



Paris, le 20 septembre 2017

INSTITUT DES ACTUAIRES

Automatisation de la chaîne de confiance lors
de déclaration de sinistre auto.

1/ Présentation : par Imen et Nathalie

- Création du Club Algo, objectifs et présentation du sujet (Nathalie)
- Principes des Réseaux de Neurones et introduction au CNN (Imen)

2/ Application des CNN à la reconnaissance marque et modèle de voiture : par Anne-Claire et Victor

- Présentation de la méthodologie retenue
- Résultats obtenus

3/ Questions

Petit-Déjeuner Club Algo

1/ Présentation du Club Algo : par Imen et Nathalie

- **Création**, objectifs et présentation du sujet (**Nathalie**)
- Principes des Réseaux de Neurones et introduction au CNN (Imen)

2/ Application des CNN à la reconnaissance marque et modèle de voiture : par Anne-Claire et Victor

- Présentation de la méthodologie retenue
- Résultats obtenus

3/ Questions

Présentation du Club Algo

- Lancé en juin 2016, dans la continuité de la formation DSA
- Le Club est un sous-groupe du « GT Big Data », de l'IA, créé en décembre 2013
- Objectif : Échanges et perfectionnement dans le domaine algorithmique pour l'assurance
- En 2016 : Cinq « uses cases »
- En 2017 : Études des algos de deep learning
- 32 membres

1/ Présentation du Club Algo : par Imen et Nathalie

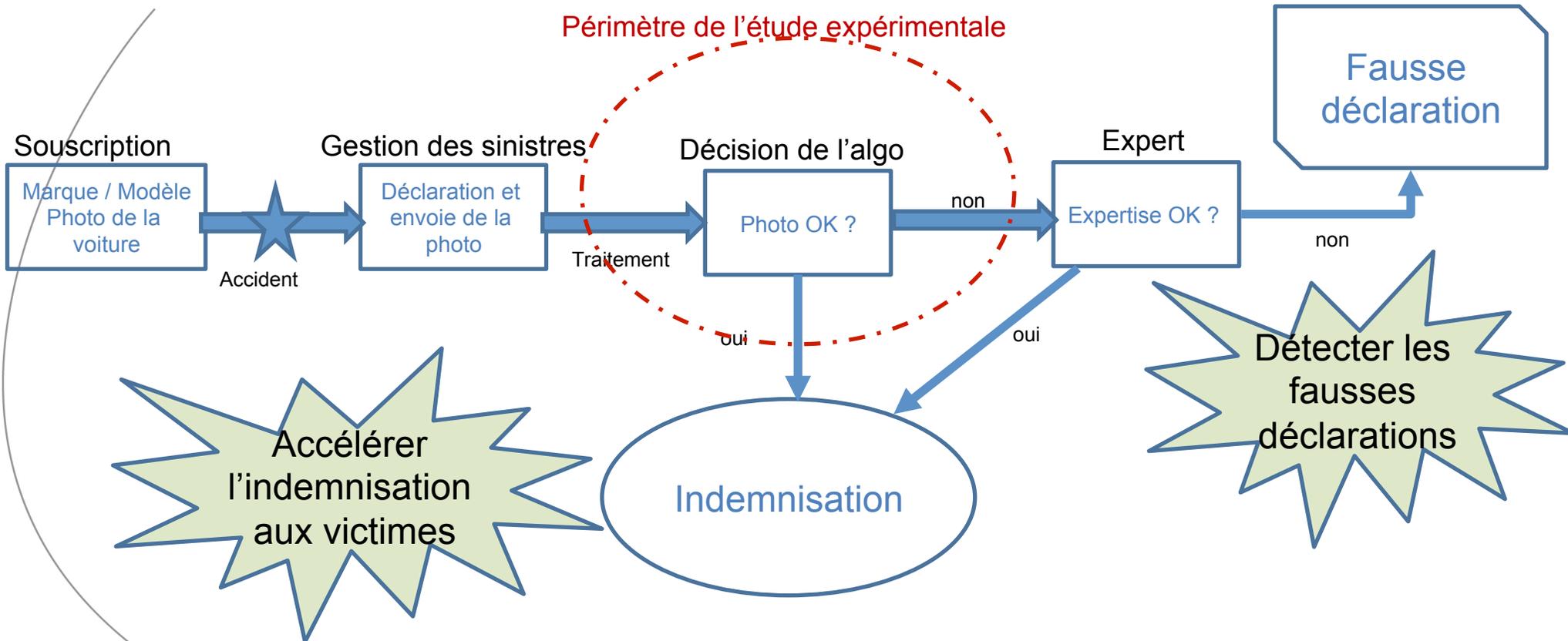
- Création, objectifs et **présentation du sujet (Nathalie)**
- Principes des Réseaux de Neurones et introduction au CNN (Imen)

