



KeyWords: age group, camera

Question Concept: age group of the person facing the camera

Caption1: The individual facing the camera in this image is likely a young child or **toddler** based on their apparent age group.

Caption2: A young child, possibly a **toddler**, is captured in this candid moment.

Caption3: The person facing the camera in the image is a young boy, likely between the ages of 2 and 5 years old.

QA1: Who is facing the camera in the image? Boy

QA2: What is the age group of the person in the image? **Toddler**

Predicted Answer: **Toddler**

Question: What state is this child most likely in?

GT: **massachusetts/boston/united states**



KeyWords: state, child

Question Concept: likely state of a child

Caption1: In the image, a young boy is standing in front of a race car with a Red Sox logo on it.

Caption2: He is wearing a baseball cap and holding a baseball glove, which suggests that he may be a fan of the **Boston** Red Sox team.

Caption3: The child appears to be a boy, possibly in the age range of 5 to 10 years old in **Massachusetts**.

QA1: What is on the race car in the image? Red Sox logo

QA2: What is the age range of the boy in the image? 5 to 10

Predicted Answer: **Massachusetts**

Question: What year was the background restaurant founded?

GT: **1955/1951/1952**



KeyWords: background restaurant

Question Concept: founding year of a background restaurant

Caption1: The founding year of the background restaurant in this image is **1955**.

Caption2: Founded in **1955**, this McDonald's location has been serving up delicious meals to the community for over six decades.

Caption3: The background restaurant in the image is McDonald's, which was founded in **1955** by Ray Kroc.

QA1: When was the background restaurant founded? **1955**

QA2: What restaurant is depicted in the image? McDonald's

Predicted Answer: **1955**

Question: When was the activity shown first invented?

GT: **1922/1920s/1950/1985**



KeyWords: activity

Question Concept: invention date of a specific activity

Caption1: Water skiing was invented in **1922**.

Caption2: The exhilarating sport of water skiing traces its roots back to the **1920s** when pioneers like Ralph Samuelson and his wooden skis revolutionized the world of water sports.

Caption3: Samuelson's invention quickly gained popularity, and by the 1930s, water skiing was a popular recreational activity.

QA1: When was water skiing invented? 1922

QA2: Who revolutionized the world of water sports with his wooden skis? Ralph Samuelson

Predicted Answer: 1922

Question: How heavy is this object?

GT: 100 lbs/very heavy/20 pounds/500 lbs



KeyWords: object

Question Concept: weight of the object

Caption1: Based on the refrigerator's appearance in the image, it appears to be a standard upright model with an estimated weight of 100 lbs.

Caption2: Based on the appearance of the refrigerator in the image it seems to be a standard upright model with 100 lbs weight.

Caption3: The refrigerator is rectangular, predominantly white, with a dark panel on the front and weight of 100 lbs.

QA1: How much does the refrigerator weigh? 100 lbs

QA2: What is the shape of the refrigerator? Rectangular

Predicted Answer: 100 lbs

Question: What is the tallest of this type of building in the world?

GT: big ben/1972 feet high/burj khalifa/ulm minster



KeyWords: building

Question Concept: tallest building of a specific type in the world

Caption1: The tallest building of its type in the world can be seen in this image, which is a church with three steeples is Burj Khalifa.

Caption2: The Burj Khalifa is not only the tallest building in the world but also holds several other records, such as the highest occupied floor, highest outdoor observation deck, elevator with the longest travel distance, and the tallest freestanding structure in the world.

Caption3: The majestic clock tower of the world's tallest church, standing tall against the sky.

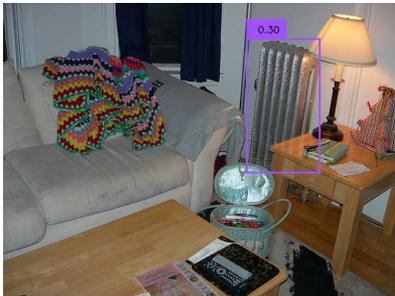
QA1: What is seen in the image? Church

QA2: What is the Burj Khalifa? Building

Predicted Answer: Burj Khalifa

Question: What is the metal pipe device in old homes that heats them called?

GT: radiator/heater/furnace



KeyWords: metal pipe device

Question Concept: name of the Heating pipe

Caption1: On the end table, a vintage-style radiator (silver color), appearing to be covered with a light-colored insulating material.

Caption2: A vintage radiator stands guard in the corner, a silent sentinel of warmth in this cozy corner of the home.

Caption3: The metal pipe device in the image is an old-fashioned radiator, which is a type of heating system commonly found in older homes.

