

13.06.2023r

## **Raport Końcowy (III i IV)**

Analiza Semantyczna Obrazu - projekt 14

Prowadzący:  
dr inż. Piotr Garbat

Zespół:

Kwarciański Michał, 310173

Podleśna Marlena, 297481

Sienkiewicz Wiktor, 297323

### **Założenia projektu**

Założeniem projektu jest uzyskanie algorytmu umożliwiającego detekcję obiektów (ludzi) na określonej scenie (stolik i okolice). Wynikiem algorytmu powinna być informacja o liczbie miejsc zajętych/wolnych przy stoliku. Algorytm działać będzie na bazie heatmapy.

### **Implementacja metod uczenia**

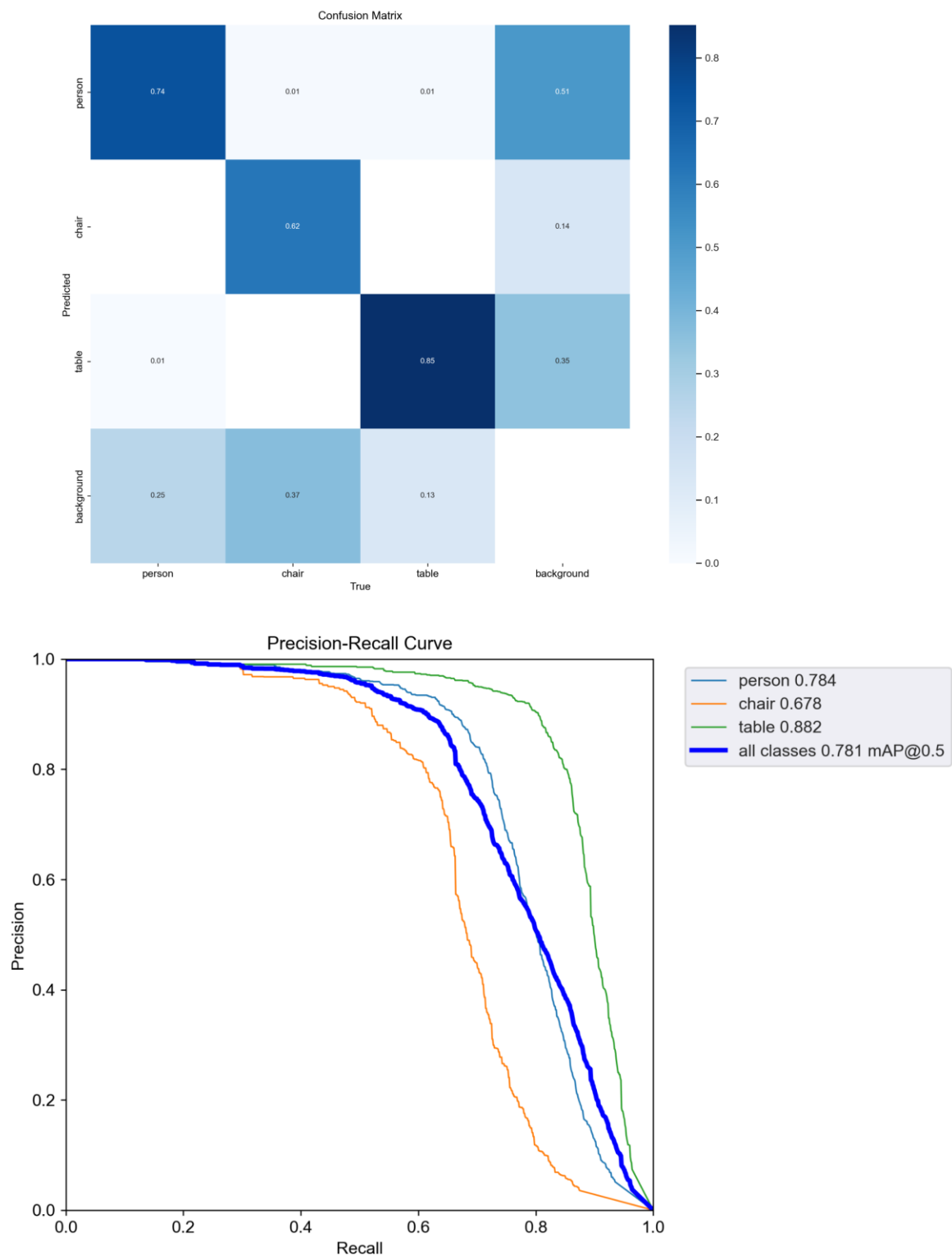
Do uczenia wykorzystano sieci yolov8, yolov5s oraz yolov4. Każda sieć została nauczona detekcji osób, krzeseł i stołów na podstawie własnego zbioru danych oraz 700 zdjęć ze zbioru COCO 2017.