2/ Application des CNN à la reconnaissance marque et modèle de voiture : par Anne-Claire et Victor

- Présentation de la méthodologie retenue
- Résultats obtenus

3/ Questions

Automatisation de la chaîne de confiance



1/ Présentation : par Imen et Nathalie

- Création du Club Algo, objectifs et présentation du sujet (Nathalie)
- **Principes des Réseaux de Neurones et introduction aux CNN (Imen)**

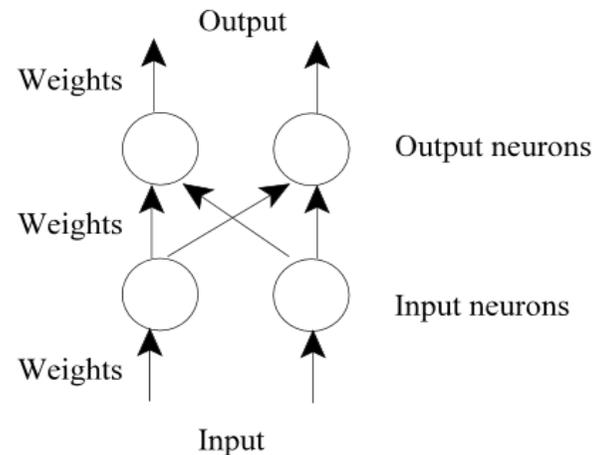
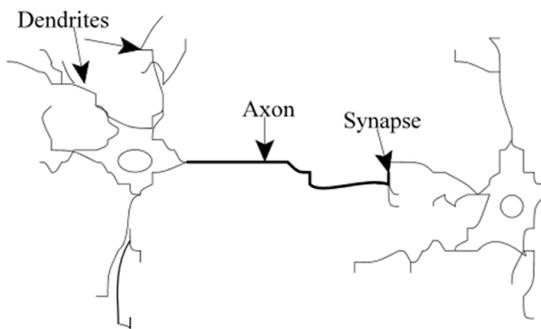
2/ Application des CNN à la reconnaissance marque et modèle de voiture : par Anne-Claire et Victor

