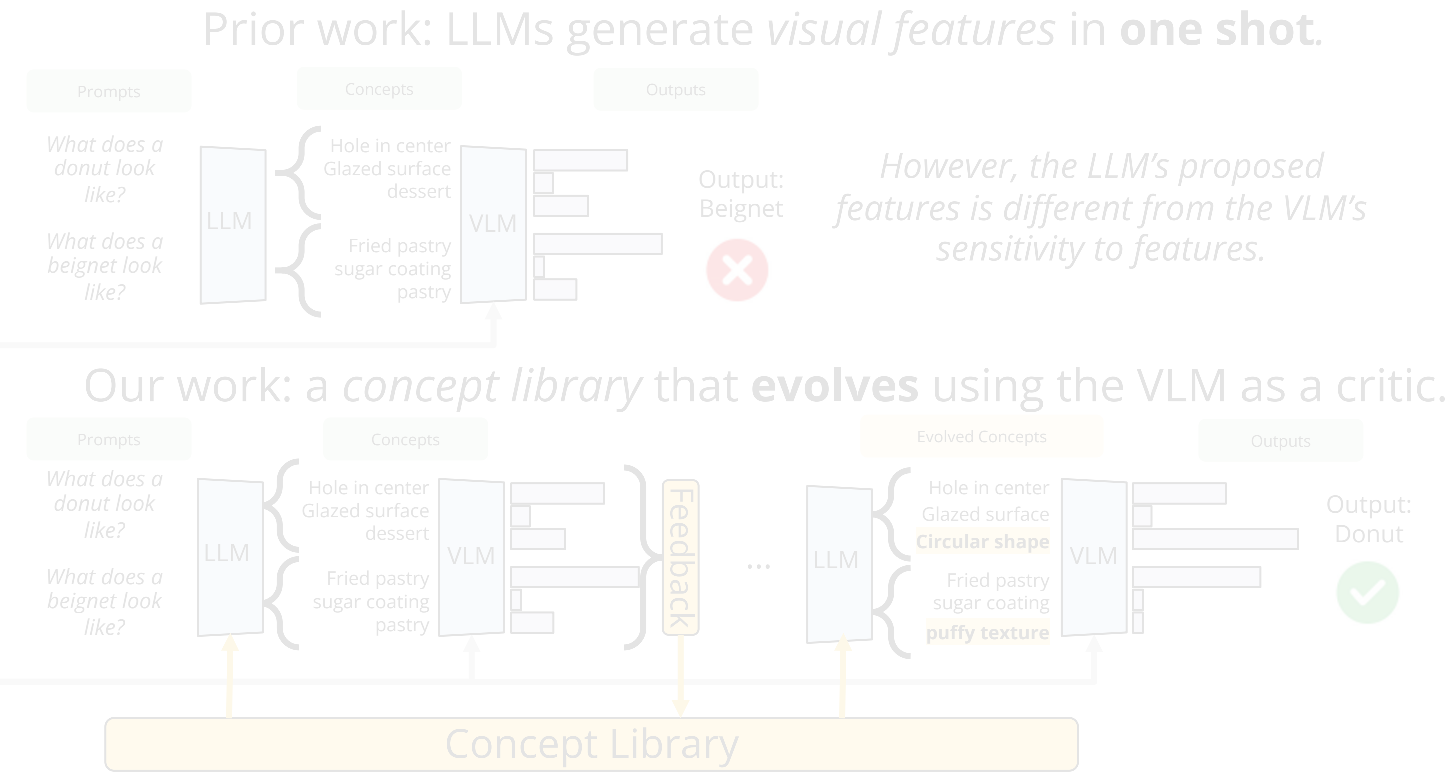
What is this a picture of ?

A donut (🍩) or a beignet (🥐)?



Prior work: LLMs generate *visual features* in **one shot**.

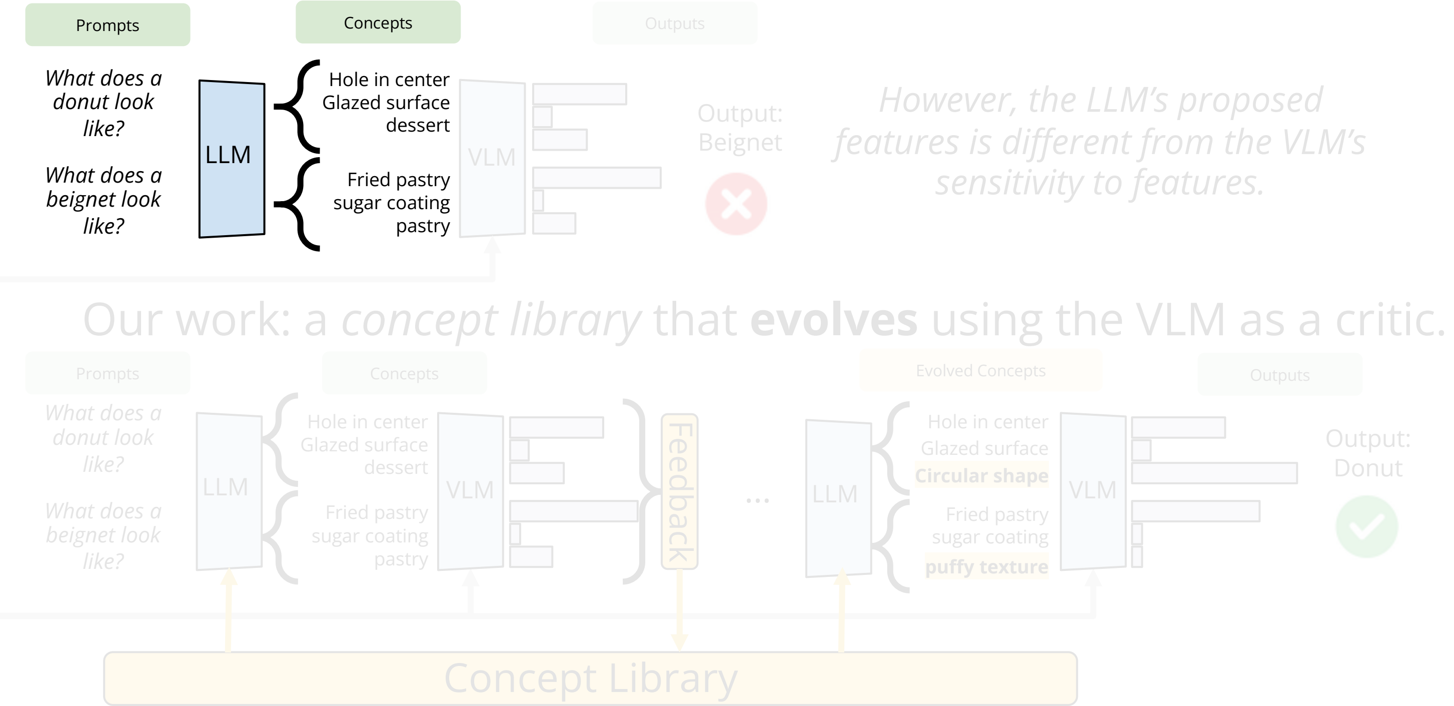
However, the LLM's proposed features is different from the VLM's sensitivity to features.

