Language Guided Concept Bottleneck Models for Interpretable Continual Learning

Supplementary Material

8. Supplementary

8.1. Concepts Preparation

Before initiating the incremental learning process, we retrieve concepts associated with specific categories through ChatGPT queries. Following the methodologies proposed in [36, 58, 61], we employ the following prompts to guide the ChatGPT interactions:

Table 5. Prompts for all benchmark datasets.

Dataset	Prompts
Coarse-grained, CUB-200	"using {num} sentenses to discribe the **appearance / color / size / shape / surroundings** of {category}"
Food101	"using {num} sentences to describe the **apperance / shape / color / texture** of a food named {category}"
Flower	"using {num} sentences to describe the **apperance / color / size / patten / texture** of a flower named {category}"
Stanford-cars	"using {num} sentences to describe the **appearance / shape / color / size / structure** of a car named {category}"

By utilizing ChatGPT with prompts mentioned above, we generate descriptive sentences containing class-specific concepts. Subsequently, we employ a T5 model, as redesigned by [61], to extract concepts from these sentences, thereby constructing a comprehensive general concept pool. Examples of extracted concepts for categories in CUB-200, ImageNet-subset, Food-101 and Flower are provided in Tables 11 to 14.

8.2. Additional Implementation Details

More Details of Concept Selection Module. Following [58], we implemented the concept selection module (CS) as a simple MLP. At the beginning of task t, the CS is trained on the training data of current task with cross-entropy loss and Mahalanobis loss, encouraging CS to construct embedding space with vision-language knowledge. We train CS for 30 epochs with a batch size of 64 and a learning rate of 0.01 . After training, the learned weight matrix of CS is leveraged to select concepts from concepts set $\mathcal{C}_t$ of task t, based on the distances between text features $f_T(\mathcal{C}_t)$ and weight matrix.

Pseudo Code. The training pipeline of our proposed method is outlined in Algorithm 1, the image encoder f_I and text encoder f_T of CLIP are both frozen during the whole training process, only the CBL, classifier and CS are trainable. For each task, we first select concepts to construct a bottleneck and extract prototypes for the categories relevant to the current task. Afterward, old prototypes are augmented using data from the current task to address catastrophic forgetting. Lastly, we calculate the CLIP concept activation matrix and train our model with the guidance of three loss functions as described in Section 4.2.

Algorithm 1 LG-CBM for Interpretable CL

Input: Incremental datasets: $\{\mathcal{D}^1, \mathcal{D}^2, \cdots, \mathcal{D}^n\}$, Task-specific concepts: $\{\mathcal{C}_1, \mathcal{C}_2, \cdots, \mathcal{C}_n\}$, Pre-trained CLIP image encoder and text encoder: f_I, f_T.
Output: Incrementally trained model with interpretability

```
for t = 1, 2, ··· , n do
    Get the training set D^t and Concepts C_t;
    Extract text feature of C_t;
    Select concepts from C_t to form bottlenecks B_t;
    Extract the prototypes of D^t as P_t;
    if t > 1 then
        Augment prototypes P_{1:t-1} via Equation (6)
        and Equation (7);
    end if
    Compute CLIP concept activation matrix
    via Equation (1);
    Optimize the CBL and classifier via Equation (9);
end for
```

8.3. More Results of Experiments

Accuracy Curve in Various Settings. We illustrate the accuracy decreasing trends of our method with other state-of-the-art baselines on $\bar{A}$ across all benchmark datasets in Figure 5. Our method outperforms other methods with higher accuracy and less forgetting in most settings, especially when the initial task contains almost half of categories of the entire datasets, the A_{last} performance of our method is also best on most benchmark datasets.

Concepts From Different ChatGPT. As shown in Table 6, we report the average incremental accuracy $\bar{A}$ on ImageNet-subset and Flower datasets with concepts generated from different ChatGPT models. The findings demonstrate that concepts derived from distinct ChatGPT models have minimal impact on the model's performance, with accuracy fluctuations of less than 0.2% on ImageNet-subset and 0.5% on Flower. For all experiments conducted with our method, we

