CVPR JUNE 18-22, 2023 VANCOUVER, CANADA

# I2MVFormer: Large Language Model Generated Multi-View Document Supervision for Zero-Shot Image Classification

WED-PM-268

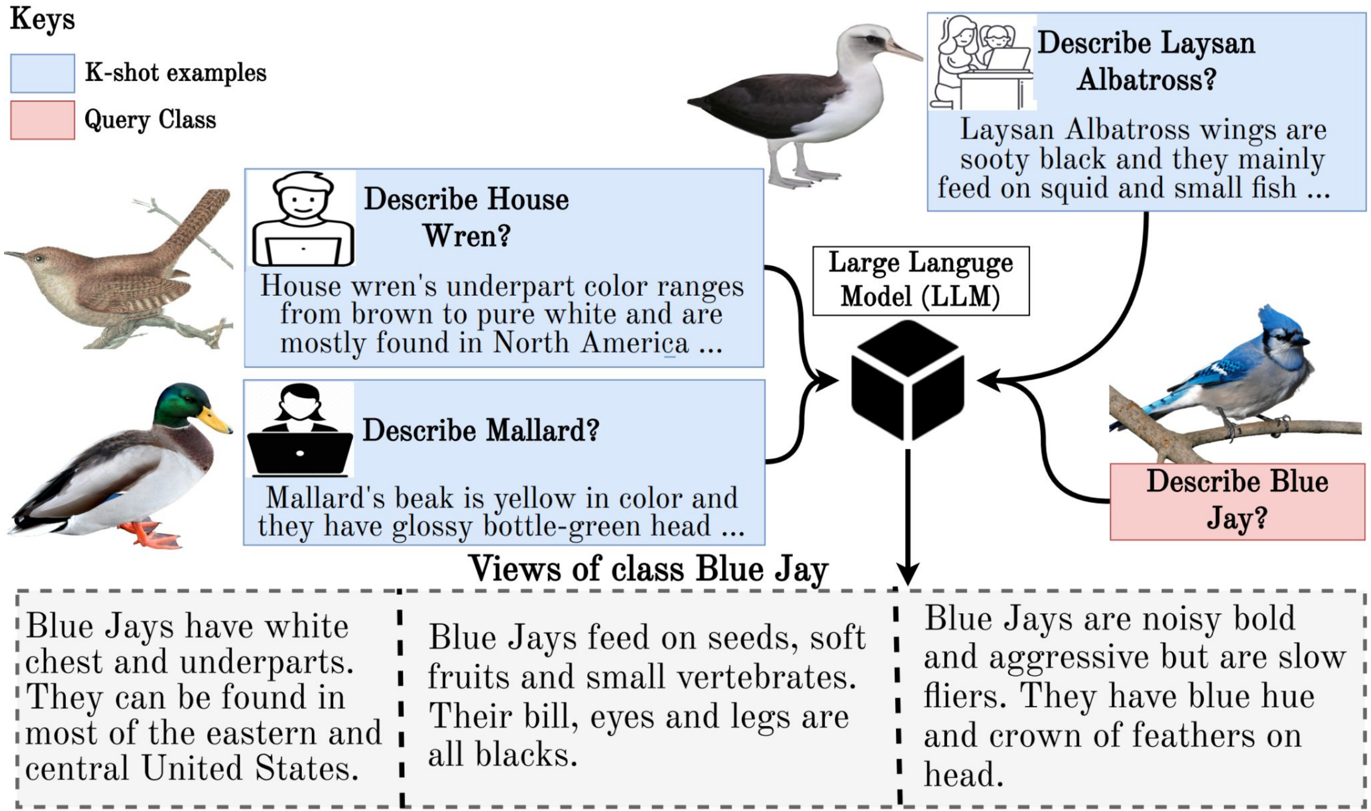
Muhammad Ferjad Naeem*, GulZain Ali*, Yongqin Xian, Zeeshan Afzal, Didier Stricker, Luc Van Gool, Federico Tombari

# Problem Statement - Zeroshot Learning

Seen Classes

Unseen Classes

# Motivation



Keys

🟦 K-shot examples

🟥 Query Class

**Describe House Wren?**

House wren's underpart color ranges from brown to pure white and are mostly found in North America ...

**Describe Mallard?**

Mallard's beak is yellow in color and they have glossy bottle-green head ...

**Describe Laysan Albatross?**

Laysan Albatross wings are sooty black and they mainly feed on squid and small fish ...

