

Rapport PW5 :

Object recognition in the wild using Convolutional Neural Networks

Département : TIC
Unité d'enseignement : ARN

Auteurs : **Laetitia Guidetti**
Tobie Praz

Professeur : **Andres Perez-Uribe**
Assistant : **Yasaman Izadmehr**

Classe : **A**
Date : **18.06.2023**

Table des matières

1	<i>Introduction</i>	2
2	<i>Problématique</i>	3
3	<i>Préparation des données</i>	6
4	<i>Création du modèle</i>	7
5	<i>Résultats</i>	9
5.1	Résultats sur le jeu de données de test	9
5.2	Résultats dans le monde réel.....	11
5.3	Analyse du grad-cam	13
5.4	Images mal classées	14
6	<i>Conclusion</i>	16

1 Introduction

L'objectif de ce travail pratique est de réaliser une application capable d'identifier un chat. Pour cela, nous utilisons le transfert d'apprentissage avec le modèle pré entraîné MobileNetV2 en y ajoutant nos propres couches.

Il existe divers cas pratiques où pouvoir différencier les chats est utile : les chats, comme les êtres humains, peuvent avoir des maladies ou des allergies nécessitant une alimentation particulière ou la prise de médicaments. Avec une application capable de reconnaître les chats, il serait possible d'identifier le chat arrivant vers un distributeur de croquette automatique et lui distribuer automatiquement le bon type de croquettes selon ses besoins. Des systèmes similaires existent déjà, mais ils utilisent des puces ou des colliers que les chats n'apprécient pas toujours de porter.

Pour réaliser le jeu de données, nous avons pris des photos de nos propres chats dans différentes positions et lieux. Nous avons ensuite réalisé un travail de préparation sur nos images que nous avons après utilisé pour entraîner les dernières couches de notre modèle. Une fois ses actions réalisées, nous avons pu tester directement notre modèle sur nos téléphones dans la vie réel.

2 Problématique

Le jeu de données comporte 5 chats :

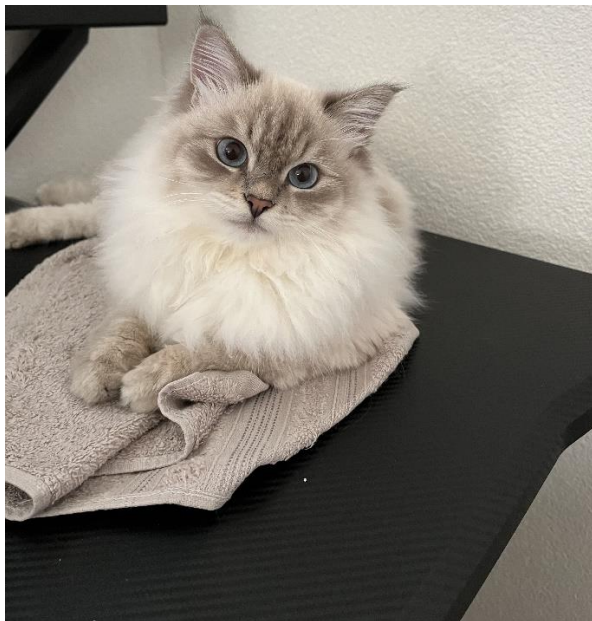
Haru



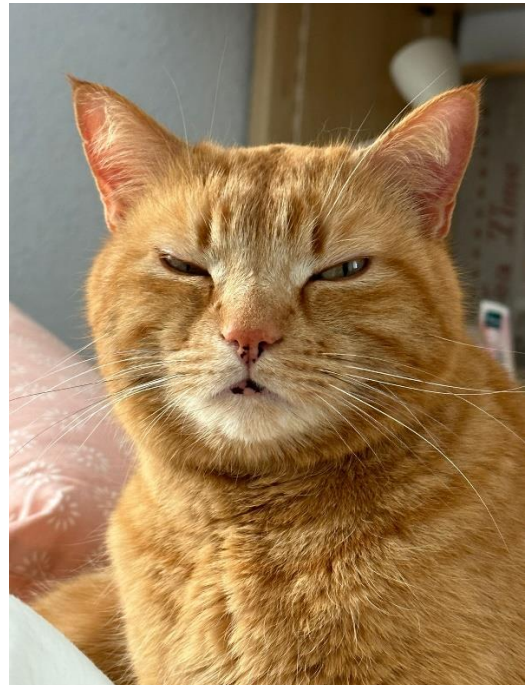
Jerry



Miyuki



Pikou



Tommy



De manière générale, les chats de notre jeu de données sont relativement différents les uns des autres.

