

Object recognition in the wild using Convolution Neural Networks

Cours : Apprentissage avec des réseaux neuronaux artificiels - ARN

Labo 5 : Transfer learning

Professeur : Andres Perez-Uribe

Assistants : Shabnam Ataee, Simon Walther

Étudiants : Julien Mühlemann, Lucas Lattion

Table des matières

- Object recognition in the wild using Convolution Neural Networks
 - Table des matières
 - 1. Introduction
 - 2. Problématique
 - Classes de détection
 - Base de données
 - Exemples d'images
 - 3. Préparation des données
 - Prétraitement
 - Augmentation des données
 - 4. Création du modèle
 - 4.1 Architecture
 - 4.1.1 Model Summary
 - 4.2 Transfer learning
 - 5. Résultats
 - 6. Conclusion

1. Introduction

Dans le cadre de ce laboratoire, nous allons suivre toutes les étapes nécessaires pour créer une application de classification d'objets. Pour ce faire, nous avons réfléchi à ce qu'il serait utile de classer et avons décidé de nous concentrer sur les différents billets de banque suisses.

En plus de nous offrir une première expérience du transfert de l'apprentissage (transfer learning), une telle application pourrait permettre la mise en place de systèmes de reconnaissance automatique pour des raisons de sécurité, une aide pour les personnes mal-voyantes, et aussi amuser les amateurs de numismatique à identifier les différents billets en circulation en Suisse.

Nous découvrirons également l'apprentissage par transfert en utilisant un modèle MobileNetV2 pré-entraîné et en ajoutant nos propres couches par-dessus. Pour comprendre quelles régions des images sont principalement utilisées par le classificateur à base de réseau neuronal pour effectuer une prédiction, nous visualiserons les images et une sorte de "carte thermique" appelée Carte d'Activation de Classe (CAM).

2. Problématique

Nous étions initialement partis dans l'idée de différencier les billets de banque internationaux, avant de nous rendre compte de la complexité de la collecte de telles données. En effet, outre le grand nombre de devises, plusieurs d'entre elles présentent des similitudes frappantes, rendant une telle classification très difficile pour une première expérience du transfert de l'apprentissage.

Nous avons donc choisi de nous concentrer sur les différents types de billets de banque suisses. Chaque billet possède des caractéristiques reconnaissables, suivant les éléments graphiques et les couleurs spécifiques à chaque dénomination. Une illustration de tout les billets Suisse, se présente ci-dessous:



Billet de
10 francs

Motif:
Le sens de
l'organisation

**Élément
principal:**
Le temps

Billet de
20 francs

Motif:
La créativité

**Élément
principal:**
La lumière

Billet de
50 francs

Motif:
L'aventure

**Élément
principal:**
Le temps

Billet de
100 francs

Motif:
La tradition
humanitaire

**Élément
principal:**
L'eau

Billet de
200 francs

Motif:
La vocation
scientifique

**Élément
principal:**
La matière

Billet de
1000 francs

Motif:
Le goût de la
communication

**Élément
principal:**
La parole

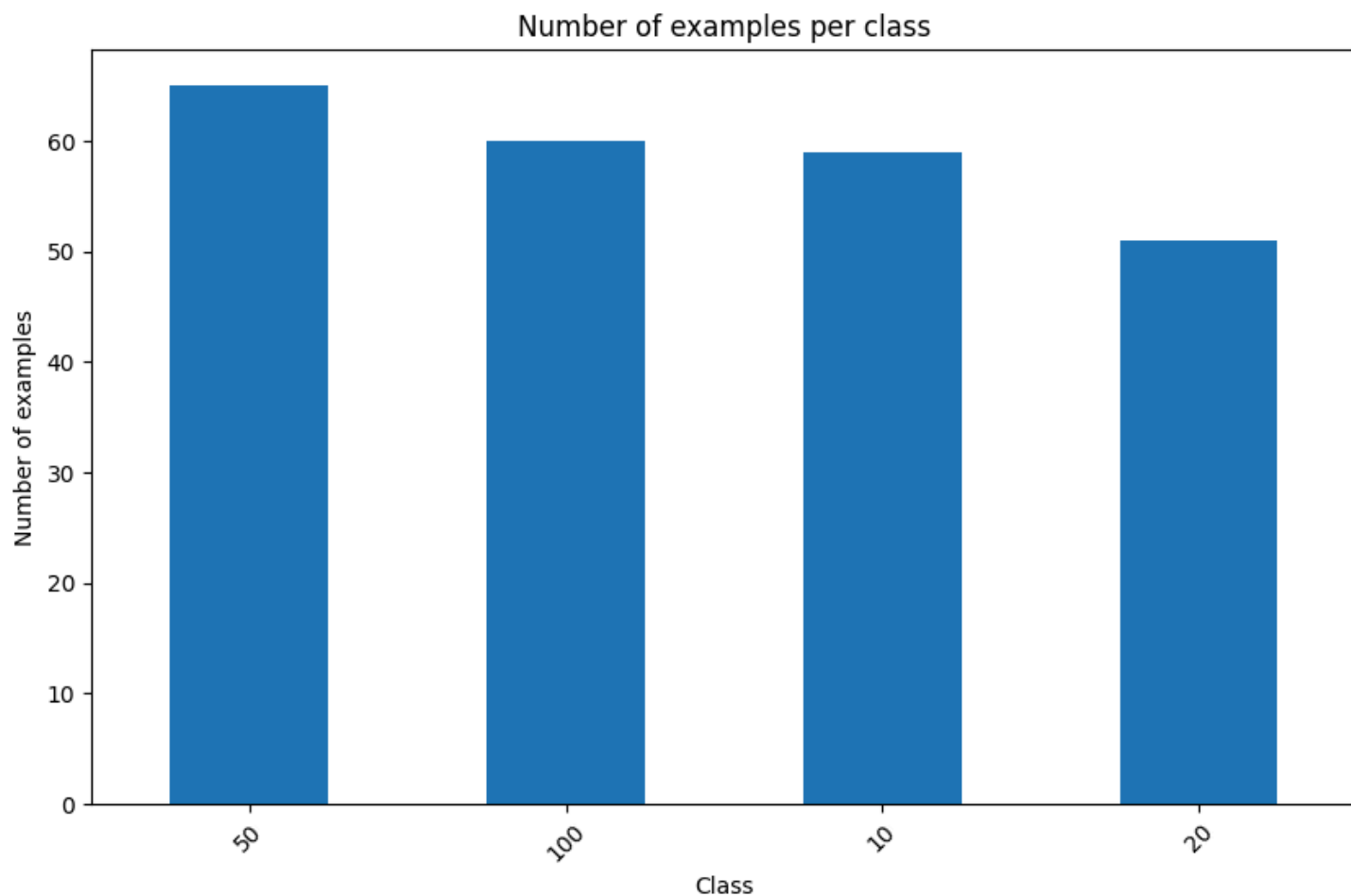
Classes de détection

Nous avons choisi les classes suivantes pour notre projet :

- Billet de 10 CHF
- Billet de 20 CHF
- Billet de 50 CHF
- Billet de 100 CHF

Base de données

Nous avons collecté des images pour chacune de ces classes. Le jeu de données contient un nombre équilibré d'images par classe. Voici une répartition des images par classe :



Exemples d'images

Voici quelques exemples d'images de notre jeu de données :



3. Préparation des données

Lors de la sélection des photos pour notre ensemble de données, nous avons conservé les photos ayant peu d'éléments perturbateurs en arrière-plan, comme d'autres objets ou des inscriptions.

Nous avons également essayé de mélanger différentes séries de billets, pour que le modèle ait une meilleure chance de reconnaître les caractéristiques spécifiques à chaque type de billet. Nous avons

sélectionné aléatoirement 1/3 des images pour notre ensemble de test, et les 2/3 restants pour l'entraînement.