QA1: What is the device shown in the image? Radiator

QA2: What type of heating system is commonly found in older homes? Radiator

Predicted Answer: Radiator

Question: What piece of ski equipment do all of these people have?

GT: skis/poles/ski stick



KeyWords: ski equipment

Question Concept: a common ski equipment item everyone has.

Caption1: People in the image are equipped with **ski poles**.

Caption2: Skiing together in the snow, united by the common thread of their **skis**, creating a vibrant winter scene.

Caption3: Everyone in the image is holding a pair of **ski poles**, a standard accessory for skiing.

QA1: What is the common piece of ski equipment that all the people have in the image? **Ski poles**

QA2: What unites the people in the winter scene? **Skis**

Predicted Answer: **Poles**

Question: What climate does the yellow fruit grow in?

GT: **tropical/warm**



KeyWords: yellow fruit, climate

Question Concept: climate where a specific yellow fruit grows

Caption1: The yellow fruit shown in this image grows in a **tropical** environment, providing optimal conditions for its cultivation.

Caption2: A **tropical** climate where the sun shines brightly and the lemons grow in abundance.

Caption3: The climate in these regions is characterized by high temperatures, low humidity, and moderate rainfall, making it conducive to the growth of citrus trees and the production of lemons.

QA1: What kind of climate is characterized by high temperatures, low humidity, and moderate rainfall? **Tropical**

QA2: What type of trees grow in a tropical climate where the sun shines brightly and the lemons grow in abundance? **Citrus**

Predicted Answer: **Tropical**

Question: What can i make with this?

GT: **smoothie/breakfast**



KeyWords: -

Question Concept: What can I create using these

Caption1: A variety of ingredients laid out on a dark-gray countertop, seemingly prepared for a **smoothie** or similar blended drink.

Caption2: They can be blended into **smoothies**, sliced and added to salads or sandwiches, or used as a topping for yogurt or oatmeal.

Caption3: A bowl of vibrant vegetables and fruits, ready to be transformed into a nutritious dish.

QA1: What can be blended with fresh ingredients to make a smoothie? **Vegetables and fruits**

QA2: What drink can the ingredients be used to make? **Smoothie**

Predicted Answer: **Smoothie**

Question: Why aren't the animals eating meat?

GT: **they eat grass/they are grass fed cattle/they are cows/they are herbivores**



KeyWords: meat, animals

Question Concept: animals' diet

Caption1: The cows seem to be enjoying their diet of grass, which is a common food source for cattle and other grazing animals.

Caption2: The presence of both cows and a dog suggests that the area is a farm or a pasture where livestock is raised and cared for.

Caption3: Both cows are eating the grass, which is a natural part of their diet.

QA1: What do the cows eat? Grass

QA2: What type of area is suggested by the presence of cows and a dog? Farm

Predicted Answer: They are herbivores

Question: What is the purpose of the white circle?

GT: time telling/clock tells time/tell time/clock



KeyWords: white circle

Question Concept: function of the white circle

Caption1: The white circle in the image is a **clock**, which serves as a **time-telling** device.

Caption2: A white circle **clock** on a concrete wall, serving as a silent sentinel of the passing hours.

Caption3: A large, round **clock** face is affixed to a light grayish-tan colored concrete wall of the building.

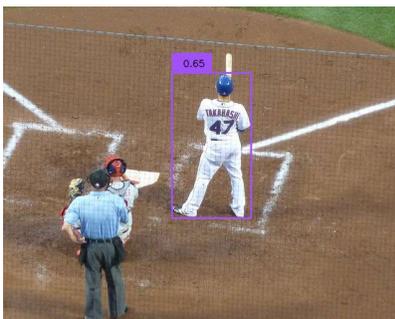
QA1: What is the white circle in the image? **Clock**

QA2: What is the white circle clock serving as? **Time-telling** device

Predicted Answer: Time telling

Question: What's the next thing this player is likely to do?

GT: swing/batting/run



KeyWords: player

Question Concept: player's next action

Caption1: The player is likely to **swinging** the bat next, attempting to hit the ball, either for a base hit or a home run.

Caption2: Takahashi prepares to **swing**, ready to make contact with the incoming pitch.

Caption3: The player is expected to **swing** the bat and attempt to hit the ball thrown by the opposing team's pitcher.

QA1: What is the player's next action likely to involve? **Swinging**

QA2: What does Takahashi prepare to do? **Swing**

Predicted Answer: Swing

Question: Why is he eating this?

GT: snack/carnival/hungry

KeyWords: eating



Question Concept: reason behind eating

Caption1: His action of eating the doughnut may stem from several factors, such as **hunger**, a desire for a **snack**, or simply enjoying the taste and texture of the doughnut.

