

# Indexation automatique d'images

RAPPORT

# Problématique

Nous travaillons pour une association de protection des animaux qui souhaitent connaître la race de ses pensionnaires canins.

Cette association ne dispose pas d'une base de données de photos de chiens déjà classifiée par races. Elle doit donc faire appel à une autre source de données pour construire son outil de reconnaissance de race.

Heureusement, une base d'images classées par race existe déjà en opendata : il s'agit **du Stanford Dogs Dataset**.

Cette base d'images labellisées va nous permettre d'entraîner un modèle de machine learning.

# Démarche proposée

---

Nous proposons à cette association d'explorer 2 modèles de classification automatique et d'en comparer les résultats :

**Un premier modèle, utilisera l'algorithme scale-invariant feature transform (SIFT)**, publié en 1999. Cet algorithme permet d'extraire de chaque image des caractéristiques visuelles spécifiques, puis d'utiliser ces caractéristiques comme features pour entraîner un modèle.

**Un deuxième modèle utilisera un réseau de neurones multi-couches (deep learning)**. Pour cette modélisation, le nombre d'images de chiens fournies par le Stanford Dogs Dataset n'est pas suffisant pour entraîner un réseau de neurones multi-couches, il faudrait disposer de centaines de milliers d'images pour obtenir un résultat utilisable. Par ailleurs l'entraînement d'un tel modèle nécessiterait un temps de calcul et des ressources de calculs inabordables pour cette association.

Heureusement des réseaux de neurones pré-entraînés sur des images de toutes sortes, pas spécifiquement canines, sont désormais disponibles et peuvent être adaptés à nos besoins spécifiques grâce à la méthode de Transfer Learning.

C'est donc la démarche que nous allons appliquer : nous utiliserons le réseau de neurones VGG16 et lui appliquerons une technique de transfert learning en utilisant les images de chiens labellisées fournies par le Stanford Dogs Dataset.

Exploration des images disponible

---

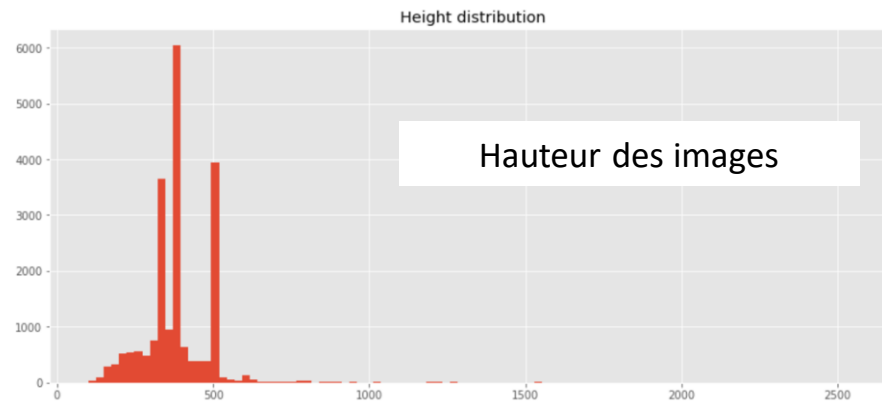
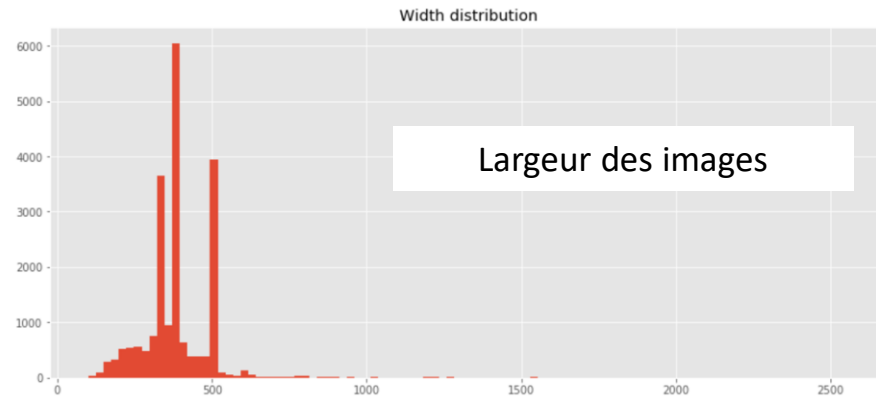
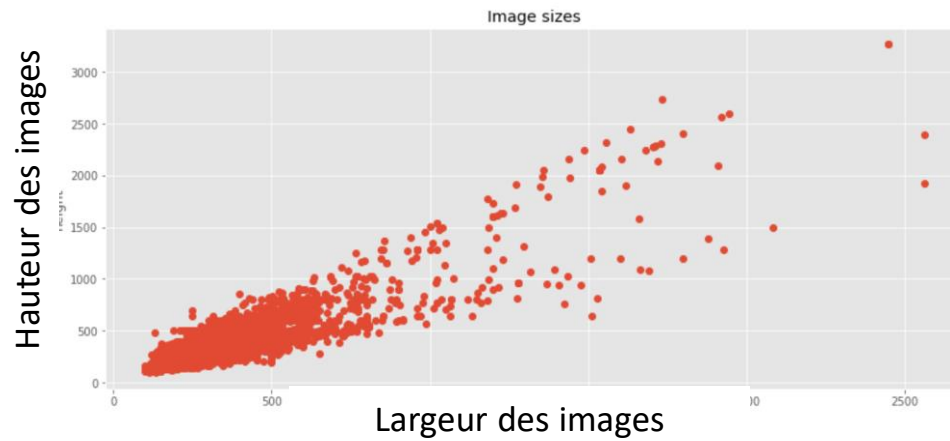
# Stanford Dogs Dataset

