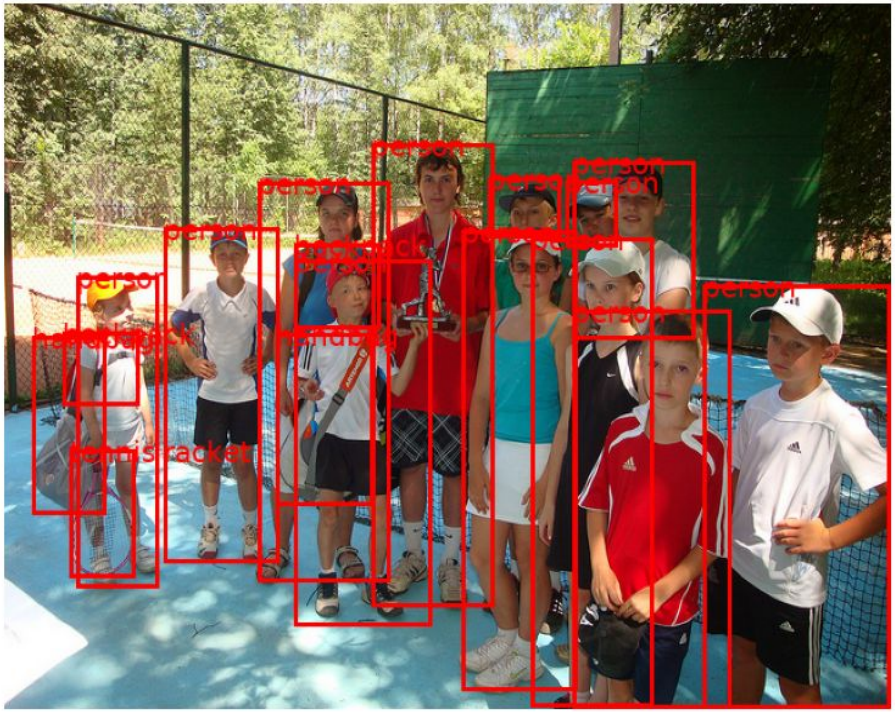
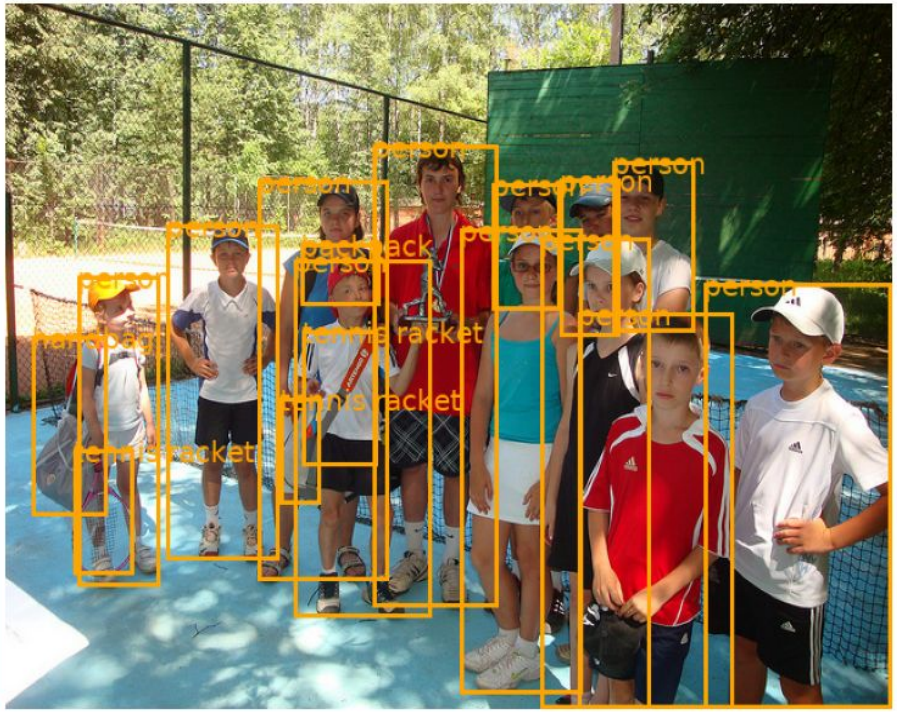
Florence 2 Analysis

