Referring to Any Person

Supplementary Material

Method	Finetuned	Attr. DF1	Pos. DF1	Inter. DF1	Rea. DF1	Cel. DF1	Avg. DF1
RexSeek	Yes	81.5	83.8	80.7	81.5	84.2	**82.3**
Qwen2.5-VL	No	54.4	52.8	53.2	40.3	80.1	56.2
Qwen2.5-VL-D	Yes	67.2	69.5	66.4	62.2	82.7	69.6
Qwen2.5-VL-B	Yes	81.4	81.3	80.2	77.3	86.8	81.4

Table 1. Performance of Qwen2.5-VL variants on HumanRef. Qwen2.5-VL-D denotes direct fine-tuning to predict coordinates. Qwen2.5-VL-B uses the box hint strategy.

A. Fine-Tuning Qwen2.5-VL on HumanRef

To better understand the impact of fine-tuning and input strategies on referring expression comprehension, we conducted further experiments using Qwen2.5-VL-7B on the HumanRef benchmark. Specifically, we explored two fine-tuning strategies: (1) Direct fine-tuning of Qwen2.5-VL to predict object coordinates based on referring expressions (denoted as Qwen2.5-VL-D), and (2) the box hint strategy (denoted as Qwen2.5-VL-B), which provides the model with bounding boxes of all candidate persons as additional inputs, following the approach used in RexSeek.

Table 1 summarizes the results. Compared to the zero-shot baseline, both fine-tuned variants show substantial performance gains. Notably, the box hint strategy yields the highest average DF1 score of 81.4, approaching RexSeek's 82.3. This demonstrates that structural priors, such as explicit spatial candidates, can significantly enhance model performance. These results suggest that the performance gap between Qwen2.5-VL and RexSeek is primarily due to differences in task-specific training and supervision rather than architectural limitations.

B. More Details of HumanRef Dataset

B.1. Detailed Definition for Each Subset

Attribute Subset: The attribute subset in HumanRef encompasses a diverse range of descriptive properties used for referring to individuals. These attributes are systematically categorized into finite-value attributes, which have a predefined set of possible values, and open-ended attributes, which allow for more flexible and detailed descriptions. As illustrated in Fig. 1, we provide a structured breakdown of these attribute categories along with representative examples. This visualization highlights the key properties that can be used to distinguish individuals, including intrinsic characteristics (e.g., gender, age, ethnicity), appearance features (e.g., hairstyle, facial expressions), clothing and accessories, poses, and actions. By incorporating both structured and descriptive attributes, HumanRef ensures a comprehensive and versatile annotation framework that better aligns with real-world referring scenarios.

Position Subset: The position subset in HumanRef captures the spatial relationships of individuals within an image. We categorize these into two main types: inner position and outer position, which represent different spatial referencing strategies. Inner position describes the relative spatial relationships between individuals, while outer position refers to absolute spatial relationships using environmental landmarks. We show some detailed examples in Table. 2.

Inner Position (Relative to Others)	**Outer Position (Relative to Environment)**
"The leftmost person in the image."	*"The person standing in the corner of the room."*
"The second person from the left."	*"The person sitting on the red chair."*
"The person at the top of the group."	*"The person standing at the left edge of the bridge."*
	"The person leaning against the wall next to a tree."
	"The person standing near the window of the room."

Table 2. Examples of Inner and Outer Position References.

Interaction Subset: The interaction subset in HumanRef captures the ways individuals interact with other people and objects within a scene. We classify interactions into two categories: inner interaction and outer interaction. Inner interaction focuses on actions between individuals, while outer interaction describes interactions between a person and objects or the surrounding environment. We show some detailed examples in Table. 3.