- 120 races de chien
- ~150 images par race
- 20,580 images

des chiens en toutes  
positions,  
sur des arrière plans de toutes  
sortes, parfois en presence  
d'autres êtres vivants.



# Des images de taille hétérogène



# Pré-processing des images

---



# Pre-processing des images

---

La pratique a démontré qu'un pré-traitement des images facilitait dans certains cas l'exploitation de l'image par un algorithme de machine learning.

Nous utiliserons un de ces pré-traitement : le **filtre gaussien** et nous comparerons les performances de notre modèle avec ou sans applications de ce filtre.

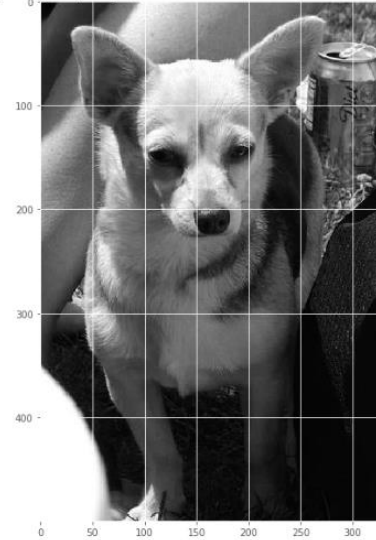
Cas particulier de l'algorithme SIFT : l'algorithme SIFT ne sait traiter que des images en niveaux de gris. Nous appliquerons un **filtre de niveau de gris**



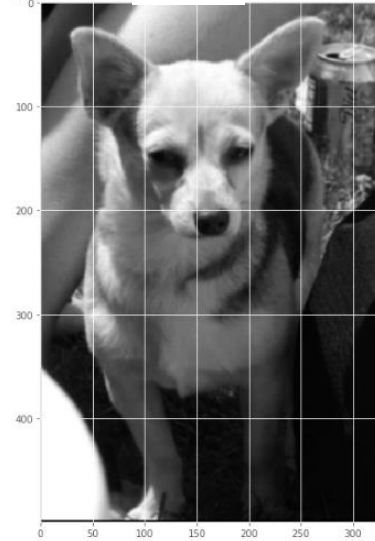
# Kernel gaussien

Image niveaux de gris

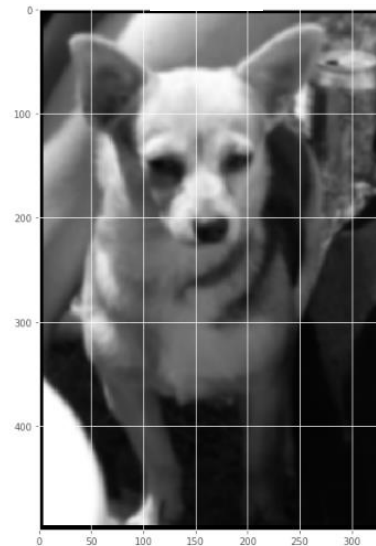
Original : ../data/Images/n02085620-Chihuahua/n02085620\_10074.jpg