Caption2: The presence of other people nearby suggests that the boy might be sharing the doughnut with others or taking a break from an outdoor activity to indulge in a quick **snack**.

Caption3: He might be enjoying a **snack** during a picnic or outdoor event, or it could be that he is participating in a **carnival** or fair where food vendors offer various types of treats like doughnuts.

QA1: What might be the reason behind the boy's action of eating the doughnut? **Hunger**

QA2: Where might the boy be eating the doughnut? **Carnival**

Predicted Answer: **Snack**

Question: Does this food taste salty sweet sour or bitter?

GT: **sweet**



KeyWords: salty sweet sour

Question Concept: taste of a food

Caption1: Indulge in the rich, chocolatey delight of this Oreo-topped cake, where every bite is a symphony of flavors and textures.

Caption2: The cake appears to be a popular dessert choice, as it is well-liked by those who have tried it.

Caption3: The chocolate chips would provide a crunchy texture and a slightly bitter taste, complementing the **sweetness** of the frosting and the cake.

QA1: What is the topping of the cake mentioned in the text? **Oreo**

QA2: What would the chocolate chips provide to the cake, according to the text? **Crunchy texture**

Predicted Answer: **Sweet**

Question: What utensil is needed to eat this meal?

GT: **fork**



KeyWords: utensil, meal

Question Concept: necessary utensil for eating this meal

Caption1: The utensil needed to eat the meal in this image is a **fork**.

Caption2: A **fork** is placed next to the plate with chicken and vegetables, indicating that it is the preferred utensil for eating that particular meal.

Caption3: A **fork** and knife are essential for this meal, as the food is served on a plate and likely needs to be cut into smaller portions for convenient eating.

QA1: What utensil is necessary for eating a particular meal in the image? **Fork**

QA2: What utensils would be necessary for eating a meal with food that is served on a plate and cut into smaller pieces? **Fork** and knife

Predicted Answer: **Fork**

Question: What is the colorful object on the man's head used for?

GT: **shade/protection/rain protection/umbrella**



KeyWords: colorful object, head

Question Concept: functionality of the colorful object

Caption1: The colorful object on the man's head is used to provide shade and protection from the sun.

Caption2: The man is wearing an umbrella hat, which is a type of headwear that features an umbrella-like structure with a colorful design.

Caption3: The purpose of the colorful object on the man's head in the image is to serve as an umbrella.

QA1: What is the purpose of the colorful object on the man's head in this image? Shade

QA2: What type of headwear is the man wearing? Umbrella

Predicted Answer: Shade

Question: What number comes after the number on the man's jersey when you count to twenty?

GT: 11/eleven



KeyWords: number, jersey

Question Concept: number that comes after the number on the man's jersey

Caption1: The number that comes after 10 in this image is 11, as the batter is ready to swing at the next pitch.

Caption2: In the image, the number that comes after the number on the man's jersey when counting to 20 is "11".

Caption3: The batter's uniform is primarily white and orange/peach, and the number "10" is visible on the front of the shirt.

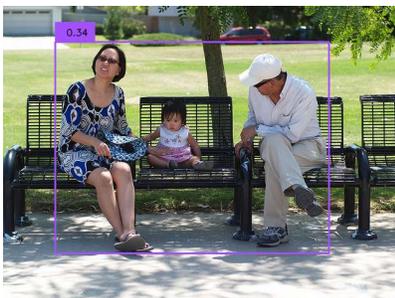
QA1: What is the number that comes after 10 in the image? 11

QA2: What is the number that comes after the number on the man's jersey when counting to 20 in the image? 11

Predicted Answer: 11

Question: In which way are the adults shown here likely related to the child?

GT: grandparents/parents



KeyWords: adults, child

Question Concept: relationship between the adults and the child

Caption1: The relationship between the adults and the child is likely that of grandparents and their grandchild, as they are spending time together in a park setting.

Caption2: The child is wearing a pink dress and appears to be enjoying the company of the adults.

Caption3: It is likely that the adults are the child's parents or caregivers, and they are spending time with the child in a relaxed and casual setting.

QA1: What is the relationship between the adults and the child likely to be? Grandparents

QA2: What is the child wearing? Dress

Predicted Answer: Grandparents

Question: The man on the left dressed in a blue uniform has what on his hand?

GT: glove/mitt/baseball mitt



KeyWords: left, blue uniform, hand

Question Concept: object on the man's hand

Caption1: In the image, the man is holding a baseball **glove**.

Caption2: The pitcher's **glove**, a crucial tool in the game of baseball, ready to catch the incoming ball.