Caractéristiques des chats :

	Adulte	Race	Couleur	Sexe
Haru	Oui	Européen	Noir et blanc	Mâle
Jerry	Oui	Européen	Noir tabby	Mâle
Miyuki	Non	Sibérien	Silver colourpoint	Femelle
Pikou	Oui	Européen	Roux	Mâle
Tommy	Oui	Européen	Roux	Mâle

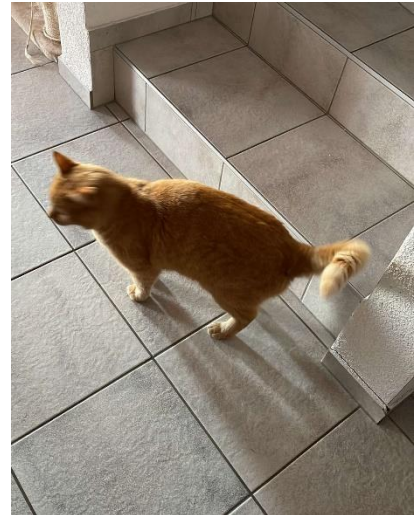
Remarque : Tommy et Jerry sont frères.

La majorité des chats ont des couleurs différentes, ce qui devrait les rendre faciles à différencier. Les plus durs à différencier devraient être Pikou et Tommy, car toutes leurs caractéristiques de base sont identiques. Il est également possible qu'il y ait certaines difficultés entre Tommy et Jerry, car hormis leur couleur, ils ont une morphologie pratiquement identique étant frères. Nous pensons que Miyuki sera la mieux prédite, car elle se différencie par rapport aux autres que cela soit au niveau de sa couleur ou de sa morphologie de chaton.

Nous avons essayé pour notre jeu de données de prendre à chaque fois les chats dans des lieux différents pour éviter que le modèle n'utilise les fonds pour reconnaître les chats. Le mieux aurait

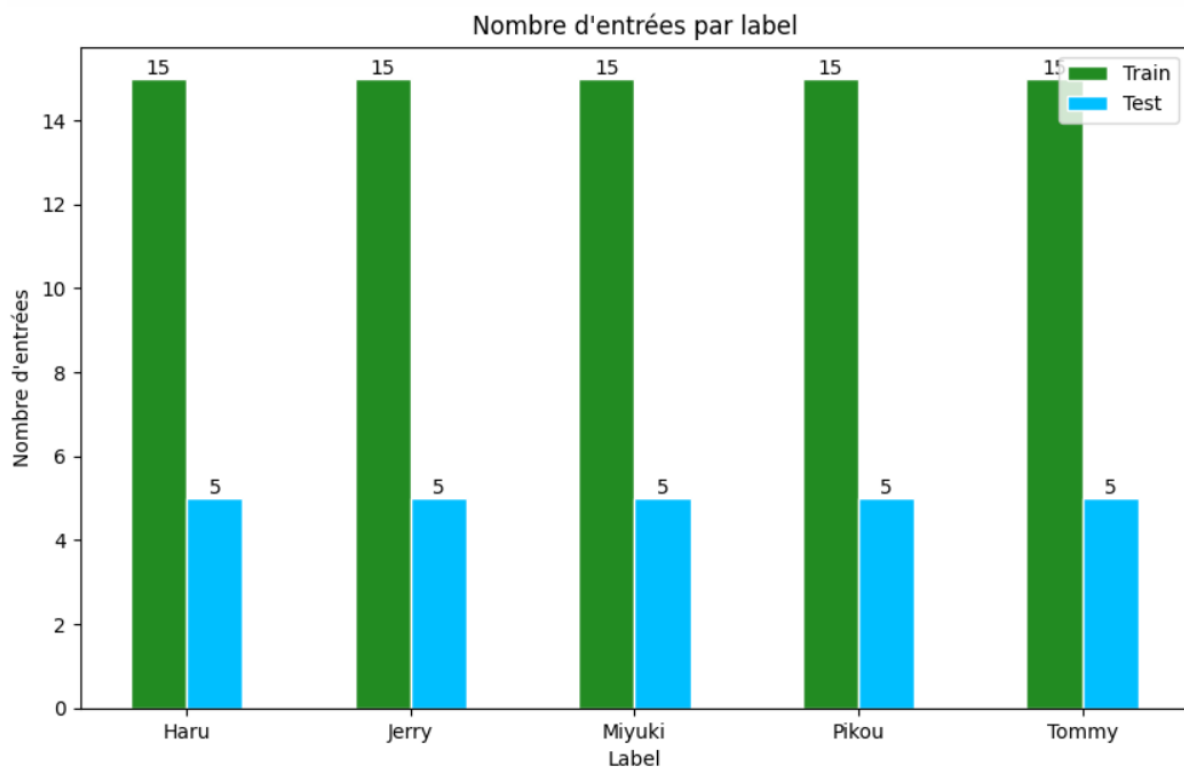
été de prendre des photos de tous les chats avec le même fonds. Cependant, cela n'est pas possible, car ses chats n'habitent pas au même endroit.