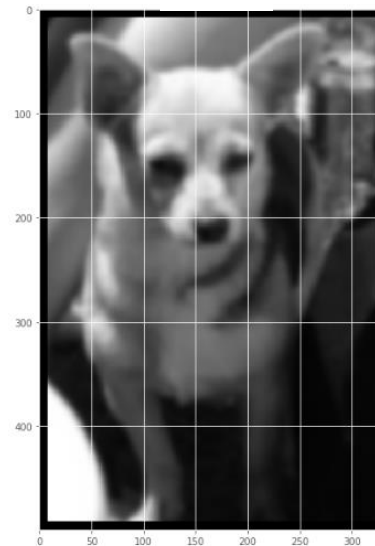
K=1



K=2



K=3



# Modélisation SIFT

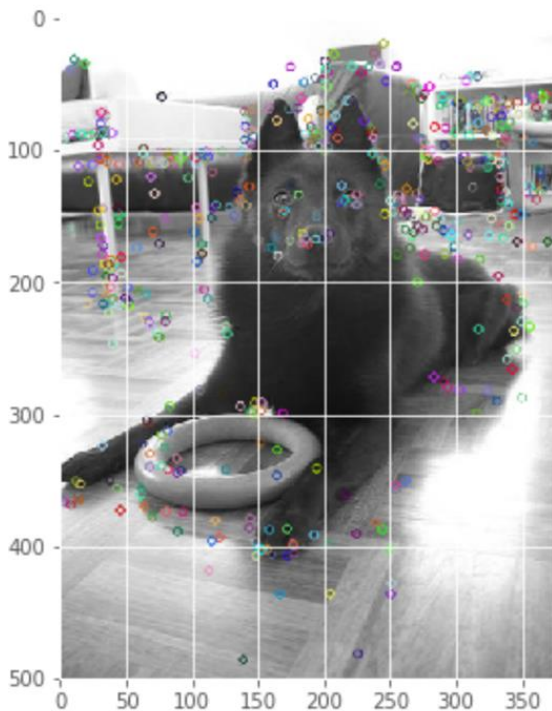
---

# Présentation de l'algorithme SIFT

Le principe de l'algorithme SIFT est d'identifier sur chaque image des points marquants, figurés par un cercle coloré sur l'image ci-contre.

Ces points peuvent être marquants pour différentes raisons, par exemple :

- Changement de couleur rapide entre un groupe de pixels et ses voisins,
  - Forme présentant un angle
  - Forme arrondie ...
- 
- Chacun de ces points marquants est décrit par 150 caractéristiques (intensité de couleur, angle de variation ...)
  - Des images avec des points marquants ou « descripteurs » similaires ont des chances de faire partie d'une même classe de chiens.



# Etapes

Regroupement des descripteurs en clusters

Construction des features : fréquence des descripteurs de chaque cluster dans l'image