Caption3: The man is crouched slightly forward in an athletic stance, suggesting focus and readiness for action.

QA1: : What is the man holding in the image? **Glove**

QA2: What is a crucial tool in the game of baseball? **Glove**

Predicted Answer: **Glove**

Examples involving counting number of objects.



Question: How many animal trunks are visible here?

GT: 2

KeyWords: animal trunks

Predicted Answer: 2

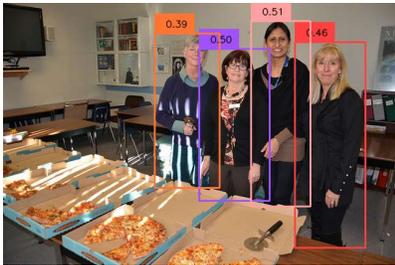


Question: How many people have their arms around the woman?

GT: 1

KeyWords: people, arms, woman

Predicted Answer: 1



Question: How many women are in the picture?

GT: 4

KeyWords: woman

Predicted Answer: 4



Question: How many vehicles have their lights on?

GT: 1

KeyWords: vehicles, lights

Predicted Answer: 1

Question: What kind of fruit is cut in half and darker than the other?

GT: grape



KeyWords: fruit

Question Concept: fruit with one darker half

Caption1: A slice of cherry, where the darker half of the fruit reveals a hidden depth of flavor

Caption2: A plate of food with a fork and knife

Caption3: The fruit in the image that is typically cut in half and has a darker color on one half compared to the other is a [grape](#).

QA1: What type of fruit is typically cut in half and has a darker color on one half compared to the other? [Grape](#)

QA2: What type of fruit is described as having a hidden depth of flavor when cut in half? Cherry

Predicted Answer: [Cherry](#)

Question: What item on the desk could help with a cold?

GT: [cough drops/tissue/ricola](#)



KeyWords: cold, desk

Question Concept: item that helps alleviate a cold

Caption1: On the desk, there is a bottle of vitamin C, which can help alleviate the symptoms of a cold

Caption2: By taking vitamin C supplements or consuming foods high in vitamin C, such as citrus fruits, berries, and leafy greens, individuals can help reduce the severity and duration of their cold symptoms.

Caption3: There are several items on the desk that could potentially help alleviate a cold in the image.

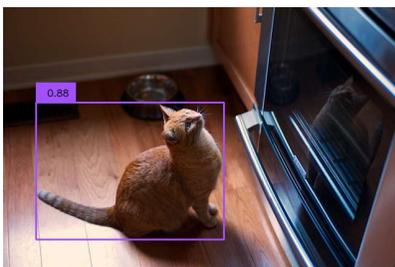
QA1: What is on the desk that can help alleviate cold symptoms? Vitamin C

QA2: What types of foods are high in vitamin C and can help reduce cold symptoms? Citrus fruits

Predicted Answer: [Vitamin C](#)

Question: What activity does the cat appear most likely to do?

GT: [watch television/eat/jump/beg](#)



KeyWords: cat, activity

Question Concept: most likely behavior of a cat

Caption1: The cat is sitting on the floor and looking up, possibly observing its surroundings or searching for something of interest.

Caption2: The cat is taking a moment to look around and explore the environment.

Caption3: The cat seems to be engaged in a behavior typical of cats when they are curious or attentive to their environment.

QA1: What is the cat likely doing in the image? Sitting

QA2: What is typical of cats when they are curious or attentive to their environment? Behavior

Predicted Answer: [Looking](#)

Question: Which is not part of specific dish?

GT: [ham/ketchup/ketchup](#)



KeyWords: ingredient, dish

Question Concept: ingredient that is not part of a specific dish

Caption1: Bacon is one of the ingredients that is not part of a specific dish in this image.

Caption2: A slice of ham on a pizza offers an unexpected savory twist to the classic dish.

Caption3: The green leafy vegetable is not included in this dish, as it is not typically used as a topping for pizza or flatbread.

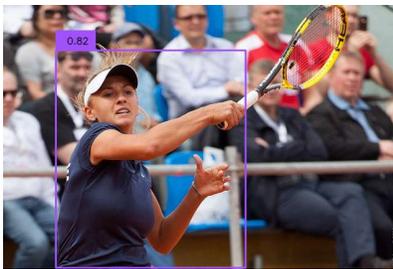
QA1: What is the unexpected flavor ingredient? Bacon

QA2: What is the green leafy vegetable? Kale

Predicted Answer: Kale

Question: Who has the same color hair as this woman?

GT: maria sharapova/swedish persons/sitting guy



KeyWords: color hair, woman

Question Concept: someone who has the same hair color as the woman in question

Caption1: A woman swinging a tennis racket.