- Présentation de la méthodologie
- Résultats obtenus

3/ Questions

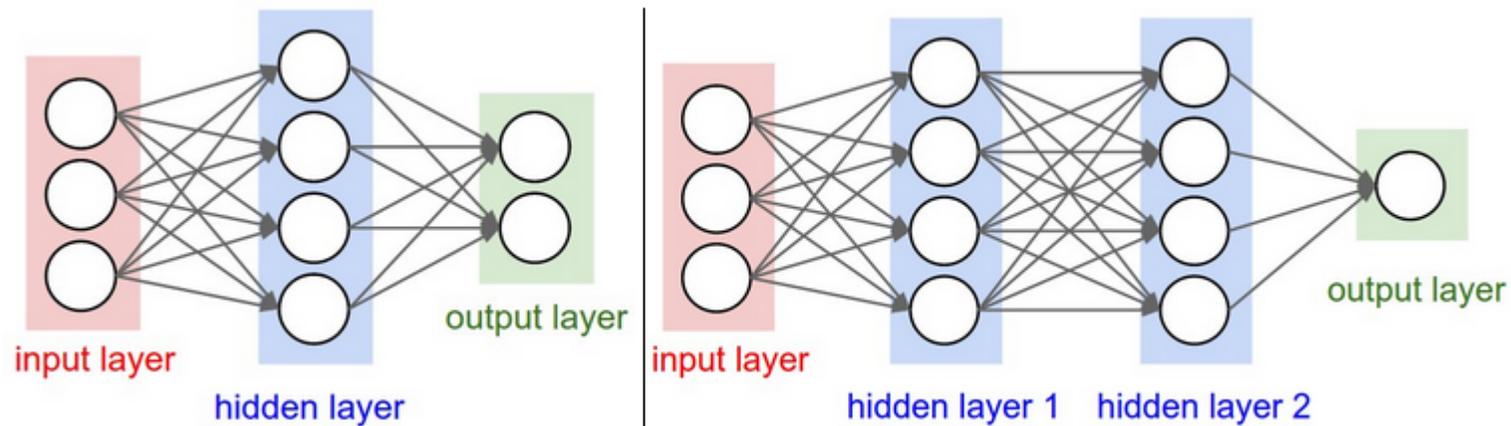
➤ Réseaux de neurones : retour sur les principes

- Les réseaux de neurones artificiels reprennent le même principe :
 - Fonction d'activation
 - Fonction output
- Exemple basique : the backpropagation algorithm (Rumelhart and McClelland, 1986)
- Ajuster les poids pour diminuer l'erreur par la méthode du *gradient descent*



➤ Réseaux de neurones : retour sur les principes

- Fully-connected layer :

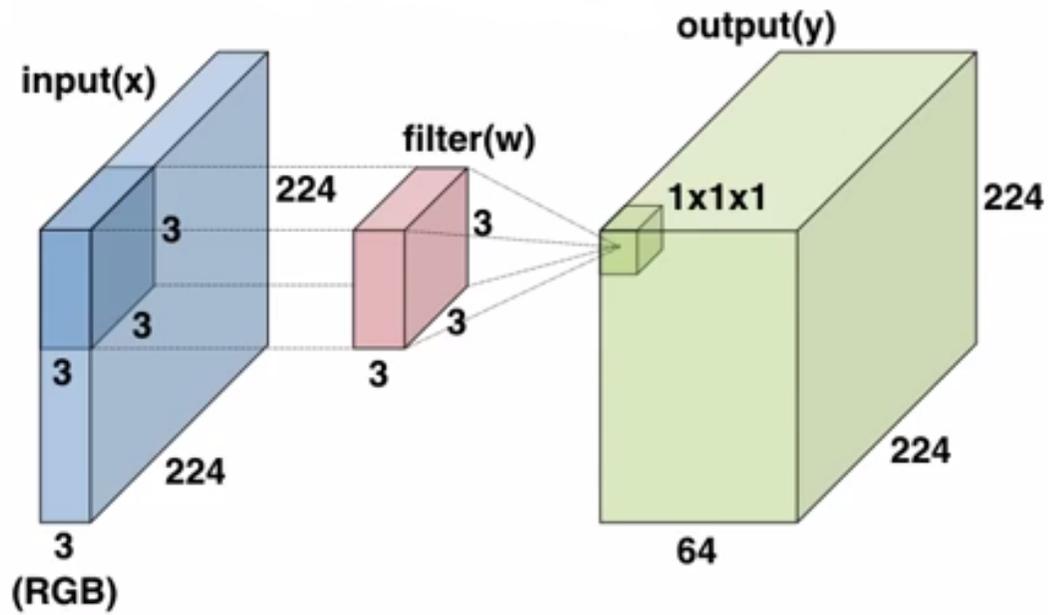


- 26 paramètres pour le premier et 41 pour le deuxième !