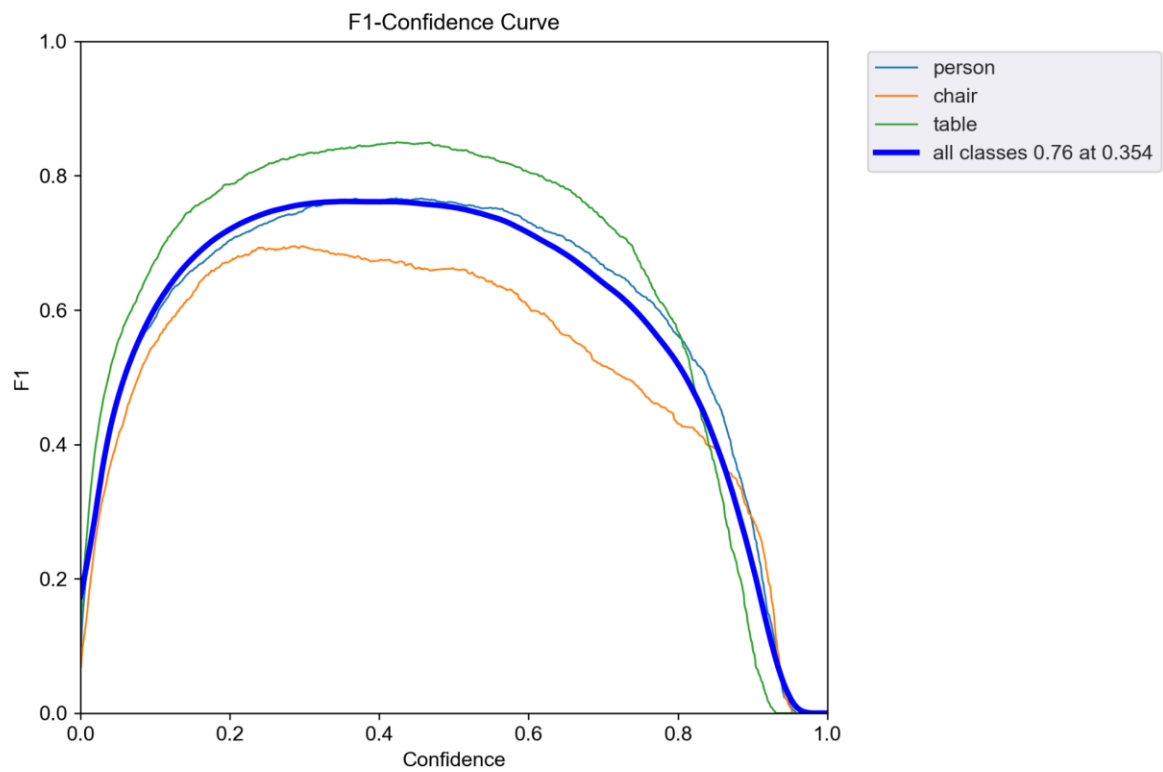
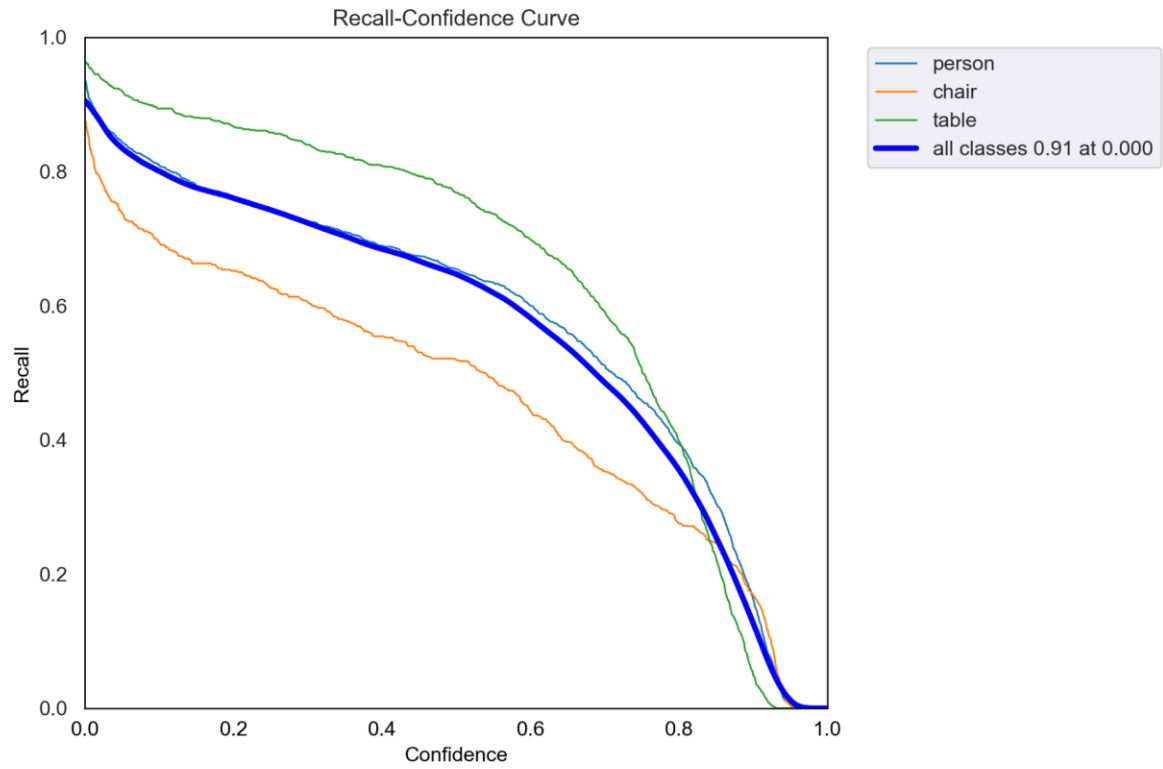
W przypadku sieci uczoney jedynie na danych własnych sieć neuronowa popełnia mniej błędów klasyfikacji. Najczęstszymi błędami w obydwu przypadkach było mylenie obiektów detektowanych z tłem. W przypadku sieci uczoney na mieszanym zbiorze danych obiekty zostały błędnie zaklasyfikowane, np. krzesło jako osoba. Detekcja jednak następuje poprawnie, co może wskazywać na błędne oznaczenia numerów klas. Podjęto próbę konwersji numerów klas na poprawne, jednak nie naprawiło to błędu.