Caption2: The woman in the image has blonde hair.

Caption3: When you're blonde, you're always ready for a good time!

QA1: What is the woman's hair color in the image? Blonde

QA2: What is often associated with being blonde, according to the quote? A good time

Predicted Answer: Blonde

Question: The top shelf has many books of the same author of what genre?

GT: thriller/thriller fiction/non fiction/mystery



KeyWords: many books, top shelf, author, genre

Question Concept: genre of books on the top shelf

Caption1: The genre of books on the top shelf in this image appears to be a mix of fiction and non-fiction genres.

Caption2: A collection of diverse reads, from classic literature to modern mysteries, the top shelf of this bookshelf offers a literary journey for every taste.

Caption3: The top shelf in the image appears to be filled with a variety of books, but without more specific information, it's difficult to determine the exact genre of each book.

QA1: What type of books are found on the top shelf in the image? Fiction and non-fiction

QA2: What kind of literary experience does the top shelf of the bookshelf offer? A literary journey

Predicted Answer: Mmix

Question: Why would we assume this surfer is not about to engage in his sport?

GT: no water/he looks tired/holding it/walking away/



KeyWords: surfer

Question Concept: Reason why the surfer will not do surfing

Caption1: A surfer's journey begins with a leap into the unknown.

Caption2: A man wearing a wetsuit and carrying a surfboard

Caption3: The person is wearing a black wetsuit, which is a common attire for surfers, indicating that they might be heading to or returning from a surfing session.

QA1: What is the surfer wearing? Wetsuit

QA2: What is the surfer carrying? Surfboard

Predicted Answer: not surfing

Question: Where might this bench be located?

GT: island/beach/oceanside/ocean



KeyWords: bench

Question Concept: possible location for a bench

Caption1: A serene park bench overlooking a lush green field, perfect for a quiet afternoon of reflection.

Caption2: The image shows two people sitting on a wooden bench in a grassy area.

Caption3: The possible location for the bench in the image is a grassy field.

QA1: What is the setting for the scene? Park

QA2: What is the material of the bench? Wooden

Predicted Answer: Park

Question: What is this wall used for?

GT: artwork/decoration art/holding paintings/hanging pictures



KeyWords: wall

Question Concept: function of the wall

Caption1: The wall in this image serves as a decorative backdrop, contributing to the overall decor of the living room.

Caption2: A wall that serves as a canvas for artistic expression, creating a harmonious and inviting atmosphere in the living space.

Caption3: The wall is used to make the living room look nice.

QA1: What is the purpose of the wall in the image? Decorative backdrop

QA2: What does the wall serve as in the living space? Canvas

Predicted Answer: Decorative backdrop

Question: What percent of the stands are full?

GT: 15/ten/twenty



KeyWords: percent, stands

Question Concept: percentage of stands that are full

Caption1: The stands are only half full, but the energy in the air is palpable as the game unfolds

Caption2: In the image, the stands appear to be approximately 50% full.

Caption3: It shows a grandstand with spectators seated on bleachers.

QA1: What percentage of the stands are filled? 50%

QA2: How full do the stands appear to be in the image? 50%.

Predicted Answer: 50

Question: What is the name of the operating system used for this device?

GT: mac/windows/apple



KeyWords: operating system, device

Question Concept: operating system used by a device

Caption1: The operating system used by the laptop in this image is not specified.

Caption2: The operating system used on this Apple's Mac computer is macOS.

Caption3: A dog's curious gaze meets the power of a laptop running on the latest operating system, ready to explore the digital world.

QA1: What is the operating system used on this Apple's Mac computers? MacOS

QA2: How is the dog looking at the computer? Curiously

Predicted Answer: MacOS

Question: Based on the clothing what year was this picture taken?

GT: 1930/1960/1913/1953



KeyWords: year, clothing

Question Concept: year a picture was taken based on the clothing

Caption1: The year a picture was taken based on the clothing depicted in the image is around the 1920s or 1930s.

Caption2: A moment frozen in time: 1945, as seen through the lens of a vintage camera.

Caption3: The image appears to be a vintage black and white photograph, likely from the early to mid-20th century, judging by the clothing and hairstyles of the individuals in the picture.

QA1: What is the likely time period of the picture? Mid-20th century

QA2: What type of photograph is the image likely to be? Black and white

Predicted Answer: 1945

Question: What decade is this?

GT: 1930/1940/1930s



KeyWords: decade

Question Concept: time period

Caption1: The time period or decade being referred to in this image is the 1920s

Caption2: A moment frozen in time, where the hustle and bustle of a bygone era unfolds.

