



Rapport sur la constitution de la base de données

Auteurs :

Mahmoud LAANAIYA

Mohamed Hamza KADRI

Léane HAUTEVILLE

Adrien NARKIEWICZ

Table des matières

1	Composition du groupe, titre du projet	3
1.1	Composition du groupe	3
1.2	Titre du projet	3
2	Description du sujet	3
3	Méthodologie pour acquérir, annoter les données	4
3.1	Acquérir les données	4
3.2	Annoter les données	4
4	Partitionnement des images en ensembles d'entraînements, de validation et de test	5
5	Pronostics	5
6	Quelques images de la base de données	6

Table des figures

1 Composition du groupe, titre du projet

1.1 Composition du groupe

Adrien NARKIEWICZ, Mahmoud LAANAIYA, Léane HAUTEVILLE, Mohamed Hamza KADRI.

1.2 Titre du projet

Classification des animaux de la savane (Eléphant, Zèbre, Lion, Guépard).

Lien vers le projet <https://github.com/Nakhaan/Projet-DeepLearning>

2 Description du sujet

Nous avons cherché un sujet qui nous permettra de classier des images avec un réseau de neuronne et pour cela nous avons choisi de classier des animaux de la savane tels que le lion, le zèbre, l'éléphant et le guépard. Nous avons opté pour cela car lors de notre jeunesse, nous avons été passionné par l'histoire du Roi Lion, un film qui a marqué notre enfance.



3 Méthodologie pour acquérir, annoter les données

3.1 Acquérir les données

Pour cette tâche nous nous sommes répartis le travail, chacun de nous a cherché les images d'un seul animal. On a employé plusieurs méthodes pour cela telles que : l'utilisation de base de données déjà constituée à travers Kaggle, récupération à la main et utilisation de l'Api Download All image sur le Google Extensions.

3.2 Annoter les données

Comme pour la tâche précédente chacun d'entre nous a utilisé une méthode pour annoter les données. Nous avons donné à chaque image un label correspondant à l'animal sur la photo. Pour cel on a utilisé un script qui parcourt les images de chaque fichier et lui attribue le label et le numéro adéquat.

```
a=1
for i in *.ExtensionFichier; do
new=$(printf "NomAnimal%03d.ExtensionFichier" "$a") #03 pad to
length of 3
mv -i - "$i" "$new"
let a=a+1
done
```

4 Partitionnement des images en ensembles d'entraînements, de validation et de test

Pour le partitionnement des images on a trois ensembles : ensemble d'entraînement, de test et enfin de validation. Puisque nous avons quatres animaux avec 250 images pour chaque animal, nous avons mis 150 images de chaque animal dans l'ensemble d'entraînement ce qui fait 600 images pour l'entraînement pour ce qui est de la validation 50 images pour chaque animal ce qui fait 200 images pour la validation, nous avons fait pareil pour le test. Ce qui nous fait un total de 1000 images dans notre base de donnée.

5 Pronostics

La complexité de notre sujet repose dans le fait qu'on ne peut pas différencier les animaux par le taille car les tailles des animaux dans les images n'est pas comme la réalité, et aussi dans le fait qu'ils ont un environnement en commun et des caractéristiques similaires(yeux,etc). C'est pour cela que dans le traitement de notre sujet il faut chercher à les différencier soit par la couleur ou quelques caractéristiques uniques à quelques animaux(la trompe de l'éléphant, les tâches noir et blanche sur le zèbre, la crinière d'un lion...).

6 Quelques images de la base de données