Exemple de la variété de photos pour Pikou :



Une variété similaire est présente pour chaque chat.

Notre jeu de données est parfaitement équilibré (15 images d'entraînement et 5 images de test par classe) et relativement petit (moins de 100 images).

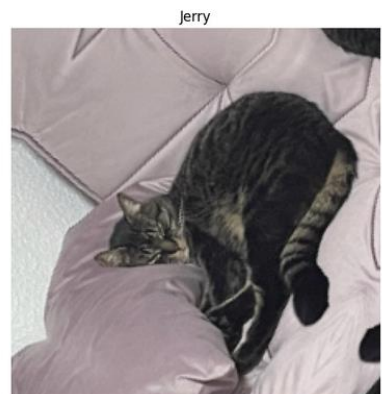
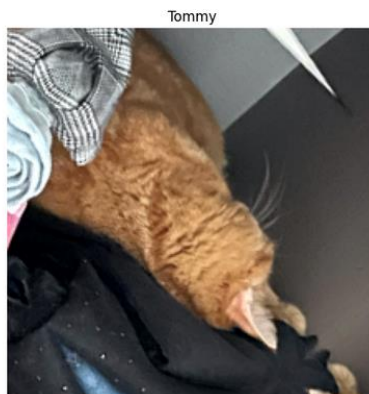
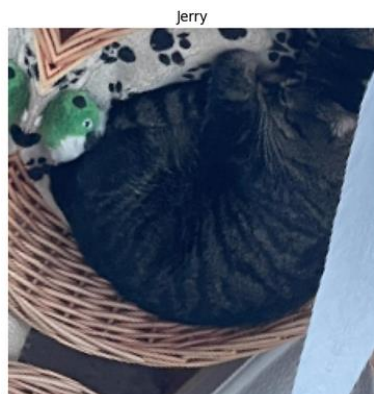
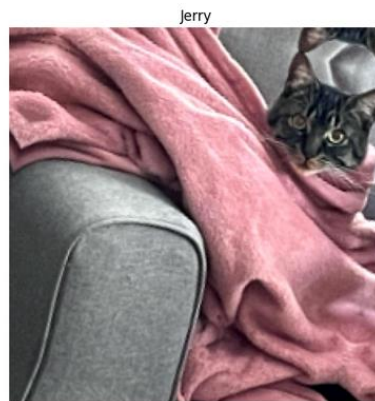


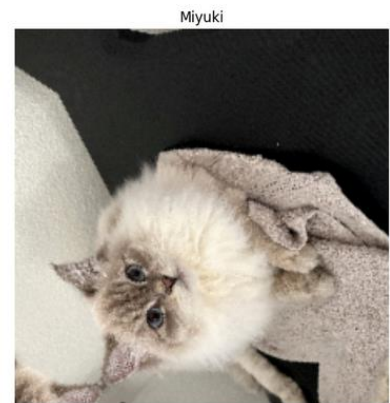
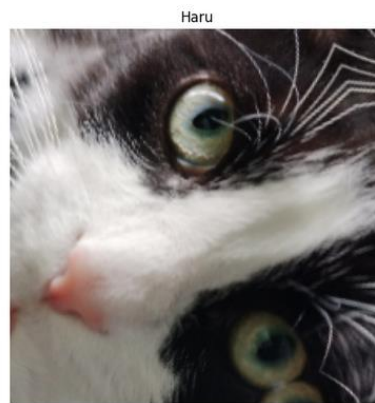
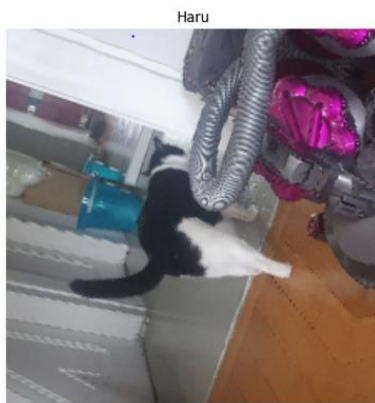
3 Préparation des données