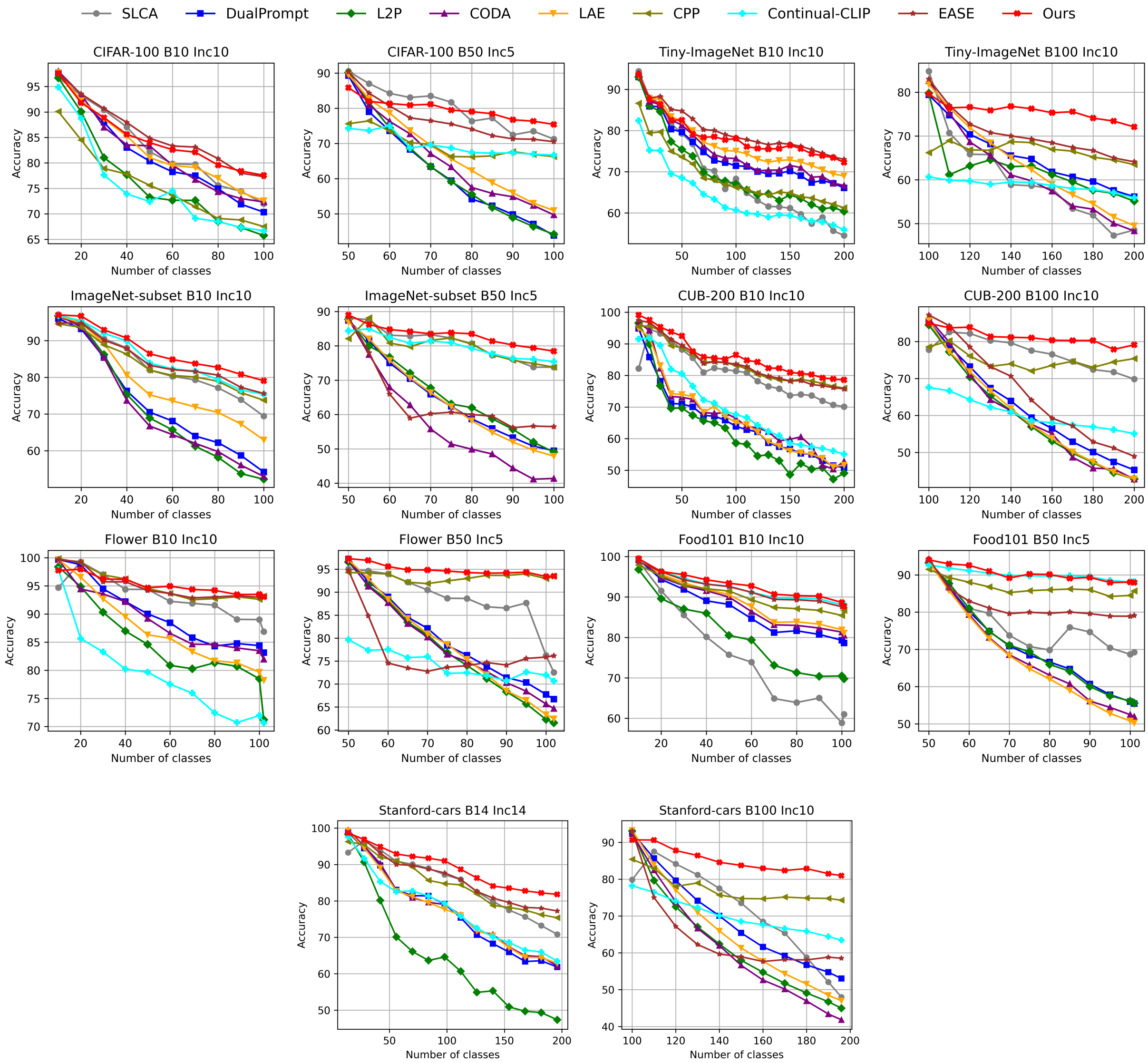

Figure 5. Average incremental accuracy $\bar{A}$ curve of our method with other state-of-the-art methods across all benchmark datasets.

obtain concepts by querying GPT-4o.

Table 6. The average incremental accuracy $\bar{A}$ on the ImageNet-subset and Flower datasets, evaluated using concepts derived from different versions of ChatGPT.

ChatGPTs	ImageNet-subset B10 Inc10	Flower B10 Inc10
GPT-3.5 turbo	87.63	94.69
GPT-4 turbo	87.50	94.76
GPT-4o	87.55	95.16

Performance in Non-continual Setting. We present the results for joint training of our method on three coarse-grained datasets in Table 7. A performance gap remains between continual learning methods and joint training.

Table 7. The performance of our method under Non-continual setting.

