

Towards General Purpose Vision Systems:
An End-to-End Task-Agnostic Vision-Language Architecture
Paper ID 4901

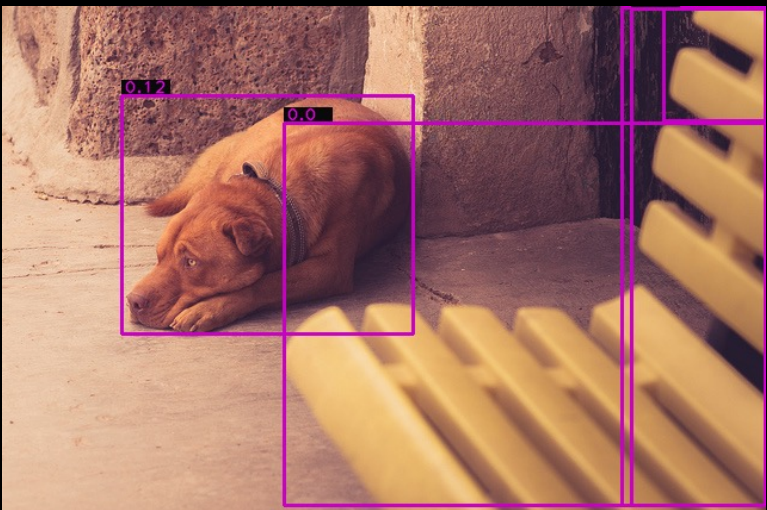
Qualitative results for GPV-I

10 randomly sampled images from the COCO-val dataset, with all associated data points across various tasks. Images sorted in decreasing order of average correctness.

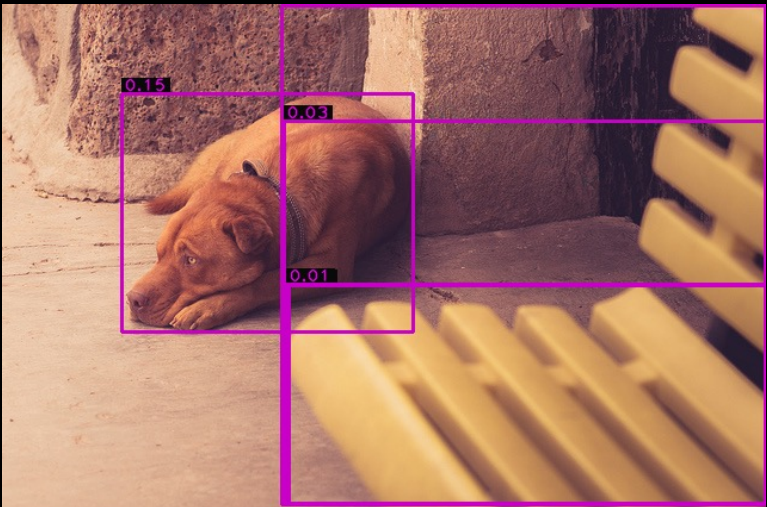
This GPV-I model was trained on the original COCO-train dataset (not COCO-SCE).

Color coding: **Correct**, **Incorrect** (with correct answer beneath it), **Partially correct** (with correct/closest answer beneath it), Unsupervised text output.

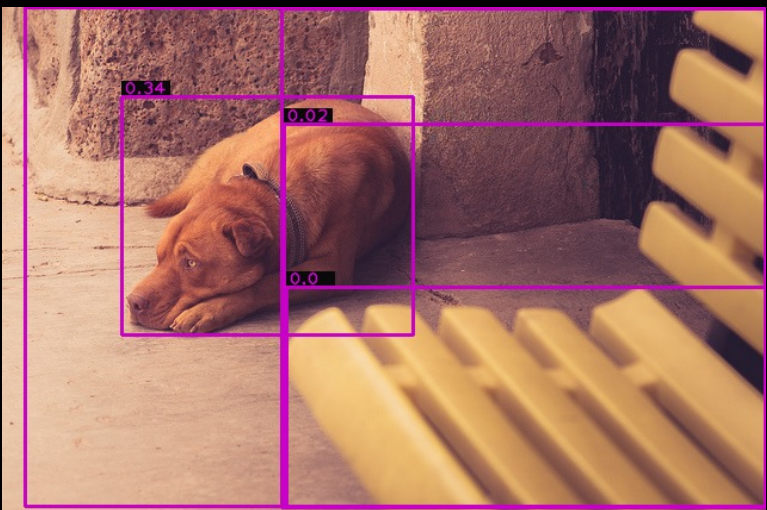
Is the dog running? **No**



Does this dog have an owner? **Yes**

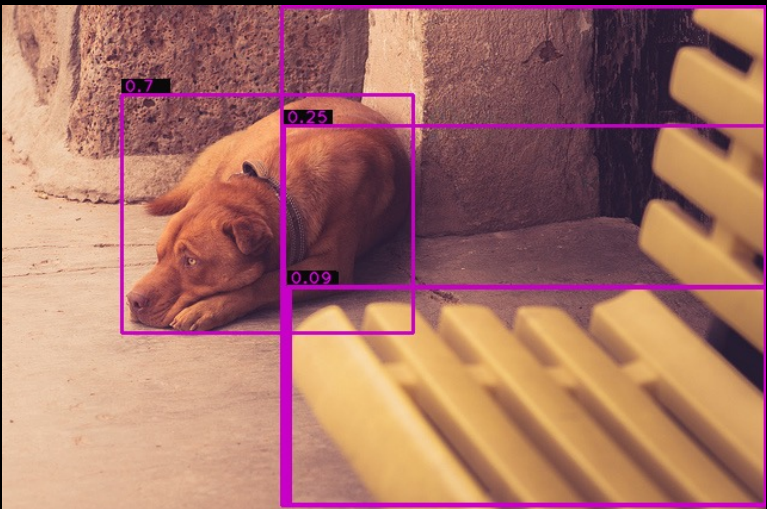


Is the dog comfortable? **Yes**



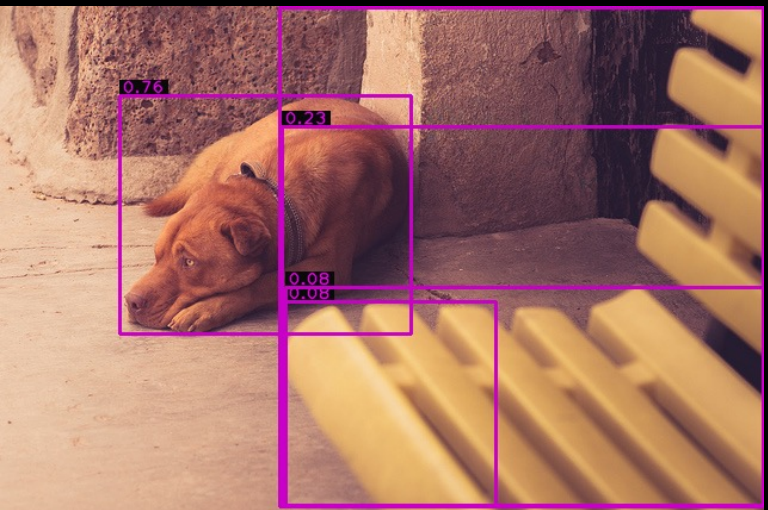
What is going on in this image.

A dog laying on a wooden floor next to a wooden bench.
(A large brown dog laying on the floor next to a wooden bench.)



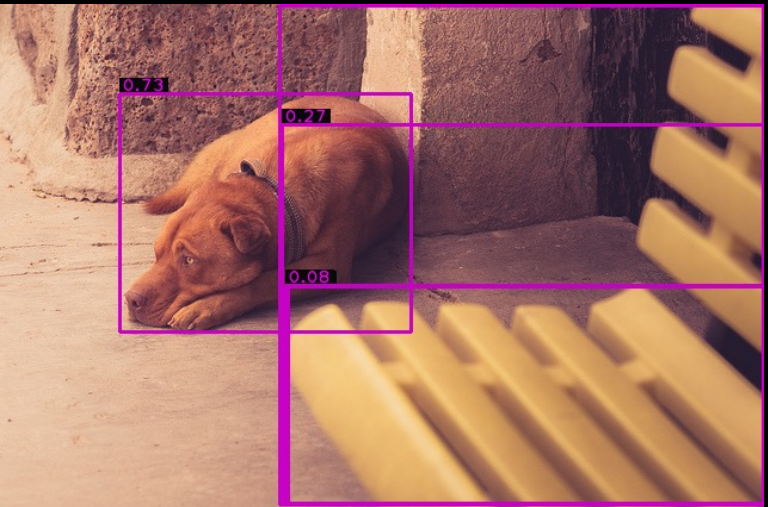
What is going on in the image.

A dog laying on a bench next to a yellow wall.
(A large brown dog laying on the floor next to a wooden bench.)



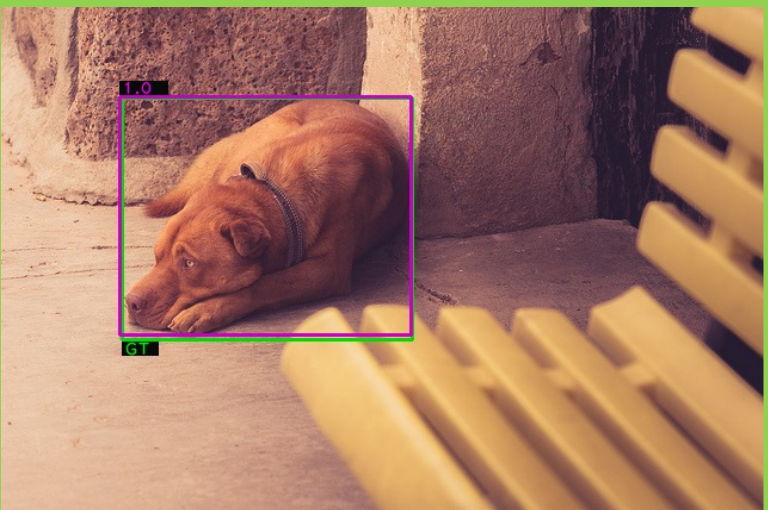
Caption the image.

A dog laying on the ground next to a bench.