Les images sont transformées pour faire une taille 224x224, car il s'agit de la taille utilisée par MobileNetV2. Nous avons ensuite réalisé quelques opérations d'augmentations des données sur nos images.

```
image_augmentations = Sequential([  
    RandomFlip(),  
    RandomRotation(0.4),  
    RandomContrast(0.2)  
])
```

Exemples d'images augmentées :





Pour l'entraînement de notre modèle, nous avons utilisé 15 images par classe et de la cross-validation (5 splits). Nous avons également réalisé 5 images par classe pour réaliser les tests une fois que le modèle sera entraîné.

4 Création du modèle

Notre modèle utilise le transfert d'apprentissage, car même si notre application est spécifique à 5 chats, il existe de nombreux modèles pré-entraînés capables de reconnaître des chats sur une image. Nous pouvons donc éviter de réaliser l'apprentissage sur cette partie en nous concentrant sur la reconnaissance des différences entre nos chats. Nous avons utilisé les poids de MobileNetV2, car ce modèle est déjà capable de différencier certaines races de chats et donc de réaliser la base de notre projet.

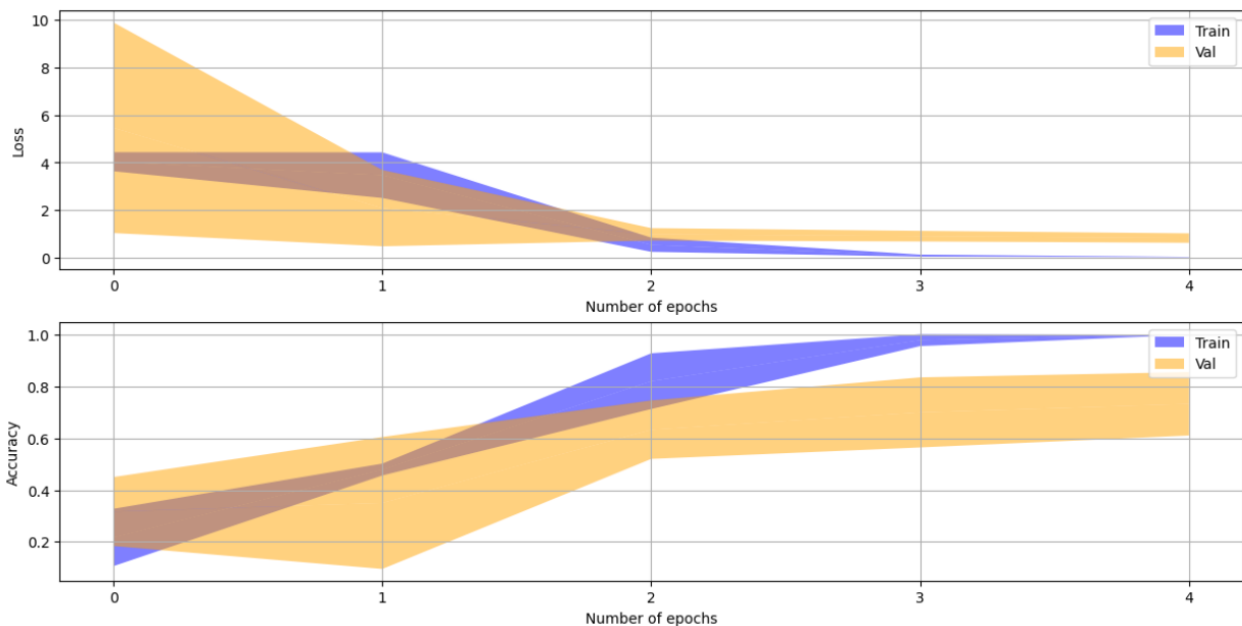
Notre modèle final utilise les poids gelés de MobileNetV2 où nous avons ajouté une couche dense de 500 neurones avec une fonction d'activation ReLU et une couche de sortie de 5 neurones (nombre de classes) avec une fonction d'activation softmax. Pour choisir la couche de 500 neurones, nous avons réalisé plusieurs tests avec 1 ou 2 couches et un nombre de neurones variable. Utiliser 2 couches n'améliorait pas les résultats par rapport à une couche, nous avons donc choisi une seule couche pour des raisons de complexité. Un faible nombre de neurones donnait un moins bon résultat. À partir des 500 neurones, il n'y avait plus d'amélioration. Nous avons alors opté pour une couche de 500 neurones. L'utilisation de dropout n'avait pas d'impact sur nos résultats.

