**Hassania Junior Entreprise**
École Hassania des Travaux Publics
Km 7 Route d'El Jadida
Casblanca, Maroc

# Annexe

**Date:** 29/04/ 2023

**Représentant HJE:** Yousra Lahbil

## Description de la prestation

### 1.1 Résumé

La présente convention a pour objet, l'annotation des interactions humains-humains-objets sur un jeu de données vidéos comptant 17 220 images.

### 1.2 Description de la prestation

La prestation consiste à annoter des vidéos, image par image, avec des informations décrivant des intéractions pré-définies entre des personnes et des objets ou d'autres personnes. Ce travail sera effectué sur la plateforme d'annotation collaborative LabelBox.

#### 1.2.1 Définition d'une intéraction

Conceptuellement, une intéraction est un évènement temporel liant 3 entités: la personne source à l'origine de l'intéraction, la personne ou l'objet cible ainsi que le verbe ou le type d'intéraction.

Les intéractions convoitées étant principalement de nature manuelle, il est possible qu'une même personne participe à deux intéractions en parallèle: la première avec la main gauche et la deuxième avec la droite. Ainsi, l'annotation précisera également de quelle main il s'agit. Par conséquent, si elles remplissent les conditions, deux annotations parallèles pourraient être nécessaires pour une même personne (une pour chaque main).

#### 1.2.2 Liste des verbes d'intéractions

Il existe 7 intéractions possibles à annoter:

- **Hold:** maintenir passivement un objet dans une main.
- **Manipulate / Operate:** effectuer une activité intentionnelle avec l'objet, ce qui implique généralement de regarder l'objet (e.g. jouer, désassembler, déformer). ●

**Give:** étendre la main envers une personne avec l'intention de lui donner un objet, même si l'objet n'est pas saisi par l'autre personne.
- **Put / Place:** poser un objet sur une surface.
- **Release:** lâcher un objet depuis la main.
- **Throw:** lancer un objet.
- **Point:** pointer du doigt un objet, une personne ou une zone si l'objet n'est pas dénombrable, e.g. un mur.

Concrétement, il faut spécifier le verbe d'intéraction (s'il y'a lieu) pour chaque main de la personne à annoter.

### 1.2.3 Processus d'annotation

L'annotation d'une intéraction sur LabelBox est décrite par l'évolution dans le temps de 2 objets d'annotations (avec des classifications imbriquées) représentés sur une séquence d'images consécutives correspondant à la durée de l'intéraction (c-à-d un segment temporel):

1. Un rectangle pour délimiter la personne source, comprenant un label pour spécifier le verbe d'intéraction (pour chaque main).
2. Un rectangle pour délimiter la personne ou l'objet cible (quand il est visible), comprenant un label pour spécifier la main de la personne source qui le manipule (c-a-d, droite ou gauche).

Les personnes à annoter dans chaque clip seront fournies. En outre, l'annotation de la position des personnes (c-a-d, le rectangle déliminant le corps de la personne), sera préalablement chargée dans l'outil LabelBox. L'idée est d'utiliser un modèle de machine learning pré-entrainé qui permet de détecter les personnes automatiquement, et pré-charger ces prédictions dans l'outil d'annotation. Ceci permettra de réduire l'effort d'annotation nécessaire pour finir le projet. Dans ce cas, l'équipe d'annotation n'aura qu'à "corriger" ou ajuster ces rectangles quand ils sont faux (c-a-d, quand la prédiction du modèle est erronée). Les rectangles délimitants les personnes seront donc annotés sur toute la durée de la vidéo, et non seulement lors d'une intéraction (contrairement aux objets cibles). En contre-partie, la spécification du verbe d'intéraction n'est requise que dans le cas où la personne est activement engagée dans une intéraction parmi la liste définie plus haut.