Methods	CIFAR-100	Tiny-ImageNet	ImageNet-Subset
Joint training	82.14±0.02	76.43±0.11	83.52±0.06

Additional Ablation Study. Semantic-Guided Prototype Augmentation (PA) is designed to enhance knowledge retention. We evaluate its impact by comparing it against "Base" (without any anti-forgetting strategy) and "Base+Proto" (classical prototype loss without augmentation). The results show the effectiveness of the PA module.

Table 8. The effectiveness of semantic-guide prototype augmentation module.

	CIFAR-100 B50 Inc5	Tiny-ImageNet B100 Inc10	ImageNet-Subset B50 Inc5
Base	11.02±1.42	12.96±3.97	23.90±8.50
Base+Proto	69.15±0.38	64.74±0.62	74.23±0.25
Base+PA	**75.91±0.50**	**71.97±0.09**	**78.21±0.29**

The Concept Alignment (CA) module helps learn human-understandable bottlenecks and enhances interpretability. We show an example of "Azalea" to analyze CA, listing the top-5 concepts in Table 9. With CA, the selected concepts are all positive and align well with human understanding, emphasizing the importance of the CA module.

Table 9. The comparison of Top-5 concepts that contribute most to classify "Azalea" with and without CA module.

Image	CA	Top-5 Concepts
	w/o CA	**NOT** radiate from the center like rays of sunshine each petal is thin and almost translucent **NOT** encase a deep pink center **NOT** striking symmetry radiating from the center stark contrast against its dark green foliage
	w/ CA	support small pink blossoms bright pink in color vibrant pink hue with a pale white margin thrive in the shaded nooks of the tree limbs stark contrast against its dark green foliage

8.4. Computational Overhead of Concept Selection Module

The computational overhead of the CS on three coarse-grained datasets is shown in Table 10. We can infer from Table 10 that the training of CS is quite efficient, with low memory usage and computation, and the training time is also acceptable.

Table 10. The training cost of the concept selection module.

Metrics	CIFAR-100	Tiny-ImageNet	ImageNet-Subset
FLOPs (M)	0.68	1.56	0.99
Time per epoch (s)	1.30	1.26	2.24
Peak Memory Usage (MB)	2571	2701	2673

8.5. More Interpretable Model Predictions.

As depicted in Figures 6 to 8, we provide additional examples of interpretable model predictions across various datasets. We can find that our method demonstrates excellent interpretability across all benchmark datasets, delivering strong performance accompanied by coherent and logical explanations.

Table 11. Example concepts of categories in CUB-200.

Category	Concepts	Category	Concepts
Black footed Albatross	“graceful, elongated body”, “long, slender wings”, “plume is primarily dark brown with subtle shades of gray”, “pale patch on its face creates an interesting visual effect”, “distinguishing feature against the rest of its body”, “large wingspan that stretches wide”, “body appears robust and streamlined”, “strong and hooked”, “overall form is well-suited for gliding over the ocean”, “endless, clear sky”, “glides over the gentle waves with ease”	Crested Auklet	“dark, slate-gray body”, “sleek look”, “head is adorned with a striking crest”, “adds to its distinctive appearance”, “around its beak”, “there is a splash of vibrant orange”, “small, bright eyes stand out against the darker feathers”, “small bird with a compact body”, “beak is short and slightly curved”, “prominent crest on its head”, “relatively short compared to its body”, “rocky coastline is covered in patches of green moss”, “waves crash against the shore under an overcast sky”
Rusty Blackbird	“dark, glossy plumage that shimmers in the sunlight”, “feathers display a subtle iridescence of blues and greens”, “rusty hue adorns its wings and patches around its eyes”, “bill is slender and pointed”, “suitable for foraging”, “medium-sized bird with a slender build”, “body appears elongated and streamlined”, “has a relatively long tail”, “straight and slightly pointed”, “bird is perched on a thin branch, surrounded by dense foliage”, “background features a calm stream reflecting the overhanging trees”	Gray Catbird	“sleek body with a predominantly gray plumage”, “head is capped with a darker, almost black hue”, “small, slender beak complements its smooth feathers”, “tail and wings show subtle, lighter shading”, “medium-sized bird with a slender body”, “overall shape is sleek and elongated”, “long tail that tapers towards the end”, “bird features a rounded head with a short, straight bill”, “gentle stream runs nearby”, “surrounded by tall reeds and vibrant autumn leaves”

Table 12. Example concepts of categories in ImageNet-subset.