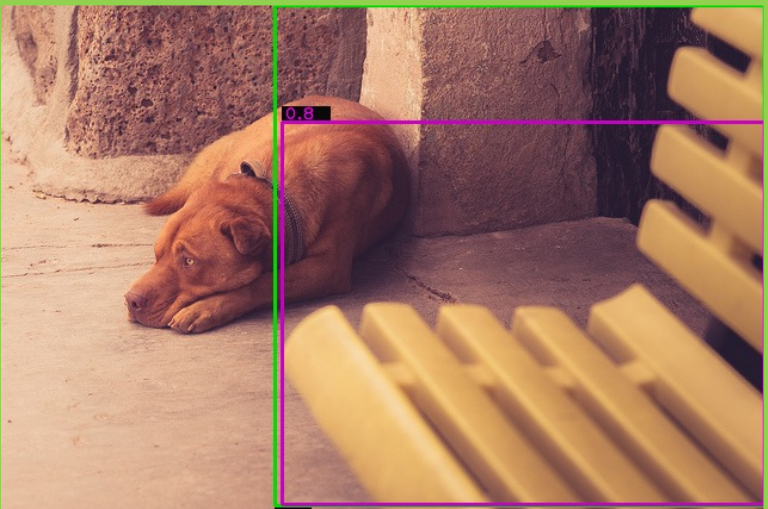
Find all instances of dog in this image.

Lab



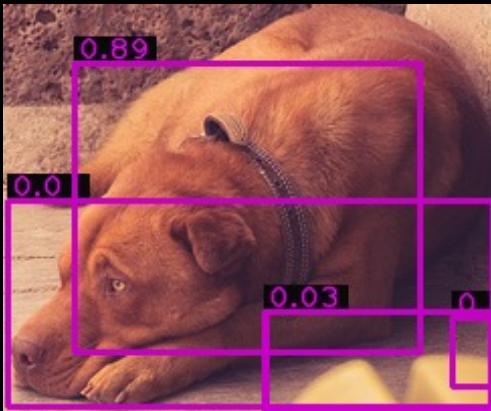
Find instances of bench in the image.

__unk__



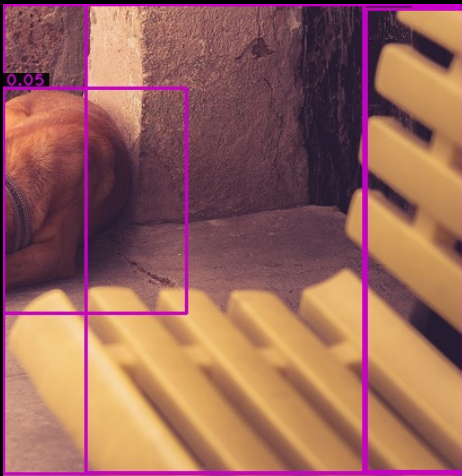
What object is this?

Dog



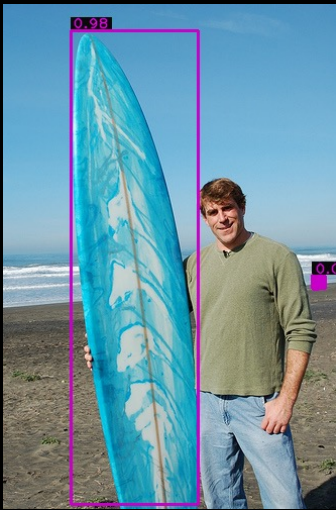
What object is this?

Bench



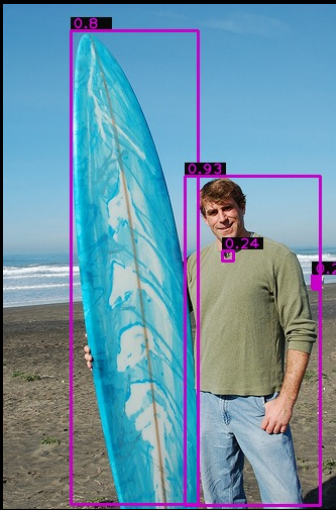
What color is the surfboard?

Blue



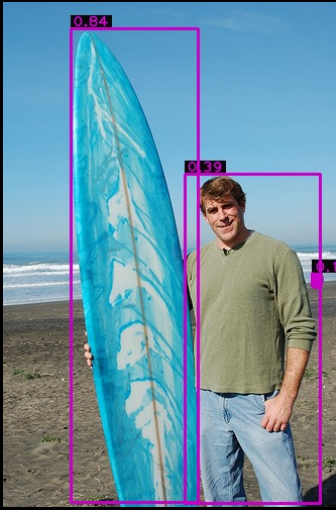
Generate a caption.

A man standing on a beach holding a surfboard.



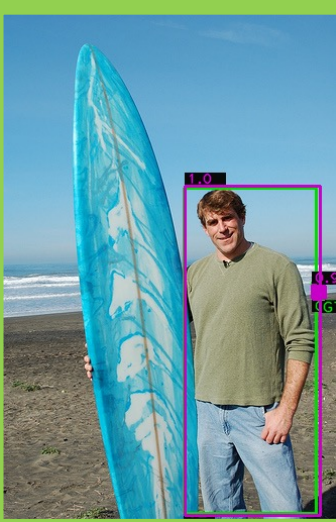
Is the man on the surfboard or next to it?

Behind
(Next to it)



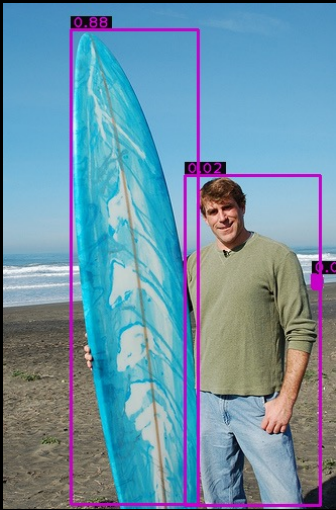
Locate all instances of person in this image.

Man standing on beach with surfboard



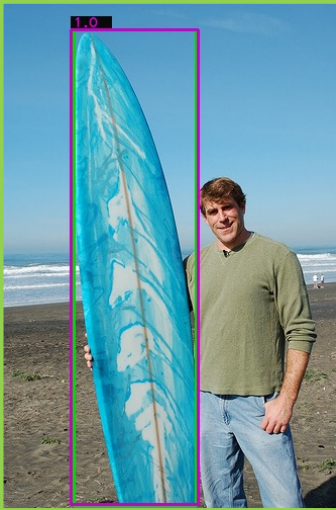
Is the man wearing the right clothes to go surfing?

No



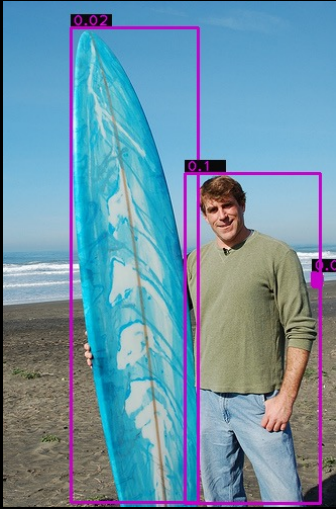
Locate surfboard in the image.

blue and white
__unk__



Is the man dressed appropriately?

Yes: 2 votes
(No: 8 votes)



What is this?

Person



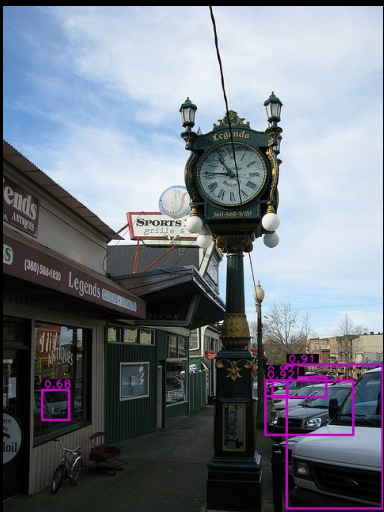
What is this?

Surfboard



Where are the cars parked?

Street: 1 vote
(On street: 2 votes)



Find all instances of chair.

A man is standing in front of a building



What time is it?

unk
(10:45)



Find all instances of car in the image.

Many people are standing outside of a building



Where is the clock?

On pole: 1 vote
(On sidewalk: 3 votes)



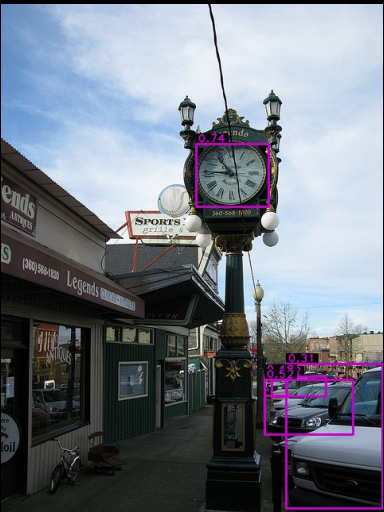
Find bicycle in this image.

1 way



Caption the image.

A clock on a pole on a street.



Locate all instances of clock in this image.

Clock tower



What is this object?

Chair



What object is this?

Car



What is this thing?

Bicycle



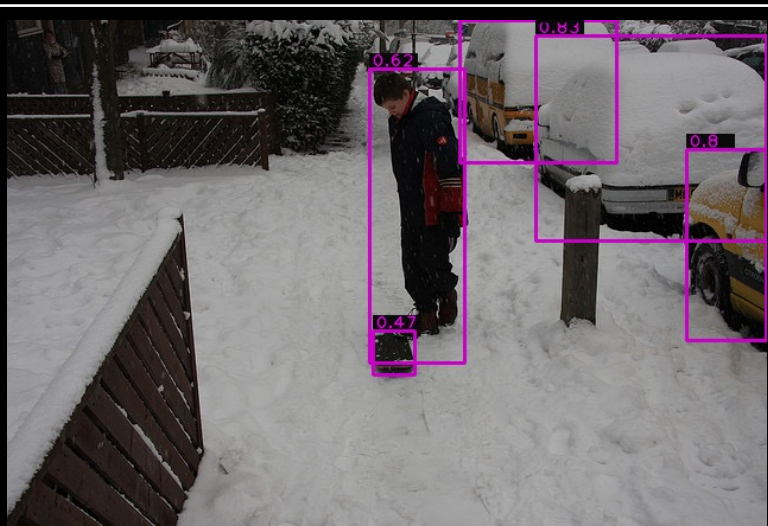
What is this thing?