Reasoning Subset: The reasoning subset in HumanRef requires models to perform multi-step inference by first identifying a reference person or object before determining the target individual. It is divided into three categories: inner position reasoning, outer position reasoning, and attribute reasoning. Inner position reasoning describes one individual using another as an anchor point, establishing a relative spatial relationship between them. This approach ensures that the model must first locate a reference person before resolving the intended target. Outer position reasoning follows a similar principle but instead uses an absolute spatial reference tied to the surrounding environment, requiring the model to integrate positional understanding of both individuals and external scene elements.

```
{
    "Gender": ["Male", "Female"],
    "Age": ["Infant", "Child", "Adolescent", "Adult", "Elderly"],
    "Ethnicity": ["White", "African", "Asian", "..."],
    "Profession": ["Doctor", "Engineer", "Teacher", "..."],
    "Appearance": {
        "Hair": {
            "Type": ["Short", "Long", "Curly", "Bald", "..."],
            "Color": ["Black", "Brown", "Blonde", "Red", "..."]
        },
        "Beard": {
            "Type": ["Full Beard", "Goatee", "Mustache", "..."],
            "Color": ["Black", "White", "Brown", "..."]
            "Detailed": ["long and black beard", "..."]
        },
        "Facial Expression": {
            "Type": ["Smiling", "Frowning", "Surprised", "..."],
            "Detailed": ["Smiling with closed eyes", "..."]
        }
    },
    "Clothing & Accessories": {
        "Upper Body": {"Type": ["T-shirt", "Shirt", "..."], "Color": ["Red", "Blue", "..."]},
        "Lower Body": {"Type": ["Jeans", "Skirt", "..."], "Color": ["Black", "White", "..."]},
        "Shoes": {"Type": ["Sneakers", "Boots", "..."], "Color": ["Red", "Blue", "..."]},
        "Hat": {"Type": ["Baseball Cap", "Knit Cap", "..."], "Color": ["Black", "White", "..."]}
    },
    "Pose": {"Type": ["Standing", "Sitting", "..."], "Details": ["Sitting on a bench", "..."]},
    "Actions": {"Type": ["Walking", "Running", "..."], "Details": ["Walking a dog", "..."]}
}
```

Figure 1. Attribute taxonomy in HumanRef. Finite-value attributes have predefined values, while open-ended attributes allow flexible descriptions.

Inner Interaction (Human-Human)	**Outer Interaction (Human-Object/Environment)**
"Two players making physical contact."	*"The person holding a dog on a leash."*
"The bride and groom walking hand in hand in the middle."	*"The person reaching out to grab the football."*
"The person raising another person's hand with their right hand."	*"The person using their right hand to pull toilet paper."*
"Two people embracing each other."	*"The person holding a transparent box containing a roll of toilet paper with their left hand."*
	"The person holding a pizza with one hand."
	"The person holding a pen in their hand."

Table 3. Examples of Inner and Outer Interaction References.

Attribute reasoning involves logical filtering within a group of candidates who share a common attribute, followed by an exclusion step based on an additional attribute with a negation rule. This forces the model to refine its selection beyond direct attribute matching, ensuring a more precise understanding of distinguishing characteristics. These three types of reasoning introduce hierarchical complexity into the referring task, strengthening the model's ability to process contextual, spatial, and attribute-based relationships in a structured manner. We show some detailed examples in Table. 4.

Celebrity Subset: The celebrity subset in HumanRef focuses on identifying well-known individuals based on their names or recognizable personas. To provide a structured classification, we divide celebrities into six categories: Character, Singer, Actor, Athlete, Entrepreneur, and Politician. This categorization ensures that the dataset covers a diverse range of public figures from different domains, each of whom may be referred to by their real name or an associated identity. In Tab. 5, we provide a detailed list of representative individuals from each category, illustrating the range of celebrities included in the dataset.

B.2. Details for Structured Property Dictionary

To facilitate the annotation process, we employ Qwen2.5-VL-7B to generate a predefined structured property dictionary for each person in the image. This dictionary serves as a reference for annotators, allowing them to construct prop-

Reasoning Type	Example Expressions
Inner Position Reasoning	*"The person to the left of the child wearing a blue-and-white striped shirt."* *"The person to the right of the man in a suit."* *"The girl to the right of the person wearing blue headphones."* *"All individuals to the right of the groom."*
Outer Position Reasoning	*"The closest person to the left of the person in the corridor."* *"The person to the left of the individual directly below the letter D."* *"The child to the right of the girl standing under the blue arched wooden door."* *"The person to the right of the individual sitting inside the shopping cart."*
Attribute Reasoning	*"The person wearing a hat but not sitting."* *"The person eating ice cream but not wearing a red top."* *"The person wearing sandals but not sitting on the bed."* *"The person on the airplane but not showing their head."*

Table 4. Examples of Reasoning-Based Referring Expressions.

erty lists based on pre-generated attribute descriptions. Figure 2 presents the prompt used to generate these structured descriptions.

