

Compte-rendu de la réunion du 01/04/2025

I Présences

- François Trahay (Tuteur)
- Alexandre Naizondard
- Nicolas Gasca
- Nicolas Nèble
- Hugo Kennedy-Martinez

II Ordre du jour

- Livrable 2
 - Présentation rapide du livrable
 - problème exécution application
- Prochaines étapes
 - Problème d'obtention d'une database suffisante

III Présentation

A Livrable 2

Demande de démonstration du livrable 2 par M. Trahay. Démonstration effectuée par Nicolas Gasca : fenêtre avec menu déroulant pour choisir la course, puis champ pour entrer son numéro de dossard. Après validation, l'image contenant ce numéro de dossard (s'il existe) apparaît dans une autre fenêtre.

Nicolas Nèble souligne qu'il est prévu d'améliorer l'affichage de l'image puisqu'à ce stade, la fenêtre n'est pas du tout dimensionnée à l'image. Alexandre souligne également que le lien entre python et la base de données n'ayant pas été encore fait, les numéros de dossards ont été rentrés à la main dans le script micro_dataset.sql.

Le groupe signale que pour le moment, seul Nicolas Gasca parvient à lancer l'application. Il y a un problème d'accès à la librairie mariadb à partir du fichier .jar. M. Trahay regarde donc avec Alexandre sur son ordinateur pour essayer de trouver la source du problème, pendant que Nicolas Gasca regarde avec Nicolas Nèble sur le sien. Au final, problème non-résolu sous vs code pour Alexandre mais résolu sous IntelliJ pour Nicolas donc Alexandre va basculer sous IntelliJ.

B Prochaines étapes

Quelles sont les prochaines étapes ?

Tout d'abord, il va falloir trouver une meilleure base de données car celle qui a été utilisée pour le prototype est largement insuffisante. En effet, elle ne contient que 30 photos avec dedans environ 150 numéros de dossards lisibles.

Se pose alors la question de comment obtenir cette base de données. On pourrait se la constituer à partir des photos de Google image, mais est ce que cela ne poserait pas des soucis quant à la non-homogénéité des dossards ? Puisqu'ils seront extraits de courses différentes. Mais cela pourrait au contraire être une force pour tester la résilience de l'algorithme de détection. Par exemple, M. Trahay montre une image d'un dossard avec une police jaune sur fond blanc, probablement dur à extraire.

Se pose aussi les questions de RGPD soulignée par Nicolas Gasca. Peut on exploiter les photos comme cela ? Il y a une course qui se déroule sur le campus le 10 avril, le Bike and Run. Cela pourrait être l'occasion de récupérer des photos de courses. Mais même si les participants ont probablement signé quelque chose qui autorise la diffusion de ces images, Nicolas dit que d'après le RGPD, les gens donnent leur autorisation pour une finalité particulière et donc l'exploitation dans le cadre d'un projet informatique n'en fait pas parti.

M. Trahay trouve très intéressant de s'être posé ces questions et que ce sera quelque chose qu'il faudra aborder lors de la soutenance. Il nous conseille également de contacter Benoît Jean, juriste de l'école, si l'on veut creuser

plus ces questions légales. Mais dans le cadre d'un projet étudiant, il nous conseille de ne pas nous soucier de cela.

Pour ce qui est du programme python d'extraction de dossard, les modules d'extractions de numéros actuels ne sont pas adaptés à notre situation donc il va falloir implémenter de nouveaux algorithmes, notamment utiliser de la détection de visage pour faciliter la détection des dossards par exemple, si on a un humain loin, en fonction de la taille de son visage, on peut créer un rectangle autour de lui.

Il y a également des soucis avec les niveaux de gris car selon les niveaux mis, l'algorithme détecte plus ou moins bien les dossards et leur numéros.

M. Trahay dit qu'il pourrait être intéressant de déterminer la qualité de la détection de dossards et se constituer un dataset pour ça. Il faudra éprouver sa résilience selon les situations. Par exemple, si on fait ça sur des courses déguisées, est ce qu'on sera capable de détecter aussi bien ? Il va donc falloir constituer un gros dataset avec des centaines de photos et l'annoter.

Nicolas Nèble pose alors la question de savoir si l'on peut exploiter l'API Mistral qui fonctionne extrêmement bien (plus de 99% des cas !!). Selon M. Trahay, il est intéressant d'exploiter plusieurs pistes. Comme la partie base de données et application ne devrait pas demander une quantité démentielle de travail, il est probable que l'on se retrouve à plusieurs sur la partie python et qu'on peut donc explorer des pistes en parallèle. Il nous conseille également d'essayer de contacter Marius PREDA pour le traitement d'image en IA.

IV Remarques et conseils de M. Trahay

- Envoyer un mail à Benoit JEAN de TSP (juriste) pour répondre à nos questions sur les soucis de RGPD
- Parler de l'aspect légal lors de la soutenance. Comment constituer un dataset et les soucis légaux liés à cette obtention et exploitation.
- Éprouver la résilience de l'algorithme d'extraction des numéros de dossards dans différentes situations telles que des courses déguisées.
- Contacter Marius PREDA pour le traitement d'image en IA.

Prochaine réunion :
Mercredi 16 Avril à 11h, D306