## 1.3 Déroulement du projet

L'annotation se fera sur la plateforme LabelBox (https://labelbox.com/). Un projet d'annotation sera créé et mis en place par le référant de l'Idiap Samy Tafasca, qui ensuite, s'occupera d'inviter les membres de l'équipe du Prestataire pour pouvoir accéder à l'outil.

Durant le projet, le référant maintiendra une communication active avec le Prestataire afin de faciliter la prise en main de l'outil, fournir plus d'informations sur la tâche, répondre aux questions, et partager du feedback sur la qualité de quelques annotations.

## 1.4 Livrables

Pour abréger la boucle de feedback et éviter des problèmes de qualité à terme, le résultat du projet sera livré de manière progressive, échelonné sur 3 livrables:
- 1er lot: un lot comprenant 5714 images, à livrer après 3 semaines de la date de début du contrat.
- 2ème lot: un lot comprenant 5714 images, à livrer après 1 mois et 3 semaines de la date de début du contrat (eu égard des 7 jours de contrôle de qualité du 1er lot).
- 3ème lot: un lot comprenant les 5792 images restantes, à livrer à la fin du contrat (après 2 mois et 3 semaines de la date de début et également en prenant en considération les 7 jours dévolus à la vérification du travail, donc en somme trois mois).

## 1.5 Contrôle de qualité et critère d'accéptation

Afin d'assurer la qualité du résultat,L'association HASSANIA JUNIOR ENTREPRISE effectuera un contrôle suite à chaque livrable pour vérifier que les annotations fournies soient conformes au standard de qualité exigé.

Vu la taille du jeu de données, une vérification manuelle de toutes les annotations ne sera pas possible. Par conséquent, le contrôle de qualité portera sur un échantillon de données aléatoirement choisi et soigneusement re-annoté par L'association HASSANIA JUNIOR ENTREPRISE . L'idée est de comparer la qualité d'annotation fournie à celle de L'association HASSANIA JUNIOR ENTREPRISE sur le même échantillon de données, et assumer cette mesure sur la totalité du livrable.

La mesure de qualité est définie par le pourcentage des intéractions valides annotées par le Prestataire parmi toutes les intéractions présentes dans l'échantillon. Cette liste exhaustive des intéractions présentes est déterminée par l'annotation de l'Idiap, et considérée comme "ground truth". Une intéraction annotée par le Prestataire est considérée comme valide si elle est comprise dans le ground truth, et si elle est complète. Une annotation est dite complète si elle comporte l'annotation de la personne source, l'objet cible (quand il est visible) et la nature de l'intéraction, sur un nombre de frames supérieur à 50% du nombre total de frames associées à cette intéraction dans le ground truth.

Par exemple, imaginons qu'après le premier livrable, l'échantillon choisi comporte des vidéos totalisant 2000 images. L'annotation de l'Idiap, dite "ground truth", a révélé que cet échantillon comprend 100 intéractions. De l'autre coté, imaginons que l'annotation du Prestataire pour ce même échantillon a aboutit à 95 intéractions. En comparant les annotations, il s'est avéré que l'intersection est de 90 intéractions (ie. intéractions valides). C'est à dire que 90 ont été correctement identifiées et matchent des deux cotés (vrais positifs), 10 intéractions sont présentes dans le "ground truth" mais n'ont pas été détectées par le Prestataire (faux négatifs), et 5 intéractions ont été identifiées par le Prestataire mais ne figurent pas dans le "ground truth" (faux positifs). Le taux d'erreur estimé de l'annotation de cet échantillon est donc de 10 (faux négatifs) / 100 (total des positifs) = 10%. Il sera ensuite considéré que 10% des intéractions sont ratées sur tout le livrable de 17 220 images.

Il est important de noter que le Prestataire n'aura pas d'accès préalable à l'échantillon aléatoire choisi pour évaluer la qualité, sinon ce dernier pourrait simplement concentrer ses efforts pour annoter correctement l'échantillon, et prêter moins d'attention au reste des données.

Le Prestataire

SIGNATURE

Le client

SIGNATURE