Large Langauge Model (LLM)

**Describe Blue Jay?**

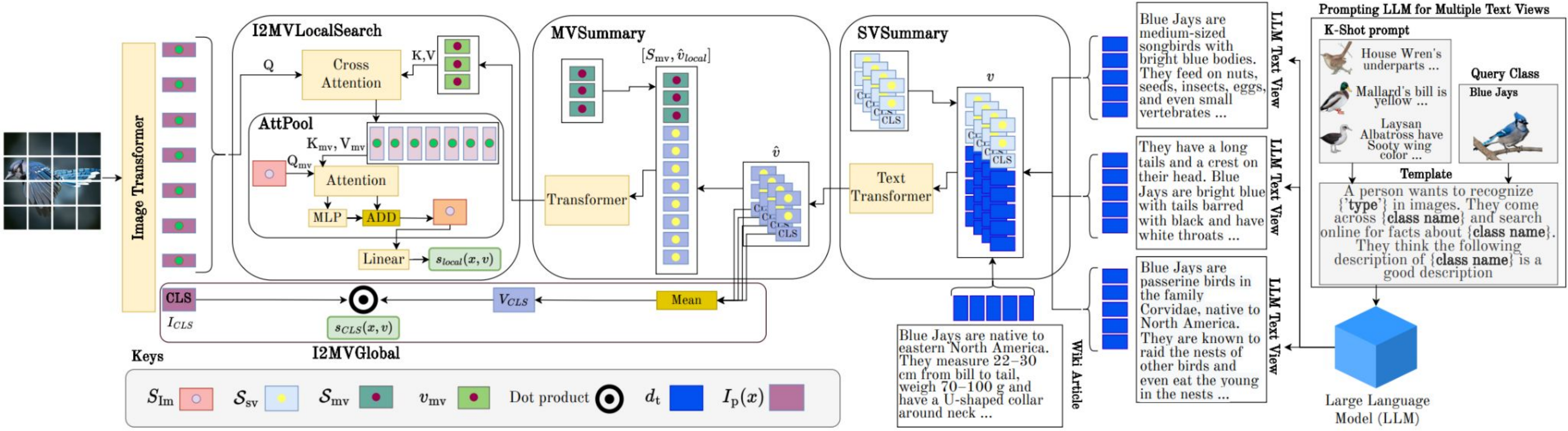
**Views of class Blue Jay**

Blue Jays have white chest and underparts. They can be found in most of the eastern and central United States.

Blue Jays feed on seeds, soft fruits and small vertebrates. Their bill, eyes and legs are all blacks.

Blue Jays are noisy bold and aggressive but are slow fliers. They have blue hue and crown of feathers on head.

# I2MVFormer



Blue Jays are native to eastern North America. They measure 22–30 cm from bill to tail, weigh 70–100 g and have a U-shaped collar around neck ...

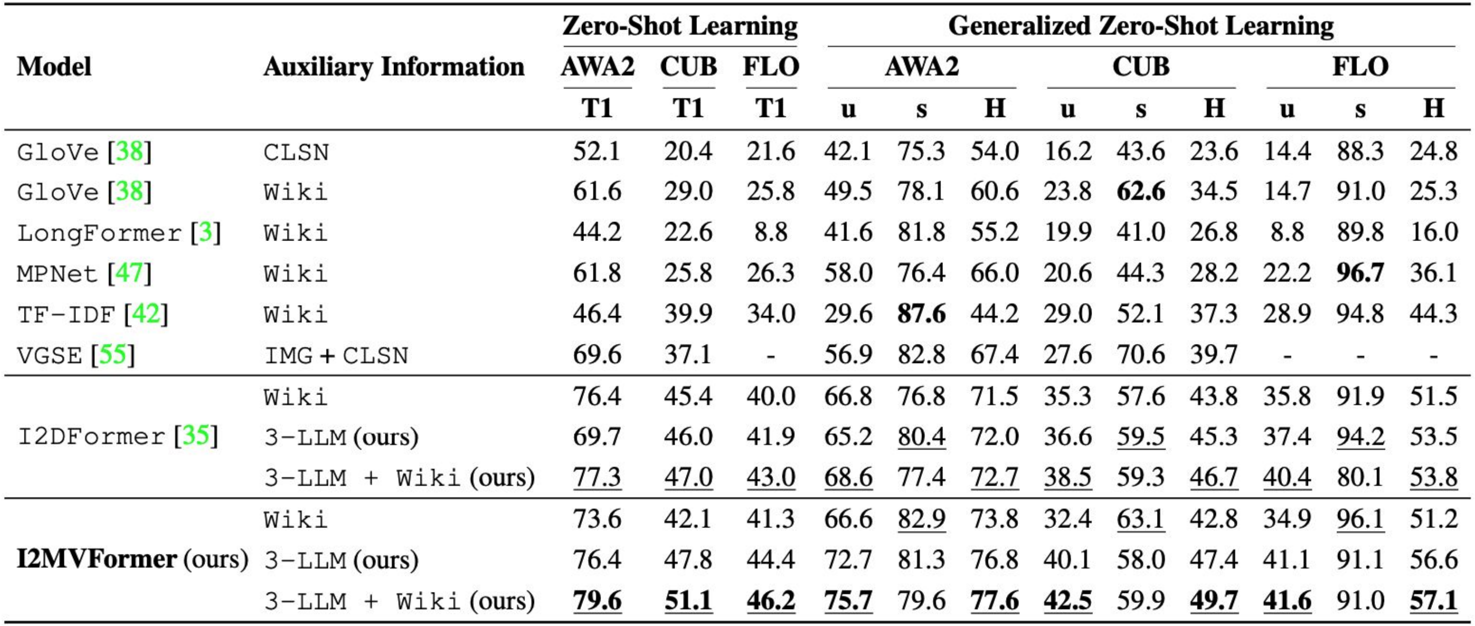
Blue Jays are medium-sized songbirds with bright blue bodies. They feed on nuts, seeds, insects, eggs, and even small vertebrates ...