Clock



Who many vehicles
are covered in
snow?

3: 1 vote
(10: 3 votes)



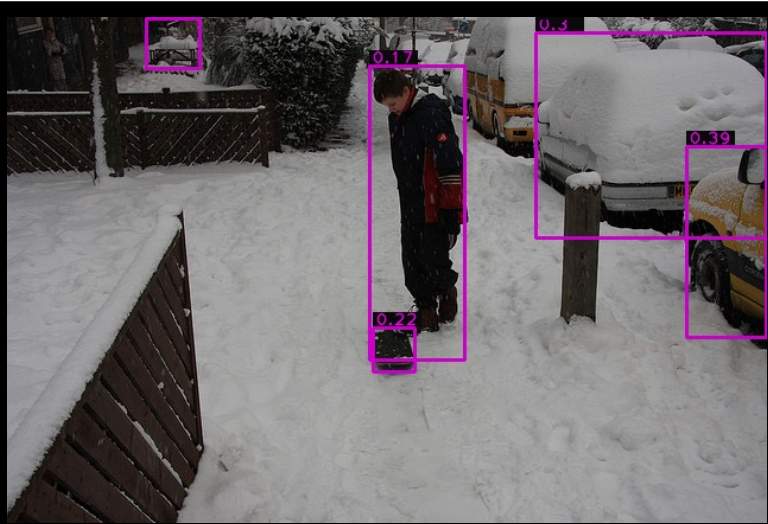
What is happening
in this image.

A person on a
snowboard on a snowy
street.



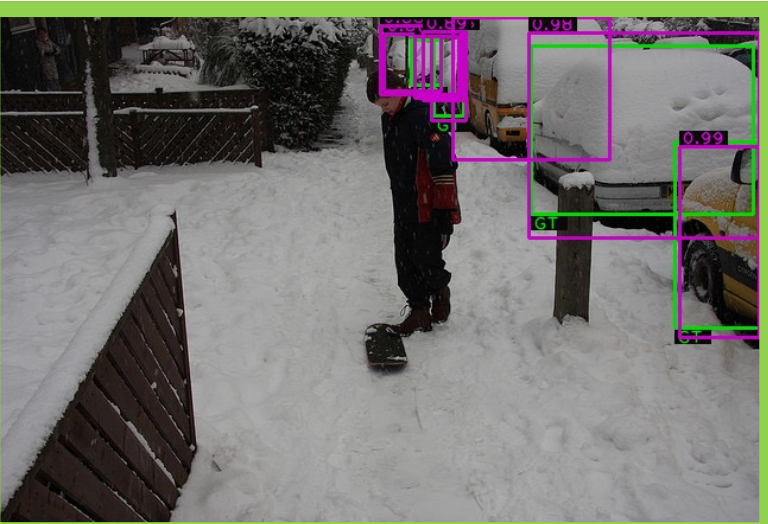
What color is the
snow?

White



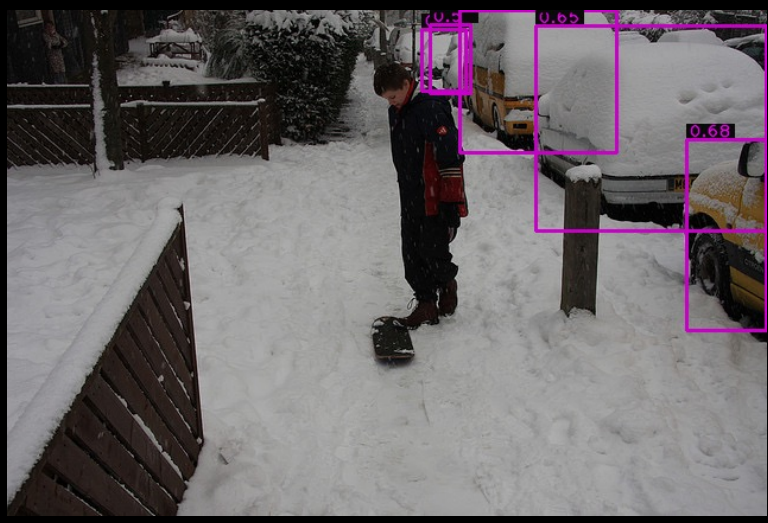
Locate instances of
car.

White __unk__ and red



What covers the car?

Snow



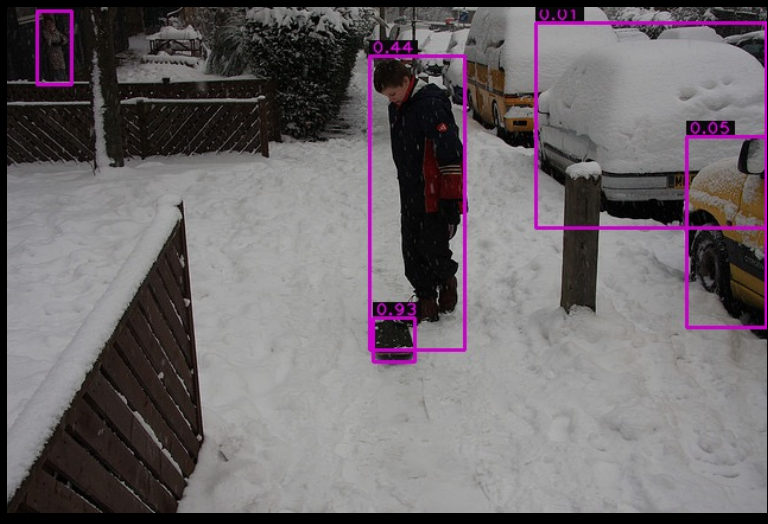
Find person in the
image.

A person riding a
snowboard on a snowy
ground



Is the boy trying to
skateboard in the
snow?

Yes



Find all instances of
snowboard.

Man's head



Locate truck

A person on a snowboard in the snow



What is this thing?

Car
(Truck)



What is this?

Car



What is this?

Person



What is this?

Snowboard



Where is the plane?

In sky
(In air)



Is the plane moving?

Yes



Is the plane landing?

No: 2 votes
(Yes: 8 votes)



How many birds are here?

0



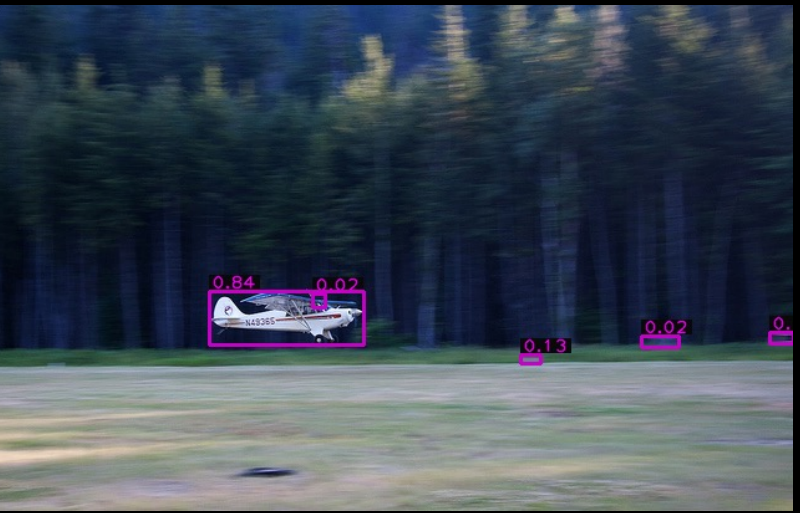
What color is the plane?

White and brown
(White and red)



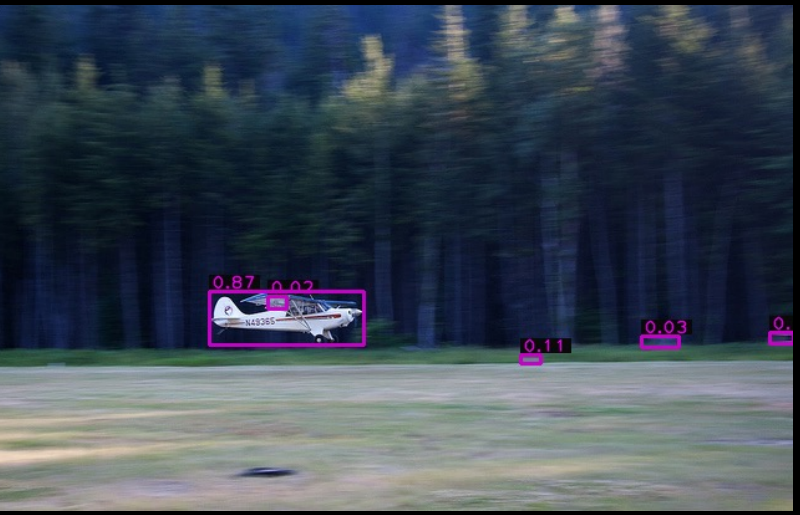
Generate a caption for this image.

A plane flying over a field of grass.



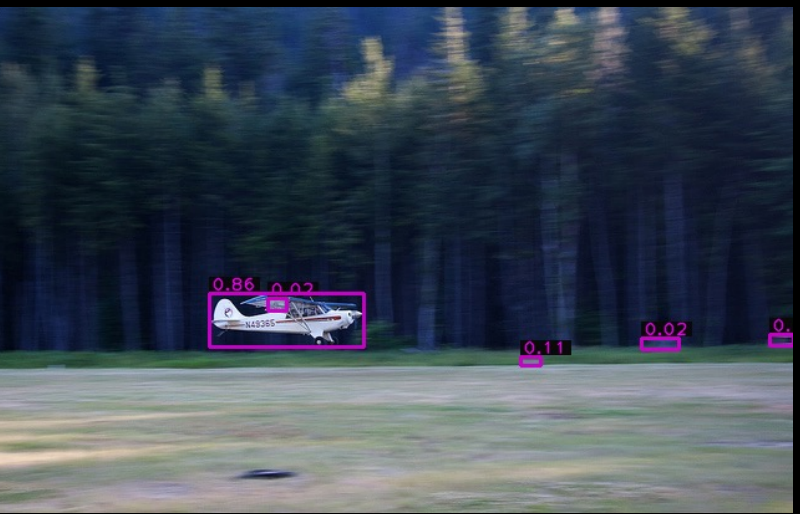
Generate a description for this image.