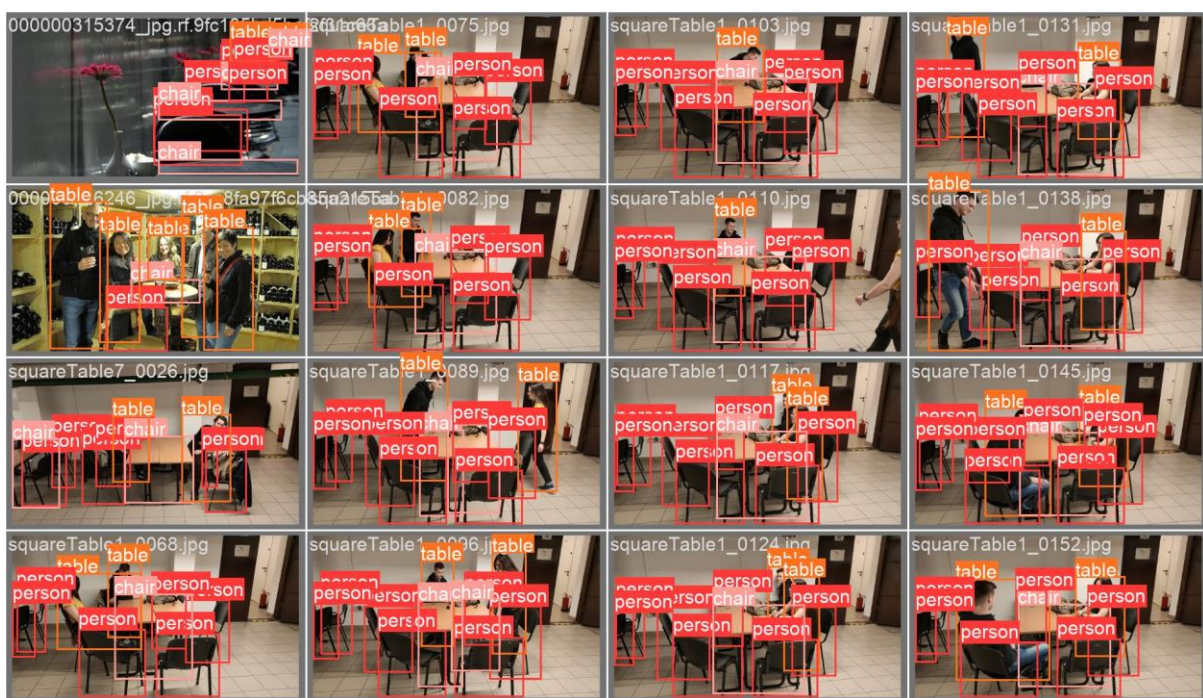
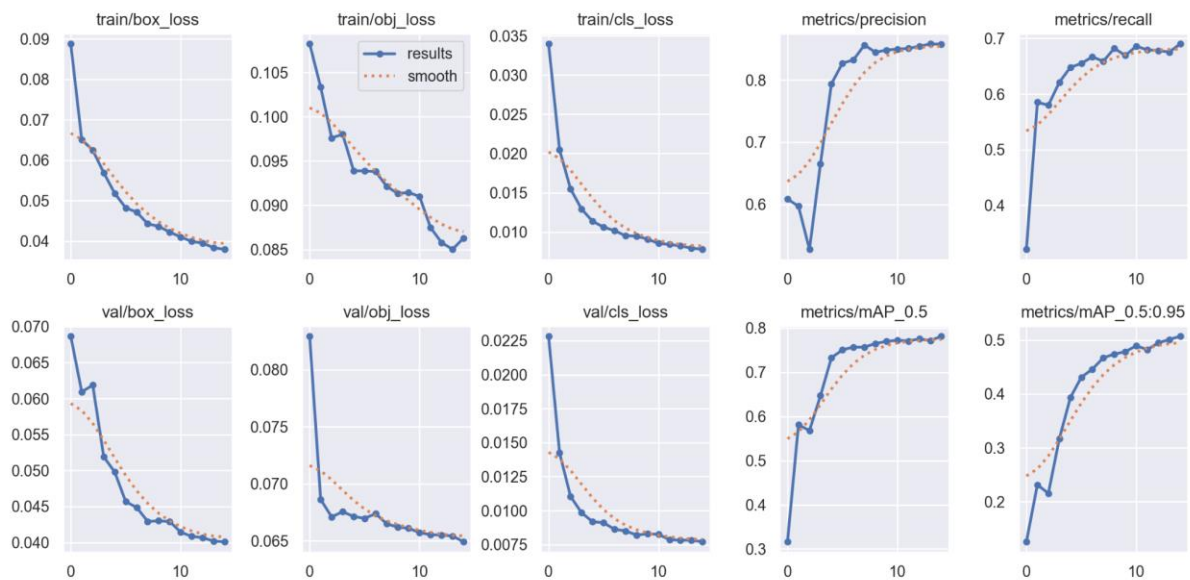
W przypadku detekcji z użyciem sieci yolov4 nauczoney na podstawie ogólnodostępnych danych treningowych ze zbioru COCO o szerokim spektrum klas

występujących obiektów możliwe było detekowanie szukanych obiektów z dużą skutecznością.