Total params: 2,900,989
Trainable params: 643,005
Non-trainable params: 2,257,984

Avec ses couches, le modèle demande tout de même l'entraînement de 643 005 paramètres.

Comme expliqué plus tôt, lors de l'entraînement, nous utilisons de la cross-validation grâce à 5 splits.

Nous avons ensuite entraîné plusieurs fois notre modèle pour déterminer le nombre d'epochs à utiliser.



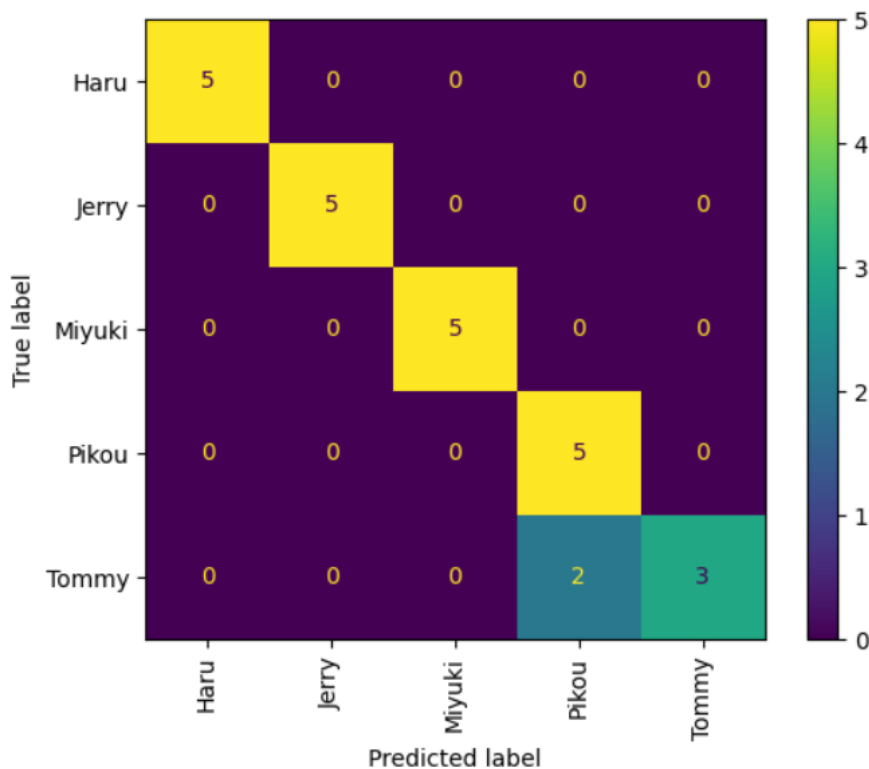
Nous avons décidé d'utiliser 5 epochs, car nous n'observons plus d'amélioration après ce nombre. Nous n'avons pas modifié les autres hyperparamètres (learning rate, optimizer, etc.), car ils n'influaient pas sur nos résultats. Ses hyperparamètres sont donc utilisés avec leur valeur par défaut.