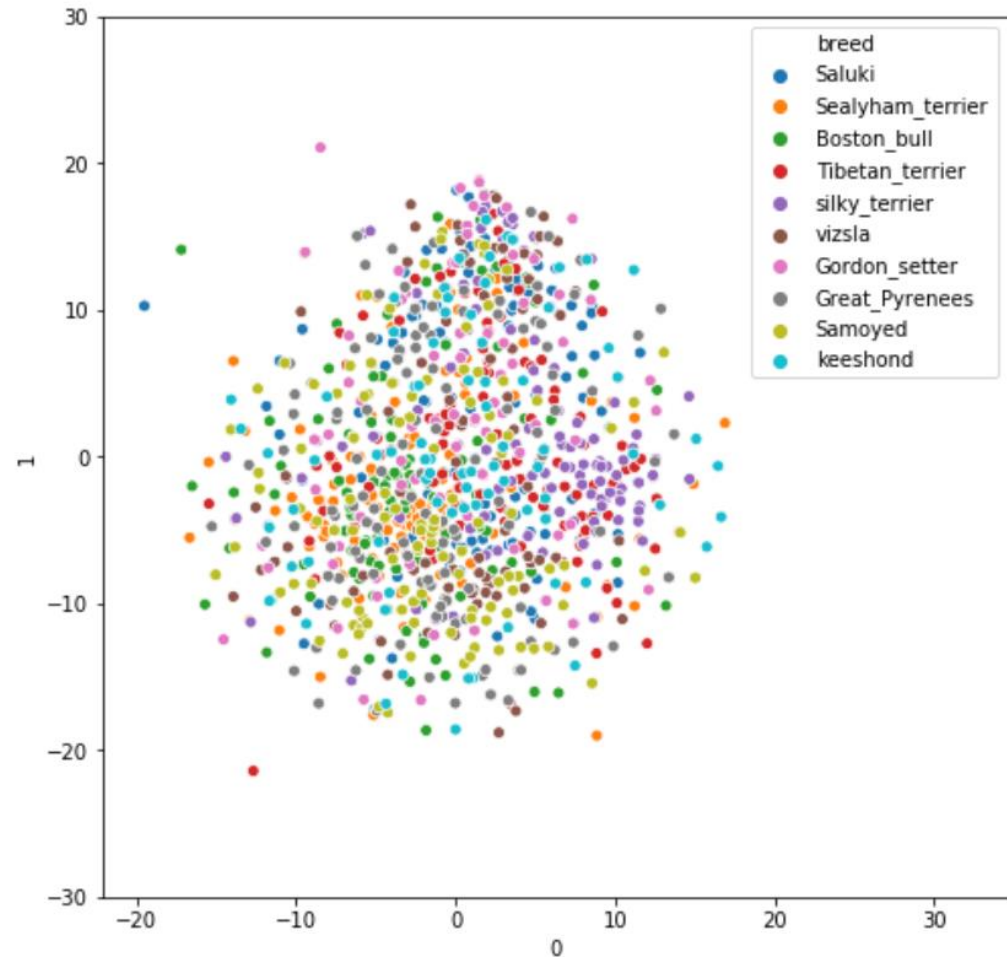
Visualisation T-SNE pour un aperçu visuel des opportunités de classification

Classification

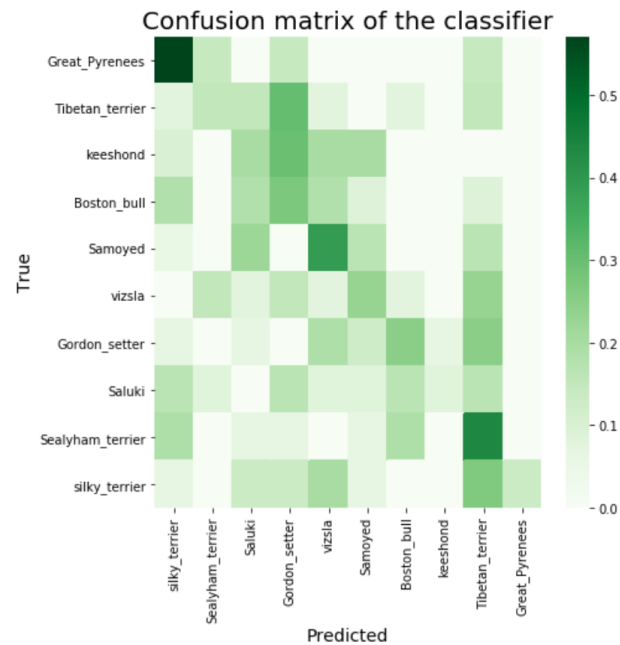
## Clustering des descripteurs

- Le nombre de points marquants étant très élevé, on cherche à les regrouper en ensembles cohérents. Cette opération est réalisée grâce à un algorithme de clustering.
- Nous avons utilisé un algorithme **Kmeans** et fixé **le nombre de clusters à 1000**.
- Pour une image donnée, on compte ensuite le nombre de descripteurs de chaque clusters présents dans l'image
- On dispose ainsi de 1000 features. Pour une image donnée, la feature n contient la quantité de descripteurs de l'image rattachés au cluster n
- Ces features nous permettent d'entraîner un modèle de classification.
- Nous avons utilisé un algorithme de classification **Random Forest**, dont les paramètres ont été optimisés à l'aide de GridSearchCV

# Visualisation R-SNE



# Classification par Random Forest : Performances



	precision	recall	f1-score	support
0	0.25	0.57	0.35	7
1	0.33	0.15	0.21	13
2	0.13	0.20	0.16	10
3	0.17	0.27	0.21	11
4	0.35	0.39	0.37	18
5	0.21	0.23	0.22	13
6	0.36	0.25	0.30	16
7	0.50	0.08	0.14	12
8	0.26	0.44	0.33	16
9	1.00	0.13	0.24	15
micro avg	0.27	0.27	0.27	131
macro avg	0.36	0.27	0.25	131
weighted avg	0.38	0.27	0.26	131