Caption3: However, without more specific details, it's difficult to pinpoint the exact decade or time period.

QA1: What time period is being referred to in this image? The 1920s

QA2: What does the image describe as part of a bygone era? Bustle

Predicted Answer: 1920s

Question: What age range do you think these people are?

GT: teenagers/20 to 50/20 to 40



KeyWords: age, people

Question Concept: age range of people

Caption1: The group of people in the image consists of both young and older individuals, with a range of ages from teenagers to adults.

Caption2: A diverse group of individuals, spanning from young adults to older teens, gathered for a casual pizza party.

Caption3: The age range of the group is not precisely defined, but it is clear that they are young adults.

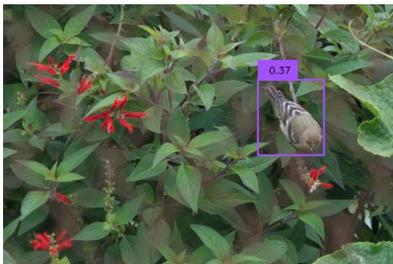
QA1: What is the age range of the group in the image? Young adults

QA2: What type of gathering did the group attend? Pizza party

Predicted Answer: young adults

Question: In this photo provide the average weight of this type of bird when they are fully grown?

GT: 1 lb/2 ounces/1 pound/half pound



KeyWords: average weight, bird

Question Concept: average weight of fully grown bird

Caption1: A lightweight feathered friend, gracefully perched on a vibrant red flower.

Caption2: The image shows a small bird, likely a finch or sparrow, with a light brownish-gray body and streaked or striped darker brown/gray wings, weighing approximately 100 grams.

Caption3: The weight of a bird can vary greatly depending on factors such as age, sex, and species.

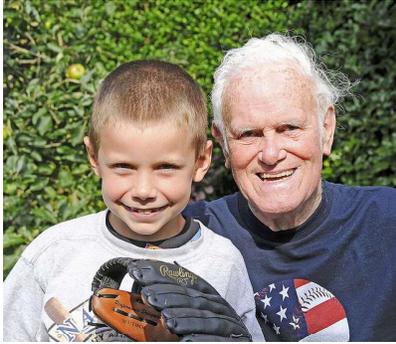
QA1: What is the bird's perch? Red flower

QA2: What is the weight of a bird? 100 grams

Predicted Answer: 100 grams

Question: What is the generational gap between these two people?

GT: 50 years/60 years/40 years/2 generations



KeyWords: people

Question Concept: difference in age

Caption1: The age gap between the two individuals highlights the importance of passing down knowledge and skills from one generation to the next.

Caption2: The two individuals in the image exhibit a notable age difference.

Caption3: The age difference between them highlights the generational gap between the two individuals, with the older man serving as a mentor or role model for the younger boy.

QA1: What is highlighted in the difference in age between the two individuals? knowledge

QA2: What does the older man serve as for the younger boy in the image? Mentor

Predicted Answer: significant

Question: How much water can the teapot in the photo hold?

GT: 2 quarts/1 gallon/2 liters



KeyWords: teapot

Question Concept: capacity of the teapot

Caption1: The teapot has a capacity of two cups.

Caption2: The teapot can hold up to two cups of tea.

Caption3: The capacity of the teapot is not clearly visible in the image, but it seems to be a standard size for a teapot, which is typically around 1 to 2 liters.

QA1: What is the capacity of the teapot mentioned in the caption? Two cups

QA2: What is the typical size range of a standard teapot? 1 to 2 liters

Predicted Answer: 1 to 2 liters

Question: How much protein is in this meal?

GT: lot/little/5 grams



KeyWords: protein, meal

Question Concept: amount of protein in a specific meal

Caption1: The meal in this image contains a moderate amount of protein, as evidenced by the presence of black beans, broccoli, and cheese.

Caption2: This meal packs a protein punch with its combination of chicken, beans, and broccoli, providing a balanced and nutritious option.

Caption3: The meal on the left would likely contain around 20-30 grams of protein per serving, while the meal on the right would contain around 15-20 grams of protein per serving.

QA1: What is the source of protein in the meal? Beans

QA2: What is the estimated amount of protein in the meal on the left? 20-30 grams

Predicted Answer: 20-30 grams

Question: How much do these boats weigh?

GT: 5 tons/4.57/1000 pounds/ton



KeyWords: boats

Question Concept: weight of the boats

Caption1: The weight of the boats in this image ranges from 10,000 to 20,000 pounds, with the largest boat weighing around 20,000 pounds and the smaller boats weighing between 10,000 and 15,000 pounds

Caption2: A trio of vessels, each with its own story, anchored in the serene waters, ready to embark on their next adventure.