➤ Réseaux de neurones : retour sur les principes

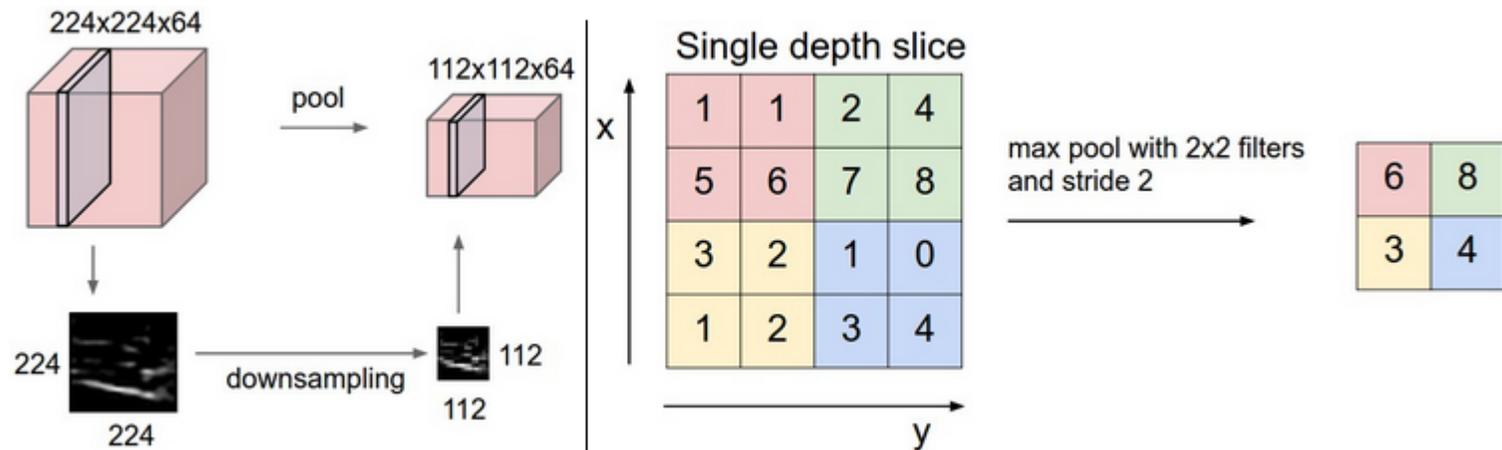
- Défi : comment paramétrer le réseau ?
 - Nombre de layers
 - Taille des layers
- Risque : sur apprentissage
- Réponse classique : diminuer le nombre de layers
- Spécificité des réseaux de neurones : autres méthodes pour traiter le bruit
 - Le dropout : probabilité de récupérer l'output à 50%
 - Permet de changer l'architecture de l'input

➤ **CNNs/ ConvNets :**



➤ Pooling layer :

- Entre deux couches de convolution, effectuer du pooling pour :
 - Réduire la dimensionnalité : downsampling
 - Par conséquent réduire le nombre de paramètres et le risque du surapprentissage
- Le pooling intervient sur les couches et non pas les vecteurs de neurones



1/ Présentation : par Imen et Nathalie

- Création du Club Algo, objectifs et présentation du sujet (Nathalie)
- Principes des Réseaux de Neurones et introduction au CNN (Imen)

2/ **Application des CNN à la reconnaissance marque et modèle de voiture : par Anne-Claire et Victor**

- **Présentation de la méthodologie retenue**
- Résultats obtenus

3/ Questions

Base de données utilisée

- La base de données initiale est constituée de **100,000** photos de voitures, de la marque et du modèle associé
- Seuls les modèles pour lesquels plus de 100 photos étaient présentes dans la base ont été conservés → conservation d'environ **150 modèles de voiture**
- La base sur laquelle ont été lancés les réseaux de neurones est finalement composée de **60,000** observations
- Nous avons récupéré des listes de plusieurs millions de photos pour affiner les détections et classifications

Sourcing des photos

- Récupération de photos publiques provenant de sites de ventes en ligne
 - ✓ La centrale
 - ✓ Le bon coin
 - ✓ watchmycar.fr (site fiabilisant les transactions en demandant des photos sous tous les angles)
- Récupération de données faiblement labellisées :
 - ✓ Utilisation de la technologie Velours pour corriger certains labels
 - ✓ Et pour avoir une bonne orientation des véhicules