## YOLOv5 uczone na danych własnych i ze zbioru COCO





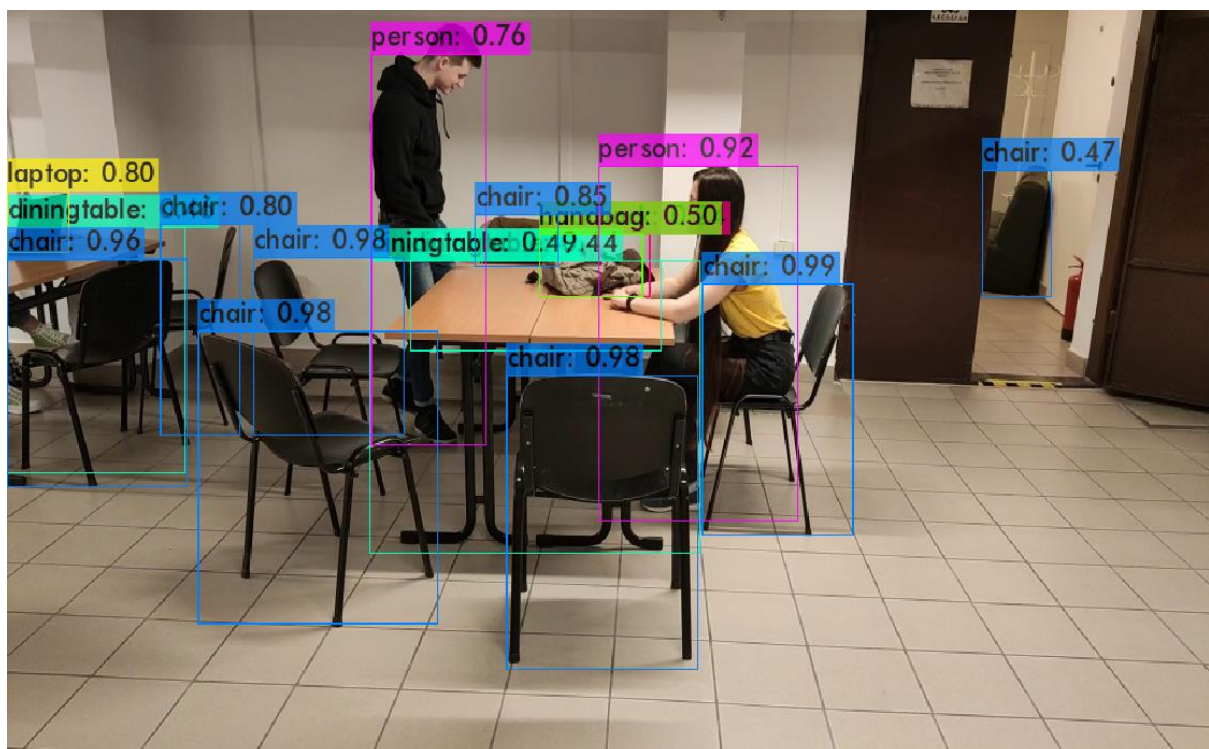


## YOLOV4 uczone na zbiorze trenującym COCO zaimplementowanym we framework *darknet*

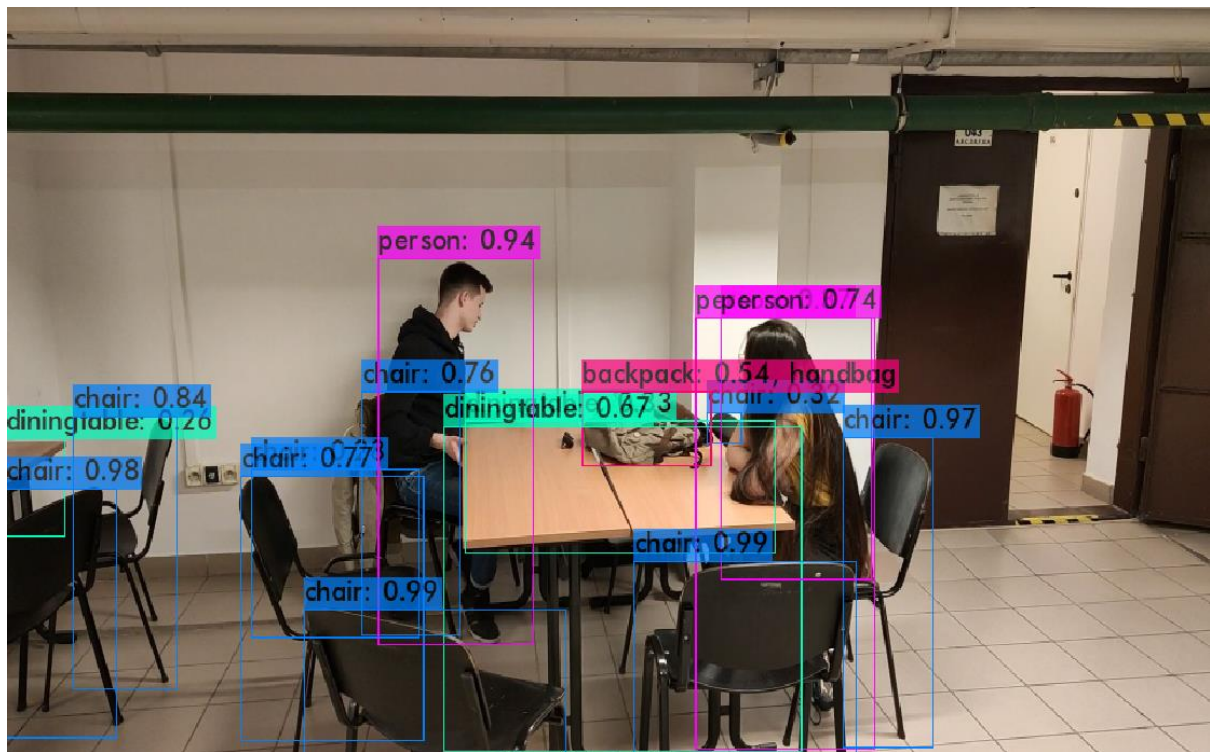
Implementacja czwartej iteracji algorytmu yolo odbyła się z użyciem wersji zaimplementowanej we frameworku *darknet*. Celem implementacji było wytrenowanie modelu na bazie materiałów własnych oraz bazy danych pobranej w ramach COCO. Niestety ze względu na brak możliwości użycia platformy CUDA uczenie sieci wymagało znacznej ilości czasu i nie dało zadowalających rezultatów skutkując wyszukiwaniem zbyt dużej ilości obiektów w trakcie detekcji z użyciem wytrenowanego algorytmu na bazie materiałów własnych.

Ze względu na trudności w implementacji tego algorytmu zdecydowano się na użycie algorytmu yolov4 i wytrenowaniu go na bazie materiałów zawierających 80 klas najczęściej występujących obiektów (materiały pochodzące z COCO), która przygotowana jest w ramach używanego frameworku.

Poniżej przedstawiono przykładowe wyniki dla detekcji obiektów na testowych materiałach własnych:

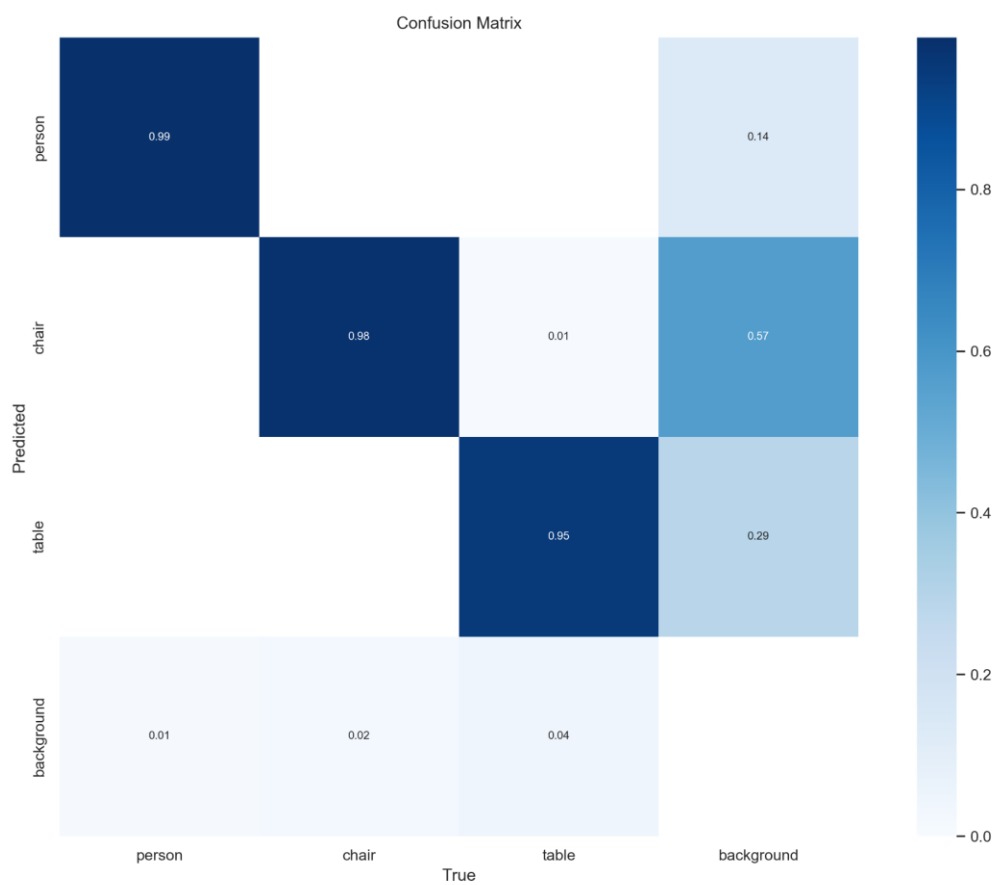


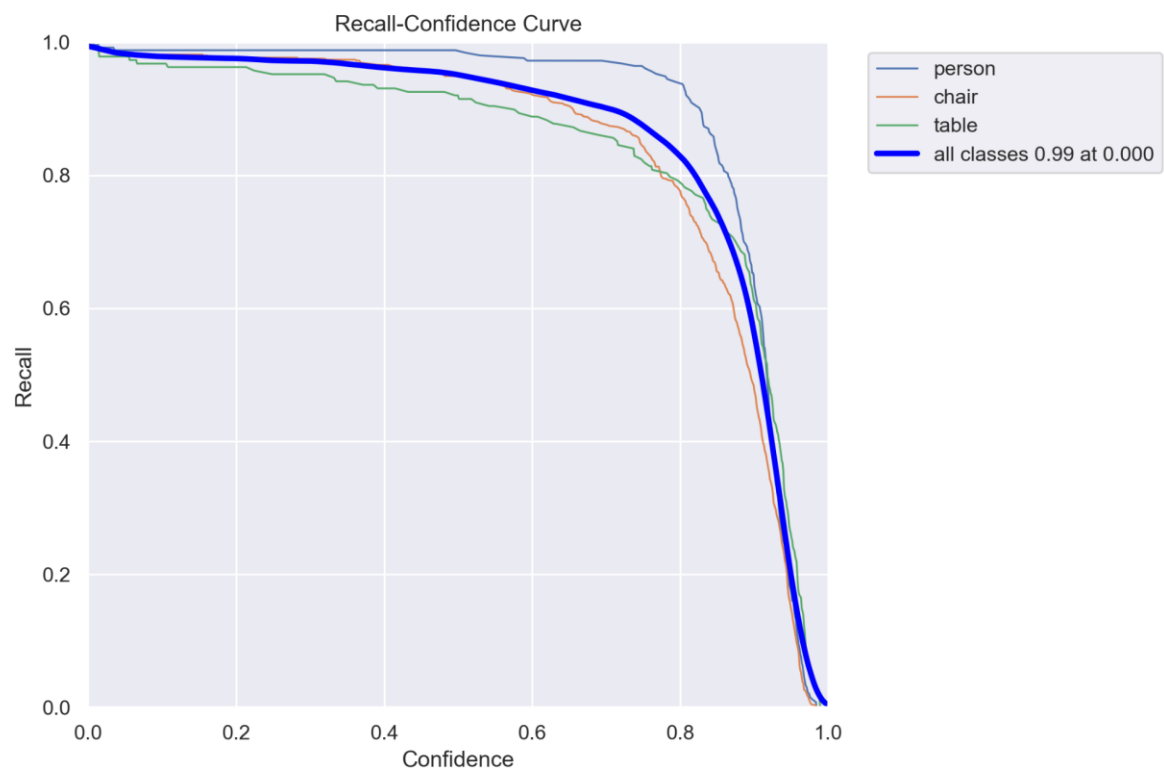
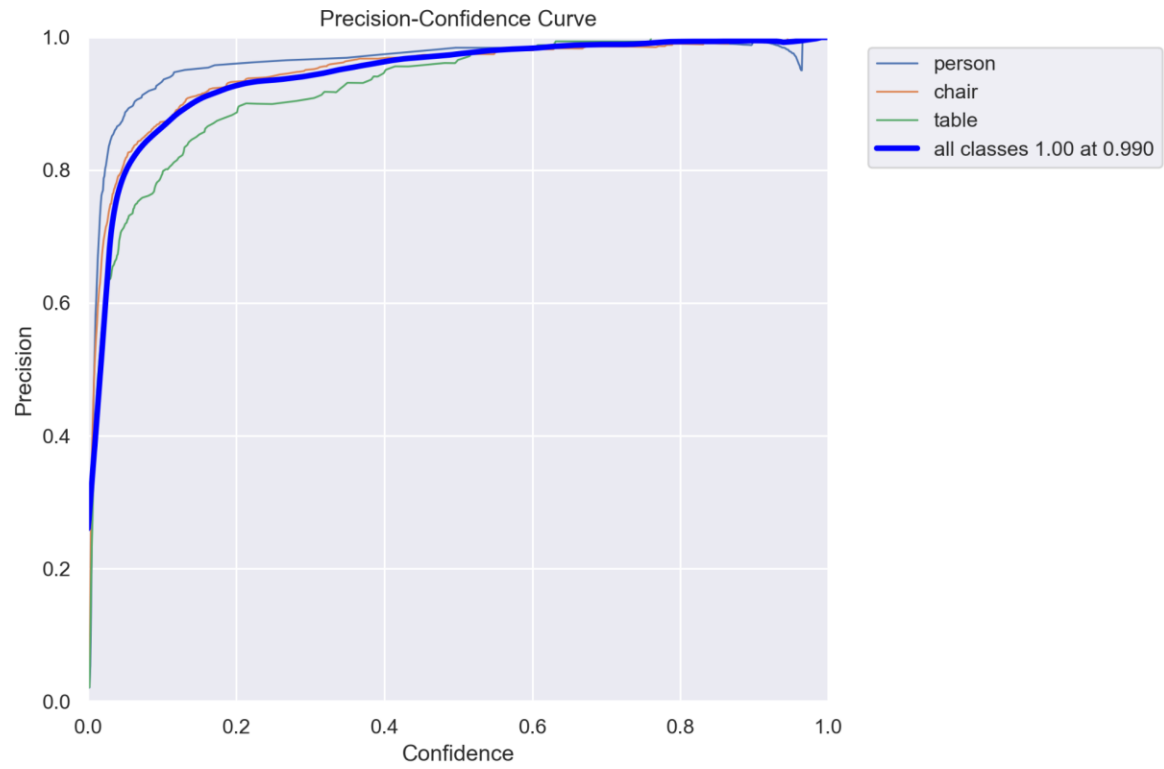