They have a long tails and a crest on their head. Blue Jays are bright blue with tails barred with black and have white throats ...

Blue Jays are passerine birds in the family Corvidae, native to North America. They are known to raid the nests of other birds and even eat the young in the nests ...

A person wants to recognize {'**type**'} in images. They come across {**class name**} and search online for facts about {**class name**}. They think the following description of {**class name**} is a good description

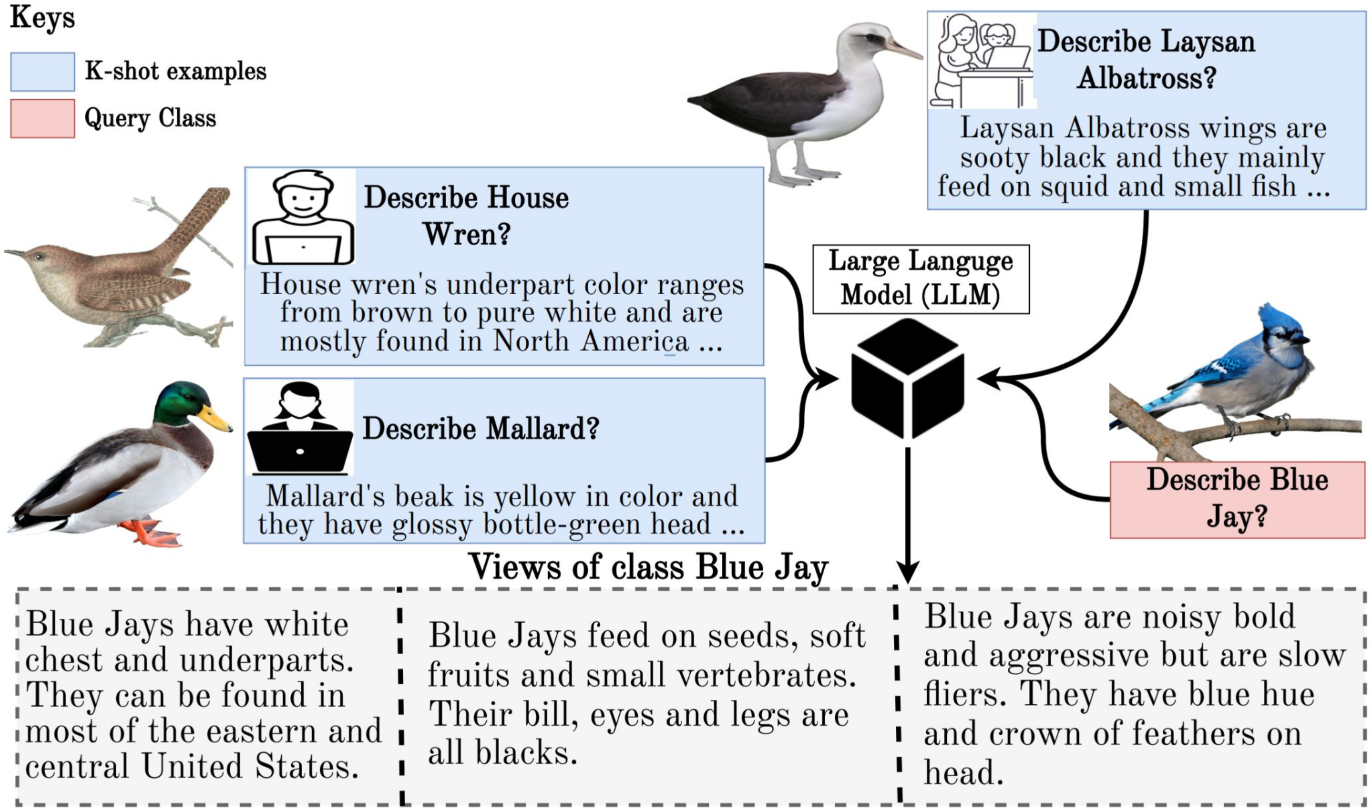
# Baseline Comparison



| Model | Auxiliary Information | Zero-Shot Learning | | | Generalized Zero-Shot Learning | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | AWA2 | CUB | FLO | AWA2 | | | CUB | | | FLO | | |
| | | T1 | T1 | T1 | u | s | H | u | s | H | u | s | H |
| GloVe [38] | CLSN | 52.1 | 20.4 | 21.6 | 42.1 | 75.3 | 54.0 | 16.2 | 43.6 | 23.6 | 14.4 | 88.3 | 24.8 |
| GloVe [38] | Wiki | 61.6 | 29.0 | 25.8 | 49.5 | 78.1 | 60.6 | 23.8 | **62.6** | 34.5 | 14.7 | 91.0 | 25.3 |
| LongFormer [3] | Wiki | 44.2 | 22.6 | 8.8 | 41.6 | 81.8 | 55.2 | 19.9 | 41.0 | 26.8 | 8.8 | 89.8 | 16.0 |
| MPNet [47] | Wiki | 61.8 | 25.8 | 26.3 | 58.0 | 76.4 | 66.0 | 20.6 | 44.3 | 28.2 | 22.2 | **96.7** | 36.1 |
