 5 Les annotations
- 6 La puissance de calcul
- 7 Conclusion

Les architectures

Mise en place

Connaissances nécessaires :

Malgré une apparente complexité, la prise en main d'une architecture neuronale est possible pour un jeune ingénieur ayant des notions de programmation. Aussi, l'utilisation d'un réseau déjà entraîné, appelée phase d'inférence, ne nécessite pas de savoir-faire particulier.

Langage de programmation :

Le langage de programmation utilisé est le Python auquel on ajoute des paquets spécifiques pour l'utilisation des réseaux de neurones.

Paquets utilisés :

En particulier on utilise l'un des paquets suivants (en fonction du code disponible en ligne) :

- Keras (le plus simple)
- TensorFlow (s'appuie sur les développements de Google)
- PyTorch (s'appuie sur les développements de Facebook)

Les architectures

Les architectures

Plusieurs architectures existent :

Type de tâche	Nom des réseaux	Temps de calcul
Classification d'image (simple)	SqueezeNet, LeNet, AlexNet, VGG-16	Faible/Moyen
Classification d'image (avancée)	Inception, ResNet	Important
Localisation et détection	YOLO, R-CNN	Faible
Segmentation d'image	U-Net, FCN, DeepLab	Important
Reconnaissance	FaceNet	Moyen

TABLE 1 – Ensemble d'architectures neuronales disponibles pour l'analyse automatique des images.

Sommaire I

- 1 Contexte
- 2 Exemples
- 3 L'apprentissage profond
- 4 Les architectures neuronales
- 5 Les annotations
 - Quel jeu de données
 - C'est quoi une annotation ?
- 6 La puissance de calcul
- 7 Conclusion

Les annotations

Quel jeu de données

Les annotations

L'analyse automatique des images par apprentissage profond a besoin d'une base de données annotée (communément appelée « base d'entraînement ») pour apprendre. Lors de la constitution de celle-ci, la phase d'annotation est cruciale et va fortement orienter les performances de l'apprentissage.

Les points de vigilances :

- Le nombre d'annotations et leurs répartitions.
- Le format des annotations.
- Le format de stockage de ces annotations.



Les annotations



Quel jeu de données

Mais de combien d'annotations a-t-on besoin ?

Les annotations



Quel jeu de données

Mais de combien d'annotations a-t-on besoin ?

En théorie il faudrait plusieurs milliers d'annotations par classe pour obtenir des résultats satisfaisants. Cependant avec la technique du *transfer learning* on peut se permettre de réduire ce nombre à **quelques centaines par classe** ($\simeq 400$).

Les annotations

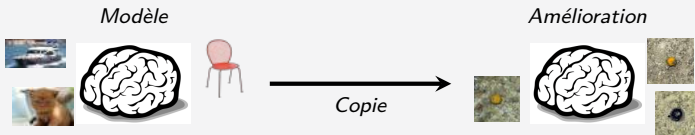
Quel jeu de données

Mais de combien d'annotations a-t-on besoin ?

En théorie il faudrait plusieurs milliers d'annotations par classe pour obtenir des résultats satisfaisants. Cependant avec la technique du *transfer learning* on peut se permettre de réduire ce nombre à **quelques centaines par classe** ($\simeq 400$).

Le *transfer learning*

C'est une technique qui exploite les poids d'apprentissage obtenus sur une autre base de données pour initialiser le réseau de neurones avant de démarrer une nouvelle phase d'apprentissage avec nos annotations.



Les annotations



Quel jeu de données

Les bases d'entraînement et de test :

Lors de la phase d'apprentissage, le réseau de neurones va « étudier » les données d'apprentissage pour régler les hyper-paramètres permettant d'associer les images à leurs labels respectifs.

Cependant, pour tester les performances du réseau il est aussi nécessaire de garder des données que le réseau « ne connaît pas » afin d'étudier son comportement sur des données nouvelles, il s'agit de la base de test.

Les annotations



Quel jeu de données

Les bases d'entraînement et de test :

Lors de la phase d'apprentissage, le réseau de neurones va « étudier » les données d'apprentissage pour régler les hyper-paramètres permettant d'associer les images à leurs labels respectifs.

Cependant, pour tester les performances du réseau il est aussi nécessaire de garder des données que le réseau « ne connaît pas » afin d'étudier son comportement sur des données nouvelles, il s'agit de la base de test.