# Transfer Learning

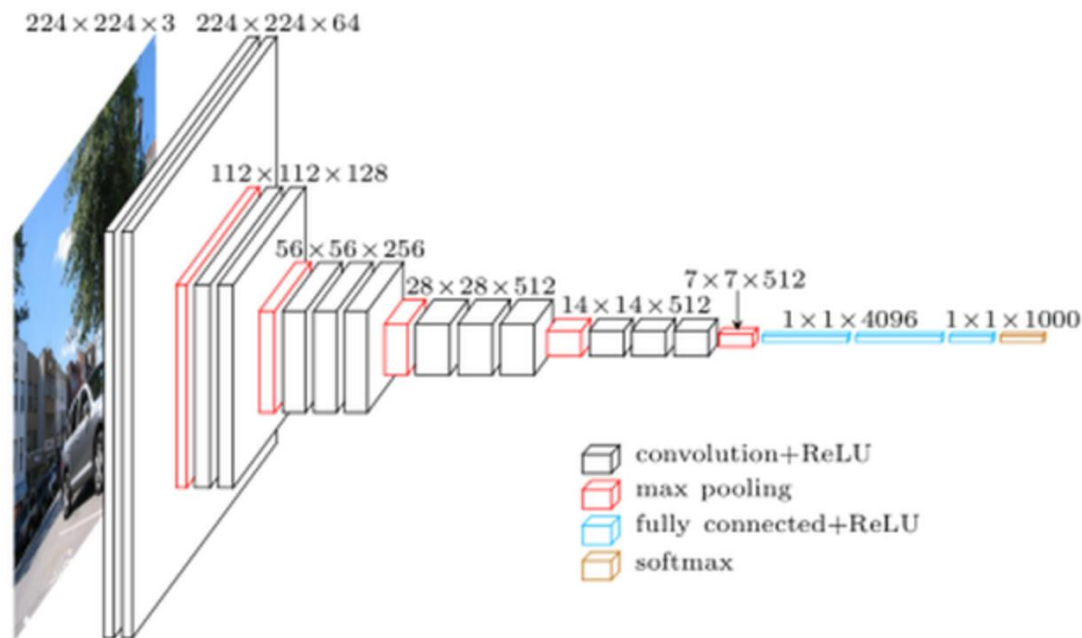
---

# Choix d'un modèle pré-entraîné

---

Nous avons utilisé le modèle VGG16.

VGG-16 est constitué de plusieurs couches, dont 13 couches de convolution et 3 *fully-connected*



Représentation 3D de l'architecture de VGG-16

# Stratégie de transfert learning

---

Nous avons choisi **d'extraire les features du réseau pré-entraîné VGG16 en re-entraînant seulement la dernière couche (fully\_connected) , softmax.**

Cette stratégie nous a semblé possible car le nombre d'image à entraîner est faible par rapport au nombre d'images total à partir duquel VGG16 a été entraîné.

Nous avons donc remplacé la dernière couche par :

- une nouvelle couche dense
- avec activation softmax pour une classification
- initialisée aléatoirement.

Pour des raisons de capacités machines, nous n'avons classifié que 20 races de chiens.

# Preprocessing

---

Le preprocessing inclut les transformations suivantes :

- Re-dimensionnement des images au format 224 x 224
- Fonction `preprocess_input` fournie par Kears spécifiquement pour VGG16