Visual identification with Ground Truth and Model preds

Ground Truth Bbox



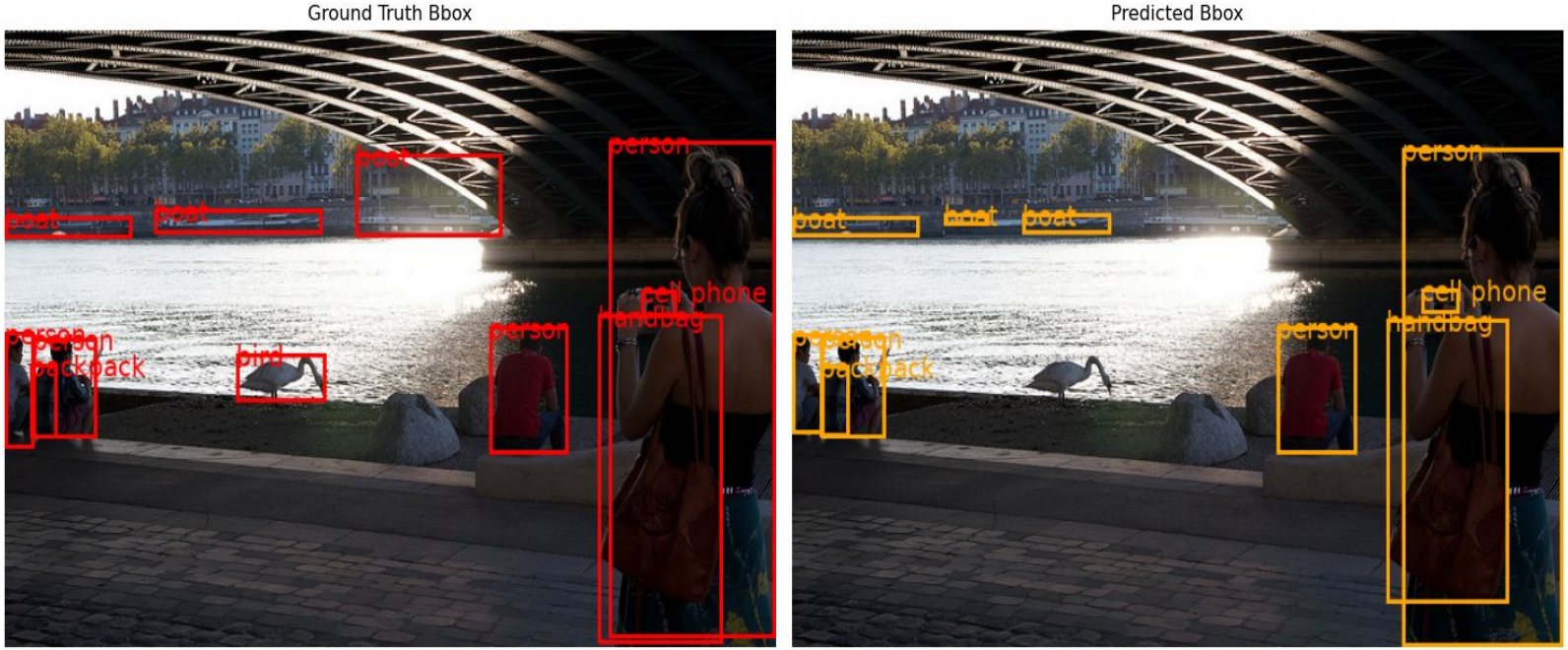
Predicted Bbox



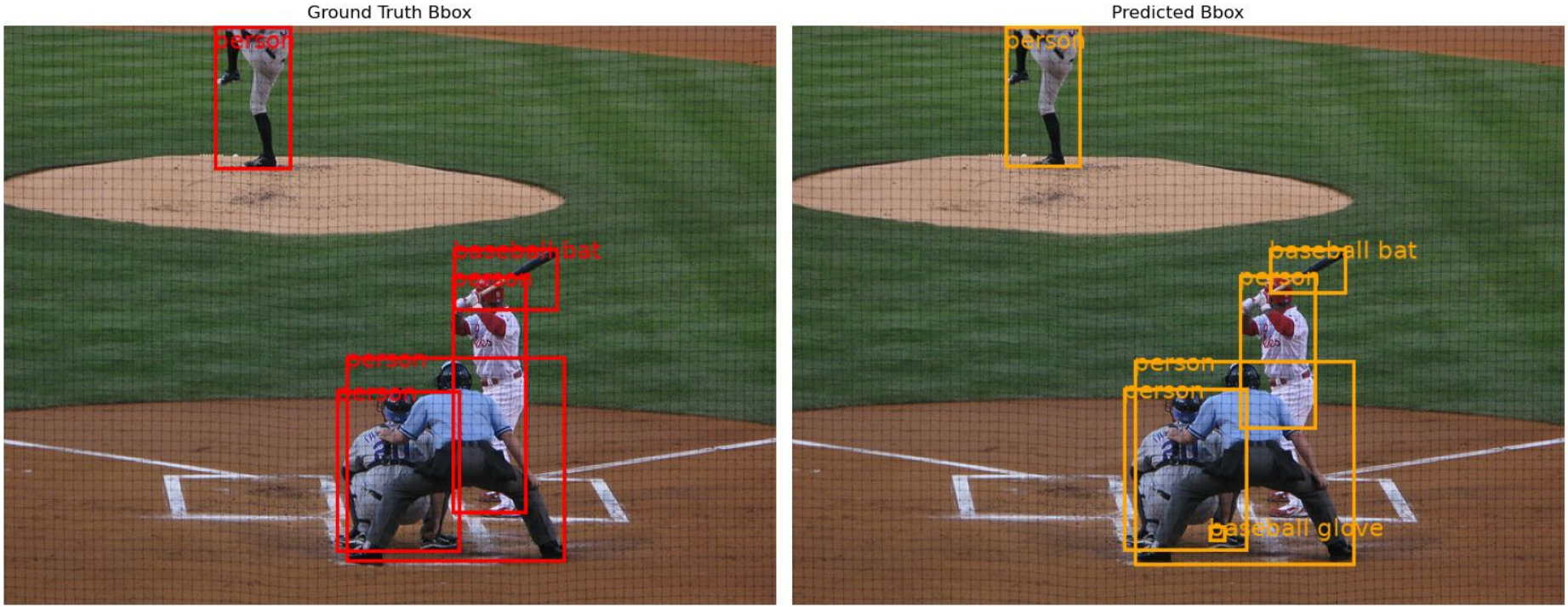
Visual identification with Ground Truth and Model preds