⇒ **On peut prendre 10 % de la base d'entraînement pour le test.**

Les annotations

Quel jeu de données

Les bases d'entraînement et de test :

Lors de la phase d'apprentissage, le réseau de neurones va « étudier » les données d'apprentissage pour régler les hyper-paramètres permettant d'associer les images à leurs labels respectifs.

Cependant, pour tester les performances du réseau il est aussi nécessaire de garder des données que le réseau « ne connaît pas » afin d'étudier son comportement sur des données nouvelles, il s'agit de la base de test.

⇒ **On peut prendre 10 % de la base d'entraînement pour le test.**

Nombre de classes et répartition :

Pour obtenir des résultats intéressants, il est conseillé :

- De ne pas dépasser 10 classes (au moins pour commencer).
- D'avoir autant d'annotations pour chacune de vos classes.

Les annotations



C'est quoi une annotation ?

C'est quoi une annotation utile ?

Étant donné le temps consacré à l'annotation d'images, il est indispensable de réaliser cette tâche correctement afin que les annotations puissent être utilisées au mieux par la suite !

Les annotations



C'est quoi une annotation ?

C'est quoi une annotation utile ?

Étant donné le temps consacré à l'annotation d'images, il est indispensable de réaliser cette tâche correctement afin que les annotations puissent être utilisées au mieux par la suite !

Quelques règles à suivre :

- Annotez des objets visibles dans leur globalité : évitez les objets tronqués, les objets se chevauchant...
- Annotez toute la variété intrinsèque aux classes.
- Enregistrez la position de l'objet avec les coordonnées de sa **boîte minimale englobante** $x_{center}, y_{center}, width, height$ (par exemple) dans le repère image.
- Annotez des objets de taille 50×50 pixels minimums et si possible supérieurs à 100×100 pixels.
- Utilisez des rectangles alignés sur le repère image comme annotation.

Les annotations

C'est quoi une annotation ?

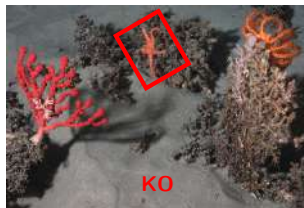
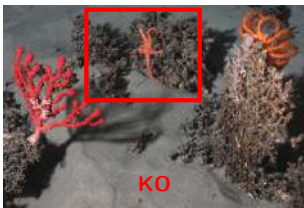


FIGURE 18 – Illustrations du processus d'annotation avec les principales erreurs à éviter.

Le calcul

Quels besoins matériels ?

Pour la phase d'apprentissage (c.-à-d. entraînement du réseau) :

C'est un processus itératif dans lequel chaque image de la base d'entraînement va devoir transiter plusieurs fois dans le réseau afin que celui-ci apprenne les meilleurs hyper-paramètres. Elle est très gourmande en ressource est nécessite un environnement particulier :



FIGURE 19 – Matériels et environnements de développement déporté pour l'entraînement d'un réseau de neurones profond.

Pour la phase d'inférence (c.-à-d. utilisation du réseau) :

C'est un processus dans lequel les images de à traiter ne passe qu'une fois à travers le réseau. Cette étape peu souvent être réalisés sur un ordinateur portable (de préférence avec un GPU *Nvidia*)

Sommaire I



- ① Contexte
- ② Exemples
- ③ L'apprentissage profond
- ④ Les architectures neuronales
- ⑤ Les annotations
- ⑥ La puissance de calcul
- ⑦ Conclusion

Conclusion

Bilan

Même si les approches utilisant l'apprentissage profond sont plus simples à mettre en œuvre, elles s'appuient tout de même sur un ensemble de ressources qu'il faut satisfaire avant de démarrer un nouveau projet.

Besoin	Disponibilité
Données brutes	Très bonne
Architectures neuronales	Bonne
Annotations des données	Mauvaise
Puissance de calcul	Bonne

TABLE 2 – État des lieux des besoins nécessaire pour la mise en place d'une analyse automatique d'images.

La phase d'annotation est cruciale est ne doit pas être négligée.

Avez vous des questions ?

Données utilisées :

- Projet PAMELA
- Projets VISISUB et UNODE

Remerciements :

- Catherine Borremans
- Karine Olu
- Melissa Hanafi-Portier
- Pierre Legeay
- Clémence Carcopino/Camille Dugard
- Alexandre Thomas/Margaux Doudet