Our work: a *concept library* that **evolves** using the VLM as a critic.

What is this a picture of ?

A donut (🍩) or a beignet ()?

Prior work: LLMs generate *visual features* in **one shot**.

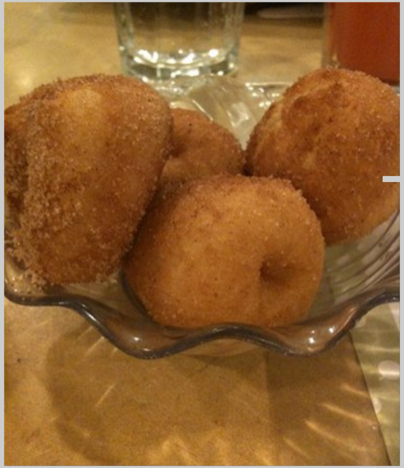


However, the LLM's proposed features is different from the VLM's sensitivity to features.

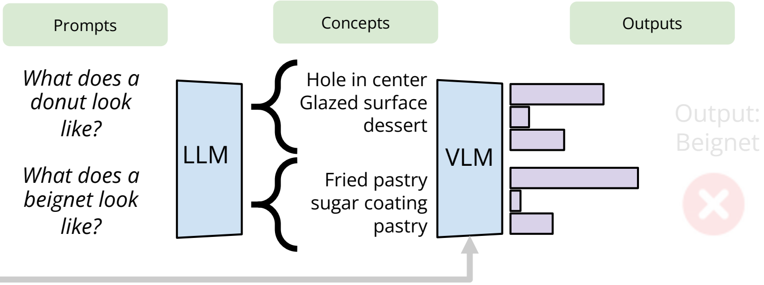
Our work: a *concept library* that **evolves** using the VLM as a critic.

Prior work: LLMs generate *visual features* in **one shot**.

What is this a picture of ?

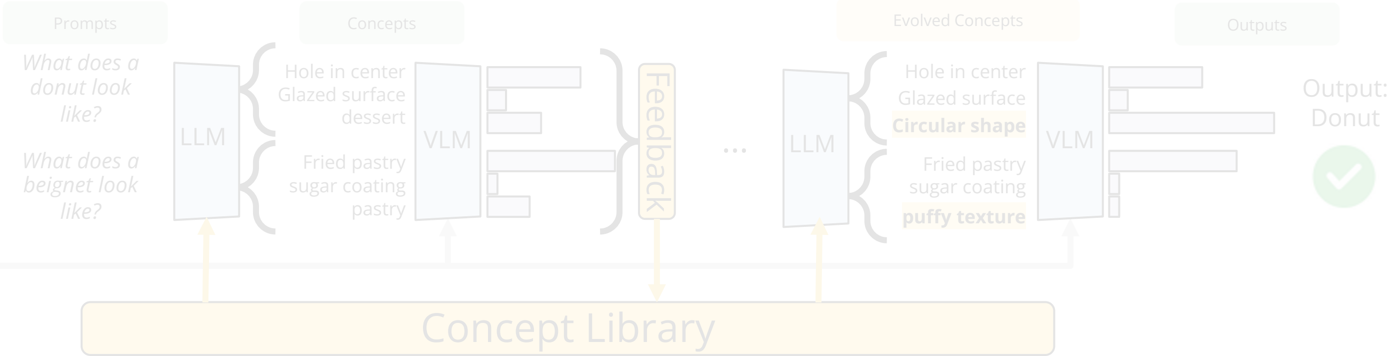


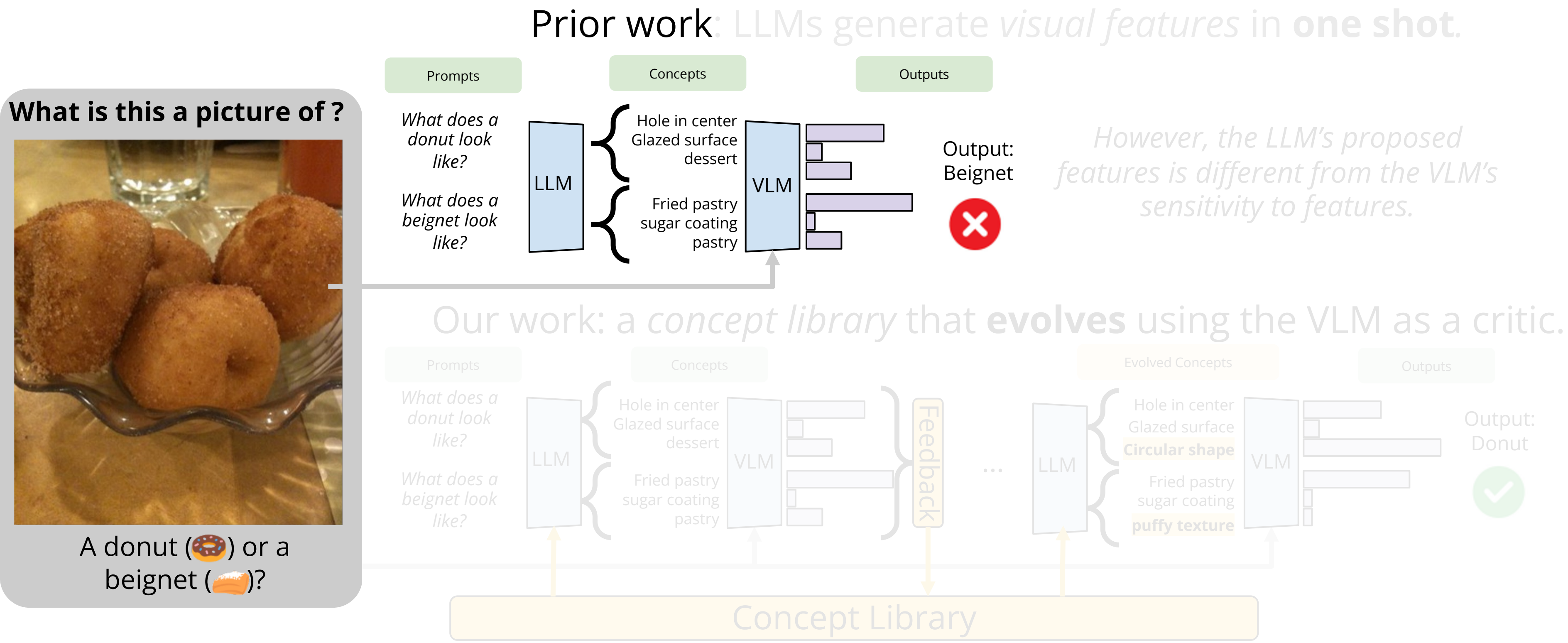
A donut (🍩) or a beignet ( )?