Category	Concepts	Category	Concepts
Goldfish	“bright orange coloration with a metallic sheen”, “fins are long and delicate”, “flowing gracefully in the water”, “sleek and oval-shaped”, “slight bulge at the sides”, “eyes are round and prominent”, “have an elongated body shape”, “typically around four to six inches long”, “fins are delicate and fan-like”, “bodies are often plump and streamlined”, “clear water surrounds them in an aquarium”, “green aquatic plants sway gently nearby”	Cock	“vibrant plumage with a mix of rich, earthy tones”, “bright red adorns its comb and wattle, adding a striking contrast”, “hues of green and blue”, “sturdy beak and strong legs complete the appearance”, “appears quite large in the photo”, “structure is straight and firm”, “surface looks smooth and uniform”, “overall, it gives an impression of solidity and symmetry”, “warm glow over the farmyard”, “illuminating the rustic wooden fence and scattered hay”, “nearby, a few chickens peck at the ground”
Tailed Frog	“brown, mottled skin”, “blends with the forest floor”, “large and sit prominently on its head”, “muscular and well-developed for jumping”, “small, tail-like appendage extends from its rear”, “small and robust with a rounded body”, “legs appear muscular and strong”, “head is wide with prominent eyes”, “slender tail extends from its back”, “small amphibian sits among the damp, moss-covered rocks”, “lush, green environment”	Agama	“vibrant body with a mixture of colors”, “head is often bright red or orange”, “body can display hues of blue or brown”, “smooth and glossy under the light”, “medium-sized reptile with a stout body”, “limbs are strong and muscular, aiding in movement”, “head is somewhat triangular, tapering towards snout”, “tail is long and slender, extending beyond the body”, “rocky terrain is dotted with patches of dry grass and scattered stones”, “clear blue sky stretches above”, “casting shadows on sunlit ground”

Table 13. Example concepts of categories in Food101.

Category	Concepts	Category	Concepts
Chocolate Cake	“appears round and sits on a flat surface”, “rich, dark brown”, “smooth, glossy finish”, “reflect off its slightly uneven edges”, “uniform throughout”, “appears dense yet tender”, “surface shows a slight sheen, suggesting a moist consistency”, “each slice reveals layers with a slightly crumbly edge;”, “smoothly glazed”, “contrasting with the softer interior”, “Tiny air pockets are scattered throughout”, “spongy nature”, “reveals a dense and moist interior”, “texture appears soft, showcasing fine crumbs”, “uniform color throughout”, “Tiny flecks may indicate the presence of fine ingredients”	Hot dog	“long, cylindrical shape with slightly rounded ends”, “nestled within a soft, oblong bun”, “bright red and mustard streaks on top”, “suggest a freshly cooked state”, “cooked evenly, its texture looks firm yet somewhat flexible”, “glisten, indicating juiciness”, “densely packed interior”, “hinting at a substantial bite”, “inside, it reveals a grilled sausage nestled within the bun”, “slightly browned”, “shows a textured casing”, “drizzled atop”, “blending into the sausage”, “soft layered”, “provide a cozy embrace”
Lasagna	“rectangular dish layered with alternating levels of ingredients”, “golden-brown with a slightly crisp texture”, “edges appear slightly darker, indicating a well-cooked surface”, “visible stripes of red and white sauce peeking through through”, “warm and inviting look”, “various shades of red, brown, and cream”, “layered structure with firm and slightly chewy pasta sheets”, “gooey and stretchy consistency from the melted cheese”, “moist and saucy texture between the layers”, “blend of soft, meaty, and creamy elements within it”, “visible layers alternate between creamy white and tangy red sauces”, “oozes over the top, creating a slightly golden crust”, “add texture and flavor”, “thin, flat pasta shapes the structure, holding everything together”, “specks of green are scattered throughout”	Ice-cream	“smooth, round shape with slight indentations”, “hue blends subtle shades of beige and light brown”, “surface appears glossy, catching light reflections”, “small, dark specks adds texture detail to its appearance”, “contrasts sharply with the dark background”, “smooth, creamy surface”, “glistens under the light”, “heaped scoops reveal tiny air pockets throughout”, “small droplets of condensation are visible on its exterior”, “soft, slightly elastic quality when scooped”, “melting edges give way to a glossy, liquid sheen”, “smooth and rich”, “Tiny air bubbles are evenly dispersed throughout the frozen treat”, “small vanilla bean specks are visible, hinting at quality ingredients”, “surface has a slightly glistening, frosty appearance”, “dense, yet soft composition”

Table 14. Example concepts of categories in Flower.