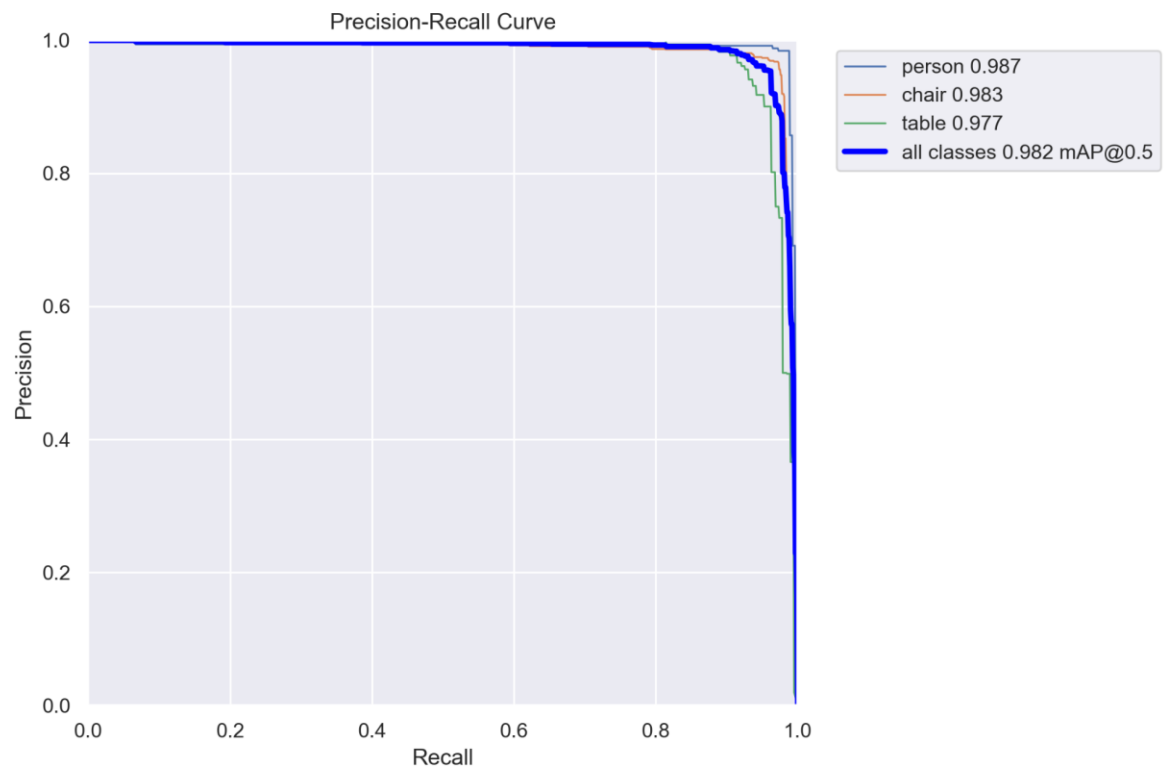
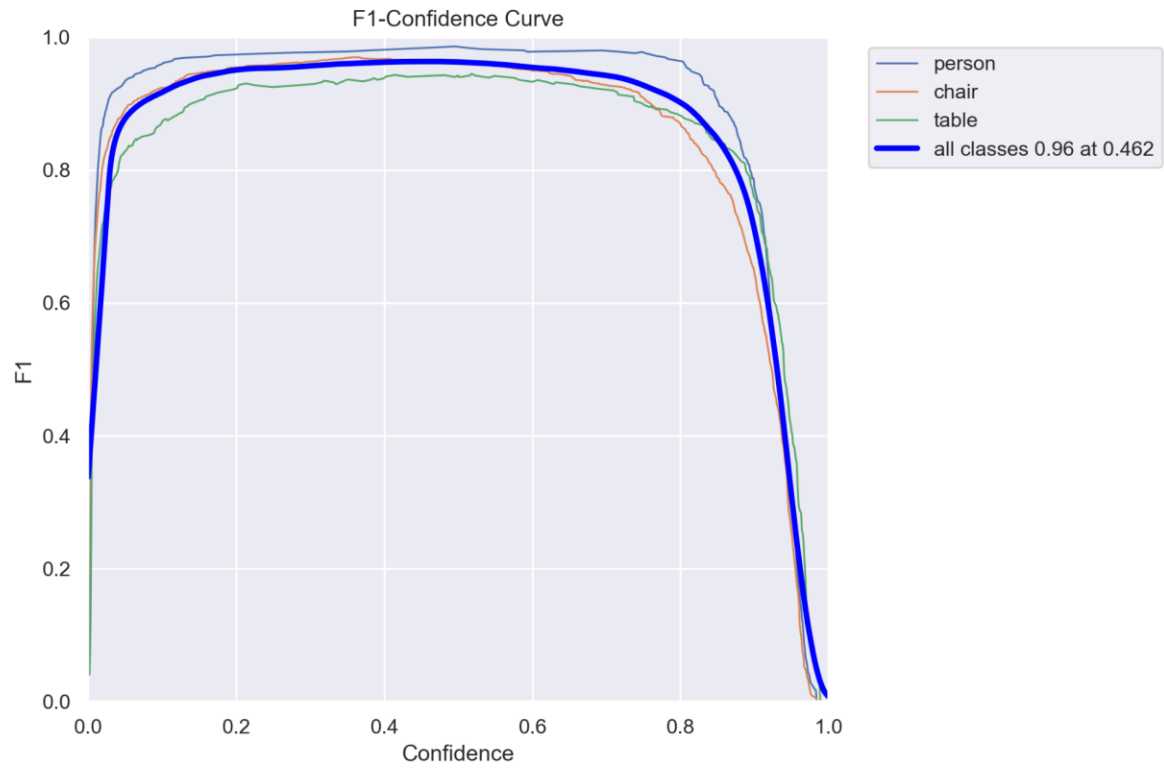


