Université de Rennes 1
ESIR2-IN - module ACI

Utilisation de CNN pré-entrainés pour la classification d'images

TRANSFER LEARNING

Le but de ce TP est d'utiliser un CNN pré-entrainé sur ImageNet pour une tache de classification d'image autre. C'est ce qu'on appelle le *Transfer Learning*. Deux façons de faire cela sont présentées dans ce TP. Il en existe d'autres.

- Méthode 1 : modification de la couche de classification uniquement (sans modification de la description). Solution intégrée (*end-to-end*).
- Méthode 2 : utilisation du CNN comme extraction de descripteurs, puis utilisation d'un classifieur quelconque.

La librairie utilisée pour manipuler les réseaux de neurones est `Keras`.
Le CNN utilisé dans ce TP s'appelle **VGG16**. Voir `https://arxiv.org/pdf/1409.1556.pdf` pour l'article introduisant le réseau.

1 Données et tache

La tache consiste à classifier si une image contient un chat ou un chien.

1.1 Données

Le jeu de données annoté est issu de : `https://www.kaggle.com/c/dogs-vs-cats/overview`, qui contient 25,000 images de chiens et de chats. Vous travaillerez sur un sous-ensemble de ces données qui est dans `cats_and_dogs_sampled`.
Regardez la façon particulière dont sont organisées les données.

1.2 Mise en place d'une chaîne d'apprentissage basique

Rappel général : un processus d'apprentissage comprend les étapes suivantes :

1. Analyse des données : identifier la quantité d'instances, le nombre et la nature des attributs, l'attribut que l'on souhaite prédire (classe). Éventuellement, sélectionner ou transformer les attributs.
2. Choix d'un classifieur : quel classifieur est le plus adapté en fonction du type et de la quantité de données ?
3. Choix de la stratégie d'évaluation : quelles sont les données d'apprentissage ? de test ? Quelles mesures d'évaluation utilise-t-on ? Faut-il prévoir des données de validation ?

4. Apprentissage : choix des hyper-paramètres, apprentissage sur l'ensemble d'apprentissage et, le cas échéant, optimisation sur l'ensemble de validation.

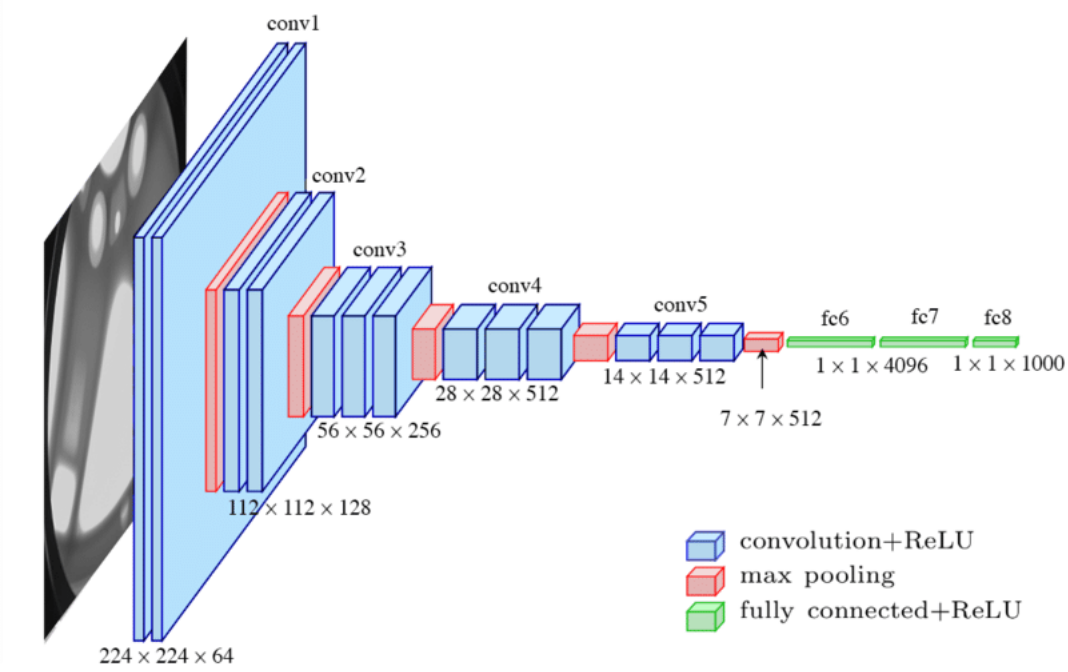
5. Évaluation du classifieur sur l'ensemble de test : quelles sont les performances du classifieur obtenu ? Analyse des résultats.