Visual identification with Ground Truth and Model preds

Visual identification with Ground Truth and Model preds



Visual identification with Ground Truth and Model preds

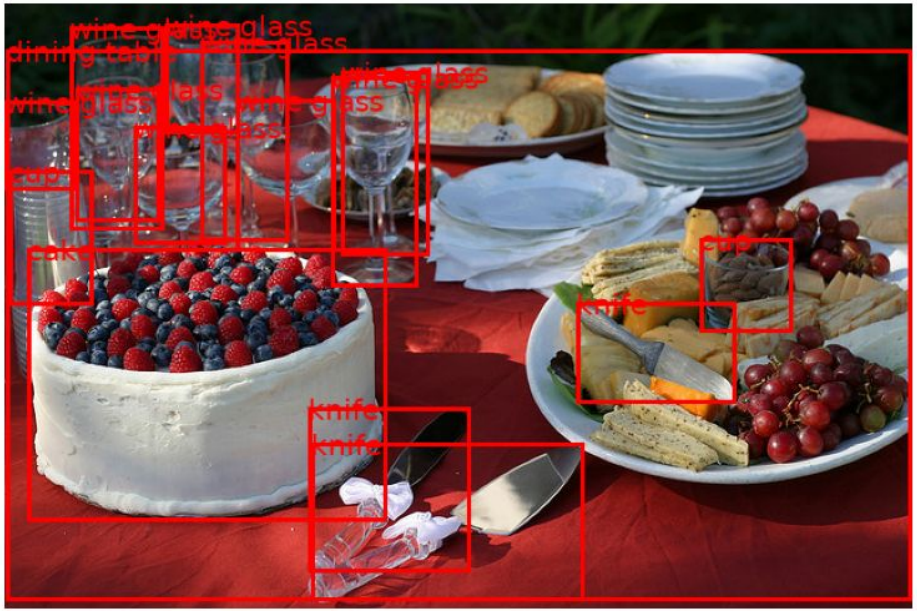


Visual identification with Ground Truth and Model preds

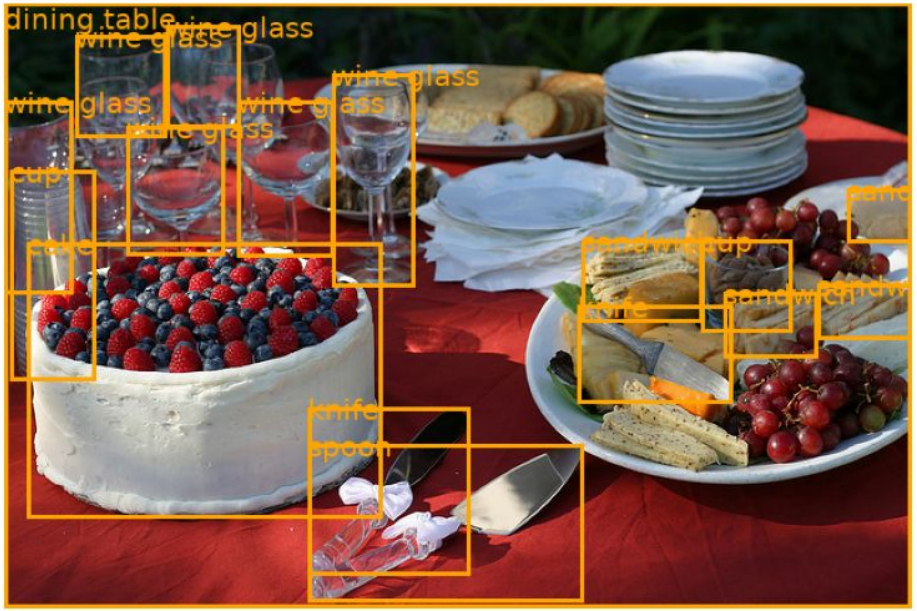
Visual identification with Ground Truth and Model preds