However, the LLM's proposed features is different from the VLM's sensitivity to features.

Our work: a *concept library* that **evolves** using the VLM as a critic.





What is this a picture of ?

A donut (🍩) or a beignet ()?

Prior work: LLMs generate *visual features* in **one shot**.

Prompts

What does a donut look like?

What does a beignet look like?

LLM

Concepts

Hole in center
Glazed surface
dessert

Fried pastry
sugar coating
pastry

VLM

Outputs

Output: Beignet ✗

However, the LLM's proposed features is different from the VLM's sensitivity to features.

Our work: a *concept library* that **evolves** using the VLM as a critic.

Prompts

What does a donut look like?

What does a beignet look like?

LLM

Concepts

Hole in center
Glazed surface
dessert

Fried pastry
sugar coating
pastry

VLM

Feedback

...

LLM

Evolved Concepts

Hole in center
Glazed surface
Circular shape

Fried pastry
sugar coating
puffy texture

VLM

Outputs

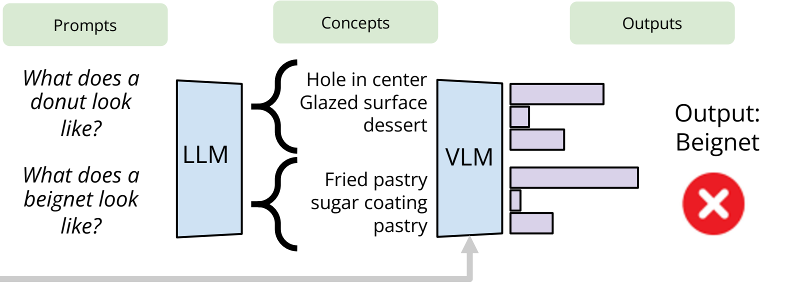
Output: Donut ✓

Concept Library

What is this a picture of ?

A donut (🍩) or a beignet (🥐)?

Prior work: LLMs generate *visual features* in **one shot**.



Prompts

Concepts

Outputs

What does a donut look like?

What does a beignet look like?

LLM

Hole in center
Glazed surface
dessert

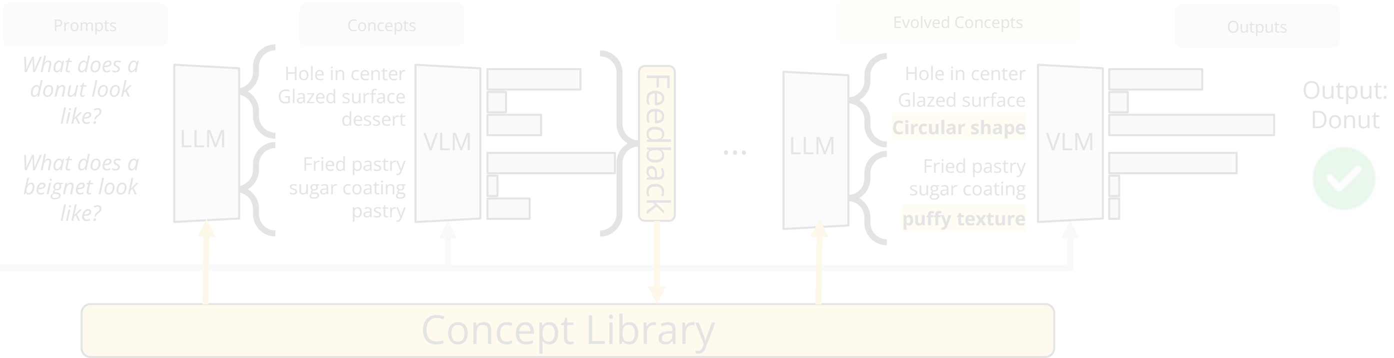
Fried pastry
sugar coating
pastry

VLM

Output:
Beignet

However, the LLM's proposed features is different from the VLM's sensitivity to features.

Our work: a *concept library* that **evolves** using the VLM as a critic.



Prompts

Concepts

Evolved Concepts

Outputs

What does a donut look like?

What does a beignet look like?

LLM

Hole in center
Glazed surface
dessert

Fried pastry
sugar coating
pastry

VLM

Feedback

...

LLM

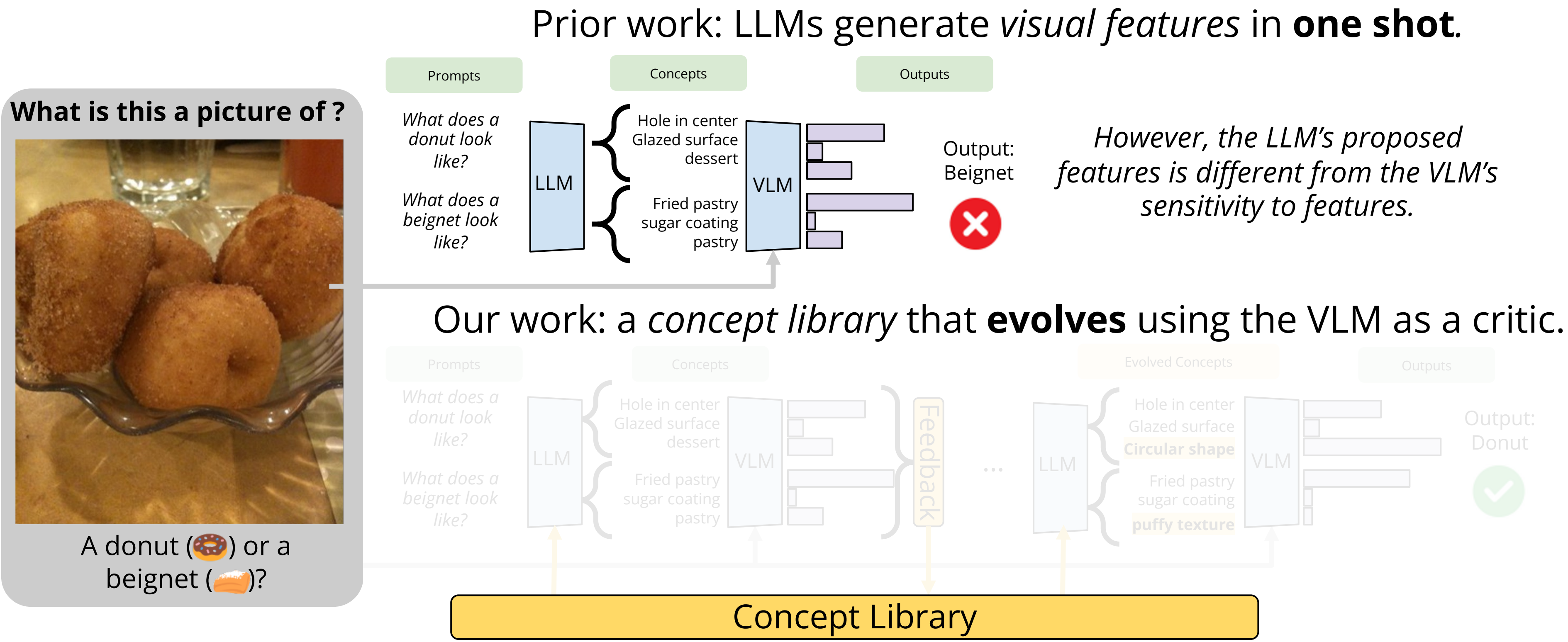
Hole in center
Glazed surface
Circular shape