Technologie utilisée

- Les calculs ont été réalisés en utilisant le package **Caffe**
- Des tests ont également été effectués en utilisant **TensorFlow** et **Keras**
- Utilisation d'une carte graphique **Titan X** de 2015 avec 2880 cœurs et 12 GB de mémoire vive sur un serveur assez basique (4 cœurs et 8GB de mémoire vive)
- Les apprentissages ont généralement pris **quelques heures**

➤ Méthodologie retenue pour la prédiction des modèles de voiture

Prise en compte la **qualité des photos** dans la prédiction du modèle grâce à deux étapes :

- **Etape 1** : Classification des photos pour distinguer celles sur lesquelles la voiture est **bien visible de l'avant des autres**
 - ✓ 1300 photos labellisées à la main : 1000 bien visibles de l'avant et 300 autres
 - ✓ Utilisation de la technologie **Velours de Fotonower** pour étendre la classification aux 60,000 photos
 - ✓ Cette technologie utilise des descripteurs d'AlexNet (couche pool5) et une classification SVM linéaire. Elle a été développée par Fotonower en 2015 grâce à l'apport de l'expertise de 3 chercheurs de l'ENPC : Matthieu Aubry, Guillaume Obonzinski et Nikos Komodakis
- **Etape 2** : Apprentissage sur **le modèle de la voiture** en distinguant les voitures bien visibles de l'avant des autres
 - ✓ 150 classes de modèle de voitures bien visibles de l'avant
 - ✓ 150 classes des mêmes modèles dans d'autres situations

Calibration sur les modèles de voiture (étape 2)

- Utilisation des poids du **VGG16** pour les premières couches du réseau convolutif
- **Calibration en cascade** des dernières couches :
 - ✓ 1^{ère} calibration **grossière** avec peu de classes et des classes grossières comme les marques de voitures
 - ✓ Calibration plus **précise**, sur les modèles

1/ Présentation : par Imen et Nathalie

- Création du Club Algo, objectifs et présentation du sujet (Nathalie)
- Principes des Réseaux de Neurones et introduction au CNN (Imen)

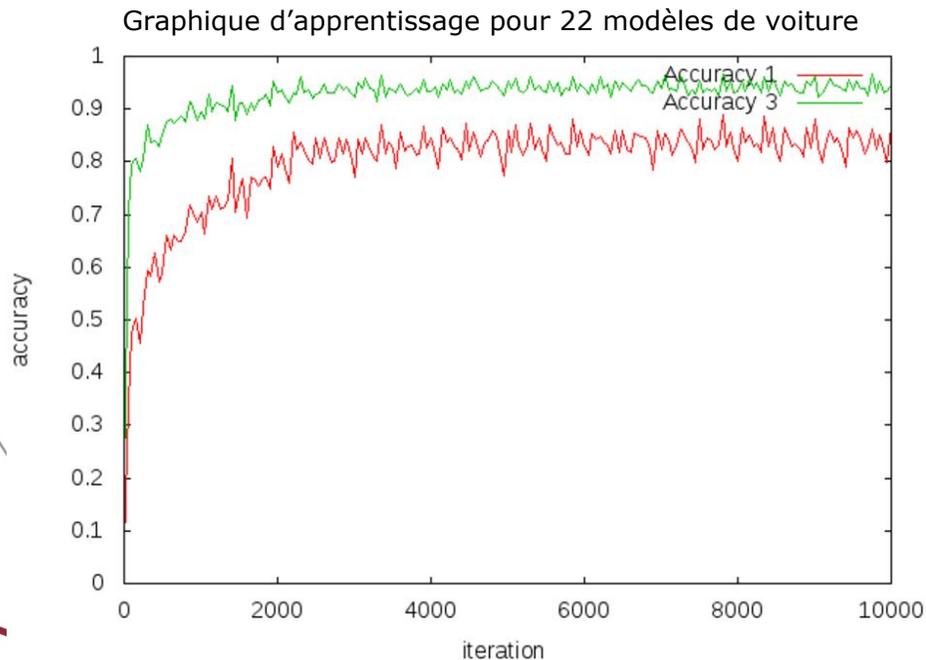
2/ Application des CNN à la reconnaissance marque et modèle de voiture : par Anne-Claire et Victor