A small plane flying over a field of grass.



What is going on in this image.

A plane flying over a city with a sky background
(A single engine plane flying low over a field.)



Locate all instances
of airplane in the
image.

1 plane



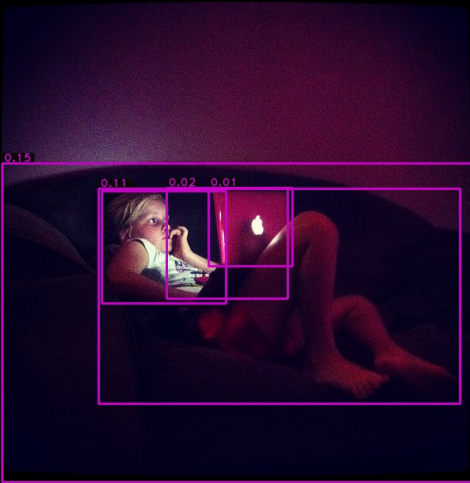
What is this thing?

Airplane



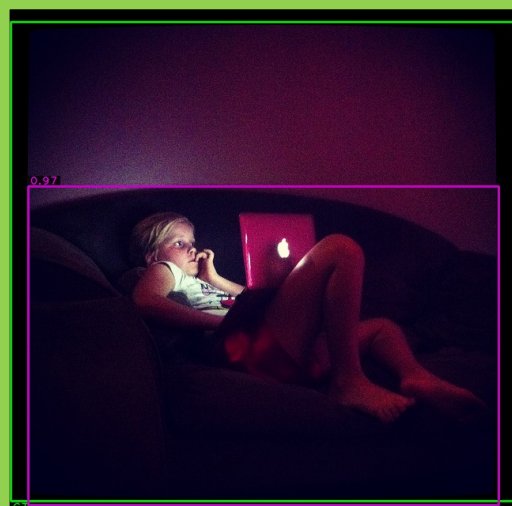
Is she tired?

Yes



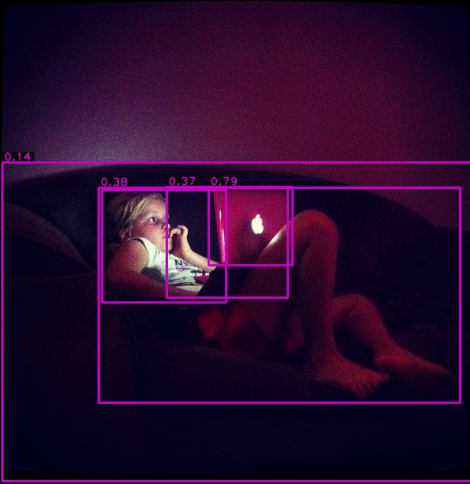
Locate all instances of couch in the image.

A man laying down on a couch



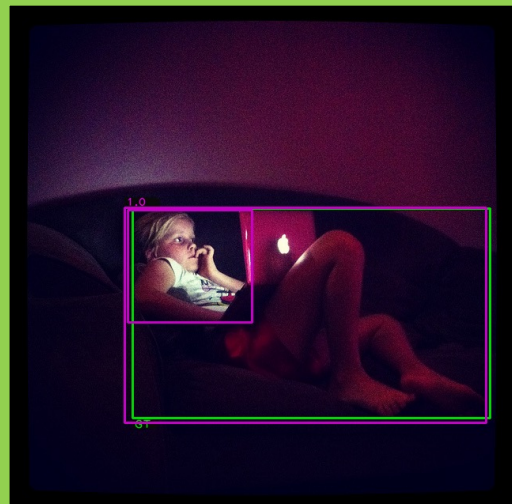
Who is using the laptop?

Woman
(Girl)



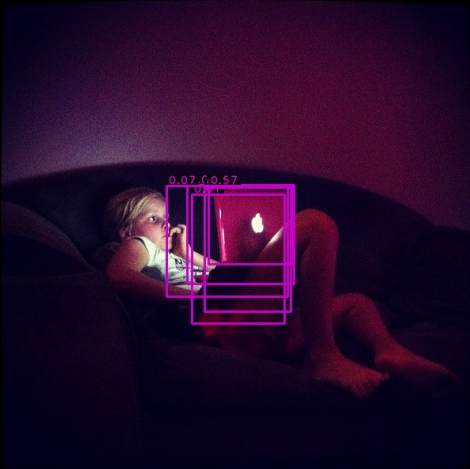
Find person in this image.

A woman laying on a couch with a laptop



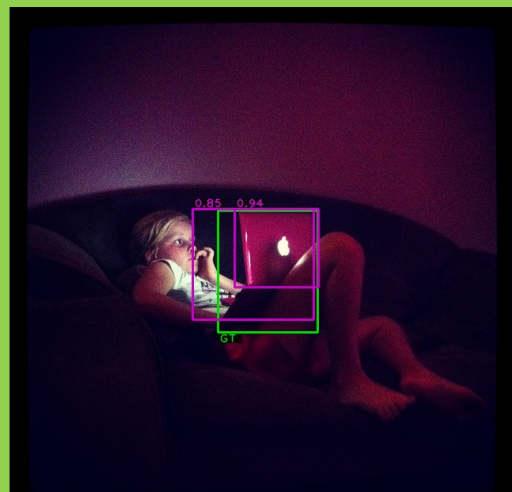
Is there a laptop in the photo?

Yes



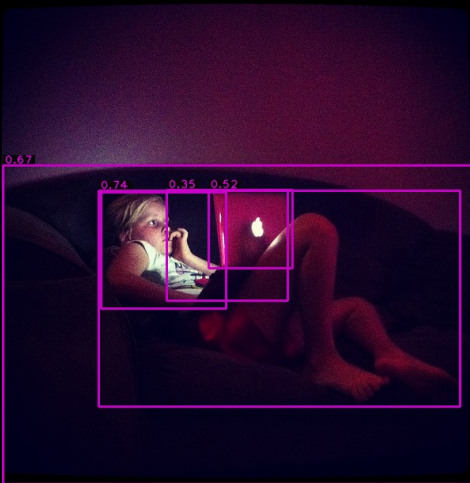
Find all instances of laptop in the image.

Apple



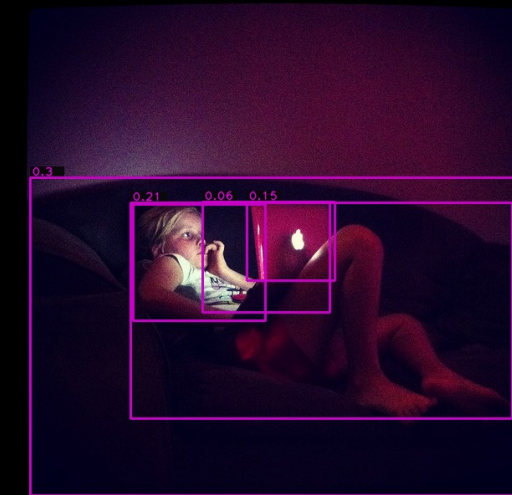
Generate a caption for this image.

A woman laying in bed while looking at a laptop.
(A little girl sits on a couch while using a laptop)



What object is this?

Bed
(Couch)



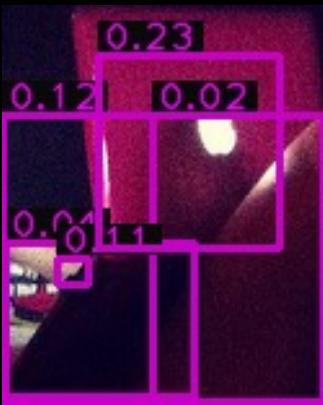
What object is this?

Person



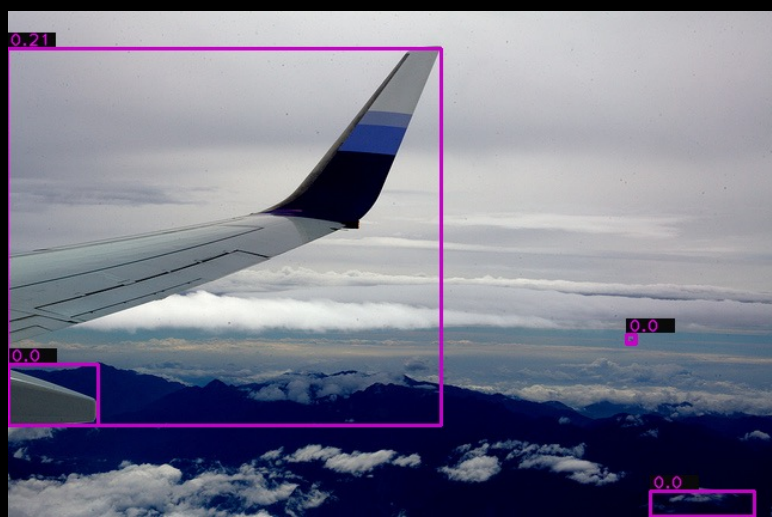
What is this thing?

Laptop



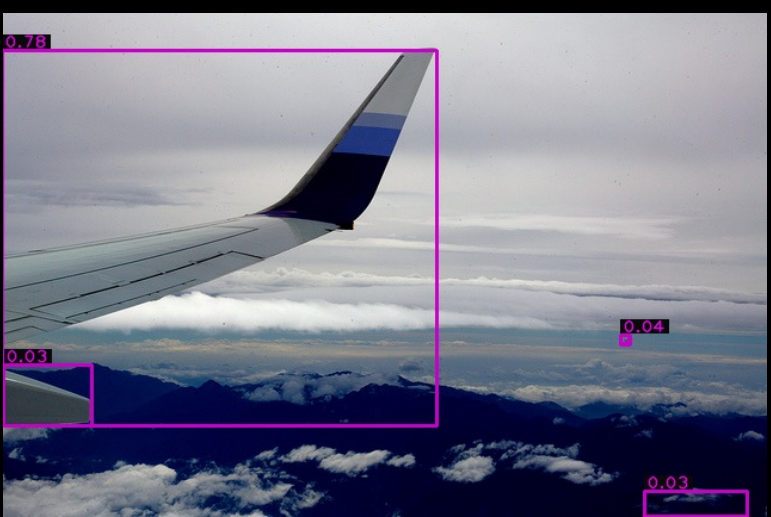
Compared to the mountains, is the plane lower or higher?