Fried pastry
sugar coating
puffy texture

VLM

Output:
Donut

Concept Library

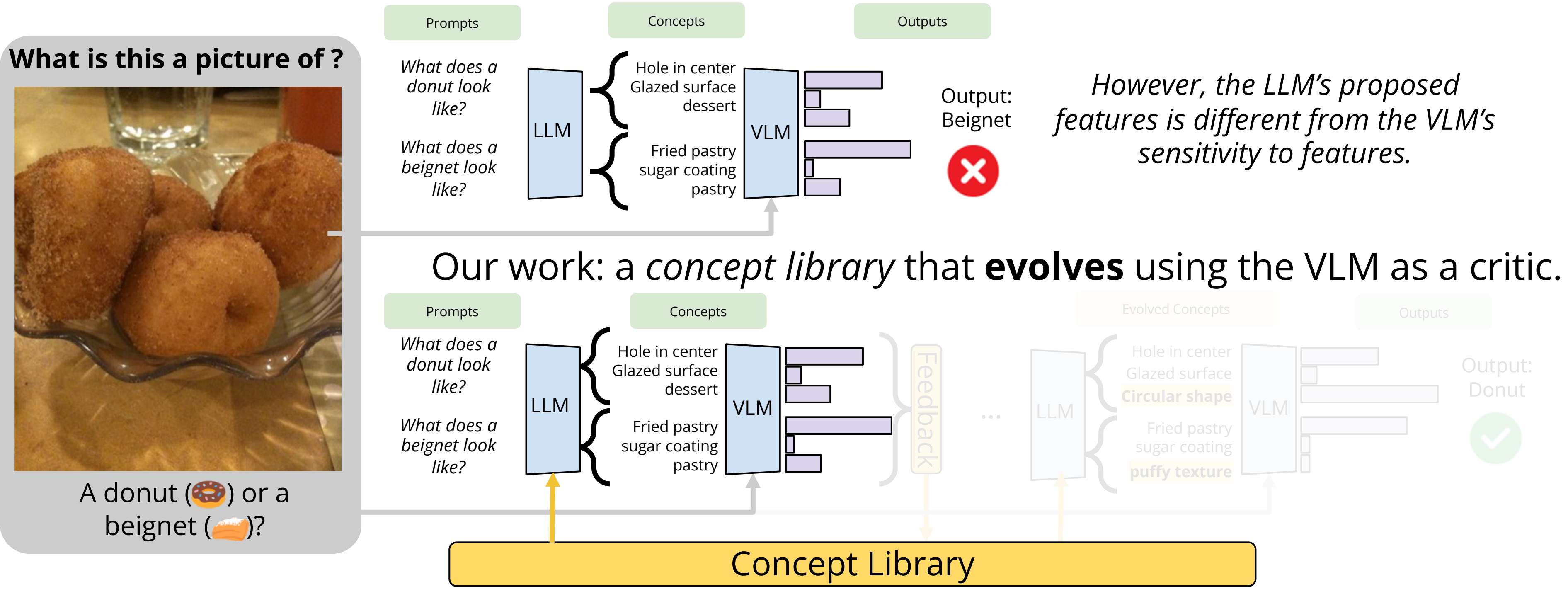


Prior work: LLMs generate *visual features* in **one shot**.

However, the LLM's proposed features is different from the VLM's sensitivity to features.

Our work: a *concept library* that **evolves** using the VLM as a critic.

Prior work: LLMs generate *visual features* in **one shot**.



What is this a picture of ?

A donut (🍩) or a beignet (🥐)?

Prompts

What does a donut look like?

What does a beignet look like?

LLM

Concepts

Hole in center
Glazed surface
dessert

Fried pastry
sugar coating
pastry

VLM

Outputs

Output: Beignet

✗

However, the LLM's proposed features is different from the VLM's sensitivity to features.

Our work: a *concept library* that **evolves** using the VLM as a critic.

Prompts

What does a donut look like?

What does a beignet look like?

LLM

Concepts

Hole in center
Glazed surface
dessert

Fried pastry
sugar coating
pastry

VLM

Feedback

...