Category	Concepts	Category	Concepts
Fire lily	“vibrant red and orange petals that flare outward”, “petals have a wavy and slightly ruffled texture”, “long, green stamens protrude prominently at the center”, “leaves are slender and gracefully arch away from the stem”, “sits atop a tall, curved stem”, “adds elegance to its appearance”, “vivid shade of red-orange”, “each petal gracefully curls backwards”, “bright yellow”, “long, slender stamens protrude from it”, “overall shape is elegant and delicate”, “vibrant orange-red petals that curve gracefully backwards”, “stamens and pistil extend prominently from the center”,	Red ginger	“stands tall with elongated, vibrant red bracts crowded closely together”, “strong green stem supports the structure and extends upwards”, “leaves are broad, glossy, and waxy with a deep green color”, “graceful arcs”, “add to its elegance”, “striking combination of vivid color and slender form”, “vibrant red color”, “stands out vividly”, “shape is elongated, resembling a cone or spike”, “smooth”, “layered neatly around the core”, “green leaves surround its base, framing it beautifully”,
Corn poppy	“vibrant red petals that catch the eye”, “dark and contrasting, almost black”, “stem is slender and green, standing tall”, “sparse and slightly jagged”, “overall appearance is delicate yet striking”, “vibrant with a bright red hue”, “petals are delicate and slightly crinkled”, “flower has a central dark spot marking its contrast”, “stands tall on a slender green stem”, “overall shape is rounded and open”, “petals are bright red and slightly crinkled”, “black spot marks the base of each petal”,	Artichoke	“stands tall with thick, green stems reaching upwards”, “leaves are large, broad”, “have a silvery hue”, “flower head is a round, dense cluster of tightly layered bracts”, “hint of purple peek through as the flower begins to bloom”, “delicate thorns edge the tips of its protective leaves”, “vibrant purple hue”, “stands tall with a robust structure”, “petals form spherical shape”, “petals overlap tightly together”, “overall appearance is spiky yet symmetrical”, “large, spiky petals”, “curl out in a vibrant display”,

