

# COMMENT CONCEVOIR DES ARCHITECTURES DE RÉSEAU DE NEURONES PERFORMANTES ET ÉCONOMES ?

Youva Addad, Alexis Lechervy, Frédéric Jurie

Normandie Univ, UNICAEN, ENSICAEN, CNRS, GREYC, Caen, FRANCE

[alexis.lechervy@unicaen.fr](mailto:alexis.lechervy@unicaen.fr)

<https://lechervy.users.greyc.fr>



## Apprendre ? Qu'est ce que c'est ?

- ▶ **Apprendre, c'est s'adapter** à des situations **nouvelles et inconnues** en prenant en compte l'expérience passée.
- ▶ Apprendre est une **propriété humaine essentielle**.
- ▶ Apprendre signifie **s'améliorer afin d'être meilleur**.
- ▶ A ne pas confondre avec la récitation par-cœur. N'importe quel ordinateur peut "apprendre" par-cœur, la difficulté est de généraliser à des situations nouvelles et inconnues.

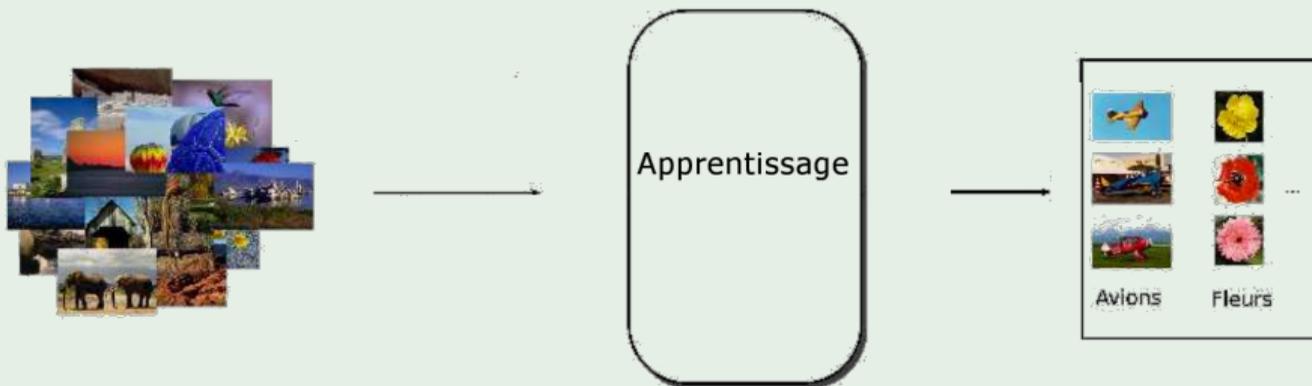
## Mais pourquoi apprendre à un ordinateur ?

- ▶ Pouvoir gérer une quantité de données très importante de manière automatique ou de manière temps réels.
- ▶ Pouvoir effectuer une action dans un contexte non prévu préalablement sans l'intervention d'un humain.
- ▶ Pouvoir prévoir des comportements ou des évolutions pour aider à la prise de décision.

## Principes

- ▶ On détermine automatiquement une règle à partir de données d'apprentissage annotées par un expert,
- ▶ Un expert a défini un ensemble de couples (donnée,label),
- ▶ Il y a un apriori sur les données,
- ▶ Les données entrées sont des couples (données collectées, observations).

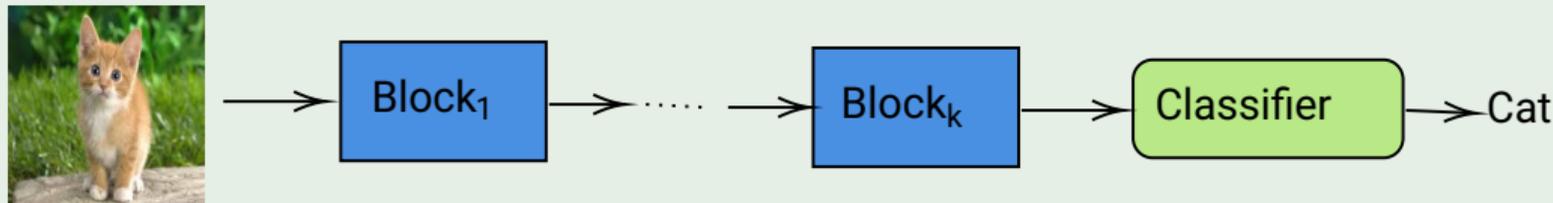
## Exemple: Catégorisation d'image



### Principes

- ▶ Un enchainement successif de blocs de calculs.
- ▶ Chaque bloc est composé de paramètres apprenables.
- ▶ Un réseau profond est composé de nombreux blocs.

### Exemple



# Les limites des approches classique

## Limite 1: Non adaptable à la puissance de calcul disponible

- ▶ Les approches classiques utilisent une puissance de calcul fixe.
- ▶ Les modèles classiques ne peuvent pas s'adapter à un changement de disponibilité des ressources.
- ▶ Il faut apprendre un modèle par puissance de calcul possible.



# Les limites des approches classique

## Limite 2: Non adaptable à la difficulté des exemples

- ▶ Les approches classiques utilisent la même puissance de calcul pour tous les exemples.
- ▶ Il n'est pas possible d'adapter les calculs à la "difficulté" de l'exemple.