2 Exercice 1 : manipulation d'un CNN pré-entrainé

Le modèle VGG16, ainsi que les poids résultants de l'apprentissage sur ImageNet (1000 classes), sont disponibles sous Keras : `https://keras.io/api/applications/vgg/`. D'autres modèles sont également disponibles.

Objectifs et savoir-faire :

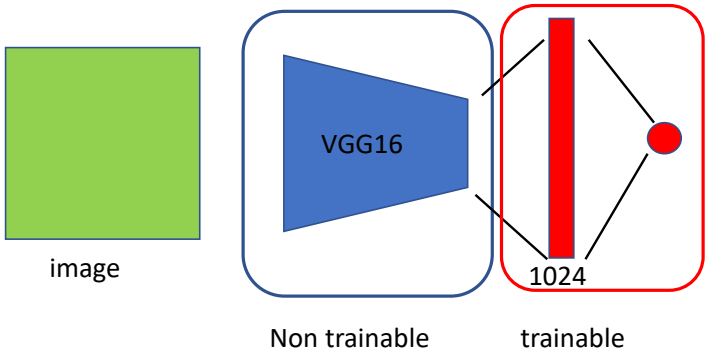
- charger un réseau pré-entrainé et analyser sa structure (architecture)
- mettre une image en entrée (chargement, pré-traitements) et récupérer la sortie (*forward propagation*)
- suppression des couches de classification (`top_layers`)
- accéder aux différentes couches du réseau : faire le lien avec l'architecture du réseau (nom des couches) et comprendre la dimension de l'objet récupéré (tenseur, vecteur...)
- transformation tenseur -> vecteur par `GobalMaxPooling`



3 Exercice 2 : méthode 1

Pour utiliser VGG16 pour la tache de classification chat/chien, on modifie ici uniquement les couches de classification, à l'origine pour les 1000 classes de ImageNet, qu'on remplace par des couches adaptées à notre tache.

On "gèle" ici le modèle : cela signifie qu'on ne va pas modifier (*fine-tuning*) les poids du réseau VGG16, mais uniquement ceux des nouvelles couches lors de l'apprentissage.



Pour l'apprentissage, on utilise les fonctionnalités de Keras permettant d'exploiter l'organisation particulière des données (répertoires) : `flow_from_directory method`.

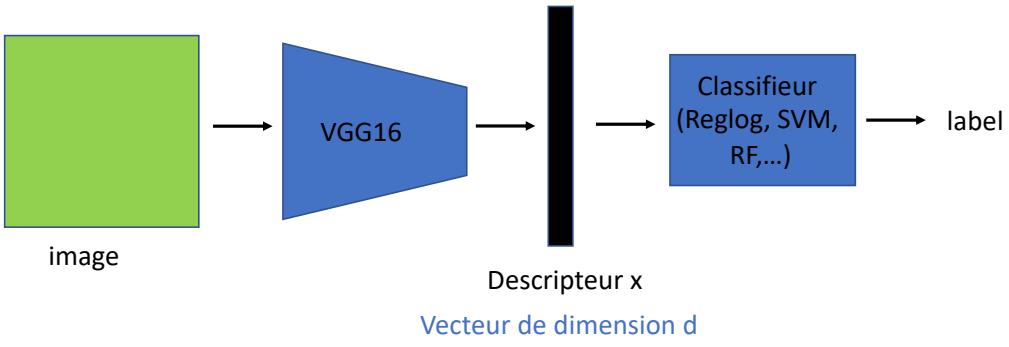
Objectifs et savoir-faire :

- Savoir modifier et entrainer les couches de classification
- Comprendre les paramètres : taille des mini-batches, nombre d'epochs, d'itérations

ATTENTION : le temps d'apprentissage sur CPU peut être très long. Vous pouvez modifier les paramètres pour ne faire que quelques itérations.