Evolved Concepts

LLM

Hole in center
Glazed surface
Circular shape

Fried pastry
sugar coating
puffy texture

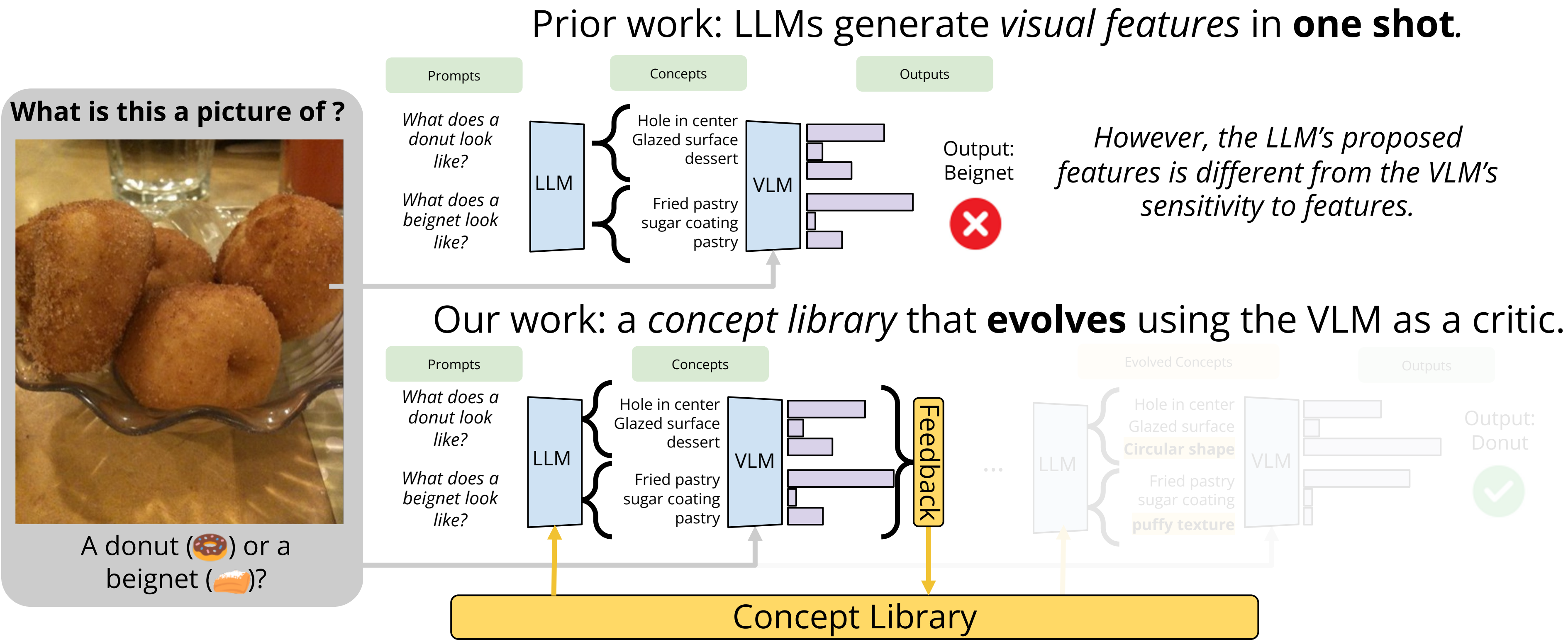
VLM

Outputs

Output: Donut

✓

Concept Library



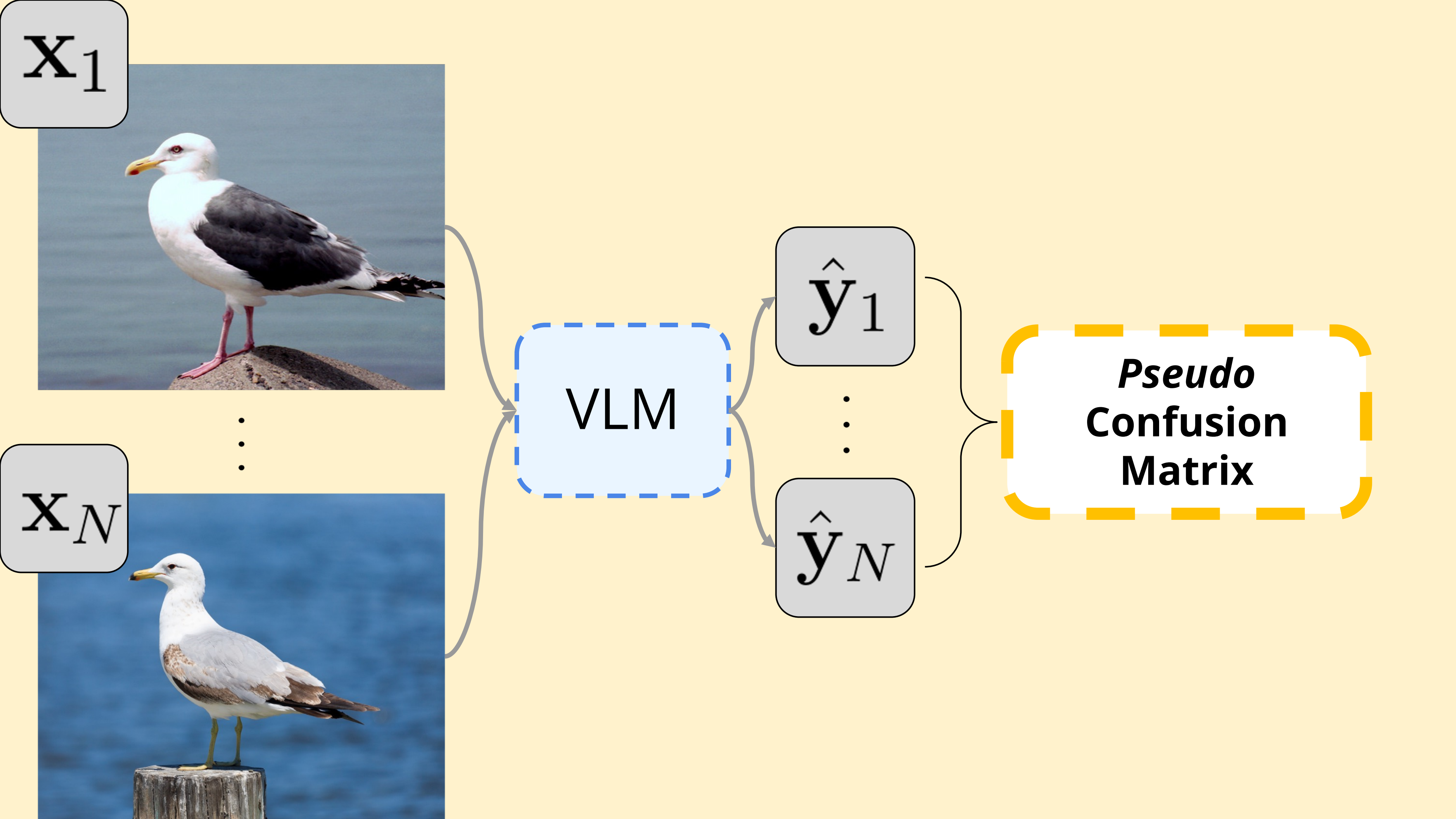
Prior work: LLMs generate *visual features* in **one shot**.

However, the LLM's proposed features is different from the VLM's sensitivity to features.

Our work: a *concept library* that **evolves** using the VLM as a critic.

What is this a picture of ?

A donut (🍩) or a beignet (🥐)?



