DSV-LFS: Unifying LLM-Driven Semantic Cues with Visual Features for Robust Few-Shot Segmentation

Supplementary Material

In the supplementary material, we provide qualitative examples with detailed class descriptions.

6. Qualitative Results

We present qualitative results from our proposed method to demonstrate its effectiveness in addressing key challenges in few-shot segmentation. The two primary challenges in this context are: (1) the misclassification of base class objects as novel classes, resulting in false positives, and (2) reliance on a limited set of support images, which often fails to capture the full range of target class variations.

The following examples illustrate challenging episodes that highlight these issues. In the examples, the input episode consists of annotated support and query images alongside the DSV-LFS output. While the support and query images are annotated to specify the object of interest, the annotation on the DSV-LFS output represents the predictions of the proposed method. Additionally, a detailed class description is provided as an input prompt for the multi-modal LLM to generate the semantic prompt.<Qimage>in the class descriptions serves as a default token assigned to the query image within the input class description. This token is subsequently replaced by the output features of the query image.

Input Episode
Motorcycle


Support Image


Query Image


DSV-LFS

Complete class description for mutli-modal LLM for motorcycle object class:

<Qimage>. this one is a query image. A motorcycle is distinctively characterized by its two-wheeled structure, which sets it apart from other vehicles. The wheels are large and typically exposed, with a prominent front wheel that often includes a visible brake disc and caliper, and a rear wheel that may have a broader tire. The frame is compact and streamlined, with a noticeable absence of a roof or any extensive enclosure. The handlebars, which are prominently situated above the front wheel, feature visible controls and mirrors extending outward. The seat is elongated and generally positioned for a straddling rider, often with a noticeable saddle shape. Beneath the seat, the engine is a dominant visual element, with its intricate metallic components like the exhaust pipes and cylinders often exposed. The fuel tank, typically located in front of the seat and above the engine, is a rounded or angular structure with a glossy finish. Additionally, motorcycles have a distinctive set of front and rear lights; the front light is usually a singular, circular or angular headlamp, while the rear includes a smaller brake light. The suspension system, including visible shock absorbers and forks, also adds to its unique visual identity. Overall, a motorcycle's open, mechanical design with exposed wheels, engine, and handlebars, along with its streamlined silhouette, distinctly sets it apart from bicycles, scooters, and other similar object categories.A motorcycle is distinguished from other similar object categories by several unique visual features. Primarily, it has two large wheels in tandem with a sleek, streamlined frame connecting them. The frame often exhibits a minimalistic, exposed design, typically featuring a prominent fuel tank situated between the handlebars and the riders seat. The handlebars, located at the front, are usually higher than the seat and are connected to a visible front fork that holds the front wheel. Unlike bicycles, motorcycles have a more robust structure with a bulkier engine block situated beneath the fuel tank, which is often exposed or partially covered. The exhaust system, consisting of one or more metal pipes, extends from the engine towards the rear. Motorcycles also feature foot pegs for the rider, often accompanied by additional pegs or a small seat for a passenger. The rear wheel is connected to the frame via a swingarm, which allows for suspension and typically houses a single large shock absorber. The design may include fairings, which are aerodynamic covers, though these are not as extensive as on scooters or other fully-enclosed vehicles. Additionally, motorcycles usually have larger, more prominent headlights and taillights compared to bicycles, often integrated into the design rather than being detachable. The tires on motorcycles are wider and more robust than those on bicycles, designed to handle higher speeds and more significant weight. These visual features collectively distinguish motorcycles from bicycles, mopeds, and scooters. This paragraph outlines the visual features of motorcycle that distinguish it from other similar categories. Please use distinguishing visual features to segment motorcycle in the query image.

Input Episode
Keyboard


Support Image


Query Image


DSV-LFS

Complete class description for mutli-modal LLM for keyboard object class:

<Qimage>. this one is a query image. A keyboard, in its distinctive visual form, is typified by its flat, elongated shape with an array of rectangular keys arranged in neat rows. Each key is typically square or slightly rectangular, often featuring rounded edges for ergonomic comfort during typing. The surface of the keys is uniformly smooth and matte or glossy, contrasting with the often darker or neutral-colored base. These keys are distinctly marked with alphanumeric characters, symbols, and functional indicators, often in contrasting colors such as white or light gray on dark backgrounds, aiding visibility and usability. Additionally, keyboards commonly include functional sections such as arrow keys, function keys (F1-F12), and a dedicated numerical keypad (on larger models), each section visually demarcated or slightly raised for tactile distinction. The overall profile of a keyboard is thin and flat, designed for ergonomic use on desks or tables, typically with a USB cable or wireless connectivity. These visual features collectively distinguish a keyboard from similar objects like calculators or remote controls, which lack the array of keys and alphanumeric layout essential for text input and control in computing environments. This paragraph outlines the visual features of keyboard that distinguish it from other similar categories. Please use distinguishing visual features to segment keyboard in the query image.