Caption3: Without specific dimensions or weight information, it's not possible to provide an exact weight for these boats

QA1: What is the range of weights for the boats in the image? 10,000 to 20,000 pounds

QA2: Where are the vessels anchored? Serene water

Predicted Answer: 10,000 to 20,000 pounds

Question: How much does it cost for city parking?

GT: 50 cents/60 cents/50 cents/0.5/0.6



KeyWords: city parking

Question Concept: city parking cost

Caption1: This suggests that the parking fees in this city are relatively low, as there are only two cars parked in the vicinity of the meter.

Caption2: A reminder of the city's parking fees, standing tall on the side of the road

Caption3: The fees are usually based on the duration of parking and can range from a few cents to several dollars per hour, depending on the location and the city's parking.

QA1: What is the basis of parking fees in the city? Duration

QA2: What is the range of parking fees in the city? A few cents to several dollars per hour

Predicted Answer: A few cents to several dollars

Ours



Q: What type of clouds are in the picture?

GT: cumulus/cumuli/nimbus

KWs: clouds

Cap1: In the image, there are various types of clouds visible in the sky, including fluffy white clouds and darker, more ominous-looking clouds.

Cap2: A sky filled with fluffy, cumulus clouds, suggesting a pleasant day with a chance of rain.

Cap3: In the given image, there are several types of clouds visible in the sky.

QA1: What is visible in the sky? Clouds

QA2: What type of clouds are fluffy and white? Cumulus

Generated Answer: cumulus

Img2LLM



Q: What type of clouds are in the picture?

GT: cumulus/cumuli/nimbus

Cap1: a cloudy - filled sky on a cloudy day over a zebras.

Cap2: the clouds are gray and full of clouds.

Cap3: there are many different clouds in this sky.

QA1: what is in the background of a photo of a zebra? sky

QA2: what type of sky is above on a cloudy day? cloudy

Generated Answer: cloud

0.85



Q: How many people can this bus carry?

GT: 50/40/39

KWs: bus, people

Cap1: A blue bus is parked on the side of the road.

Cap2: Arriving at Ashington, the X22 bus is ready to transport passengers to their destinations.

Cap3: There are several people visible in the scene, some standing near the bus and others walking along the sidewalk.

QA1: What is parked on the side of the road? bus

QA2: Where is the X22 bus ready to transport passengers? Ashington

Generated Answer: 50



Q: How many people can this bus carry?

GT: 50/40/39

Cap1: a passenger bus traveling on a street side.

Cap2: blue commuter bus with parked on the side of the road.

Cap3: a bus that says aradara rides down the street.

QA1: what color bus is driving down the street? blue

QA2: what is making it's way down the street? bus

Generated Answer: many

Ours



Q: What type of shot is the woman about to hit?

GT: forehand/tennis shot/swing

KWs: shot, woman

Cap1: The woman in the image is swinging a tennis racket at a tennis ball, preparing to hit it during a game of tennis.

Cap2: In the image, a woman is captured in the midst of a powerful forehand shot on a tennis court.

Cap3: She is holding a tennis racket and has just hit the ball with it.

QA1: What is the woman holding in her hand? Tennis racket

QA2: What is the woman about to do on the court? Swing

Generated Answer: Swing

Img2LLM



Q: What type of shot is the woman about to hit?

GT: forehand/tennis shot/swing

Cap1: tennis player is hitting a tennis ball with her racket.

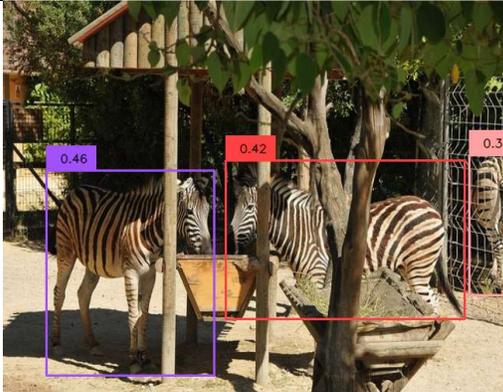
Cap2: a woman in pink outfit hitting a tennis ball.

Cap3: a woman in a cropped top and pants swinging a tennis racket.

QA1: what is a tennis player doing with a tennis racket? swinging

QA2: who is swinging a tennis racket at a tennis ball? woman

Generated Answer: volley



Q: How many zebras are there?

GT: 3

KWs: zebras

Generated Answer: 3



Q: How many zebras are there?

GT: 3

Cap1: two zebras walking by a feeder full of food.