5 Résultats

5.1 Résultats sur le jeu de données de test

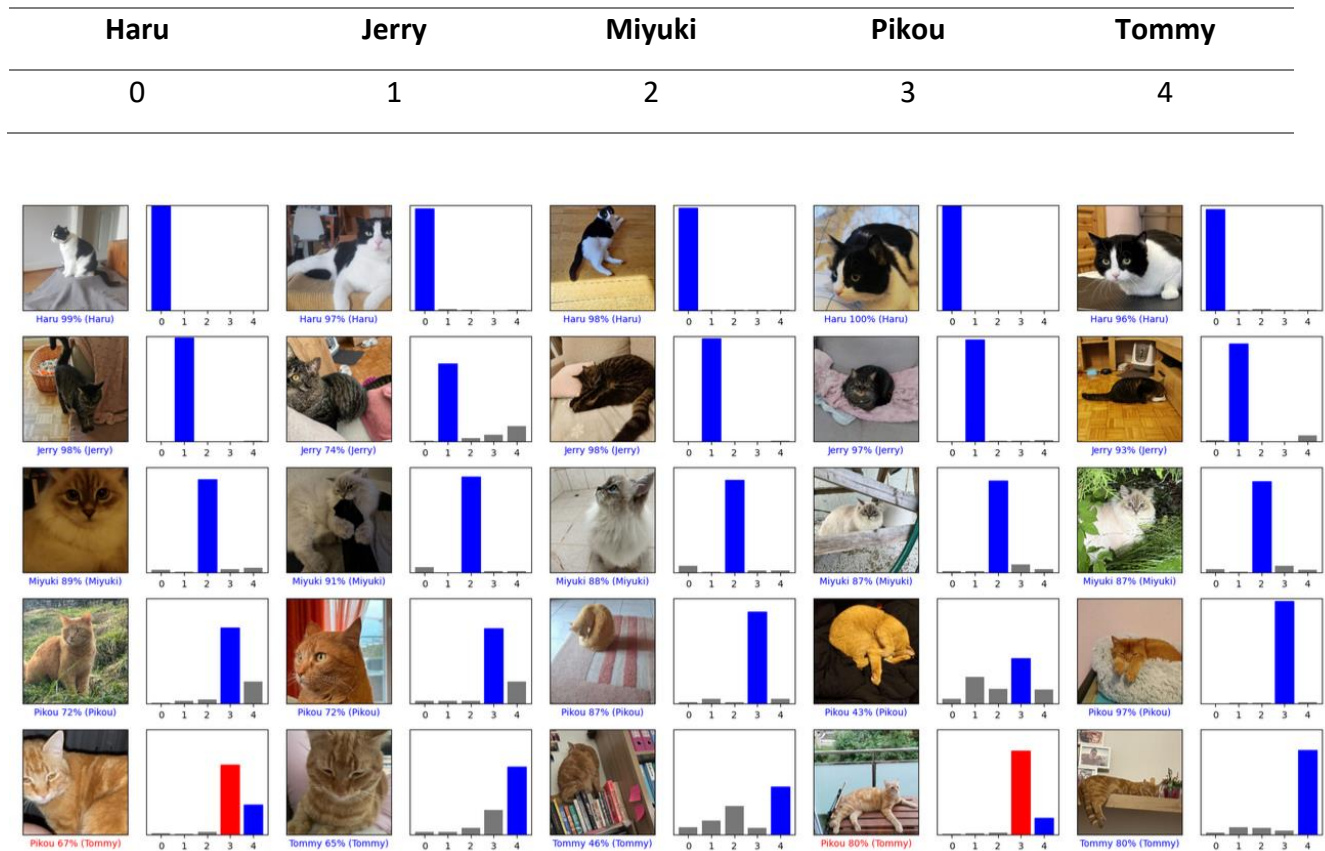
Voici les résultats sur notre jeu de données de test :

		precision	recall	f1-score	support
Haru	0	1.00	1.00	1.00	5
Jerry	1	1.00	1.00	1.00	5
Miyuki	2	1.00	1.00	1.00	5
Pikou	3	0.71	1.00	0.83	5
Tommy	4	1.00	0.60	0.75	5
accuracy				0.92	25
macro avg		0.94	0.92	0.92	25
weighted avg		0.94	0.92	0.92	25



Nos résultats semblent plutôt bons avec une accuracy de 0.92. Grâce au f1-score et à la matrice, nous voyons directement que certaines classes sont parfaitement prédites. Il s'agit de Haru, Jerry et Miyuki. Cela n'est pas étonnant, rappelons que de base ses 3 chats ont une couleur de pelage unique par rapport aux autres. Les difficultés sont rencontrées pour 2 autres chats, Pikou et Tommy. Encore une fois, cela n'est pas étonnant, ses deux chats ont la même couleur de pelage et beaucoup de caractéristiques en commun.

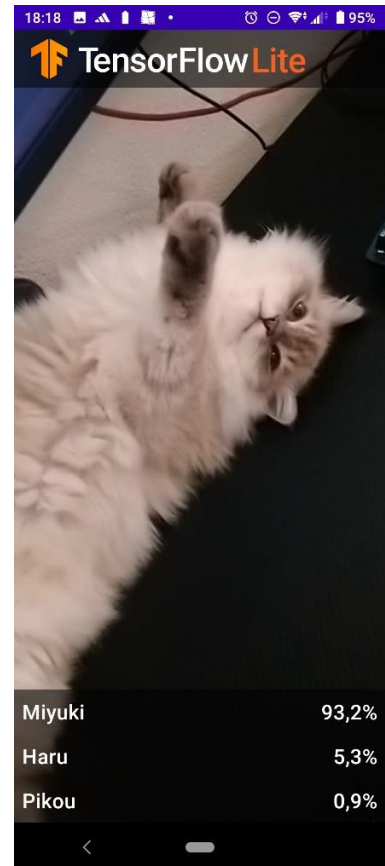
Résultats détaillés de chaque image :



Note : bleu = bon chat, rouge = mauvais chat

De manière générale, on voit que les 3 premiers chats sont déterminés sans la moindre ambiguïté. Cependant, dès qu'il s'agit de Pikou et Tommy, il y a de nombreux doutes entre les deux. Mais dès qu'il y a une erreur, le bon chat est la seconde option. Les résultats correspondent à nos suppositions initiales.