$\mathbf{x}_1$

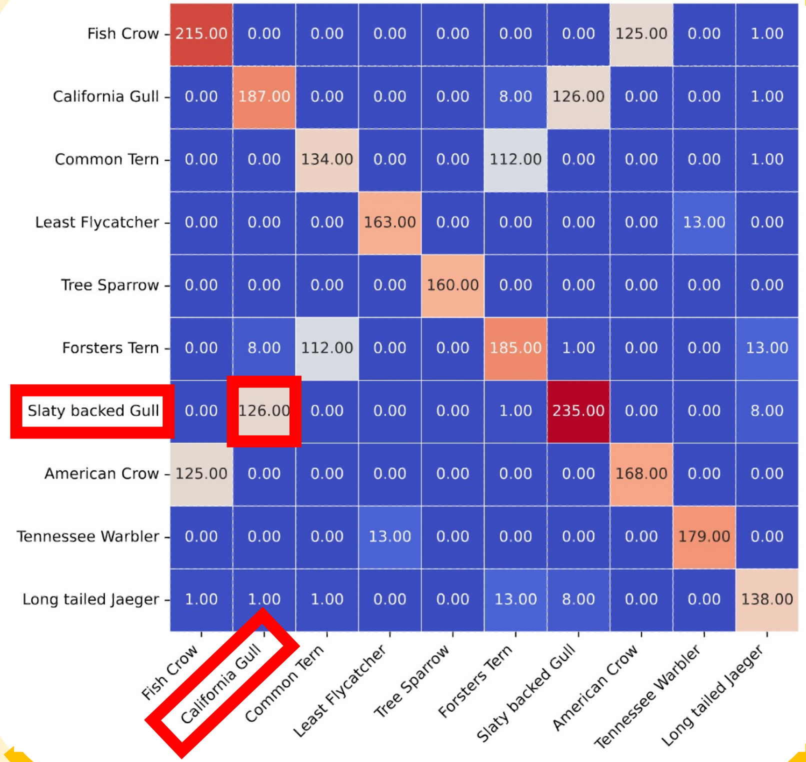
$\mathbf{x}_N$

VLM

$\hat{\mathbf{y}}_1$

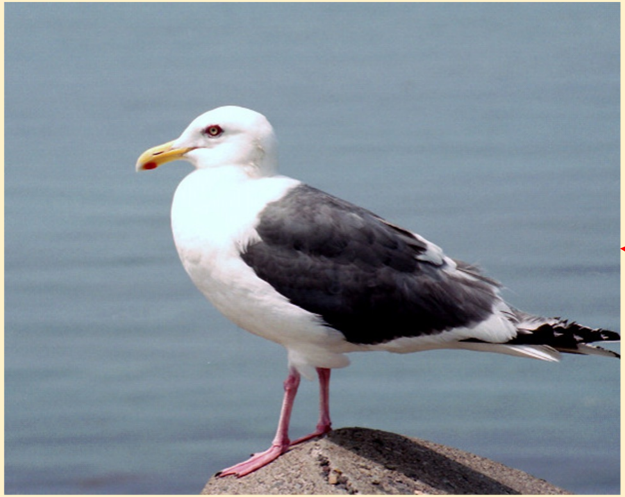
$\hat{\mathbf{y}}_N$

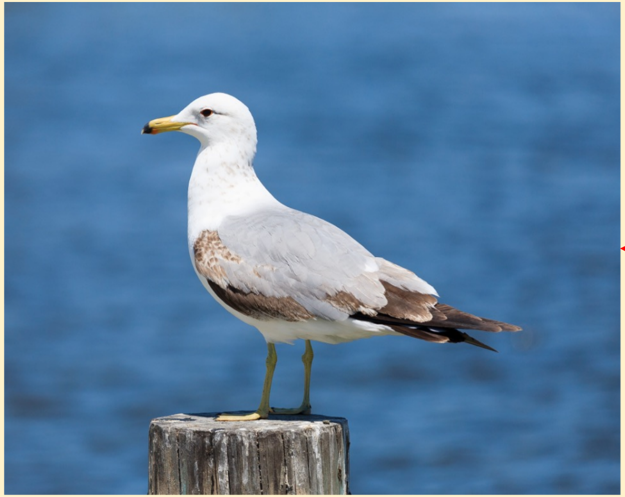
Pseudo **Confusion Matrix**

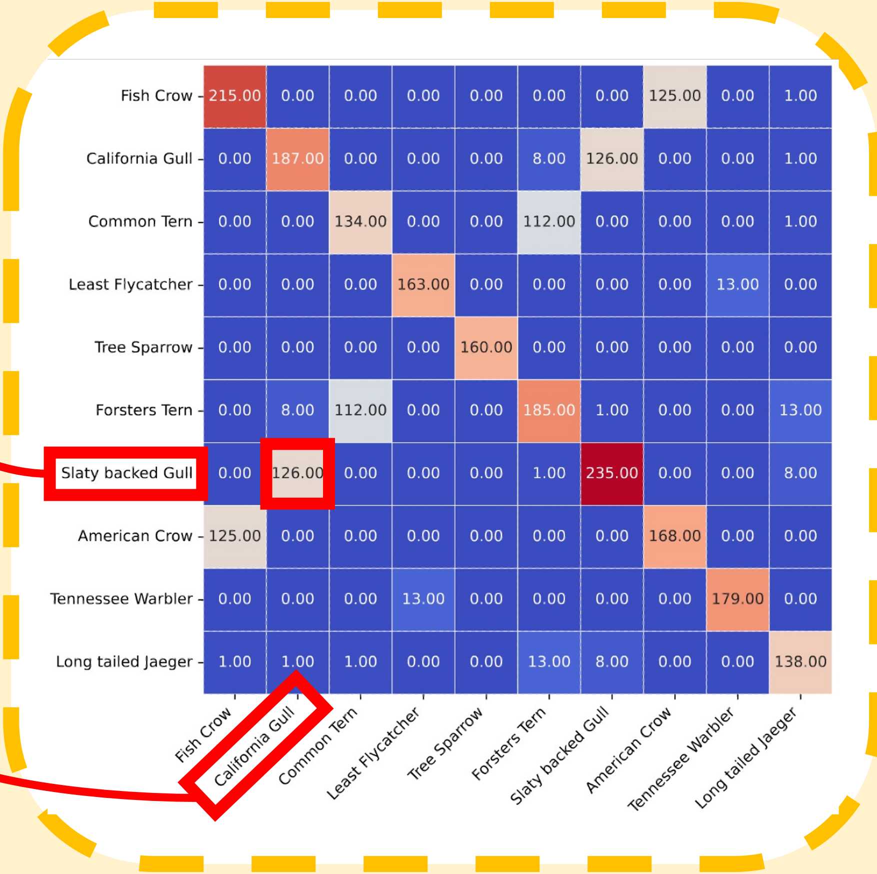


	Fish Crow	California Gull	Common Tern	Least Flycatcher	Tree Sparrow	Forsters Tern	Slaty backed Gull	American Crow	Tennessee Warbler	Long tailed Jaeger
Fish Crow	215.00	0.00	0.00	0.00	0.00	0.00	0.00	125.00	0.00	1.00
California Gull	0.00	187.00	0.00	0.00	0.00	8.00	126.00	0.00	0.00	1.00
Common Tern	0.00	0.00	134.00	0.00	0.00	112.00	0.00	0.00	0.00	1.00
Least Flycatcher	0.00	0.00	0.00	163.00	0.00	0.00	0.00	0.00	13.00	0.00
Tree Sparrow	0.00	0.00	0.00	0.00	160.00	0.00	0.00	0.00	0.00	0.00
Forsters Tern	0.00	8.00	112.00	0.00	0.00	185.00	1.00	0.00	0.00	13.00
Slaty backed Gull	0.00	126.00	0.00	0.00	0.00	1.00	235.00	0.00	0.00	8.00
American Crow	125.00	0.00	0.00	0.00	0.00	0.00	0.00	168.00	0.00	0.00
Tennessee Warbler	0.00	0.00	0.00	13.00	0.00	0.00	0.00	0.00	179.00	0.00
Long tailed Jaeger	1.00	1.00	1.00	0.00	0.00	13.00	8.00	0.00	0.00	138.00