Visual identification with Ground Truth and Model preds

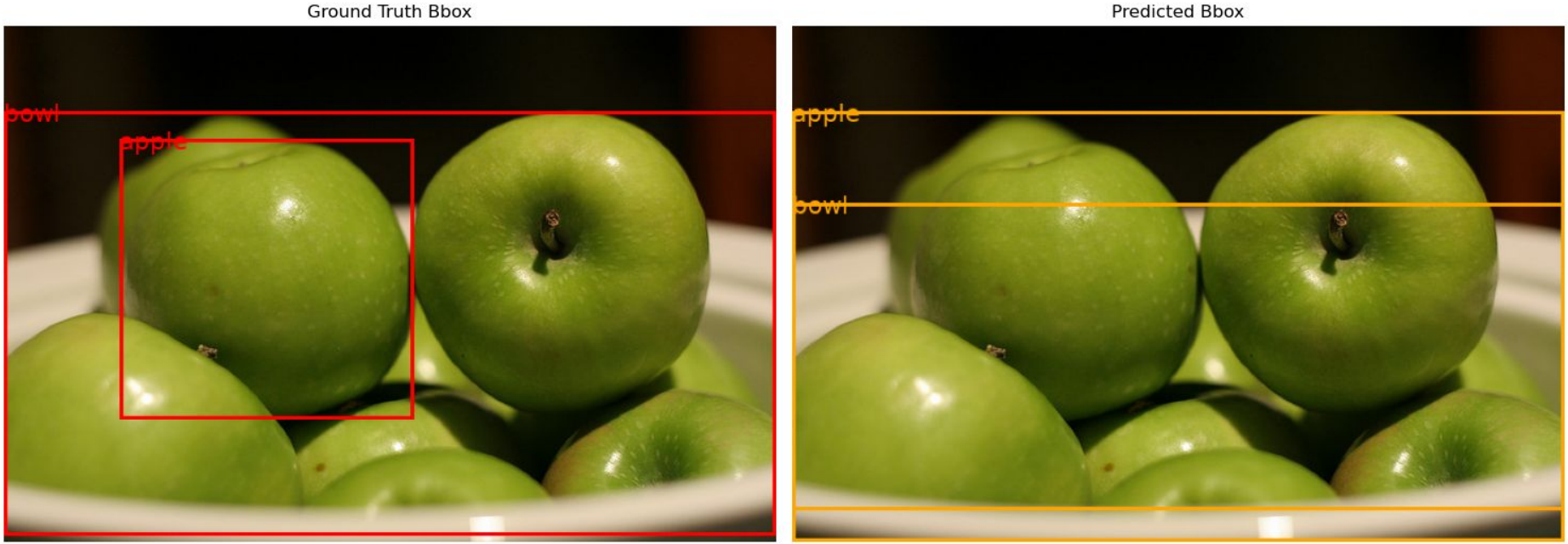
Ground Truth Bbox



Predicted Bbox



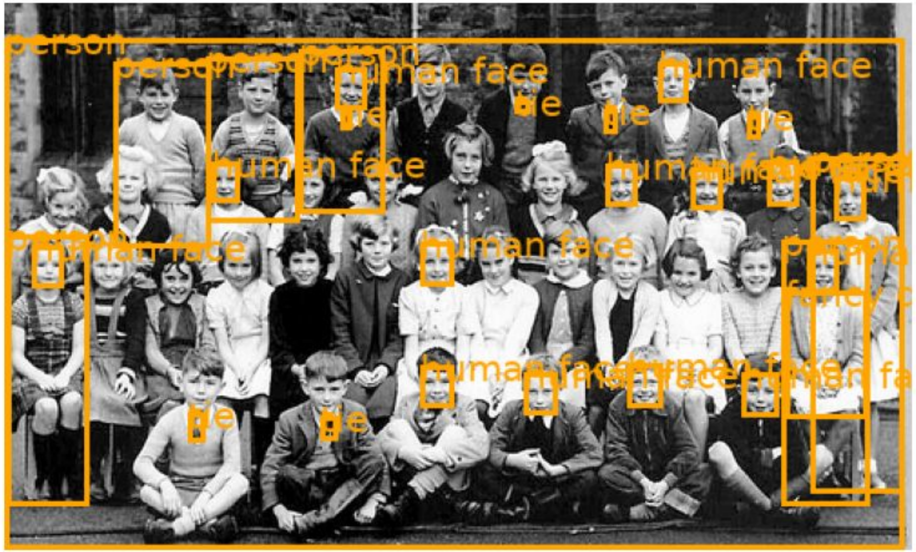