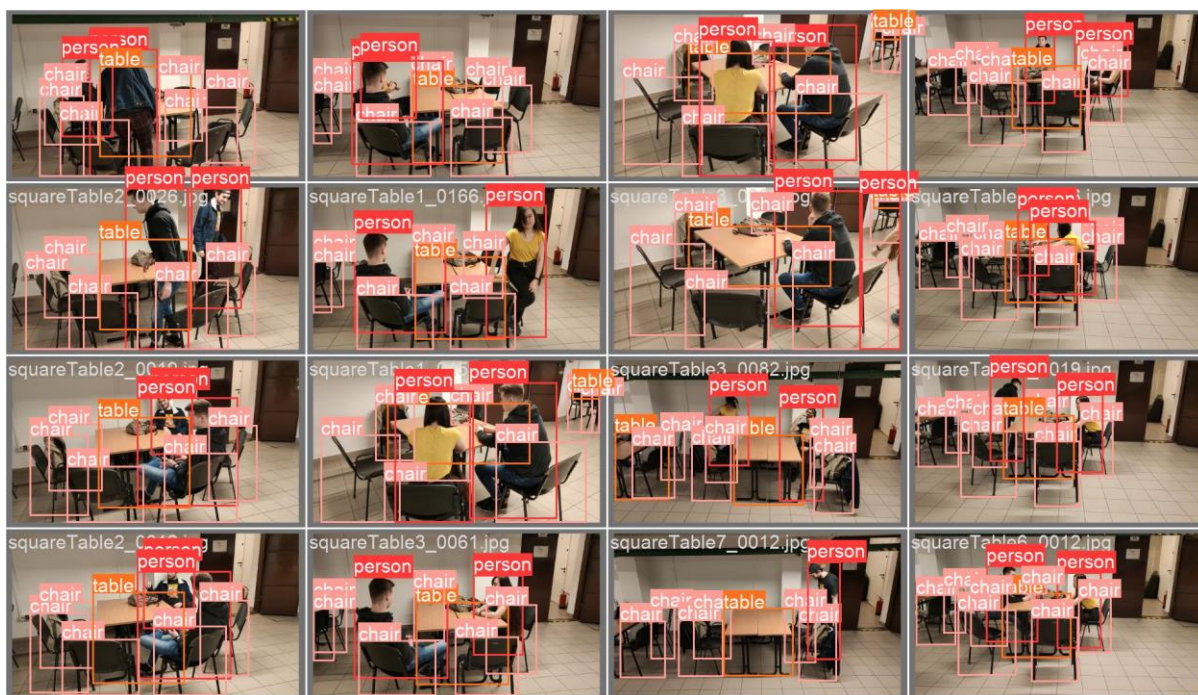
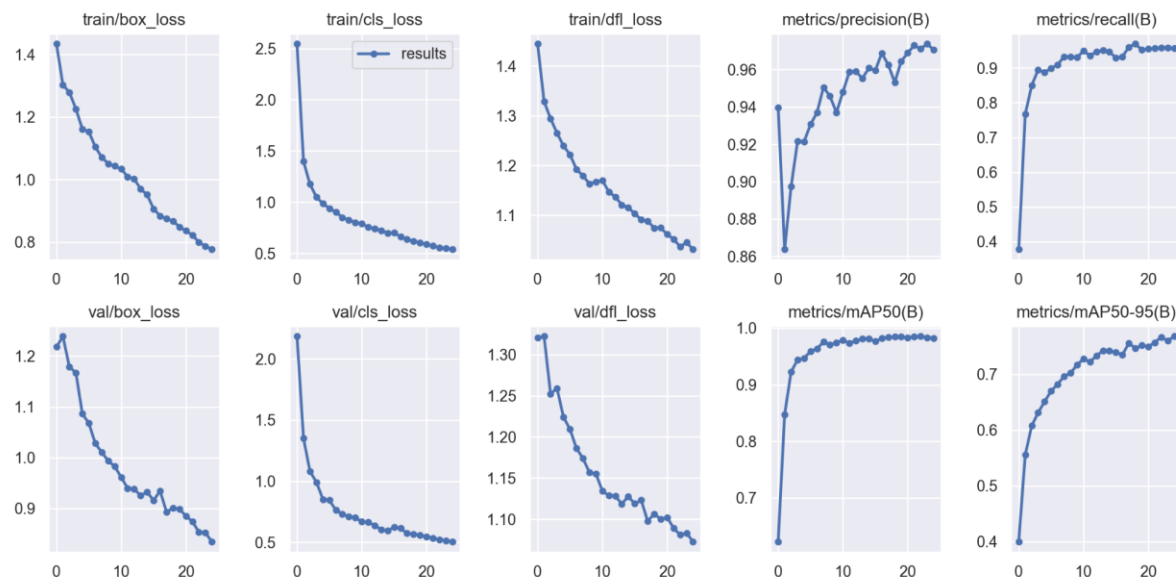
## YOLOV8 (uczone tylko danych zrobionych przez nas)











## Implementacja heat mapy

Sieć neuronowa wykrywa: osoby, stoły i krzesła. Na podstawie ich lokalizacji zwracanych przez sieć działa algorytm:

1. Znajdź osoby w okolicy krzeseł
2. Znajdź krzesła w okolicy stołów
3. Policz ile ludzi jest na krzesłach
4. Zsumuj ile ludzi jest na krzesłach podlegający pod dany stół

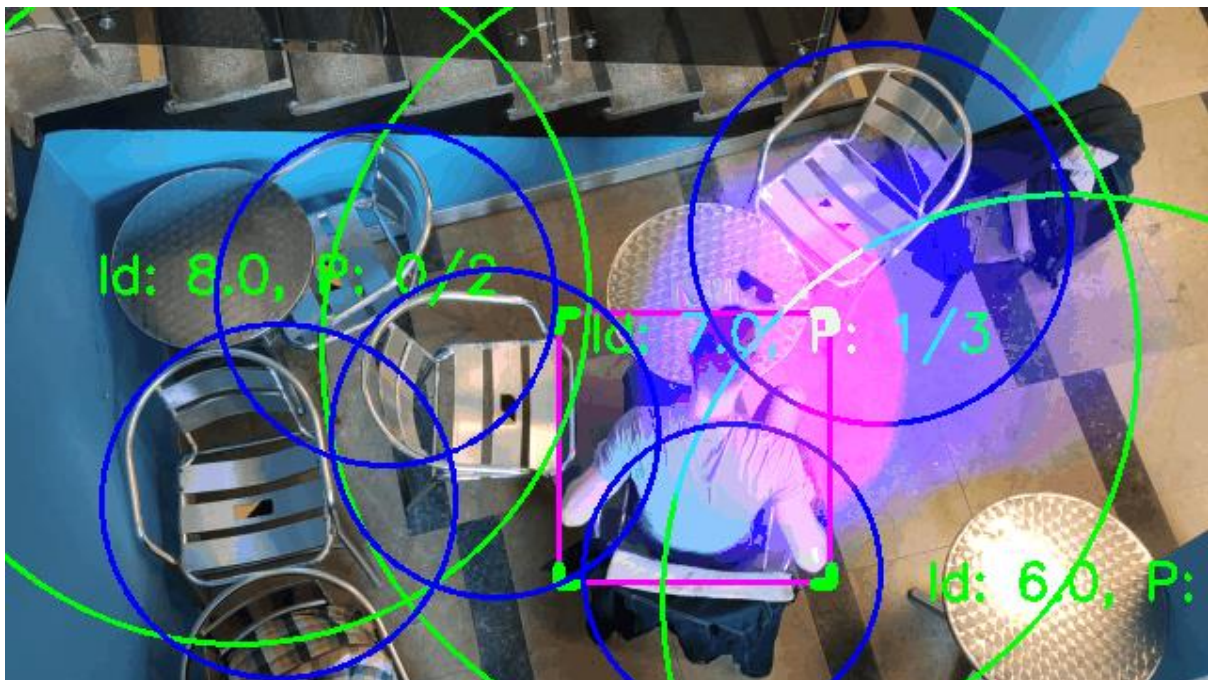
W algorytmie używany jest algorytm śledzenia obiektów. Śledzone są stoły i krzesła. Chodź ich ruch i tak nie jest duży na filmach.