# Hyper paramètres

---

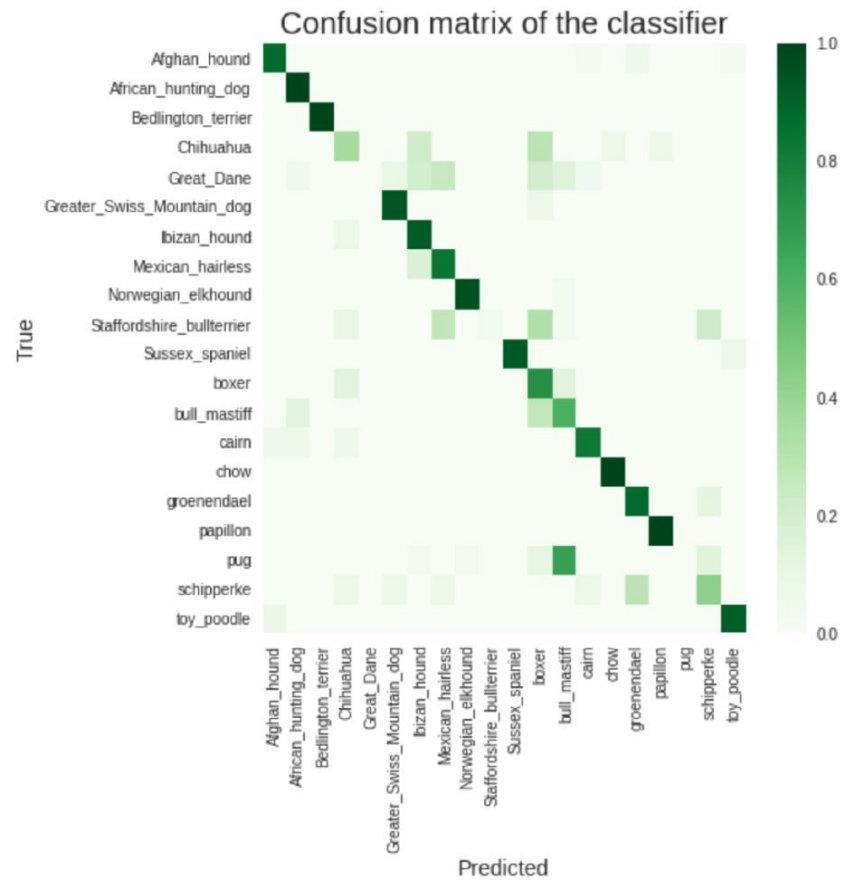
Le modèle a été entraîné avec les hyper paramètres suivants :

```
loss="categorical_crossentropy", optimizer=optimizers.SGD(lr=0.0001,  
momentum=0.9), metrics=["accuracy"]
```

# Performances

---

	precision	recall	f1-score	support
Afghan_hound	0.94	0.88	0.91	33
African_hunting_dog	0.79	1.00	0.88	15
Bedlington_terrier	1.00	1.00	1.00	16
Chihuahua	0.42	0.36	0.38	14
Great_Dane	0.00	0.00	0.00	20
Greater_Swiss_Mountain_dog	0.83	0.94	0.88	16
Ibizan_hound	0.57	0.92	0.71	13
Mexican_hairless	0.29	0.83	0.43	6
Norwegian_elkhound	0.96	0.96	0.96	24
Staffordshire_bullterrier	1.00	0.05	0.09	22
Sussex_spaniel	1.00	0.93	0.97	15
boxer	0.32	0.73	0.45	15
bull_mastiff	0.26	0.60	0.37	15
cairn	0.82	0.82	0.82	17
chow	0.95	1.00	0.98	20
groenendael	0.71	0.88	0.79	17
papillon	0.94	1.00	0.97	15
pug	0.00	0.00	0.00	27
schipperke	0.35	0.43	0.39	14
toy_poodle	0.85	0.92	0.88	12
micro avg	0.68	0.68	0.68	346
macro avg	0.65	0.71	0.64	346
weighted avg	0.67	0.68	0.63	346