Visual identification with Ground Truth and Model preds



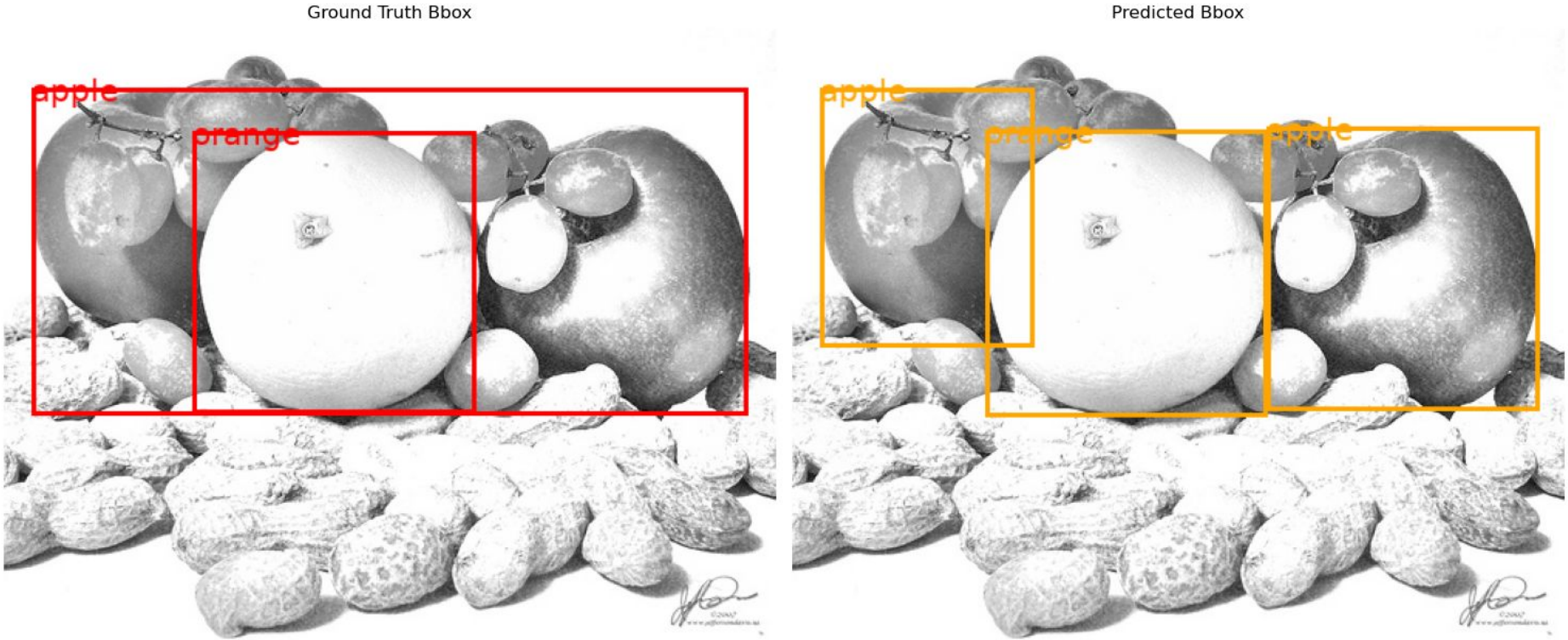
Visual identification with Ground Truth and Model preds