C. Details for RexSeek model

C.1. Model Architecture

We utilize a dual-encoder design for vision processing. The low-resolution visual encoder is based on the CLIP ViT-Large-14-336 model, while the high-resolution visual encoder leverages the LAION ConvNeXt-Large-320 model. The input resolution is set to 336×336 for the low-resolution encoder and 768×768 for the high-resolution encoder, allowing the model to capture both coarse and fine-grained visual details effectively.

C.2. Training Details

During the pretraining stage (Stage-1), we use a batch size of 32 per device, resulting in a total batch size of 256 across all devices. The instruction-tuning stage (Stage-2, Stage-3, Stage-4) employs a reduced batch size of 16 per device, with a total batch size of 128. The learning rate is initialized at 1e-3 for pretraining and adjusted to 2e-5 during instruction tuning to ensure stable fine-tuning and convergence.

Character	Eleven from the Strange Things, Obi-Wan Kenobi, Captain America, Queen Maeve, Buzz Lightyear, Mary Poppins, John Rambo, Hermione Granger, Rick Grimes, Ferris Bueller, Sheldon Cooper, Sarah Connor, Negan Smith, Amy Farrah Fowler, Professor Severus Snape, Shrek, Ada Shelby, Palpatine, Jorah Mormont, Thomas Shelby, Cersei Lannister, Luke Skywalker, Maximus Decimus Meridius, Sansa Stark, John Wick, Atticus Finch, Harry Potter, Wolverine, Mr. Bean, Ronald Weasley, Melisandre, Ross Geller, Bilbo Baggins, Samwise Gamgee, Maggie Greene, Jon Snow, Wednesday Addams, Hans Gruber, Rocky Balboa, Michael Corleone, Leonard Hofstadter, Chandler Bing, Littlefinger, Barbossa, Rachel Green, Howard Wolowitz, Jaime Lannister, Legolas, Axel Foley, James T. Kirk, Kevin McCallister, Margaery Tyrell, Leia Organa, Forrest Gump, Marty McFly, Han Solo, Billy Butcher, Mr. Kesuke Miyagi, Professor Albus Dumbledore, Steve Harrington, Sirius Black, Gus Fring, Jack Sparrow, Inigo Montoya, Professor Minerva McGonagall, Homelander, Arthur Shelby, Superman, Indiana Jones, Beetlejuice, Polly Gray, Hughie Campbell, Peter Venkman, Sandor Clegane, Soldier Boy, Joey Tribbiani, Rajesh Koothrappali, Gollum, Vito Corleone, Daryl Dixon, Aragorn, Andy Dufresne, Jesse Pinkman, Penny Hofstadter, Jean-Luc Picard, Lord Voldemort, Oberyn Martell, Luna Lovegood, Carol Peletier, Monica Geller, Alfred Pennyworth, Bernadette Rostenkowski-Wolowitz, Darth Maul, Thor, Neo, John McClane, Phoebe Buffay, Spock, Dorothy Gale
Singer	Michael Bublé, Demi Lovato, Aerosmith, Drake, ZAYN, Jennifer Lopez, Olivia Rodrigo, Kali Uchis, Flo Rida, Doja Cat, Skrillex, Chris Brown, Ellie Goulding, 50 Cent, Katy Perry, Gucci Mane, Charli XCX, Avril Lavigne, Shawn Mendes, Lil Wayne, J. Cole, Linkin Park, Migos, Taylor Swift, DJ Khaled, Red Hot Chili Peppers, Justin Bieber, Calvin Harris, 2 Chainz, Frank Ocean, Jason Mraz, Alicia Keys, Miley Cyrus, Childish Gambino, Meghan Trainor, Tyga, Usher, SZA, Bad Bunny, Labrinth, Diplo, Jack Johnson, Halsey, Young Thug, Bon Jovi, Post Malone, Christina Aguilera, The Kooks, John Mayer, The 1975, Akon, G-Eazy, Panic! At the Disco, Eminem, Ed Sheeran, Maroon 5, Ne-Yo, Zedd, Dr. Dre, Queen, Nelly Furtado, Steve Lacy, Imagine Dragons, Sia, Mac DeMarco, Big Sean, Martin Garrix, Camila Cabello, The Rolling Stones, Khalid, Harry Styles, Charlie Puth, Kanye West, The Weeknd, Kendrick