- Présentation de la méthodologie retenue
- **Résultats obtenus**

3/ Questions

Résultats obtenus sur l'ensemble de la base

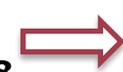
- La précision est globalement de **75%** pour le **1^{er} choix** et de **90%** si on tolère de considérer les **3 premiers choix** proposés par le modèle



Exemples de résultats



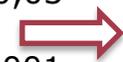
Xsara : 0,40
307 : 0,14
C4 : 0,13
206 : 0,08
Laguna : 0,04



**Mauvaise
prédiction**



Laguna : 0,96
Megane : 0,03
Clio : 0,01
Scenic : 0,001
Autres : 0,0001



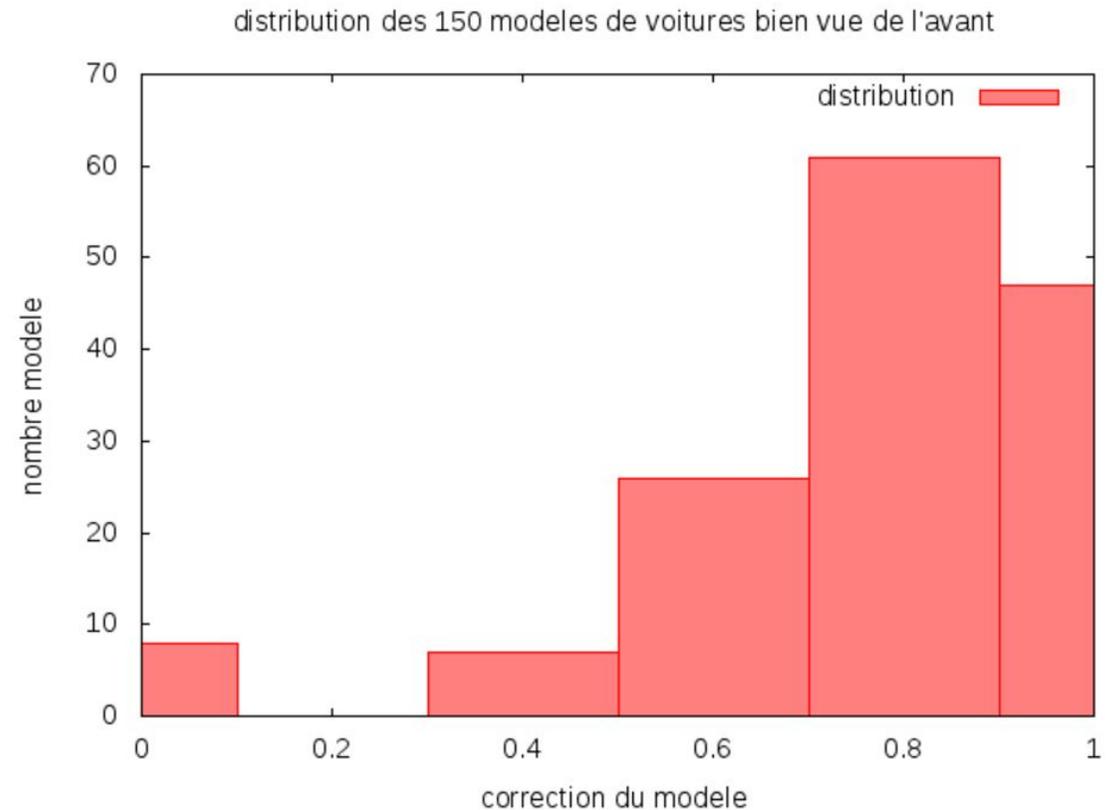
**Bonne
prédiction**

Résultats inégaux entre les deux classes de véhicules

- La précision sur le 1^{er} choix est de :

