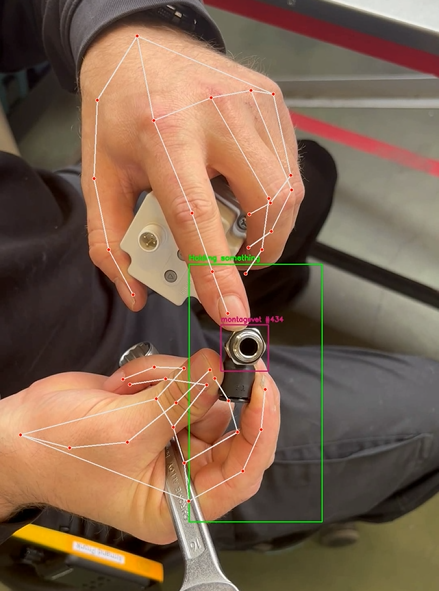
# Robustness

In deze paragraaf wordt de robuustheid van het CAKE-framework geëvalueerd aan de hand van een analyse van de bottlenecks in de pipeline. Het framework bestaat uit meerdere AI-modellen die op elkaar voortbouwen. Een fout in een eerder stadium kan gevolgen hebben voor de uiteindelijke genereerde antwoorden. In deze paragraaf worden per component de mogelijke fouten besproken.

## Objectdetectie

De eerste mogelijke bottleneck in de pipeline is het objectdetectiemodel. Dit model kan verschillende soorten fouten maken die gevolgen hebben op het eindantwoord van het model. Zo kunnen objecten verkeerd geclassificeerd worden, al gebeurt dit niet vaak. Op zich is een verkeerde classificatie niet gelijk problematisch, tenzij het object gekoppeld wordt aan een handeling of hand. In Figuur … is een voorbeeld te zien waarbij dit wel tot een fout leidt. Het label “montagevet” wordt gekoppeld aan de handeling “holding something”, terwijl het in werkelijkheid om een kniekoppeling gaat. Deze foutieve koppeling wordt vervolgens doorgegeven aan het LLM dat de triplets genereert.



Een andere mogelijke fout is het detecteren van objecten die er niet zijn. In Figuur … worden twee objecten gedetecteerd die er niet zijn. Bij de wielgoot heeft dit geen gevolgen, aangezien dit object niet verder gekoppeld wordt. Bij het montagevet is dit wel een probleem, omdat dit object gekoppeld wordt aan een hand. De toppen van de duim en wijsvinger vallen namelijk binnen de bounding box, wat leidt tot een koppeling. Ook deze incorrecte combinatie wordt doorgegeven aan het LLM dat triplets genereerd.

Afbeelding met persoon, overdekt, bureau, tekst

Door AI gegenereerde inhoud is mogelijk onjuist.

Een andere mogelijk fout is dat het objectdetectie model geen objecten detecteert die er in de werkelijkheid wel zijn. In de praktijk is dit zelden een probleem, omdat het object meestal in latere frames alsnog gedetecteerd wordt.

## HAR

Een ander model waar het mogelijk fout kan gaan is het HAR model. Een grote tekortkoming van dit model is dat veel handelingen niet worden gedetecteerd. Zo worden in de video over de stabilizer pressure control maar vijftien handelingen gedetecteerd, terwijl er visueel duidelijk meer aanwezig zijn. Een voorbeeld van een niet gedetecteerde handeling is te zien in Figuur …, waar niet gedetecteerd wordt dat de luchtdruk control unit wordt vastgehouden.

Afbeelding met overdekt, persoon, kleding, verbruiksartikelen voor kantoor

Door AI gegenereerde inhoud is mogelijk onjuist.

Een ander voorkomende fout is een onlogisceh plaatsing van de boundingbox door het HAR model. In Figuur … wordt dit weergegeven. De bounding box van de handeling “holding something” is voornamelijk over de arm geplaatst in plaats van over de hand waarmee het object wordt vastgehouden. Met deze rede is ook de IOU drempel bij de koppeling van de objectdetectie en HAR relatief laag gekozen.

Afbeelding met overdekt, kleding, persoon, Kinderkunst

Door AI gegenereerde inhoud is mogelijk onjuist.

## Handdetctie

Het handdetectiemodel lijkt over het algemeen vrij robuust en maakt zelden fouten. In enkele gevallen worden handen niet herkend, of treedt er verwarring op wanneer twee handen dicht bij elkaar zijn, zoals weergegeven in Figuur ….

Afbeelding met tekst, persoon, handschrift, verbruiksartikelen voor kantoor

Door AI gegenereerde inhoud is mogelijk onjuist.

## Audio

Een andere bottleneck is het transcriptiemodel, dat niet altijd de juiste tekst genereert. Een voorbeeld hiervan is het woord “silencer”, dat verkeerd wordt getranscribeerd als “zijlens”. Dit leidt tot een mismatch tussen de transcriptie en het gedetecteerde object bij het genereren van de triplets.