Below
(Higher)



What is happening in the image.

A plane flying over a mountain range.



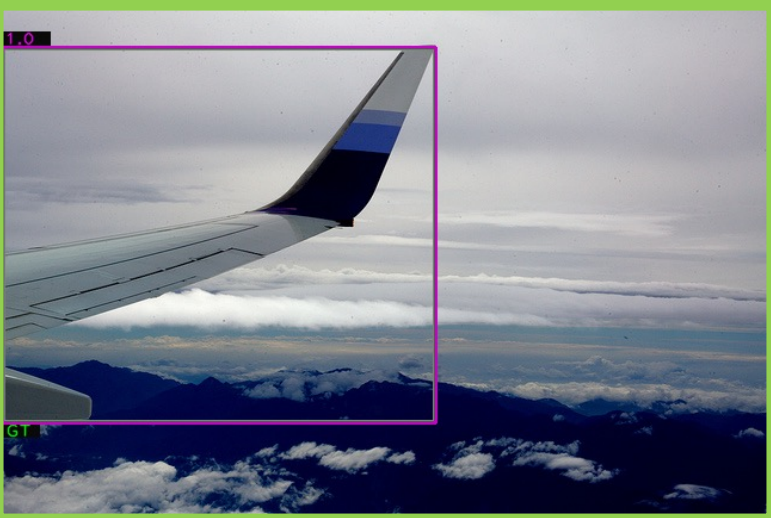
Is this plane in the sky?

Yes



Find instances of airplane.

Airplane



Do planes travel through the earth's atmosphere?

Yes



What is this thing?

Airplane



Generate a description for this image.

A view of a plane wing from the ground.
(A wing of an air plane flying in the air)



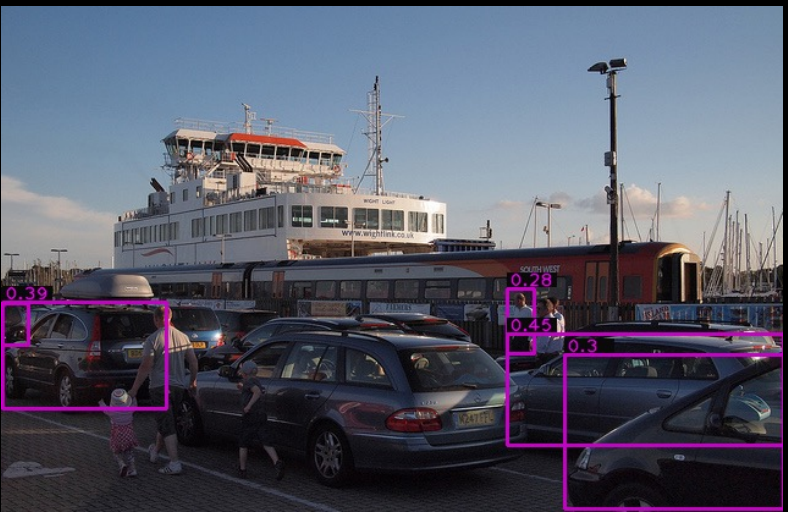
Are these cars older models?

No



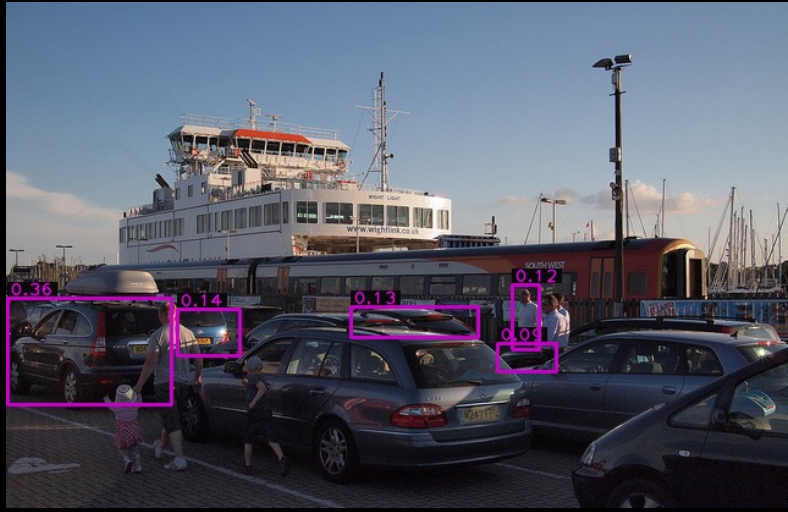
What event is the car going too?

Takeoff
(Cruise)



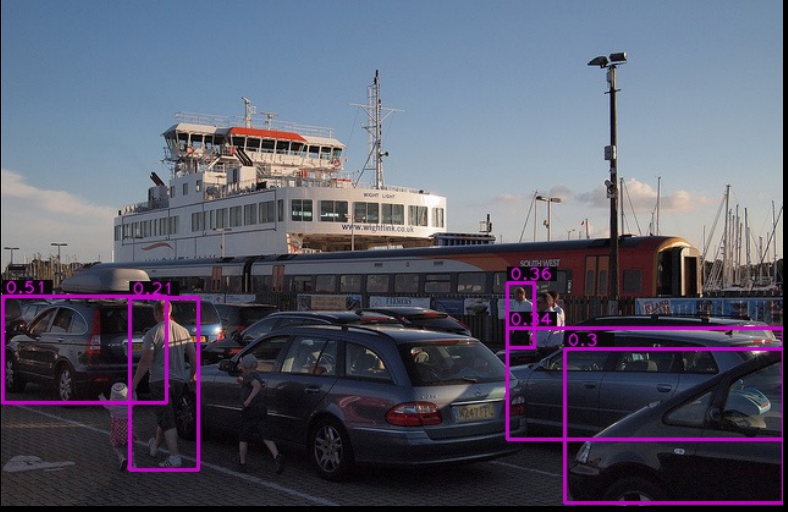
What color is the van?

Blue: 2 votes
(Gray: 3 votes)



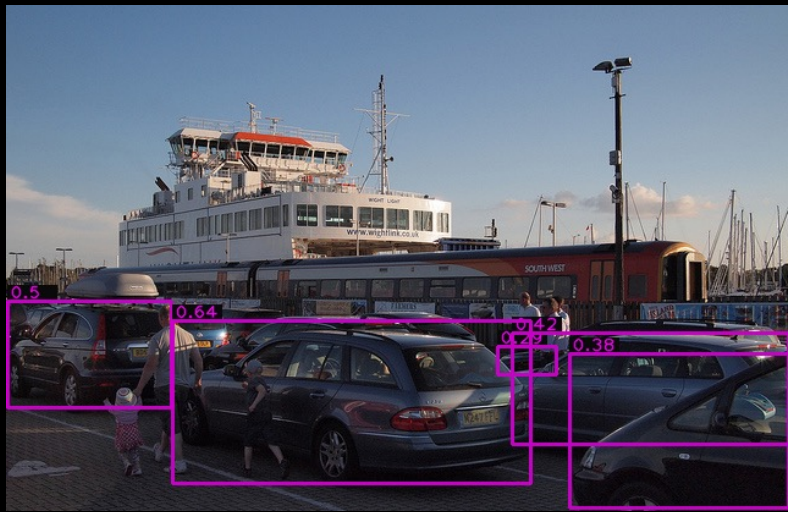
Is this a busy street?

Yes



Is there a car on the boat?

No



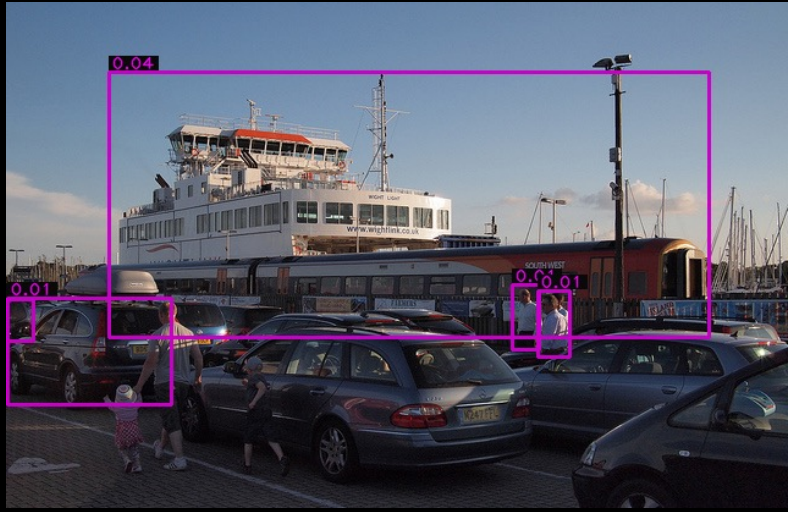
How many cars are in this scene?

6
(11)



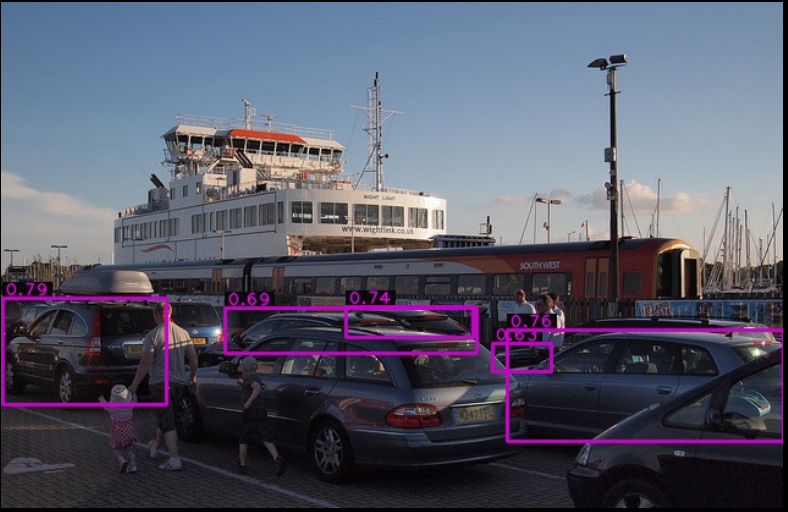
Is there water here?