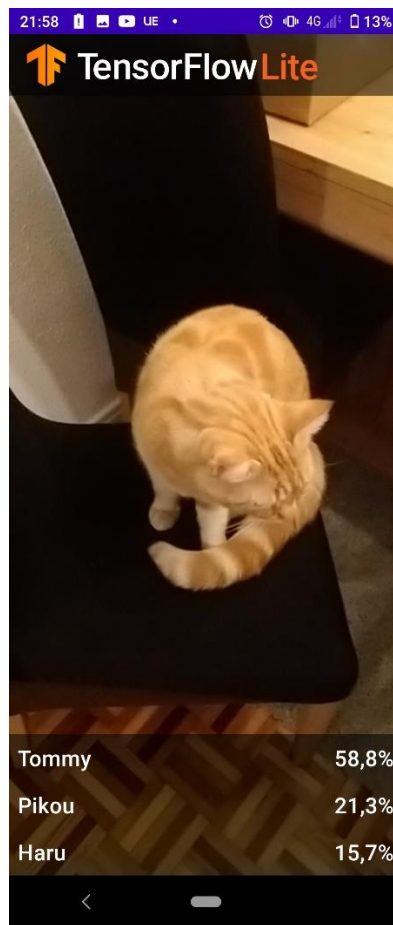
5.2 Résultats dans le monde réel



Pour Haru, Jerry et Miyuki, nos résultats sont identiques au jeu de données de test. Les chats sont parfaitement identifiés sans le moindre doute.



Bonne prédiction



Bonne prédiction



Mauvaise prédiction, il s'agit de Pikou

Pour Pikou et Tommy, on voit encore une fois certaines difficultés à les différencier probablement en raison de leur couleur de pelage identique. Ces 3 exemples montrent également que le pelage des chats peut ressortir différemment selon la luminosité et donc complexifier encore la tâche d'identification.

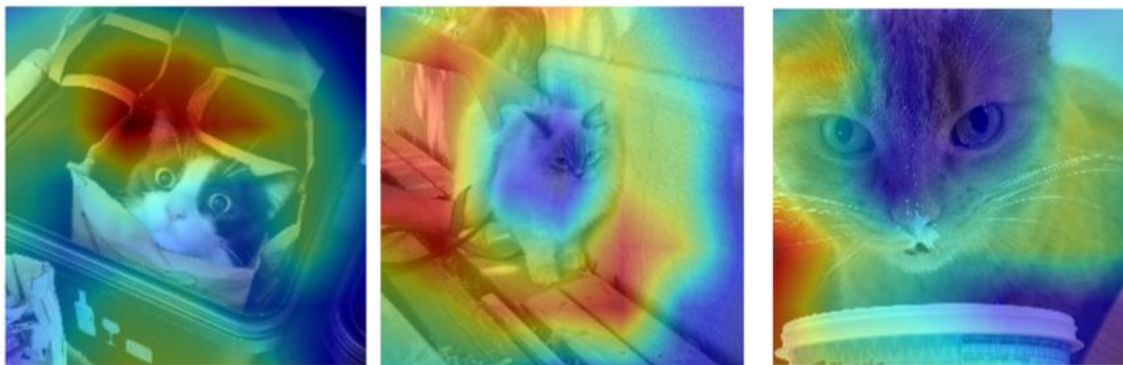
Nous voyons directement que nos résultats dans le monde réel sont identiques à ceux du jeu de données de test. Cependant, cela n'a absolument rien d'étonnant, l'inverse serait même étrange. Notre jeu de données d'entraînement et celui de tests ont été réalisés en photographiant nos chats durant leur vie quotidienne et aucune photo n'a été prise sur Internet. Les tests dans le monde réel ont donc naturellement été réalisés dans les mêmes conditions. Par conséquent, il est normal que les résultats soient identiques.