Pseudo Confusion Matrix





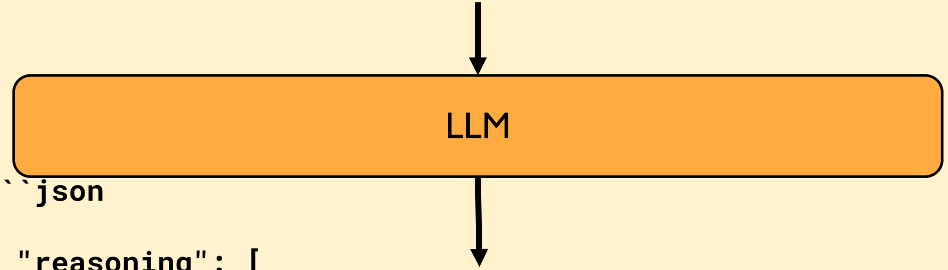


	Fish Crow	California Gull	Common Tern	Least Flycatcher	Tree Sparrow	Forsters Tern	Slaty backed Gull	American Crow	Tennessee Warbler	Long tailed Jaeger
Fish Crow	215.00	0.00	0.00	0.00	0.00	0.00	0.00	125.00	0.00	1.00
California Gull	0.00	187.00	0.00	0.00	0.00	8.00	126.00	0.00	0.00	1.00
Common Tern	0.00	0.00	134.00	0.00	0.00	112.00	0.00	0.00	0.00	1.00
Least Flycatcher	0.00	0.00	0.00	163.00	0.00	0.00	0.00	0.00	13.00	0.00
Tree Sparrow	0.00	0.00	0.00	0.00	160.00	0.00	0.00	0.00	0.00	0.00
Forsters Tern	0.00	8.00	112.00	0.00	0.00	185.00	1.00	0.00	0.00	13.00
Slaty backed Gull	0.00	126.00	0.00	0.00	0.00	1.00	235.00	0.00	0.00	8.00
American Crow	125.00	0.00	0.00	0.00	0.00	0.00	0.00	168.00	0.00	0.00
Tennessee Warbler	0.00	0.00	0.00	13.00	0.00	0.00	0.00	0.00	179.00	0.00
Long tailed Jaeger	1.00	1.00	1.00	0.00	0.00	13.00	8.00	0.00	0.00	138.00

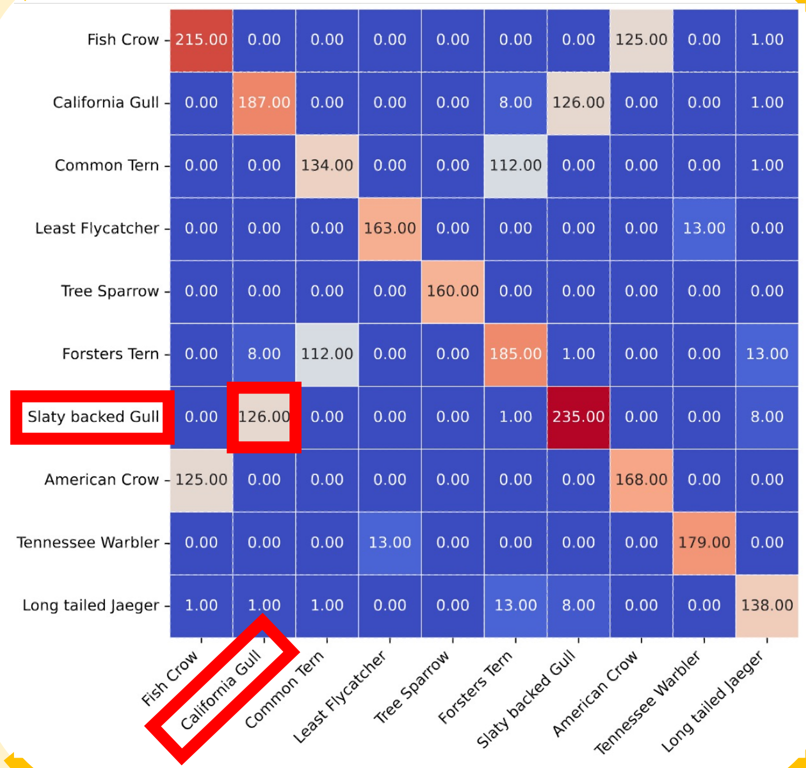
***Pseudo* Confusion Matrix**

Please suggest visual features that distinguish **a slaty backed gull** from **a california gull**.

The current descriptors for **slaty backed gull** are: `{cls1_concepts}`.
The current descriptors for **california gull** are: `{cls2_concepts}`.

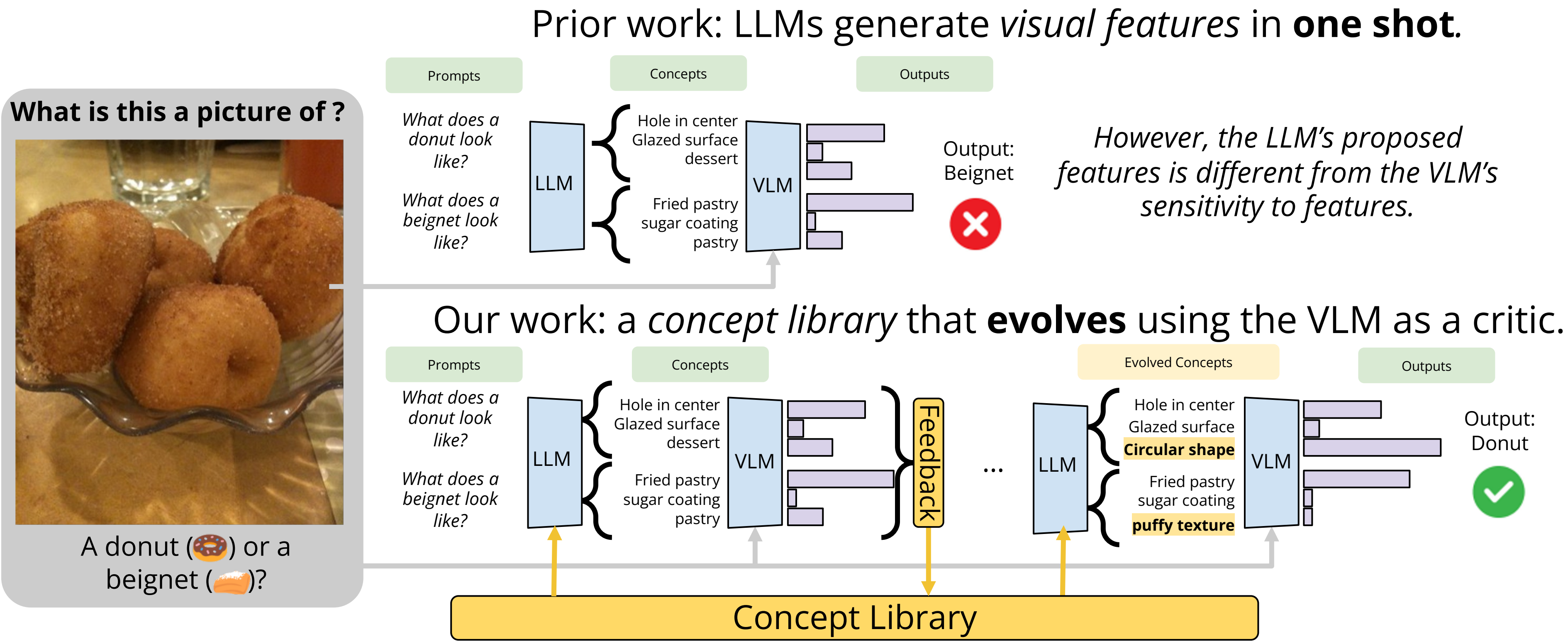


~~~
```json
{
  "reasoning": [
            "Slaty-backed gulls have a darker, slate-
gray back and wings, whereas California gulls have a
lighter gray back and wings, making the back coloration
a key distinguishing feature.",
            "Slaty-backed gulls exhibit prominent white
'mirrors' on their primary feathers, which are larger
and more distinct compared to those of the California
gull.", ...
            ],
  "features": [
            "Slate-gray back and wings",
            "Large white mirrors on primary feathers",
...
            ]
}
```
~~~