No



What car brand is in the photo?

Toyota



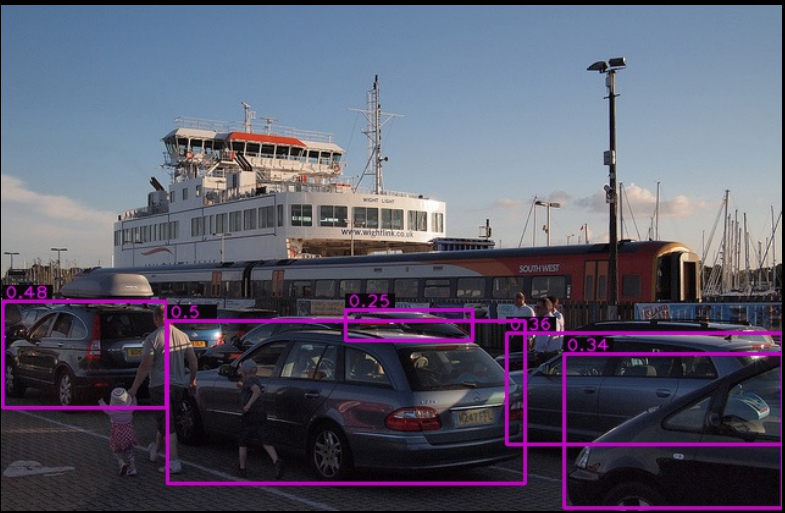
What type of boat is this?

Ferry



What kind of vehicles are these?

Boats
(Cars, train and boat)



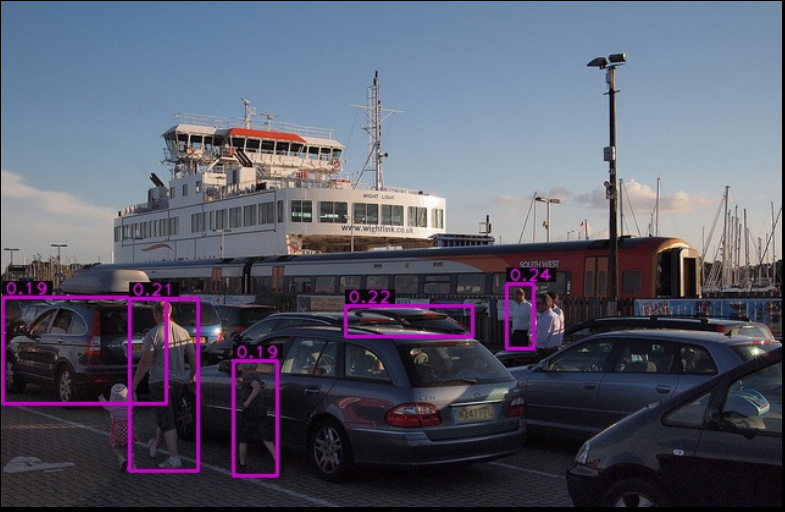
Is there a body of water in this picture?

No



Is the owner of this car a couch potato?

No



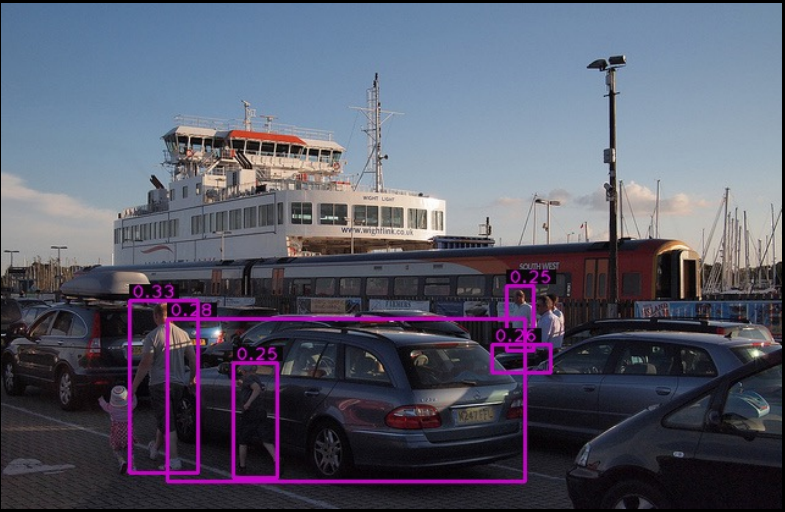
Are clean windshields important today?

Yes



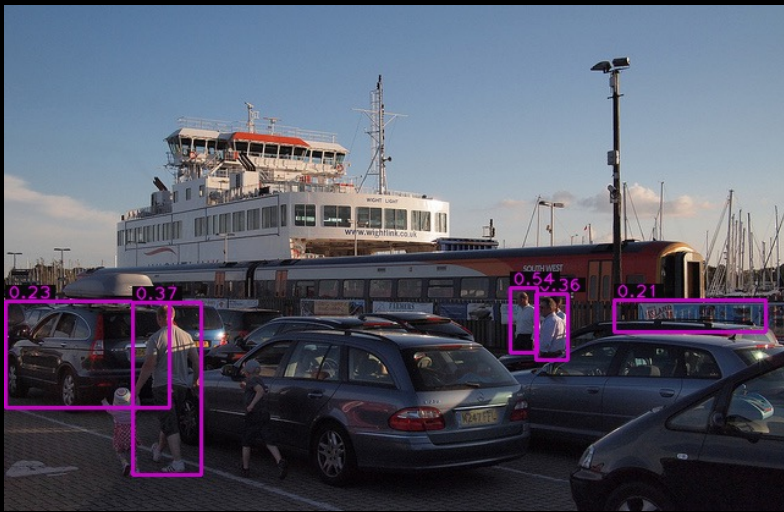
Where are the kids riding?

Parking lot
(Car)



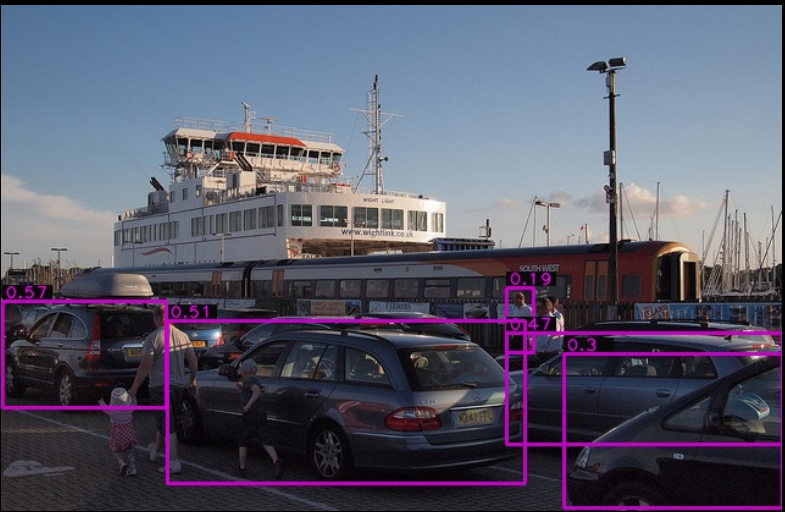
Would you rather walk or take the bus here?

Wait
(Walk)



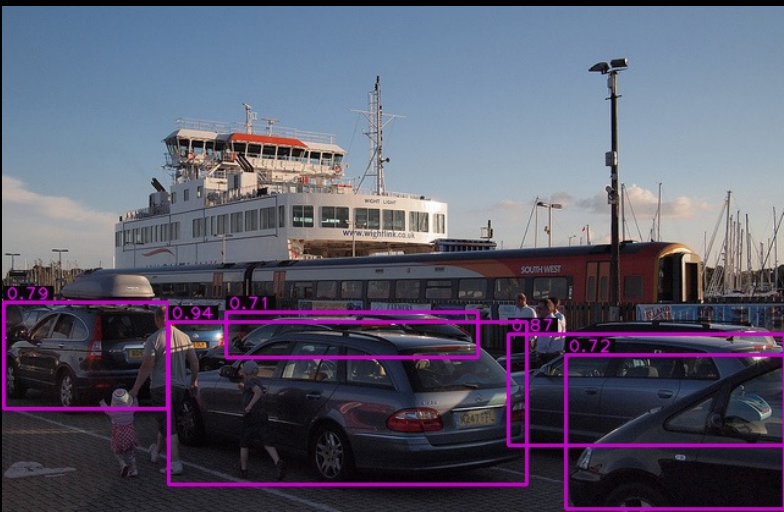
Are the cars moving?

No



What kind of car is parked?

Sedan
(Hatchback, wagon, sedan)



How many children are there?

3
(2)



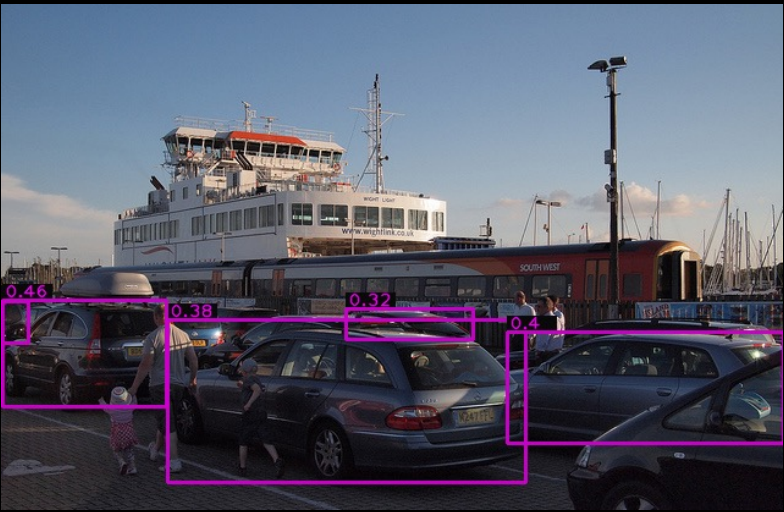
Is a train going over the traffic?