4 Exercices 3 et 4 : méthode 2

On utilise ici le CNN comme extraction de descripteur des images. Les descripteurs obtenus sont ensuite utilisés en entrée d'une méthode de classification. La méthode de classification utilisée ici est la régression logistique. Il y a donc 2 étapes dans cette approche.



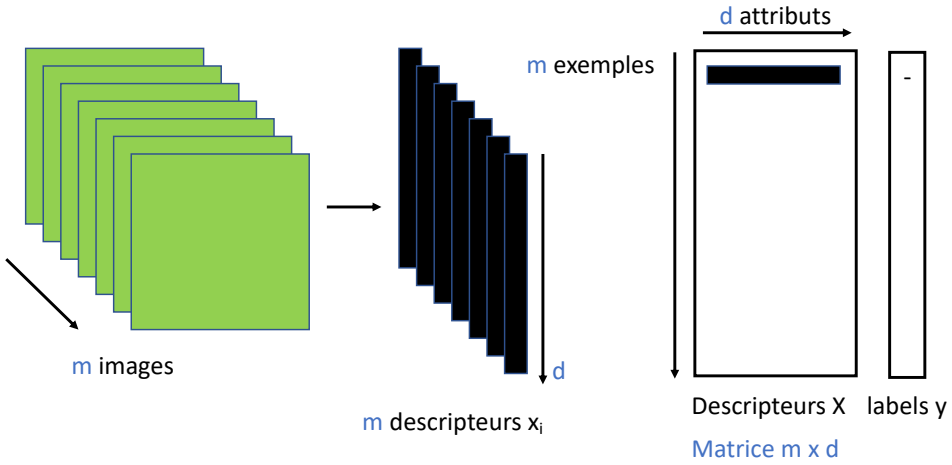
Objectifs et savoir-faire :

- Mise en oeuvre d'un processus d'apprentissage (voir 1.2)
- Utilisation d'un CNN comme extraction de descripteur
- Utilisation de scikit-learn : préparation des donnés (formatage), apprentissage, prediction et evaluation

4.1 Etape 1 : extraction des descripteurs (exo 3)

Cette étape nécessite d'extraire et de **sauvegarder** les descripteurs de toutes les images de la base de données, afin de pouvoir les utiliser dans l'étape 2. Il s'agit de faire passer toutes les images dans le CNN après avoir choisi la couche que l'on souhaite extraire.

Il faut également formater correctement ces données pour l'utilisation ultérieure. En général, cela consiste à stocker les descripteurs de tous les exemples dans une matrice $\boldsymbol{X}$ et les labels dans un vecteur $\boldsymbol{y}$.



4.2 Etape 2 : apprentissage du classifieur (exo 4)

Une fois les descripteurs mis en forme, ils sont mis en entrée d'un classifieur. Pour la tache d'apprentissage, on utilise la librairie python `scikit-learn` : `http://scikit-learn.org/stable/index.html`. La documentation est disponible en ligne sur le site : `http://scikit-learn.org/stable/documentation.html`.

Voir en particulier :

- User guide
- API

Rappels : (voir TP sur la regression logistique)

- hyper-paramètres de la régression logistique

- méthode de selection des meilleurs hyper-paramètres avec ensemble de validation
- spécificité `scikit-learn` = GridSearchCV : principe + cross-validation

Remarques :

- On pourrait aussi utiliser un SVM (ou tout autre classifieur) plutôt que la regression logistique. Le code pour les SVMs vous est fourni, il pourra vous être utile pour la suite du module, mais vous ne devez pas le faire tourner pour ce TP.
- De même dans le code, la **normalisation** des données d'entrée, qui peut être primordiale en fonction de vos données et du classifieur, est implémentée en commentaires . Cette étape est facultative à ce stade, mais je vous invite à comprendre en quoi elle consiste et à la tester.