Cap2: pair of zebras standing together at water trough in zoo.

Cap3: the zebras are eating out of a feeder box.

QA1: how many zebras are standing next to each other? two

QA2: what are the zebras doing? eating

Generated Answer: 2

Ours

Q: How many buses are in the picture?

GT: 8

KWs: buses

Generated Answer: 8

Img2LLM

Q: How many buses are in the picture?

GT: 8

Cap1: a lot of buses sit parked in a line in front of a hill.

Cap2: a group of purple passenger buses all in a row.

Cap3: a row of purple bus buses next to each other.

QA1: how are the buses parked? a line

QA2: what color buses are parked in front of each other?
purple

Generated Answer: several



Q: This dish is suitable for which group of people?

GT: vegetarian/vegan/family

KWs: dish

Cap1: This dish is suitable for a variety of people, including vegetarians and those who enjoy a simple, hearty meal.

Cap2: The target group of people for the pasta and broccoli dish in this image is likely health-conscious individuals who are looking for a low-calorie, nutritious meal option.

Cap3: The target group of people for this dish could include families, individuals looking for a quick and easy meal, vegetarians, and those who enjoy a balance of carbohydrates and vegetables in their diet.

QA1: What is the target group of people for the pasta and broccoli dish? Vegetarians

QA2: What type of meal is the pasta and broccoli dish?
Low-calorie

Generated Answer: Vegetarians



Q: This dish is suitable for which group of people?

GT: vegetarian/vegan/family

Cap1: a pasta dish sitting on top of a white plate.

Cap2: a broccoli pasta dish that has very pasta.

Cap3: a dish of pasta with noodles and tomato sauce.

QA1: what vegetable is on a white plate? broccoli

QA2: what color is a plate of pasta with broccoli on it?
white

Generated Answer: children

Ours



Q: Why is the woman holding the umbrella?
GT: shade/sun protection/get shadow
KWs: woman, umbrella
Cap1: The umbrella provides shade and protection from the sun for both the woman and the dogs.
Cap2: The umbrella is open and provides shade for the woman and the two dogs she is interacting with.
Cap3: The woman is holding the umbrella over her head, which indicates that she is using it to shield herself and the dogs from the weather.
QA1: What is the woman using to shield herself and the dogs from the weather? Umbrella
QA2: What is the woman doing with the umbrella?
 Holding
Generated Answer: Shade

Img2LLM



Q: Why is the woman holding the umbrella?
GT: shade/sun protection/get shadow
Cap1: a young woman and the umbrella are on an orange blanket.
Cap2: a woman's umbrella and two dogs under an umbrella.
Cap3: a woman holding an umbrella is getting some light under her umbrella.
QA1: who is holding an umbrella while her dog sits under it? woman
QA2: : what is a woman holding and a dog under it? an umbrella
Generated Answer: to protect herself from the sun



Q: What is in front of the monitor?
GT: chair/keyboard/webcam
KWs: monitor, front
Cap1: The computer features a black monitor and a keyboard, which are placed next to each other on the desk.
Cap2: A white keyboard and mouse, ready for productivity
Cap3: A clean and organized computer desk with a black monitor and a white keyboard.
QA1: What color is the keyboard on the desk? White
QA2: What is placed next to the monitor? Keyboard
Generated Answer: Keyboard



Q: what is in front of the monitor?
GT: chair/keyboard/webcam
Cap1: a corner table with computer computer on the desk.
Cap2: a computer on the small desk in a small office area.
Cap3: view of a computer monitor in a light lit room.
QA1: what is a computer sitting on in a corner of a room? desk
QA2: how big is the desk in the corner? small
Generated Answer: desk

Ours



Q: What is in the bottles?

GT: alcohol/liqueur/baileys

KWs: bottles

Cap1: The two bottles of beer are positioned close to the bowl of salad, indicating that they may have been used to accompany the meal.

Cap2: A pair of dark beverages, possibly coffee or tea, served alongside a fresh salad, ready to complement the meal with a warm or cool sip.

Cap3: There are two bottles of dark liquid, possibly coffee or tea, with a white cup of dark liquid in the image.

QA1: What are placed on the table? Bottles

QA2: What could be the contents of the bottles? Beer

Generated Answer: Beer

Img2LLM



Q: what is in the bottles?

GT: alcohol/liqueur/baileys

Cap1: a sandwich on a plate with a glass of beer bottle.

Cap2: a table that has a sandwich, beer, and beer on it.

Cap3: a sandwich on a plate with a glass of beer bottle.

QA1: What is next to a sandwich and a beer? bottle

QA2: where is a sandwich with a beer and beer on a plate? table

Generated Answer: beer