Lamar, Travis Scott, Kesha, Nelly, Tyler, The Creator, Billie Eilish, Metro Boomin, Gwen Stefani, Sean Paul, Vampire Weekend, Jay-Z, Kelly Clarkson, Stevie Wonder, Adele, Arctic Monkeys, Lorde, Britney Spears, Selena Gomez, Daddy Yankee, 21 Savage, David Guetta, J Balvin, The Cure, Bruno Mars, Dua Lipa, Bruce Springsteen, Snoop Dogg, B.o.B, OutKast, Lady Gaga, Hozier, Wiz Khalifa, Foo Fighters, Lana Del Rey, Beyoncé, Madonna, Shakira, John Legend, Mark Ronson, Sam Smith, Billy Joel, Jeremih, Paramore, Chance the Rapper, DJ Snake, Sabrina Carpenter, Kid Cudi, Trey Songz, Kings of Leon, Enrique Iglesias, Pharrell Williams, Arcade Fire, Jessie J, Lil Uzi Vert, Bob Dylan
Actor	Tom Wilkinson, Al Pacino, Kevin Costner, Franco Nero, Philip Seymour Hoffman, Alan Rickman, Leonardo DiCaprio, Ben Affleck, William Hurt, Mark Wahlberg, Jonah Hill, Shia LaBeouf, Don Cheadle, Orlando Bloom, Jeff Goldblum, Denzel Washington, Alec Baldwin, Bradley Cooper, Ed Harris, Jason Clarke, Mahershala Ali, Viggo Mortensen, Owen Wilson, Alan Arkin, James Caan, Nicolas Cage, Samuel L, David Strathairn, Matt Damon, George Clooney, Giovanni Ribisi, Jared Leto, Kevin Spacey, Matthew McConaughey, Gary Sinise, Pete Postlethwaite, Keanu Reeves, Timothy Spall, Harry Dean Stanton, John Carroll Lynch, Chiwetel Ejiofor, Woody Harrelson, Ryan Gosling, Joaquin Phoenix, Donald Sutherland, Paul Dano, Chris Hemsworth, David Oyelowo, Tom Hardy, Barry Pepper, Kurt Russell, Christian Bale, Jeff Daniels, Ben Whishaw, Sterling Hayden, Edward Norton, Sam Shepard, Andy Garcia, Harvey Keitel, Benicio Del Toro, Gene Hackman, Bruce Willis, Guy Pearce, Jonathan Pryce, Michael Fassbender, James Stewart, Zach Galifianakis, Forest Whitaker, Vincent Cassel, Michael Sheen, Tom Berenger, Jim Carrey, Steve Buscemi, Joe Pesci, Christian Berkel, Rutger Hauer, Mel Gibson, Elliott Gould, Tim Robbins, Daniel Craig, Jeffrey Wright, Matthew Modine, Domhnall Gleeson, Brendan Gleeson, John Hurt, Michael Stuhlbarg, Hugo Weaving, John Goodman, Mark Hamill, Colin Farrell, Ken Watanabe, Clint Eastwood, Ralph Fiennes, Val Kilmer, John Hawkes, Ben Kingsley, Seth Rogen, Robert Duvall, Brad Pitt, Max von Sydow, Stanley Tucci, Tom Cruise, Christopher Lloyd, Tommy Lee Jones, Jason Statham, Michael Caine, Paul Giamatti, Josh Hutcherson, Adrien Brody, Michael J, Jeremy Renner, Liam Neeson, Mark Ruffalo, Terrence Howard, John Cleese, Harrison Ford, Clive Owen, Jake Gyllenhaal, Will Smith, Danny DeVito, Elijah Wood, Sean Connery, Tom Sizemore, Stellan Skarsgård, Robin Williams, Hugh Jackman, John Lithgow, Benedict Cumberbatch, Mykelti Williamson, John Malkovich, Gary Oldman, Johnny Depp, Jeff Bridges, Hugh Grant, Jean Reno, Aaron Eckhart, Michael Madsen, Jude Law, J.K, Jon Voight, Casey Affleck, Robert Pattinson, Daniel Brühl, Billy Bob Thornton, Russell Crowe, Ewan McGregor, Christopher Walken, Morgan Freeman, Josh Brolin, Richard Harris, Shea Whigham, Bill Murray, Christoph Waltz, Jamie Foxx, Christopher Plummer, Ethan Hawke, Albert Finney, Miles Teller, Don Johnson, Javier Bardem, Bill Paxton, Robert De Niro, Timothée Chalamet, Sam Rockwell, Kevin Bacon, Simon Pegg, Sean Penn, Ving Rhames, Tom Hanks, Anthony Hopkins, Heath Ledger, Tim Roth, Martin Sheen, Michael Keaton, Joseph Gordon-Levitt, Kyle Chandler, John Travolta, Bruce Dern, Steve Carell, Dustin Hoffman, Oscar Isaac