Prétraitement

Les images sont redimensionnées et normalisées pour être compatibles avec le modèle MobileNetV2 :

```
from tensorflow.keras.layers import Resizing, Rescaling
from tensorflow.keras import Sequential

IMG_HEIGHT = 224
IMG_WIDTH = 224

image_preprocesses = Sequential([
    Resizing(IMG_HEIGHT, IMG_WIDTH, crop_to_aspect_ratio=True),
    Rescaling(1. / 255)
])
```

Augmentation des données

Pour augmenter la diversité de notre jeu de données, nous avons appliqué des transformations aléatoires comme des retournements horizontaux et des rotations :

```
from tensorflow.keras.layers import RandomFlip, RandomRotation

image_augmentations = Sequential([
    RandomFlip("horizontal"),
    RandomRotation(0.1),
])
```

4. Création du modèle

4.1 Architecture

1. MobileNetV2 comme base

- Nous avons utilisé MobileNetV2 comme base, sans inclure la couche de classification finale (`include_top=False`). Cela signifie que nous avons uniquement utilisé les couches

convolutionnelles de MobileNetV2, qui sont responsables de l'extraction des caractéristiques des images.

2. Couches ajoutées

- **GlobalAveragePooling2D** : Cette couche effectue un pooling global moyen sur les cartes de caractéristiques issues de MobileNetV2. Cela réduit chaque carte de caractéristiques à un seul nombre, créant ainsi un vecteur de caractéristiques de taille fixe, indépendamment de la taille des cartes de caractéristiques.
- **Dense (100 unités, activation 'relu')** : Cette couche dense entièrement connectée comporte 100 unités et utilise la fonction d'activation ReLU (Rectified Linear Unit). Elle introduit une non-linéarité et permet au modèle d'apprendre des combinaisons complexes de caractéristiques.
- **Dropout (0.7)** : La couche Dropout est utilisée pour régulariser le modèle et prévenir le surapprentissage. Elle désactive de manière aléatoire 70% des unités de la couche précédente à chaque étape d'entraînement.
- **Dense (nombre de classes, activation 'softmax')** : La dernière couche dense comporte autant d'unités que de classes dans notre jeu de données et utilise la fonction d'activation softmax. Cette couche produit une probabilité pour chaque classe, indiquant la probabilité que l'image d'entrée appartienne à chaque classe.

3. Compilation du modèle

- **Optimizer** : Nous avons utilisé l'optimiseur RMSprop, connu pour son efficacité dans le traitement des grands jeux de données.
- **Loss Function** : La fonction de perte utilisée est `sparse_categorical_crossentropy`, adaptée aux problèmes de classification multi-classes.
- **Metrics** : Nous avons utilisé `accuracy` comme métrique pour évaluer les performances du modèle pendant l'entraînement et la validation.

4.1.1 Model Summary

Total params: 2386488 (9.10 MB)

Trainable params: 128504 (501.97 KB)

Non-trainable params: 2257984 (8.61 MB)

4.2 Transfer learning

Pour effectuer le transfert de l'apprentissage, nous avons utilisé MobileNetV2 avec les poids d'ImageNet. Étant donné la petite taille de notre dataset, tous les paramètres d'ImageNet ont été figés, ne laissant que la couche dense comme étant paramétrable. Notre dataset n'étant toutefois pas totalement similaire aux objets utilisés pour ImageNet, nous aurions probablement pu laisser une