5.3 Analyse du grad-cam

Nous avons ensuite voulu vérifier la pertinence du système via les grad-cam. De manière générale, le point important de chaque image est le chat comme voulu.



Cependant, nous avons des cas où le focus n'est pas du tout réalisé sur la bonne zone et des points du décor sont utilisés à la place du chat.



Il s'agit probablement de la cause de la mauvaise classification de certaines images.

Pour améliorer notre modèle et régler ce problème, le mieux serait de photographier tous les chats avec le même fond. Cependant, cela est très difficile à réaliser. Il s'agit d'êtres vivants ne vivant pas au même endroit et nous ne voulons pas stresser en les réunissant en un lieu. Il faudrait également penser à la difficulté technique de déplacer 5 chats à la fois. Une autre idée qui permettrait d'améliorer les résultats seraient probablement d'agrandir le jeu de données avec plus de photos de chaque chat. Cela forcerait le modèle à trouver de nouvelles similitudes entre les images.

5.4 Images mal classées

Voici les images mal prédites pour notre jeu de tests :

Prédit : Pikou, Réalité : Tommy

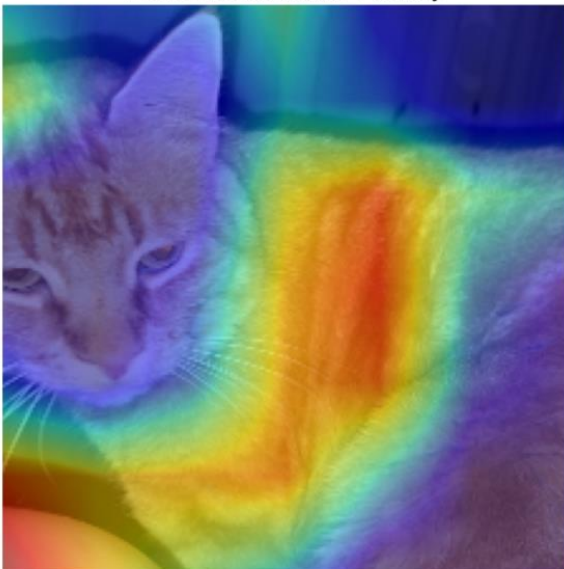


Prédit : Pikou, Réalité : Tommy



Nous avons essayé de comprendre la raison de cet échec via le grad-cam.

Prédit : Pikou, Réalité : Tommy



Prédit : Pikou, Réalité : Tommy



On remarque que pour la première image, le pelage est le seul point utilisé pour déterminer le chat. La différence entre Pikou et Tommy est plutôt sur la forme de leur visage, il n'est donc pas étonnant que les deux soient confondus pour cette image. Pour la seconde image, on remarque qu'un élément du décor est fortement utilisé pour reconnaître le chat, ce qui n'est évidemment pas pertinent. L'erreur provient sans doute de ce point.

Pour le monde réel, nous avons un exemple de cas mal prédit.



Il s'agit encore une fois d'une confusion entre Pikou et Tommy, mais on remarque que comme pour notre jeu de données de test, le deuxième choix est la bonne réponse.

6 Conclusion

Pour conclure, les résultats de notre modèle sont excellents pour les chats ayant une couleur unique. Cependant, des couleurs identiques peuvent rapidement créer de la confusion lors de la classification. Il n'y a aucune différence entre notre jeu de données de tests et les tests dans le monde réel, car notre jeu de données est composé de photo prise en condition réelle.

Cependant, ce système reste limité, le nombre de chats utilisés est faible et pourtant, il rencontre des difficultés à différencier deux chats. Il serait cependant pertinent pour notre cas d'utilisation de base : la reconnaissance de chats via une gamelle connectée. Le nombre de chats dans une habitation dépasse très rarement les 5 et il serait possible d'avoir un fond identique pour chaque chat. Le modèle serait alors encore plus performant.

De manière générale, nous n'avons pas rencontré de grandes difficultés lors de la réalisation de ce laboratoire. Le notebook fourni pour la première partie était très clair pour comprendre le fonctionnement du transfert de connaissance. Nous avons même été surpris de la performance du modèle pour un temps d'entraînement très court (quelques minutes) et un petit jeu de données. Particulièrement pour notre jeu de données, car il n'y avait pas toujours les mêmes parties du corps du chat qui était photographiée et donc une grande variété intra classe.

Nous avons apprécié réaliser ce laboratoire, car il nous permettait de choisir notre problématique et de découvrir les difficultés liées à la réalisation d'un bon jeu de données.