Input Episode
Toilet


Support Image


Query Image

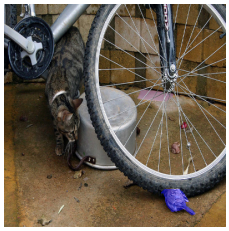

DSV-LFS

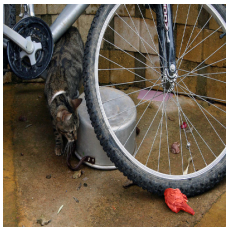
Complete class description for mutli-modal LLM for toilet object class:

<Qimage>. this one is a query image. A toilet is a distinct bathroom fixture characterized by several unique visual features. It typically has a bowl-shaped seat made of porcelain or ceramic, with a rounded or oval opening that slopes inward. The bowl is often connected to a pedestal or base, which is relatively narrow compared to the bowl itself, giving it a recognizable silhouette. Attached to the back of the bowl is a water tank, which is usually rectangular and taller than it is wide, designed to hold flushing water. The toilet seat, often made of plastic, is hinged at the rear and can be lifted or lowered. This seat usually has a lid that matches in material and design. The bowl's interior is smooth and glazed, facilitating easy cleaning and often features a water-filled trap at the bottom, visible when the lid and seat are raised. The flush handle or button is typically located on the side or top of the water tank, which distinguishes it from other fixtures like bidets or urinals that lack such a tank. Overall, the combination of the bowl's shape, the attached water tank, the hinged seat and lid, and the flush mechanism make the toilet visually distinct from similar bathroom objects.This paragraph outlines the visual features of toilet that distinguish it from other similar categories. Please use distinguishing visual features to segment toilet in the query image.

Input Episode
Bird


Support Image


Query Image


DSV-LFS

Complete class description for mutli-modal LLM for bird object class:

<Qimage>. this one is a query image. Birds are characterized by their distinctive features, which set them apart from other similar object categories. Birds possess a unique feather covering, often brightly colored or patterned, providing insulation and aiding in flight. Their beaks vary in shape and size depending on their diet, from sharp, curved beaks in birds of prey to flat, broad ones in filter feeders. They have lightweight, streamlined bodies adapted for flight, with a high degree of symmetry and hollow bones. Their wings, a key identifier, exhibit a range of shapes and sizes, from long and narrow in soaring birds to short and rounded in those requiring rapid takeoff. The presence of a tail, often fan-shaped and used for steering during flight, further distinguishes them. Birds also have distinctive legs and feet, with variations such as webbed feet for swimming or strong talons for hunting. Their eyes are generally large and positioned on the sides of their heads, offering a wide field of vision. These combined features create a visual profile unique to birds, setting them apart from other animal categories. This paragraph outlines the visual features of bird that distinguish it from other similar categories. Please use distinguishing visual features to segment bird in the query image.

Input Episode
Cow


Support Image


Query Image


DSV-LFS

Complete class description for mutli-modal LLM for cow object class:

<Qimage>. this one is a query image. Cows possess several distinguishing visual features that set them apart from similar object categories. They have a large, robust body with a pronounced rectangular shape, supported by four sturdy legs ending in cloven hooves. Their heads are relatively large, with broad, flat foreheads and distinctive long, broad snouts. A cow's eyes are large and round, usually positioned on the sides of their head, giving them a wide field of vision. They have large, prominent ears that can be either upright or slightly drooping. One of the most notable features is their pair of horns, which can vary in size and shape but are typically curved and symmetrical, though some cows may be naturally polled (hornless). Their tails are long and thin, ending in a tuft of hair, used to swat away insects. The skin of cows is covered in short hair, with color patterns that can vary significantly, including solid colors, spots, and patches in hues of black, white, brown, or a combination thereof. Unlike other similar animals, cows have a prominent udder with visible teats, particularly in dairy breeds, which is a key distinguishing feature. Additionally, cows have a distinctive gait and posture, often appearing more slow-moving and deliberate compared to other livestock. This paragraph outlines the visual features of cow that distinguish it from other similar categories. Please use distinguishing visual features to segment cow in the query image.

Input Episode
Hair dryer


Support Image


Query Image


DSV-LFS

Complete class description for mutli-modal LLM for hair dryer object class:

<Qimage>. this one is a query image. A hair dryer can be visually distinguished from similar objects primarily by its specific design features. Typically, a hair dryer consists of a cylindrical or slightly tapered body with a prominent handle and a nozzle at one end. The body often features a perforated grill or vents for airflow, which is essential for its function. The handle is ergonomically designed for grip and control, often contrasting in texture or color from the main body to enhance usability and visibility. On the body, there are frequently control buttons or switches for adjusting heat and airflow settings, which are clearly marked and distinct in appearance. The nozzle itself is narrow and elongated, sometimes with a distinct shape or curvature depending on the model, facilitating directional airflow during use. These visual characteristics collectively differentiate a hair dryer from other similar objects like handheld vacuum cleaners or electric razors, which have different body shapes, nozzle configurations, and control mechanisms tailored to their respective functions. This paragraph outlines the visual features of hair dryer that distinguish it from other similar categories. Please use distinguishing visual features to segment hair dryer in the query image.