Jeśli jakaś osoba została przypisana do krzesła (tzn. jej środek znajduje się odległości mniejszej niż 1.5 odległości między środkiem krzesła a jego narożnikiem), zakładamy, że owo krzesło jest zajęte. Przypisywanie krzeseł również działa w podobny sposób tylko w tym przypadku jeśli "obszary przyciągania" kilku stołów nakładają się ze sobą to krzesło przypisywane jest do tego stołu, do którego ma najmniejszą odległość.

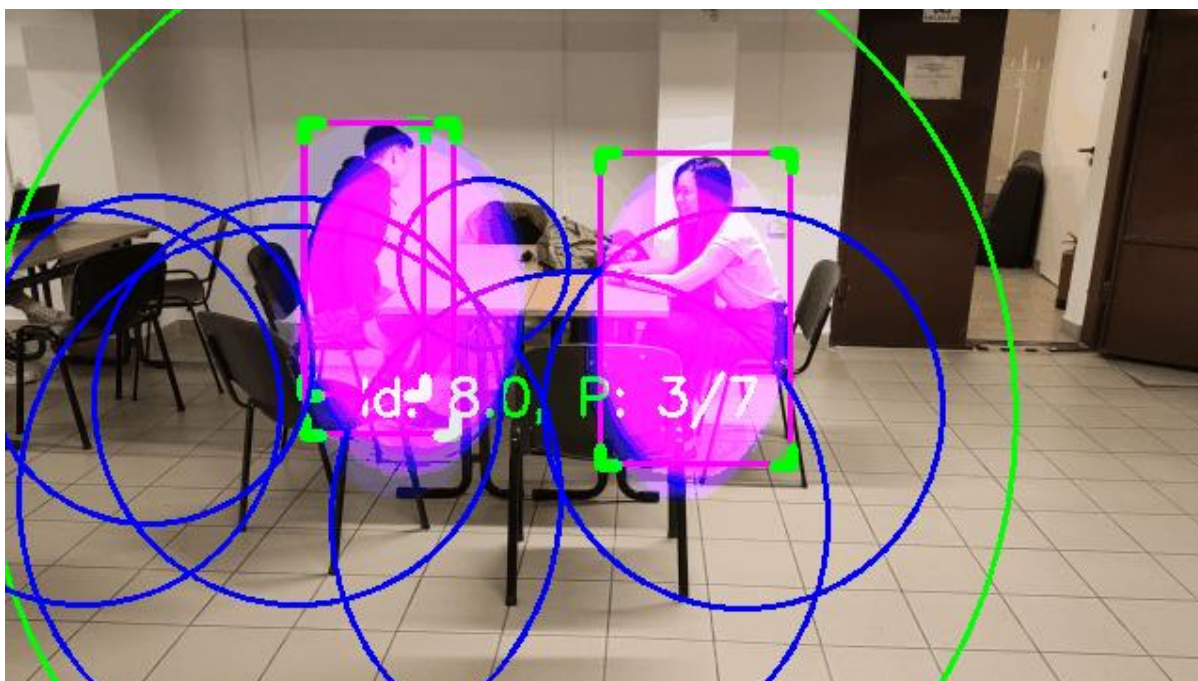
Heat mapa jest generowana w krokach:

1. Dla ramki człowieka (przypisanego do krzesła) jest wyznaczana jest eliptyczna maska (środek w środku ramki i jej krawędzie są styczne do środków boków ramki) aproksymująca człowieka w niej się znajdującego.
2. Maski ma rozkład wartości od maksymalnego w środku do minimalny na krańcach
3. Każda maska jest dodawana do głównej maski
4. Główna maska jest normalizowana w zakresie od 0 do 255
5. Na podstawie wartości głównej maski tworzone są 2 kanały jeden niebieski dla wartości  $< 127.5$  i drugi czerwony dla reszty wartości

Ostateczny wynik to analiza danej klatki i dodana do niej ostateczna kolorowa maska.







## Testowanie opracowanego rozwiązania

Testowanie opracowanego rozwiązania zostało oparte o różne zbiory materiałów przedstawiających sceny zawierające obiekty, mające być detekowane przez przygotowane modele. Na zbiory testujące składały się materiały filmowe/zdjęciowe przygotowane przez zespół w ramach projektu oraz ogólnodostępne materiały ze zbioru COCO.

Testy odbywały się poprzez dostarczenie jako zestawu trenującego zdjęć w celu detekcji obiektów na scenie (krzesła, ludzie, stoły), następnie dokonaniu niezbędnych obliczeń (liczba osób siedzących przy stole, liczba krzeseł) i sprawdzeniu zajętości poszczególnych miejsc przy stole, co w konsekwencji pozwala na określenie ilości wolnych miejsc przy danym stoliku.

Wyniki testów były analizowane pod kątem poprawności wykrywania obiektów, jakości obliczeń oraz dokładności określania ilości wolnych miejsc przy stolikach. Na podstawie tych wyników można było ocenić skuteczność opracowanego rozwiązania a w szczególności pozwoliło na dostrzeżenie różnicy pomiędzy wersjami “yolo” na bazie którego opracowane zostały badane modele.