✓ **86%** pour les véhicules bien visibles de face

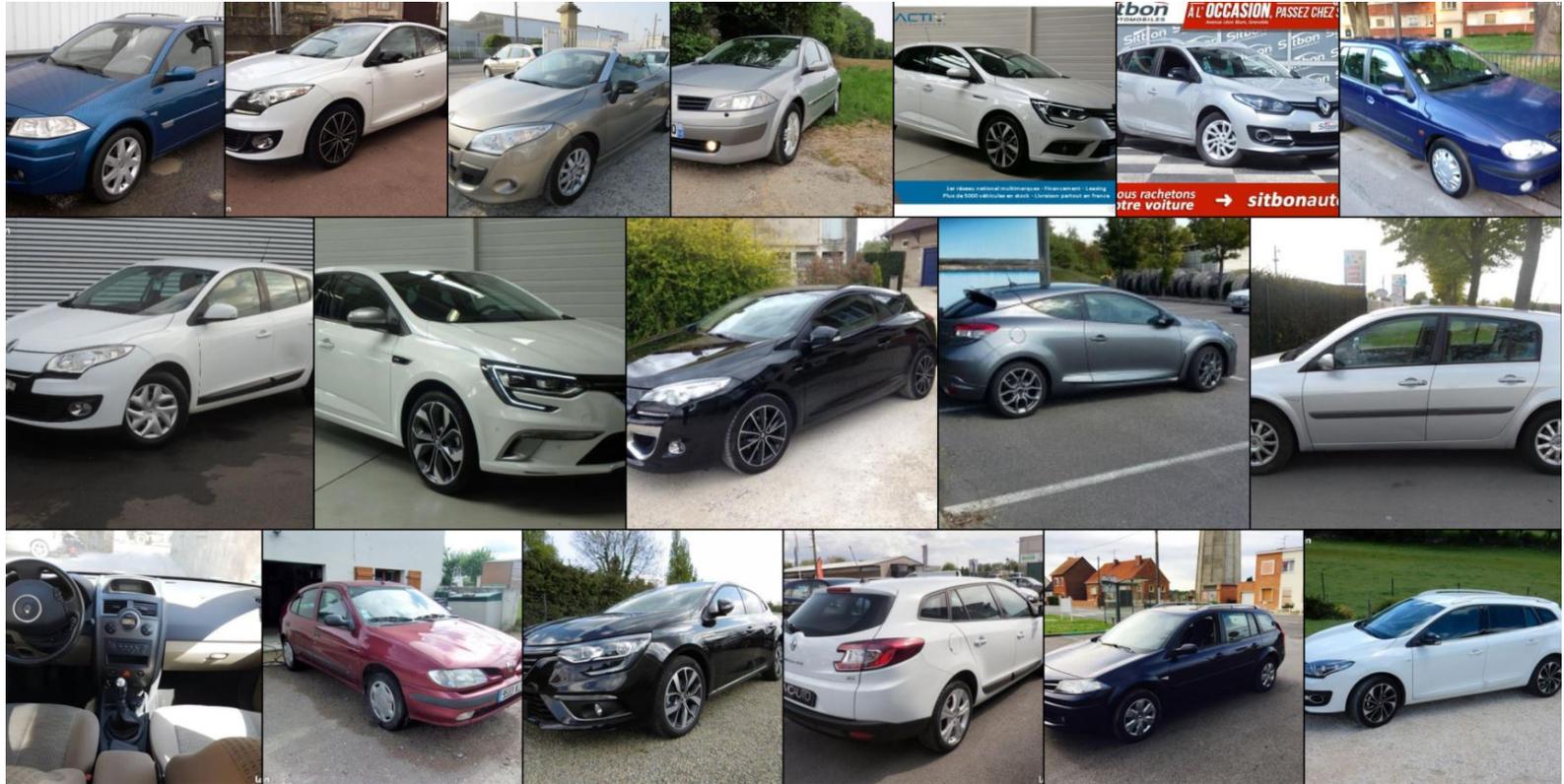
✓ **46%** pour les autres véhicules



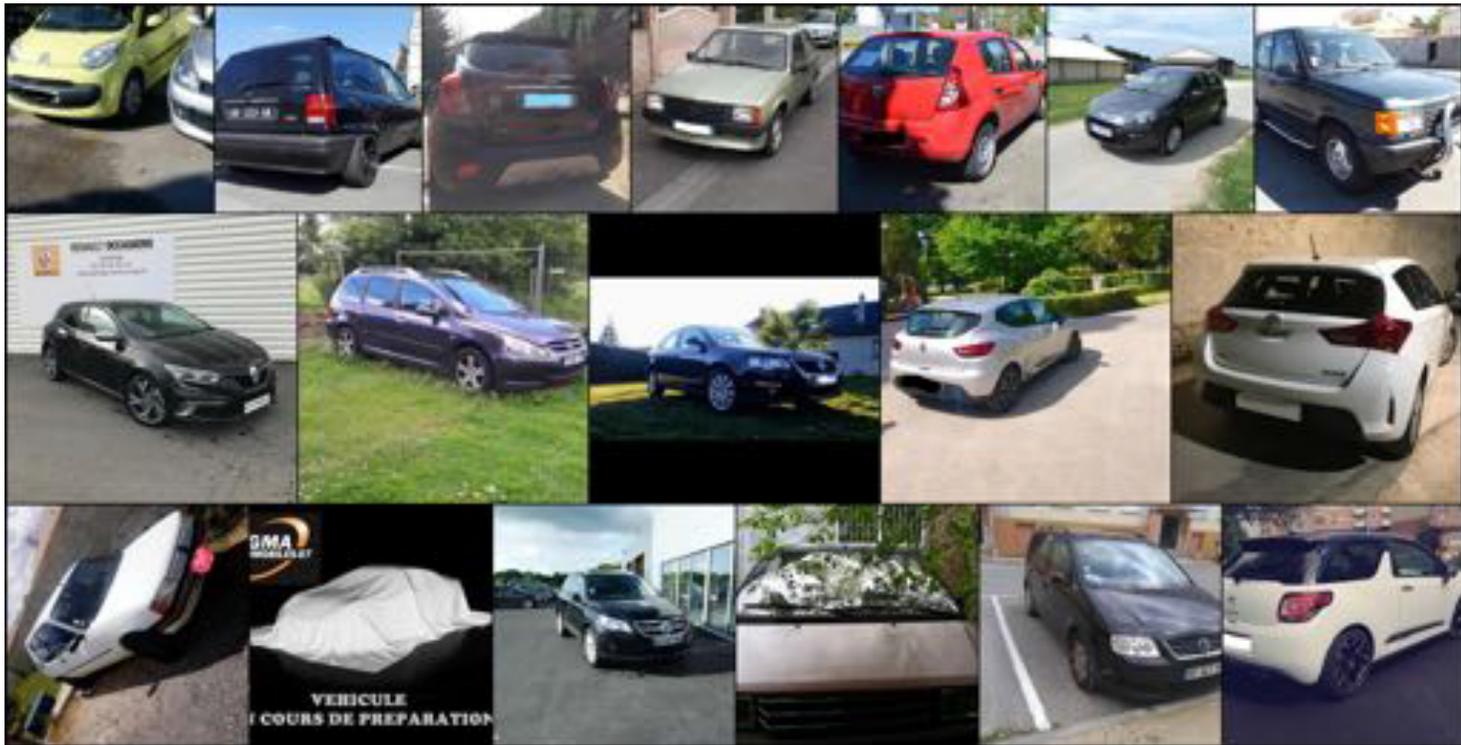
Exemples de résultats : 1^{er} choix de détection correct



Exemples de résultats : bon choix dans les 3 premières propositions



Exemples de résultats : aucune des 3 premières propositions n'est correcte



Exemple : autonower.site

QUESTIONS



Merci de votre attention !

Merci à Dataiku !

Pour nous contacter :

club-algo@institutdesactulaires.com

LinkedIn