No



What type of vehicle is shown?

Boat: 1 vote
(Car: 4 votes)



Is anyone holding an umbrella?

No



Is this a real train?

Yes



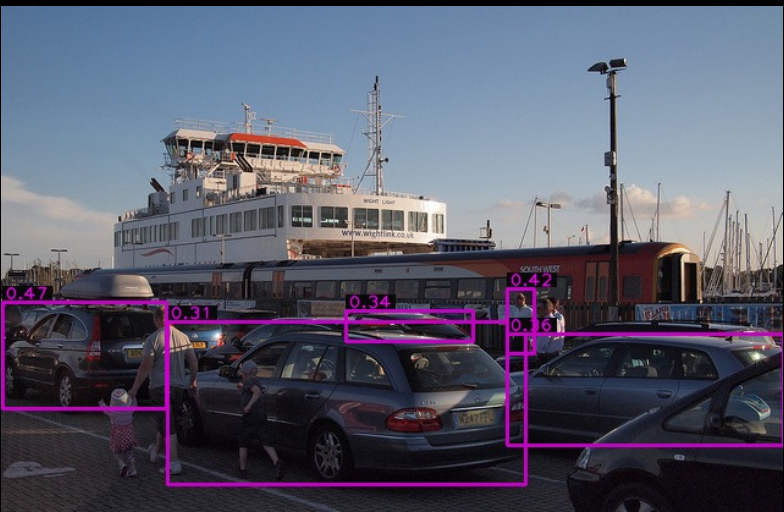
What type of transportation is this?

Boat



What is on the roof of the SUV?

Boat: 1 vote
(Luggage: 2 votes)



Who is wearing an all pink outfit?

Woman
(Girl)



How many people are here?

6



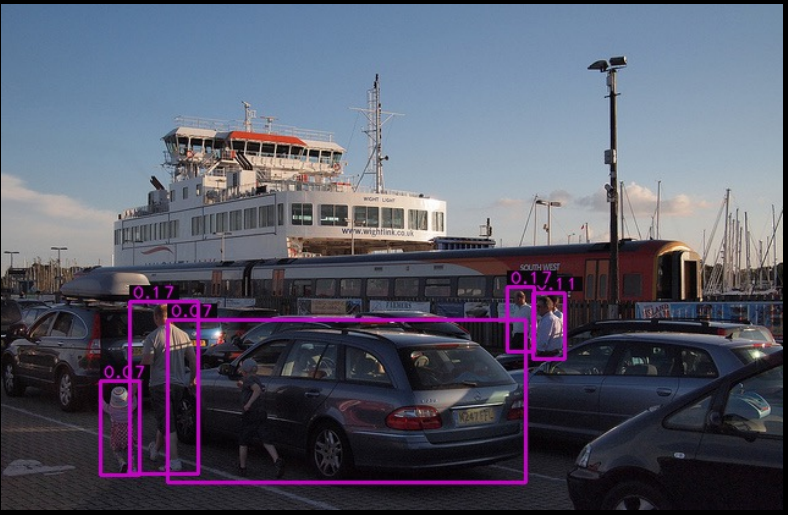
What is in the background of this picture?

Boat



Is the picture in color?

Yes



Are they at an intersection?

No



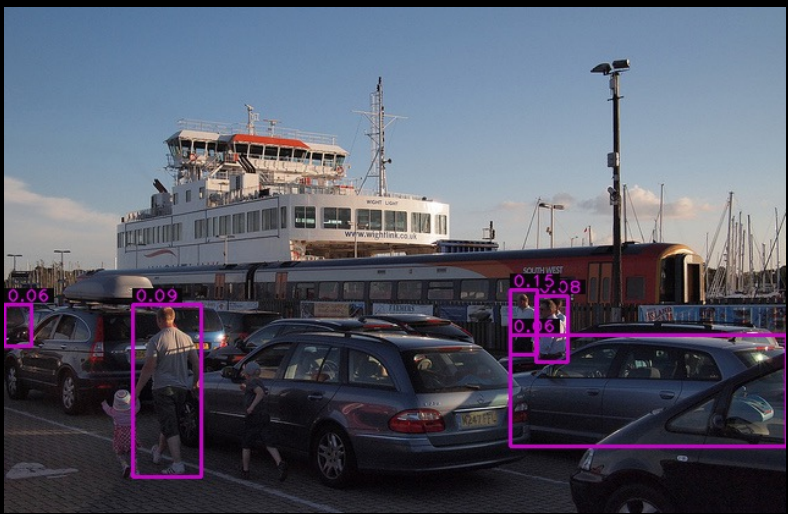
How many different types of transportation?

4: 1 vote
(2: 5 votes)



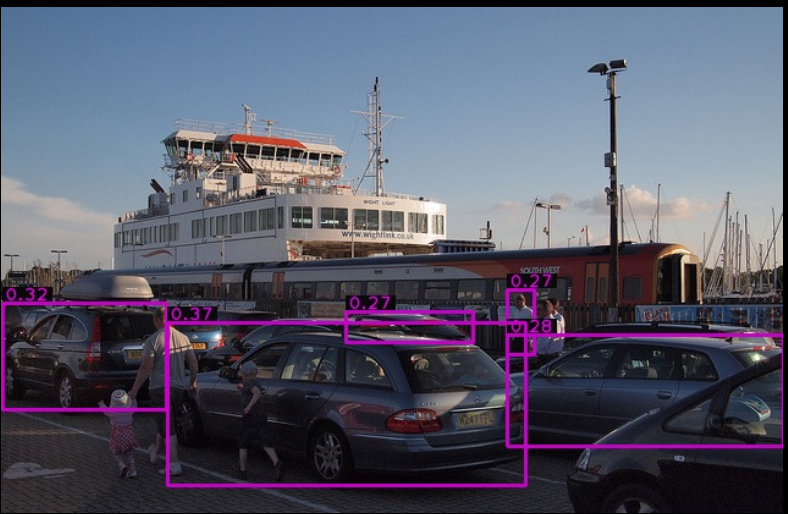
Could this be a house of worship?

No



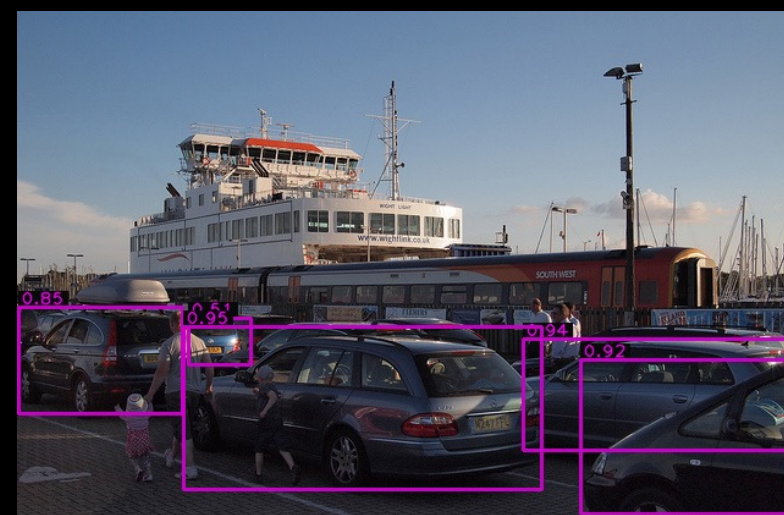
What type of vehicle is this?

Boat



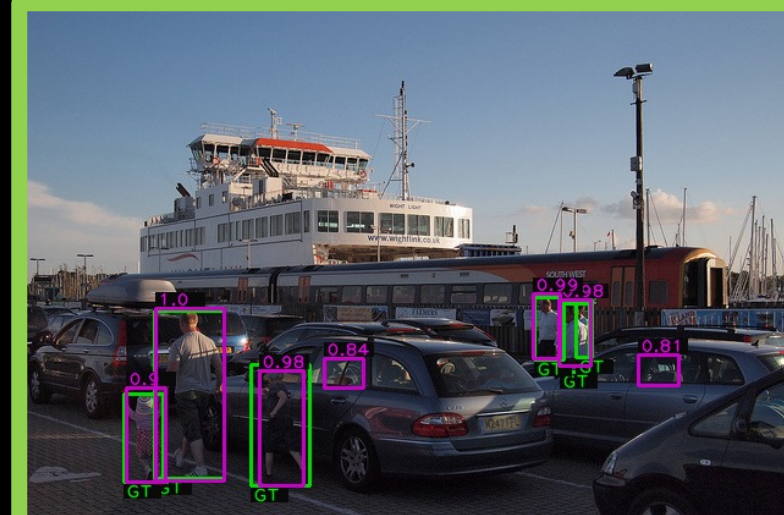
How many vehicles
are in the picture?

8



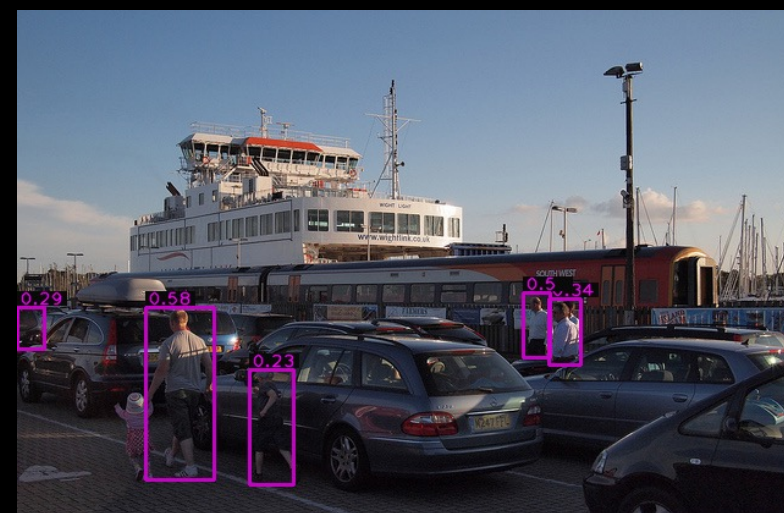
Find person.

*People are walking on
the sidewalk near cars
and trains.*



What is in the background?