Athelete	Dirk Nowitzki, Mia Hamm, Diana Taurasi, Allyson Felix, Sheryl Swoopes, David Ortiz, James Harden, Mike Trout, Mookie Betts, Chris Paul, Aitana Bonmati, Roger Federer, Faker, Annika Sorenstam, Thierry Henry, Jimmie Johnson, Tom Brady, Randy Moss, Shohei Ohtani, Kevin Durant, Zinedine Zidane, Calvin Johnson, Novak Djokovic, LeBron James, Alexia Putellas, Albert Pujols, Max Scherzer, Mariano Rivera, Ichiro Suzuki, Patrick Mahomes, Aaron Donald, Steve Nash, Georges St-Pierre, Giannis Antetokounmpo, Stephen Curry, Floyd Mayweather, Clayton Kershaw, Manny Pacquiao, Andrés Iniesta, Barry Bonds, Tim Duncan, Lauren Jackson, Luka Modric, Shelly-Ann Fraser Pryce, Bryce Harper, Ray Lewis, Simone Biles, Rafael Nadal, Derek Jeter, Shaun White, Michael Schumacher, Peyton Manning, Candace Parker, Nikola Jokic, Serena Williams, Jason Kidd, Andy Murray, Mikaela Shiffrin, Lewis Hamilton, Lisa Leslie, Bernard Hopkins, Kobe Bryant, Justin Verlander, Tamika Catchings, Alex Rodriguez, Jon Jones, Tiger Woods, Dwyane Wade, Kohei Uchimura, Michael Phelps, Xavi Hernandez, Cristiano Ronaldo, Usain Bolt, Max Verstappen, Kawhi Leonard, Venus Williams, Katie Ledecky, Kylian Mbappé, Maya Moore, Alex Ovechkin, Sidney Crosby, Phil Mickelson, Adrian Beltré, Kevin Garnett, Miguel Cabrera, Lionel Messi
Entrepreneur	Brian Chesky, Garrett Camp, Kevin Systrom, Sam Walton, Larry Ellison, Ratan Tata, Ritesh Agarwal, Ted Turner, Steve Jobs, Jack Ma, Richard Branson, Jeff Bezos,
Politician	Che Guevara, Mike Pence, Li Ka-shing, Woodrow Wilson, Dwight D. Eisenhower, Abdel Fattah el-Sisi, Franklin D. Roosevelt, Yasser Arafat, Haruhiko Kuroda, Donald Trump, Bill Clinton, Stephen Schwarzman, Narendra Modi, Gianni Infantino, Masayoshi Son, Bernard Arnault, Hui Ka Yan, Benjamin Netanyahu, Winston Churchill, Vladimir Putin, John F. Kennedy, Theodore Roosevelt, Nelson Mandela, Sergey Brin, Margaret Thatcher, Xi Jinping, Ronald Reagan, Golda Meir, Recep Tayyip Erdogan, Charles de Gaulle, Jim Yong Kim, Warren Buffett, Qamar Javed Bajwa, Jerome H. Powell, Wang Jianlin, Lech Wałęsa, Michel Temer, Doug McMillon, Mohammed bin Salman Al Saud, Lloyd Blankfein, Lee Hsien Loong, Jawaharlal Nehru, Shinzo Abe, Michael Bloomberg, Tony Blair, Li Keqiang, Rodrigo Duterte, Justin Trudeau, Hu Jintao, Bob Iger, Mario Draghi, Khalifa bin Zayed Al-Nahyan, Bashar al-Assad, Ayatollah Khomeini, Ali Hoseini-Khamenei, Indira Gandhi, Deng Xiaoping, Kim Jong-un, Ma Huateng, Joseph Stalin, Mikhail Gorbachev, Moon Jae-in, Mary Barra, Mahatma Gandhi, Christine Lagarde, Jokowi Widodo, Mao Zedong, Ken Griffin, Mustafa Kemal Atatürk, Theresa May, Aliko Dangote, Darren Woods, Jiang Zemin, Rupert Murdoch, Fidel Castro, Jean-Claude Juncker, Robert Mueller, Enrique Pena Nieto, Carlos Slim Helu, Tim Cook, Robin Li, Antonio Guterres, Larry Fink
Scientist	lbert Einstein, Fei-Fei Li, Jennifer Doudna, Yann LeCun, Thomas Edison, Gregor Mendel, Geoffrey Hinton, John von Neumann, Max Planck, Sam Altman, Tim Berners-Lee, Andrew Ng, Rosalind Franklin, Andre Geim, Marie Curie, Ilya Sutskever, Kip Thorne, James Watson, Srinivasa Ramanujan, Nikola Tesla, Niels Bohr, Enrico Fermi, Rachel Carson, Yoshua Bengio, Alan Turing, Stephen Hawking, Francis Crick, Linus Pauling, Demis Hassabis, Werner Heisenberg, Barbara McClintock