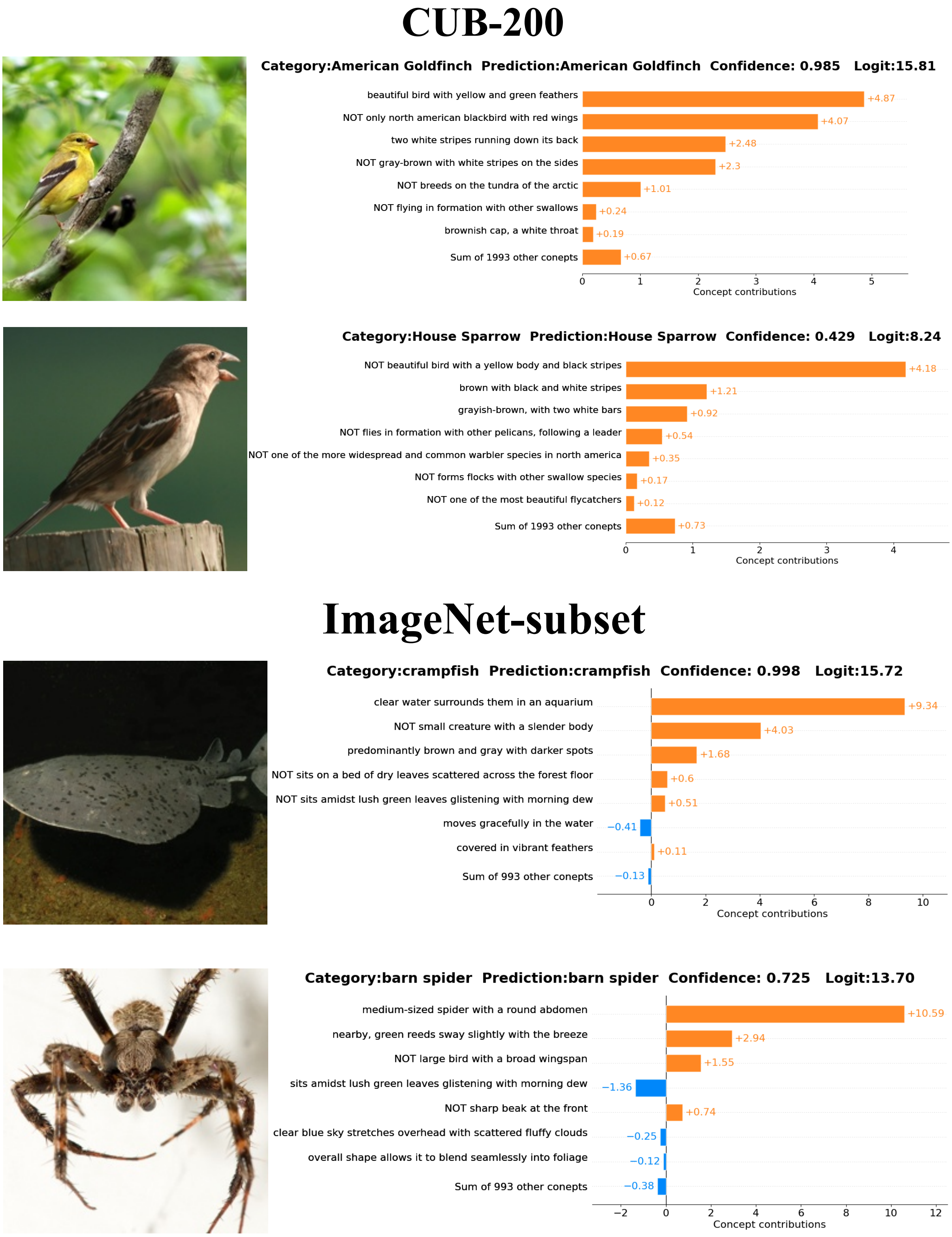

Figure 6. Contribution Visualization after training on CUB-200 B10 Inc10 and ImageNet-Subset B10 Inc10.