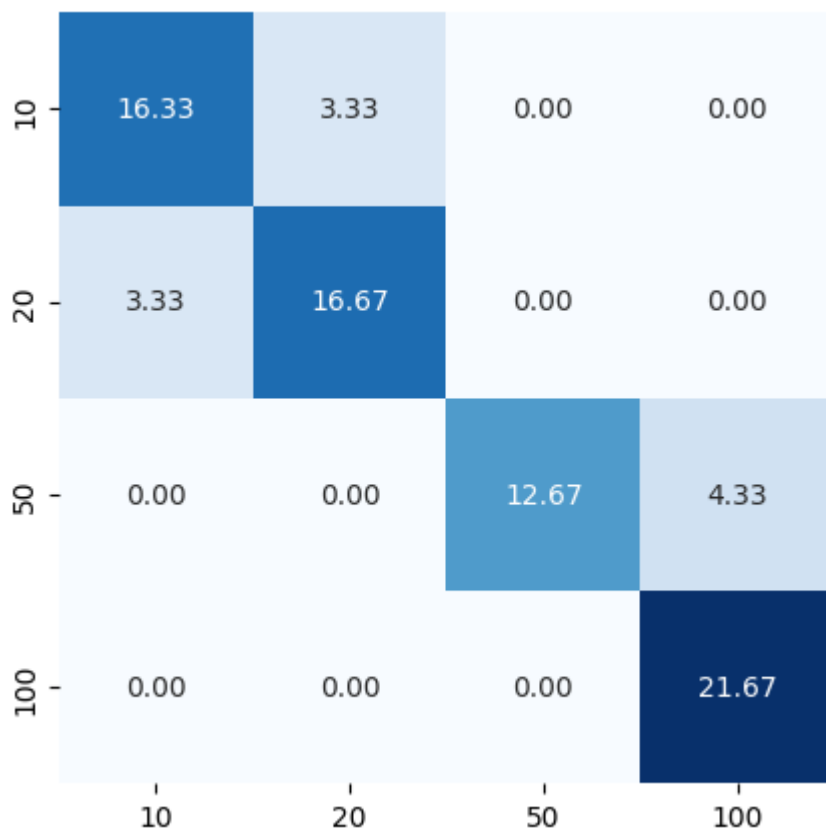
partie des paramètres du modèle entraîna**bles**, mais nous avons décidé de nous concentrer principalement sur la modification des hyper-paramètres.

Le transfert de l'apprentissage dans notre cas est très utile. En effet, c'est un moyen efficace d'obtenir un modèle performant capable de différencier certaines caractéristiques d'une photo et d'analyser de manière plus précise un petit ensemble d'images. Un modèle convolutif classique entraîné sur notre dataset aurait donné des résultats de très mauvaise qualité, avec un overfitting inévitable.

5. Résultats

La performance de l'entraînement est donnée dans la figure suivante:

Mean F1 Score across all folds: 0.860



Comme nous pouvons l'observer, le modèle a plus de difficultés avec les prédictions de certaines classes de billets. Cela semble logique car certaines dénominations peuvent avoir des similarités graphiques. Toutefois, nous sommes plutôt satisfaits de la performance de notre modèle, qui affiche un f-score relativement bon pour plusieurs classes de billets. Cela nous rassure quant à la capacité du modèle à séparer les caractéristiques spécifiques aux différents types de billets.

En application réelle, le modèle affiche quelques incertitudes, mais une fois la position de la caméra stabilisée, les pourcentages de classification sont relativement stables. Les essais réels ont montré quelques difficultés avec certains types de billets, notamment ceux de petites dénominations souvent confondus avec des billets de dénominations proches.

6. Conclusion

Ce laboratoire a été très intéressant. Un ensemble de classes plus facile et plus proche des objets reconnus par ImageNet aurait probablement causé moins de problèmes dans la classification. Nous avons rencontré plusieurs difficultés lors de la mise en place du dataset, il nous a fallu reprendre de nouveaux échantillons pour obtenir des résultats intéressants.

Nous sommes satisfaits des résultats obtenus et pensons que l'expérience nous aura formés sur l'utilisation du transfert de l'apprentissage ainsi que sur l'importance de la sélection initiale du dataset.

Nous pensons que les résultats valident notre problématique initiale, et, considérant que ImageNet est entraîné sur des objets ordinaires du quotidien, la classification finale est satisfaisante pour une première expérience sur le transfert de l'apprentissage et les réseaux convolutifs complexes.

La classification telle qu'effectuée ici présente des limites. En effet, les types de billets autres que ceux sélectionnés ne pourront pas être classifiés correctement. L'utilisation de ce modèle reste donc restreinte dans l'état actuel. Un travail futur pourrait partir des observations effectuées en incorporant un set de données plus grand et avec un nombre plus élevé de types de billets, que nous n'avons pas pu inclure dans notre dataset. Il pourrait également être utile de fournir une catégorie supplémentaire « autre » pour les objets qui ne sont pas des billets.

Ce rapport présente une vue d'ensemble complète de notre approche et des résultats obtenus, fournissant ainsi une base solide pour des améliorations futures.