Boat



Find boat in the image.

Ferry



Generate a caption.

A large boat is parked
on the side of a road.



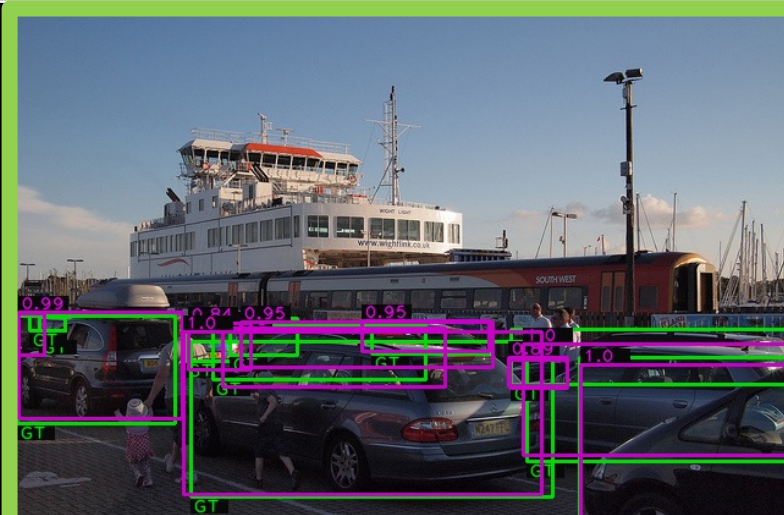
Locate all instances
of train in the image.

1 __unk__



Locate car in this image.

unk



What is this thing?

Car



What is this?

Person



What is this thing?

Boat



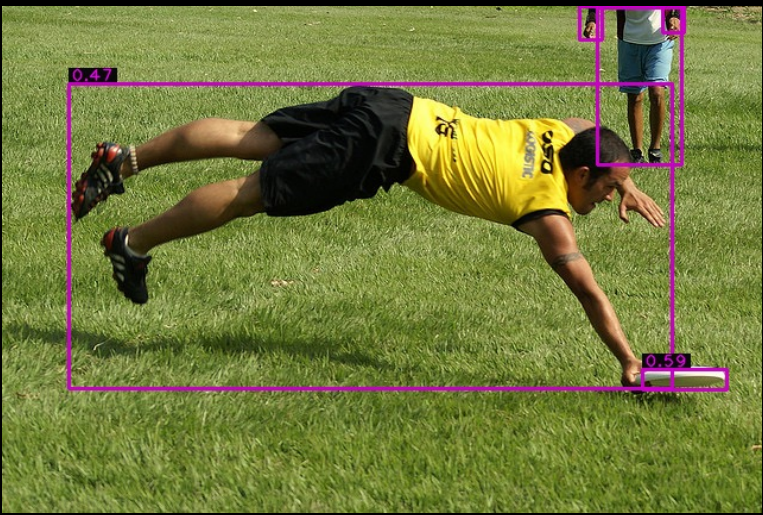
What is this thing?

Train



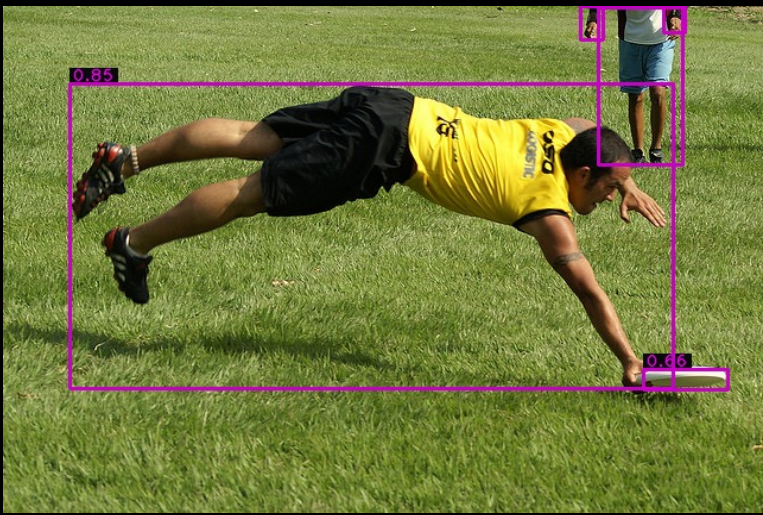
What is the man catching?

Frisbee



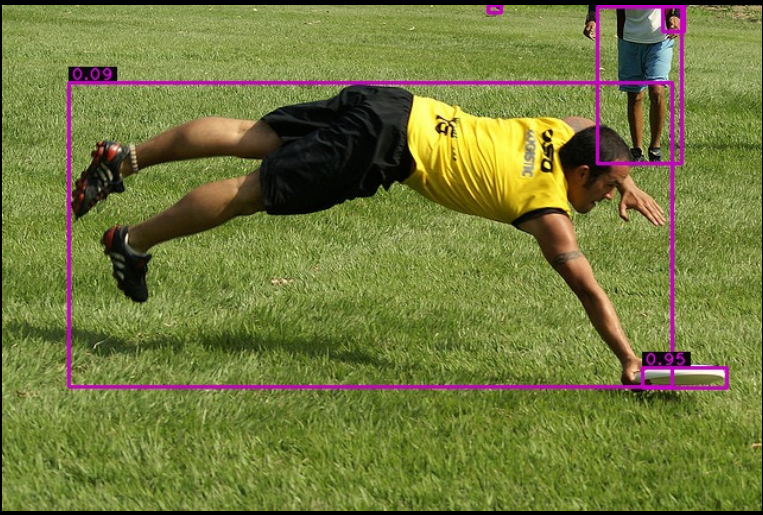
Generate a description for this image.

A man dives to catch a frisbee in a grassy field.



Which man has the Frisbee?

Right
(Yellow shirt)



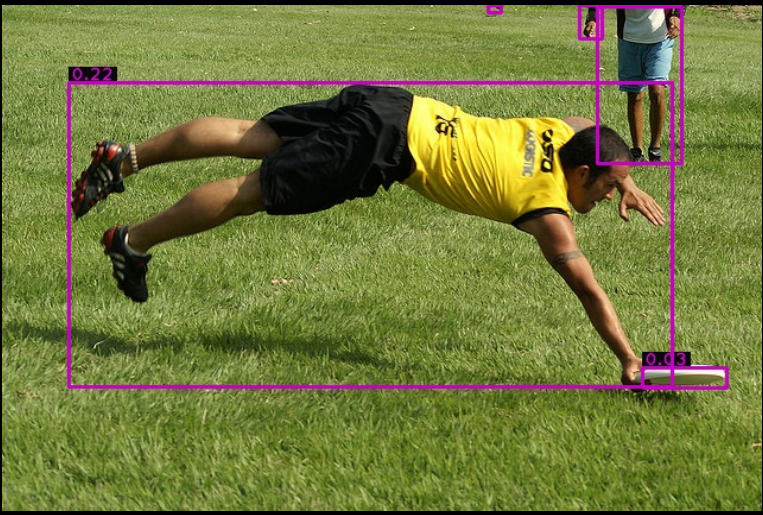
Locate all instances of person in the image.

Woman



Will the man hit the ground?

No
(Yes)



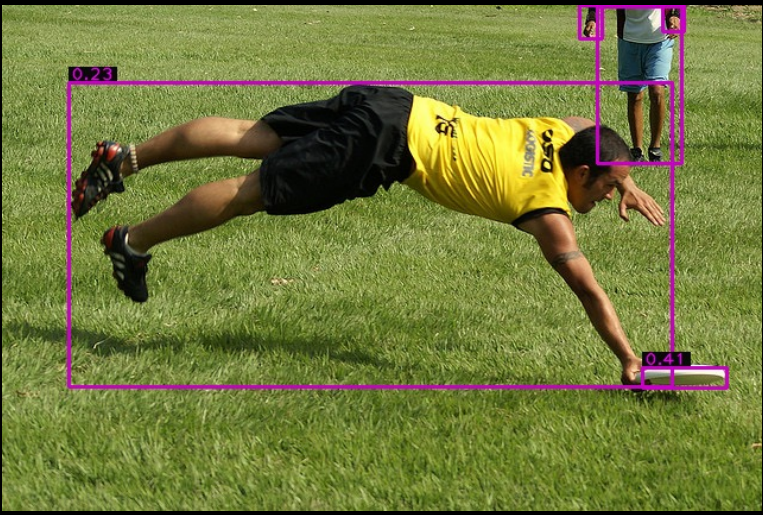
Find all instances of frisbee in this image.

1



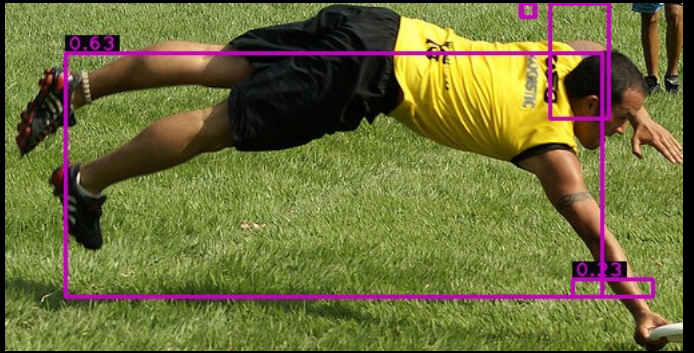
What sport is being played?

Frisbee



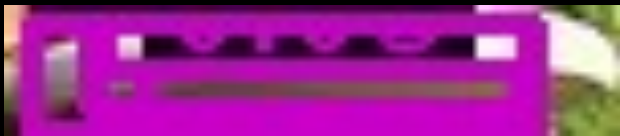
What is this?

Person



What object is this?

Frisbee



Is this a good place
to sleep?

Yes

(No)



What is on top of
this ugly bed?

Bunk

(Sheets)



What is the
headboard made of?

Wood

(Metal)



Generate a caption
for this image.

A bed with a tarp and
a tarp on it.

(A bed covered in
heaps of rags and dirty
bedding)



Describe this image.

A bed and a small
truck in a room.

(There is a very large
room with only a bed
in it)



Locate instances of
bed in this image.

A bed room with a
bed and a bed



What is this object?

Bed