	Fish Crow	California Gull	Common Tern	Least Flycatcher	Tree Sparrow	Forsters Tern	Slaty backed Gull	American Crow	Tennessee Warbler	Long tailed Jaeger
Fish Crow	215.00	0.00	0.00	0.00	0.00	0.00	0.00	125.00	0.00	1.00
California Gull	0.00	187.00	0.00	0.00	0.00	8.00	126.00	0.00	0.00	1.00
Common Tern	0.00	0.00	134.00	0.00	0.00	112.00	0.00	0.00	0.00	1.00
Least Flycatcher	0.00	0.00	0.00	163.00	0.00	0.00	0.00	0.00	13.00	0.00
Tree Sparrow	0.00	0.00	0.00	0.00	160.00	0.00	0.00	0.00	0.00	0.00
Forsters Tern	0.00	8.00	112.00	0.00	0.00	185.00	1.00	0.00	0.00	13.00
Slaty backed Gull	0.00	126.00	0.00	0.00	0.00	1.00	235.00	0.00	0.00	8.00
American Crow	125.00	0.00	0.00	0.00	0.00	0.00	0.00	168.00	0.00	0.00
Tennessee Warbler	0.00	0.00	0.00	13.00	0.00	0.00	0.00	0.00	179.00	0.00
Long tailed Jaeger	1.00	1.00	1.00	0.00	0.00	13.00	8.00	0.00	0.00	138.00

***Pseudo* Confusion Matrix**



Prior work: LLMs generate *visual features* in **one shot**.

Prompts

Concepts

Outputs

What does a donut look like?

What does a beignet look like?

LLM

Hole in center
Glazed surface
dessert

Fried pastry
sugar coating
pastry

VLM

Output: Beignet ✗

However, the LLM's proposed features is different from the VLM's sensitivity to features.

What is this a picture of ?

A donut (🍩) or a beignet (🥐)?

Our work: a *concept library* that **evolves** using the VLM as a critic.

Prompts

Concepts

Evolved Concepts

Outputs

What does a donut look like?

What does a beignet look like?

LLM

Hole in center
Glazed surface
dessert

Fried pastry
sugar coating
pastry

VLM

Feedback

...

LLM

Hole in center
Glazed surface
Circular shape

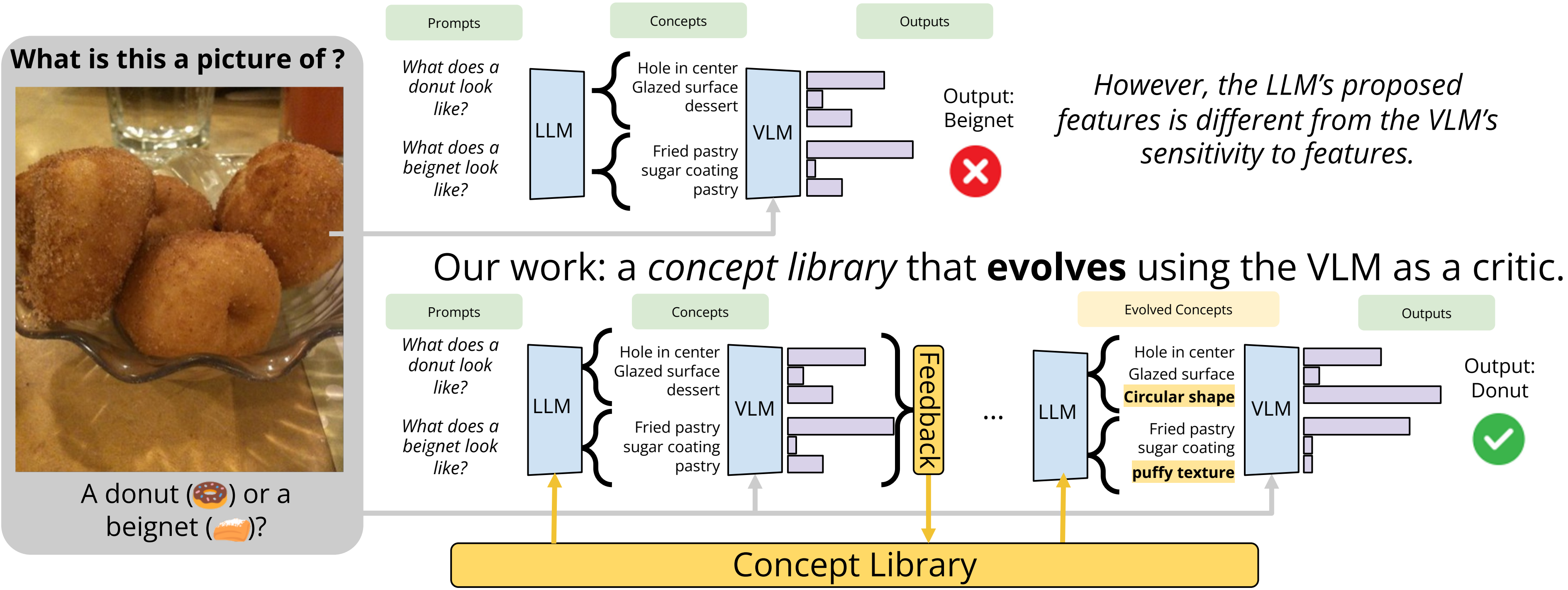
Fried pastry
sugar coating
puffy texture

VLM

Output: Donut ✓

Concept Library

Prior work: LLMs generate *visual features* in **one shot**.



What is this a picture of ?

A donut (🍩) or a beignet ()?

However, the LLM's proposed features is different from the VLM's sensitivity to features.

Our work: a *concept library* that **evolves** using the VLM as a critic.

Key insight: The concept library is simultaneously bootstrapping the VLM inputs and the LLM inputs.