Ground Truth Bbox

Predicted Bbox



Visual identification with Ground Truth and Model preds



Metrics_df for 1k images

	file	TP	FP	FN	precision	recall	accuracy	f1_score
0	000000000139.csv	12	7	8	0.631579	0.600000	0.600000	0.615385
1	000000000285.csv	1	0	0	1.000000	1.000000	1.000000	1.000000
2	000000000632.csv	6	10	12	0.375000	0.333333	0.333333	0.352941
3	000000000724.csv	2	0	2	1.000000	0.500000	0.500000	0.666667
4	000000000776.csv	1	0	3	1.000000	0.250000	0.250000	0.400000
...	...	...	...	...	...	...	...	...
995	000000118594.csv	2	0	0	1.000000	1.000000	1.000000	1.000000
996	000000118921.csv	2	0	1	1.000000	0.666667	0.666667	0.800000
997	000000119038.csv	3	0	0	1.000000	1.000000	1.000000	1.000000
998	000000119088.csv	4	0	0	1.000000	1.000000	1.000000	1.000000
999	000000119233.csv	3	1	6	0.750000	0.333333	0.333333	0.461538

1000 rows × 8 columns

Accuracy, precision, recall for MS coco dataset

```python
total_TP = metrics_df["TP"].sum()
total_FP = metrics_df["FP"].sum()
total_FN = metrics_df["FN"].sum()
```

Python

```python
acc = total_TP/(total_TP+total_FP+total_FN)
p = total_TP/(total_TP+total_FP)
r = total_TP/(total_TP+total_FN)

acc,p,r
```

Python

```
(np.float64(0.5429718875502008),
 np.float64(0.7579689251962197),
 np.float64(0.6568573014991671))
```

Metrics	Value
accuracy	0.542
precision	0.757
recall	0.656

GT captions

000000000139.jpg, A living area with a television and a table
000000000139.jpg, A person standing at a table in a room.
000000000139.jpg, A woman stands in the dining area at the table.
000000000139.jpg, A woman standing in a kitchen by a window
000000000139.jpg, "A room with chairs, a table, and a woman in it."

Uploaded Image

Generated Caption:

A woman standing in a kitchen next to a table.

GT captions

000000000632.jpg,A bed and a mirror in a small room.
000000000632.jpg,a bed room with a neatly made bed a window and a bookshelf
000000000632.jpg,This room has a bed with blue sheets and a large bookcase
000000000632.jpg,"Bedroom scene with a bookcase, blue comforter and window."
000000000632.jpg,A bedroom with a bookshelf full of books.

Uploaded Image

Generated Caption:

A bedroom with a bed, dresser, mirror and bookcase.

GT captions

000000011149.jpg,"Parked bikes on a beautiful, picture perfect, sunny day."
000000011149.jpgA man sitting on a motorcycle near several bicycles with a partially visible person standing nearby.
000000011149.jpgTwo parked bikes on a sidewalk with a person riding a motor bike.
000000011149.jpgA person standing by a bicycle as a motorcycle drives by.
000000011149.jpgA man standing next to a bikes and a motorcycle.

Uploaded Image

Generated Caption:

A man standing next to a motorcycle and two bicycles.