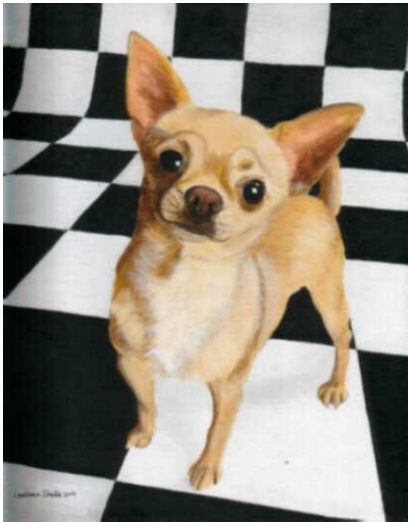


# Races mal identifiées

---

## Races mal identifiées

Chiouaoua



## Confusions de races

Pug confondu avec le bullmastif



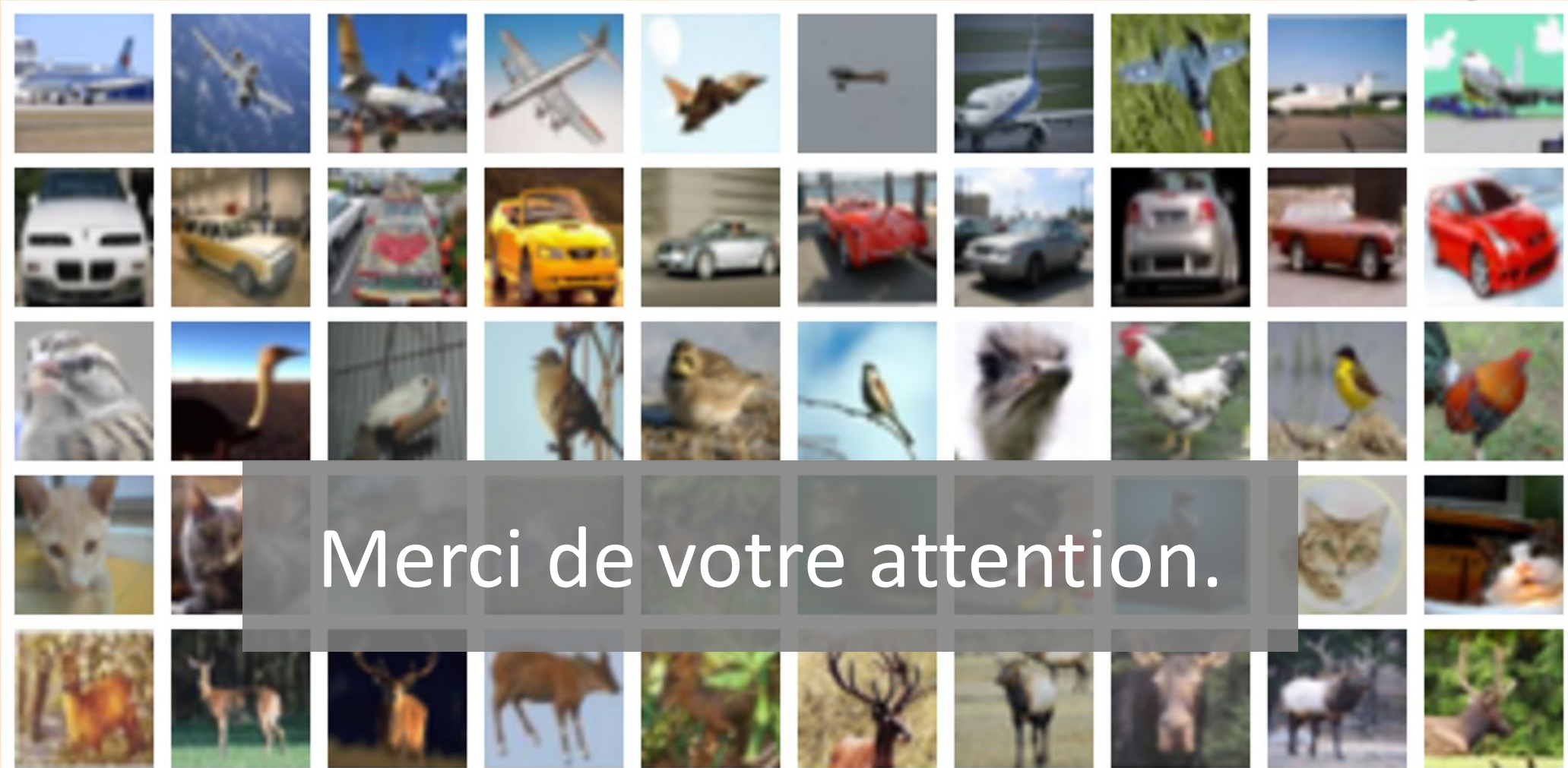
# Conclusions

Le Transfer Learning est une méthode très puissante qui permet de ré-utiliser les acquis de travaux qui ont nécessité des mois de collecte de données et d'entraînement.

**Le classifieur obtenu ici a des performances intéressantes qui ont pu être obtenu dans un environnement gratuit :** Google Colab avec 11 Go de GPU.

Avec des ressources supérieures, le modèle pourrait être amélioré, et étendu à l'ensemble des races du Stanford Dog Dataset.

Il pourrait être intéressant aussi de **Ré-entraîner les 3 dernières couches** denses du VGG16 avant la couche softmax.



contact : [brigitte.maillere@Machine-Learn.it](mailto:brigitte.maillere@Machine-Learn.it)