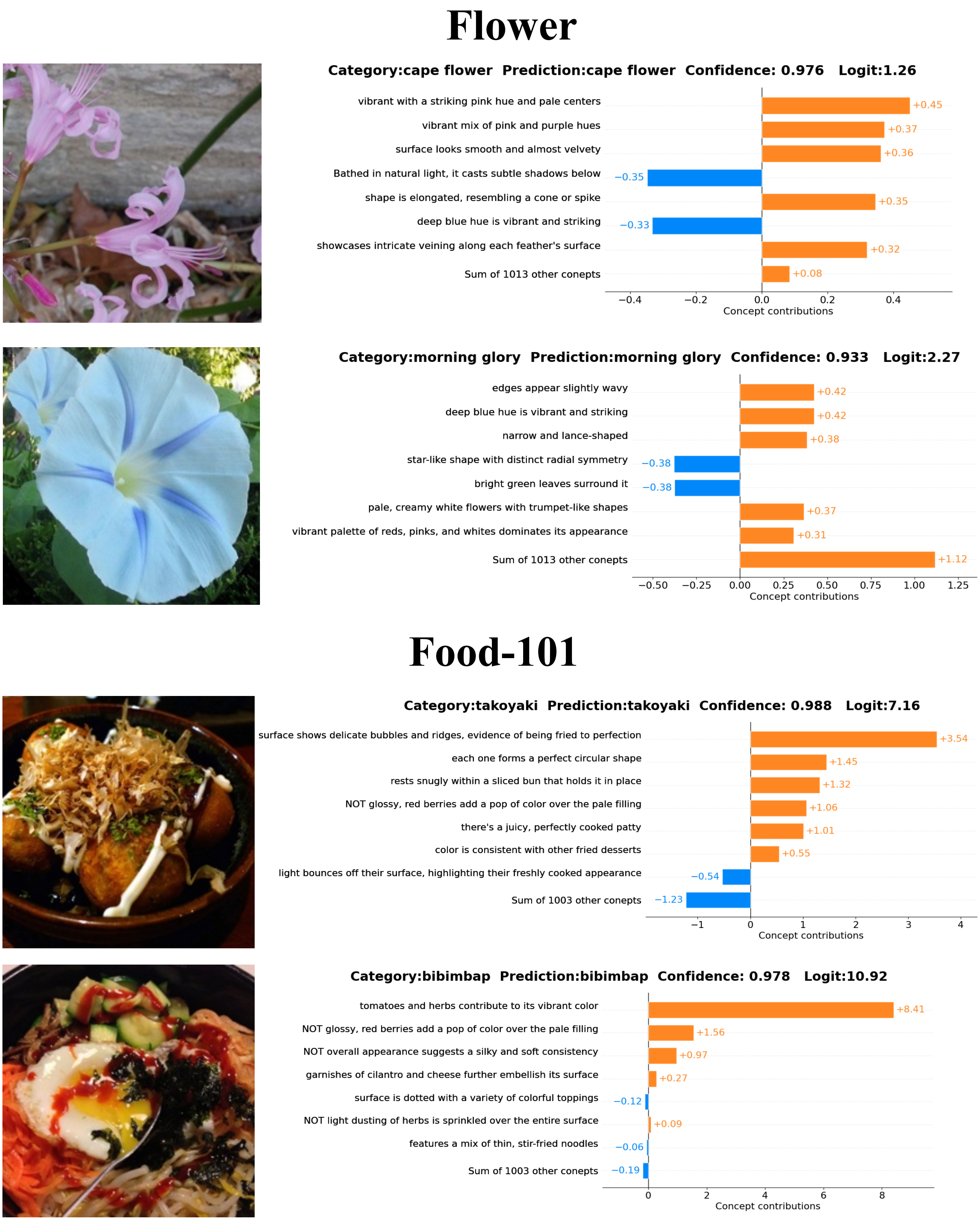

Figure 7. Contribution Visualization after training on Flower B10 Inc10 and Food-101 B10 Inc10.

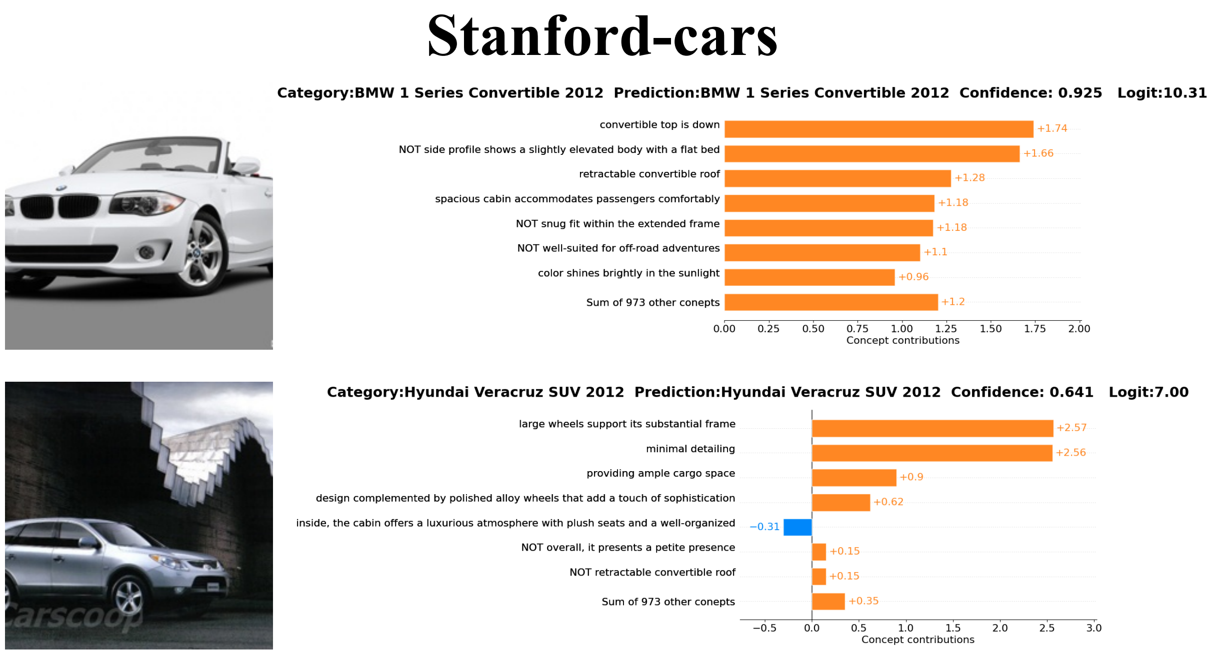

Figure 8. Contribution Visualization after training on Stanford-cars B14 Inc14.