Table 5. Names for each sub-domain of the celebrity recognition subset.

```
{
    "Instruction": "I will provide you with an image of a person. Please list as many detailed attributes as possible based on the following categories. Below are some examples you can refer to:",
    "Gender": ["Male", "Female", "Unknown"],
    "Age": ["Infant", "Child", "Teenager", "Adult", "Elderly", "Unknown"],
    "Ethnicity": ["Caucasian", "African", "Asian", "...", "Unknown"],
    "Occupation": ["Doctor", "Engineer", "Teacher", "...", "Unknown"],
    "Appearance": {
        "Hair": {
            "Type": ["Short", "Long", "Curly", "Straight", "Bald", "Ponytail", "Unknown"],
            "Color": ["Red", "Blue", "Green", "Yellow", "Black", "White", "...", "Unknown"],
            "Description": ["Short black hair with dotted pattern", "...", "Unknown"]
        },
        "Beard": {
            "Type": ["Full Beard", "Goatee", "Mustache", "Unknown"],
            "Color": ["Red", "Blue", "Green", "Yellow", "Black", "White", "...", "Unknown"],
            "Description": ["Black full beard", "...", "Unknown"]
        },
        "Expression": {
            "Type": ["Smiling", "Frowning", "Surprised", "Angry", "Sleeping", "Crying", "Laughing", "...", "Unknown"],
            "Description": ["Smiling with closed eyes", "...", "Unknown"]
        }
    },
    "Clothing & Accessories": {
        "Clothing": {
            "Upper": {
                "Type": ["T-shirt", "Shirt", "Blouse", "Unknown"],
                "Color": ["Red", "Blue", "Green", "Yellow", "Black", "White", "...", "Unknown"],
                "Description": ["Purple hooded puffer jacket", "...", "Unknown"]
            },
            "Lower": {
                "Type": ["Jeans", "Skirt", "Shorts", "Unknown"],
                "Color": ["Red", "Blue", "Green", "Yellow", "Black", "White", "...", "Unknown"],
                "Description": ["Black pants with dotted patterns", "...", "Unknown"]
            },
            "Shoes": {
                "Type": ["Sneakers", "Boots", "Sandals", "Barefoot", "Unknown"],
                "Color": ["Red", "Blue", "Green", "Yellow", "Black", "White", "...", "Unknown"],
                "Description": ["White sneakers with red stripes", "...", "Unknown"]
            }
        },
        "Accessories": {
            "Hat": {
                "Type": ["Baseball Cap", "Beanie", "Fedora", "Unknown"],
                "Color": ["Red", "Blue", "Green", "Yellow", "Black", "White", "...", "Unknown"],
                "Description": ["Black beanie with white stripes", "...", "Unknown"]
            },
            "Glasses": {
                "Type": ["Sunglasses", "Prescription Glasses", "Goggles", "Unknown"],
                "Color": ["Red", "Blue", "Green", "Yellow", "Black", "White", "...", "Unknown"],
                "Description": ["Black sunglasses with red frame", "...", "Unknown"]
            }
        }
    },
    "Posture": {
        "Type": ["Standing", "Sitting", "Lying", "Arms Crossed", "Arms Raised Overhead", "...", "Unknown"],
        "Description": ["Sitting on a bench", "Standing on one leg", "...", "Unknown"]
    },
    "Actions": {
        "Type": ["Walking", "Running", "Jumping", "Sitting", "Standing", "Sleeping", "...", "Unknown"],
        "Description": ["Walking a dog", "Reading a book with a red cover", "...", "Unknown"]
    }
}
```

Figure 2. Prompt used to generate structured property dictionary.