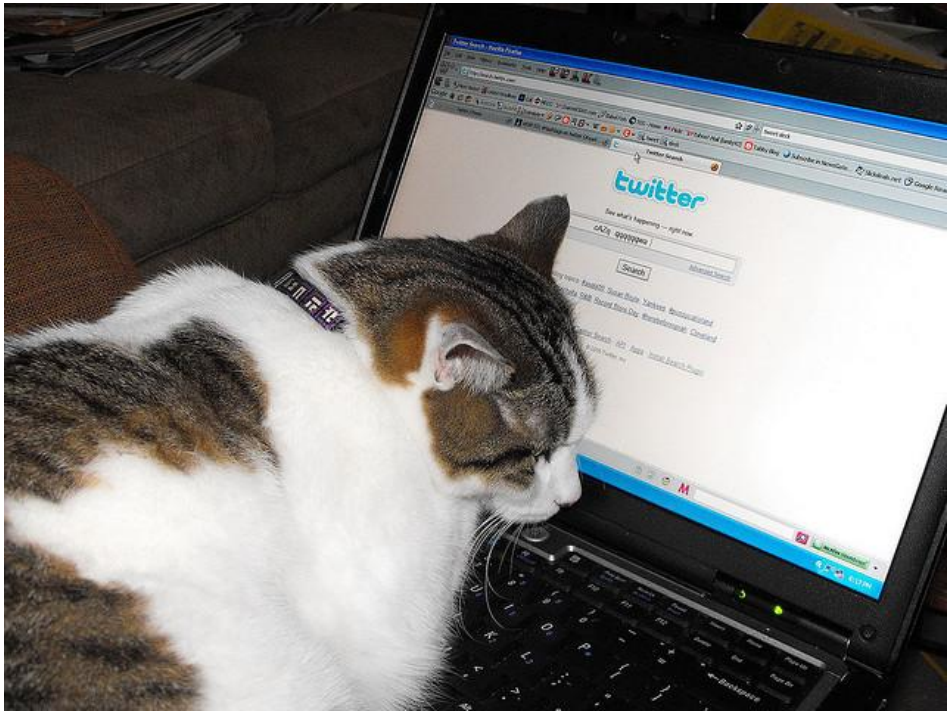
GT captions

000000016451.jpg, "Surf boards, towels chairs and bags are lying on a beach."
000000016451.jpg, Several items including two surf boards on the beach
000000016451.jpg, two surf boars on a beach near the water
000000016451.jpg, A bunch of surfboards that are in the sand.
000000016451.jpg, A sandy beach with surf boards sitting on top of it

Uploaded Image

Generated Caption:

Two surfboards and a towel on a beach.



GT captions

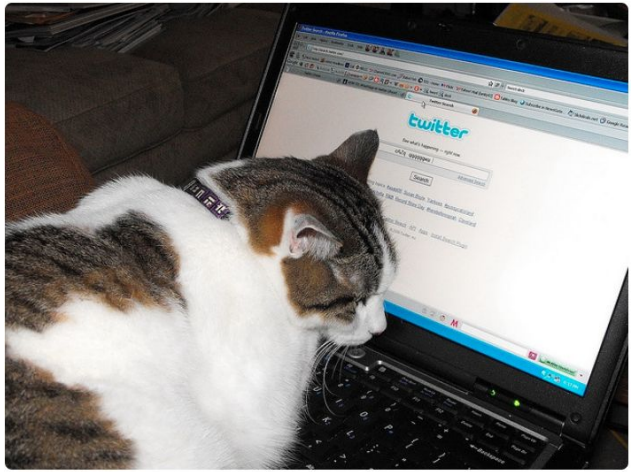
000000051008.jpg, A cat looking like it is using a laptop

000000051008.jpg, A cat that is sitting on a laptop.

000000051008.jpg, A cat sitting in front of a laptop computer.

000000051008.jpg, A cat that is checking into its Twitter account to post a tweet.

000000051008.jpg, A large cat is sitting on a laptop computer.



Uploaded Image

Generated Caption:

A cat is looking at a computer screen.

GT captions

000000052413.jpg,A person holding up a toy cell phone.
000000052413.jpg,A photo of someone holding a fake pink cell phone.
000000052413.jpg,"A child holding a pink toy cell phone with ""Princess Aurora"" inside it."
000000052413.jpg,there is a adult that is holding a pink princess phone
000000052413.jpg,Someone is holding a pick cellular flip phone.

Uploaded Image

Generated Caption:

A person holding a pink cell phone with a cartoon character on it.