### Images de toiles d'araignées

#### Facile



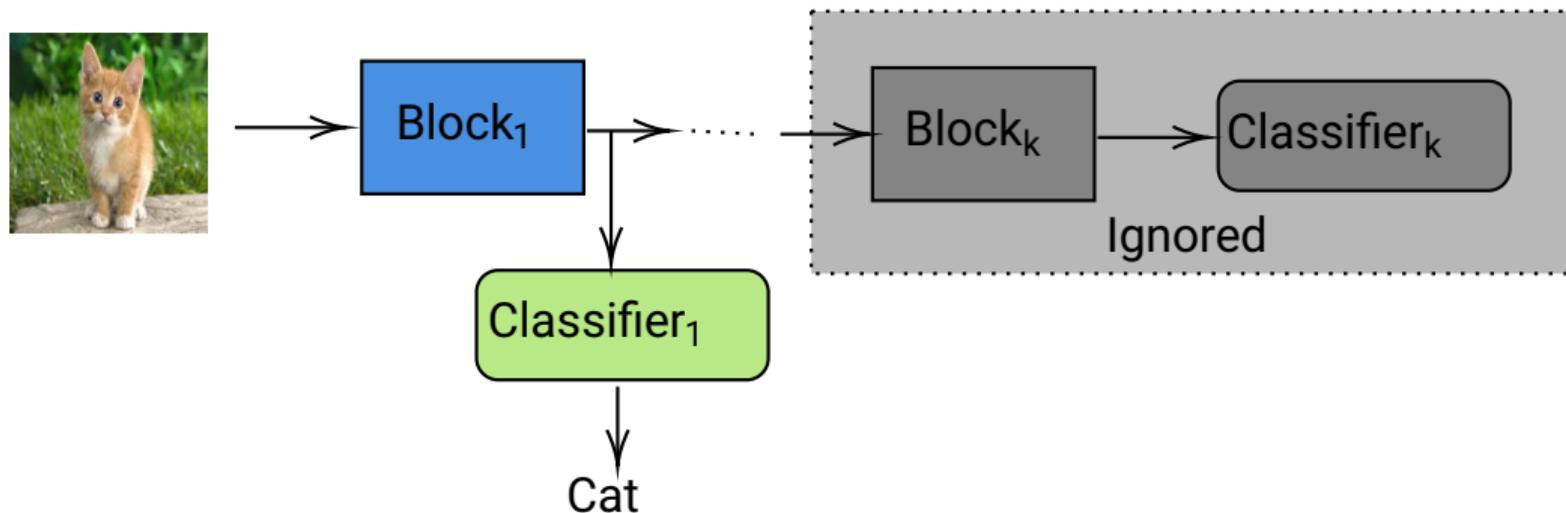
### Images de toiles d'araignées

#### Difficile



# L'apprentissage multi-sorties

Sortie anticipée dynamique:



## Idées

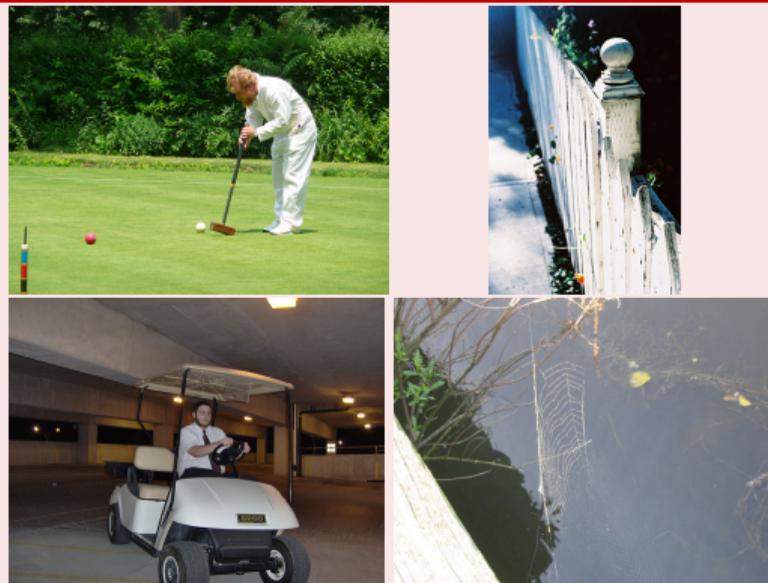
- ▶ Ajouter des sorties anticipées dans l'architecture.
- ▶ Avoir un mécanisme de mesure de l'incertitude de la décision au niveau des sorties.

## Exemples "faciles"



Temps de traitement : 15ms pour 2048 images

## Exemples "difficiles"



Temps de traitement : 218ms pour 2048 images

## Cifar10 / Cifar100

- ▶ 2 nœuds avec 2 GPU P100 par nœuds,
- ▶ 2h pour l'apprentissage du réseau de neurones,
- ▶ 2s pour le traitement de 10.000 images de tailles  $32 \times 32$  de tests,
- ▶ 12 expériences pour trouver les meilleurs paramètres,
- ▶ Total: 24h par base.

## ImageNet

- ▶ 4 nœuds avec 2 GPU P100 par nœuds,
- ▶ 87h pour l'apprentissage du réseau de neurones,
- ▶ 19s pour le traitement de 50.000 images de tailles  $224 \times 224$  de tests,
- ▶ 8 expériences pour trouver les meilleurs paramètres,
- ▶ Total: 24 jours.

- ▶ Développement d'une nouvelle architecture neurale efficace pour la classification d'images.
- ▶ Prise en compte du **compromis Performance/Ressources de calcul**.
- ▶ **Excellente performance** en classification pour un budget de calcul fixé.
- ▶ **Possibilité d'adapter la consommation** de ressources à la difficulté des exemples ou à la puissance de calcul disponible **sans réapprendre le modèle**.

# [Merci pour votre attention]

Avez-vous des questions ?

Pour aller plus loin:

*Multi-Exit Resource-Efficient Neural Architecture for Image Classification with Optimized Fusion Block*, Youva Addad, Alexis Lechervy, Frédéric Jurie ; ICCV 2023 Workshop.



`alexis.lechervy@unicaen.fr`  
`https://lechervy.users.greyc.fr`