| TF-IDF [42] | Wiki | 46.4 | 39.9 | 34.0 | 29.6 | **87.6** | 44.2 | 29.0 | 52.1 | 37.3 | 28.9 | 94.8 | 44.3 |
| VGSE [55] | IMG + CLSN | 69.6 | 37.1 | - | 56.9 | 82.8 | 67.4 | 27.6 | 70.6 | 39.7 | - | - | - |
| | Wiki | 76.4 | 45.4 | 40.0 | 66.8 | 76.8 | 71.5 | 35.3 | 57.6 | 43.8 | 35.8 | 91.9 | 51.5 |
| I2DFormer [35] | 3-LLM (ours) | 69.7 | 46.0 | 41.9 | 65.2 | 80.4 | 72.0 | 36.6 | 59.5 | 45.3 | 37.4 | 94.2 | 53.5 |
| | 3-LLM + Wiki (ours) | 77.3 | 47.0 | 43.0 | 68.6 | 77.4 | 72.7 | 38.5 | 59.3 | 46.7 | 40.4 | 80.1 | 53.8 |
| | Wiki | 73.6 | 42.1 | 41.3 | 66.6 | 82.9 | 73.8 | 32.4 | 63.1 | 42.8 | 34.9 | 96.1 | 51.2 |
| **I2MVFormer** (ours) | 3-LLM (ours) | 76.4 | 47.8 | 44.4 | 72.7 | 81.3 | 76.8 | 40.1 | 58.0 | 47.4 | 41.1 | 91.1 | 56.6 |
| | 3-LLM + Wiki (ours) | **79.6** | **51.1** | **46.2** | **75.7** | 79.6 | **77.6** | **42.5** | 59.9 | **49.7** | **41.6** | 91.0 | **57.1** |

# Thank you!

# Motivation



Keys

🟦 K-shot examples

🟥 Query Class

Describe House Wren?

House wren's underpart color ranges from brown to pure white and are mostly found in North America ...

Describe Mallard?

Mallard's beak is yellow in color and they have glossy bottle-green head ...

Describe Laysan Albatross?

Laysan Albatross wings are sooty black and they mainly feed on squid and small fish ...

Large Languge Model (LLM)

Describe Blue Jay?

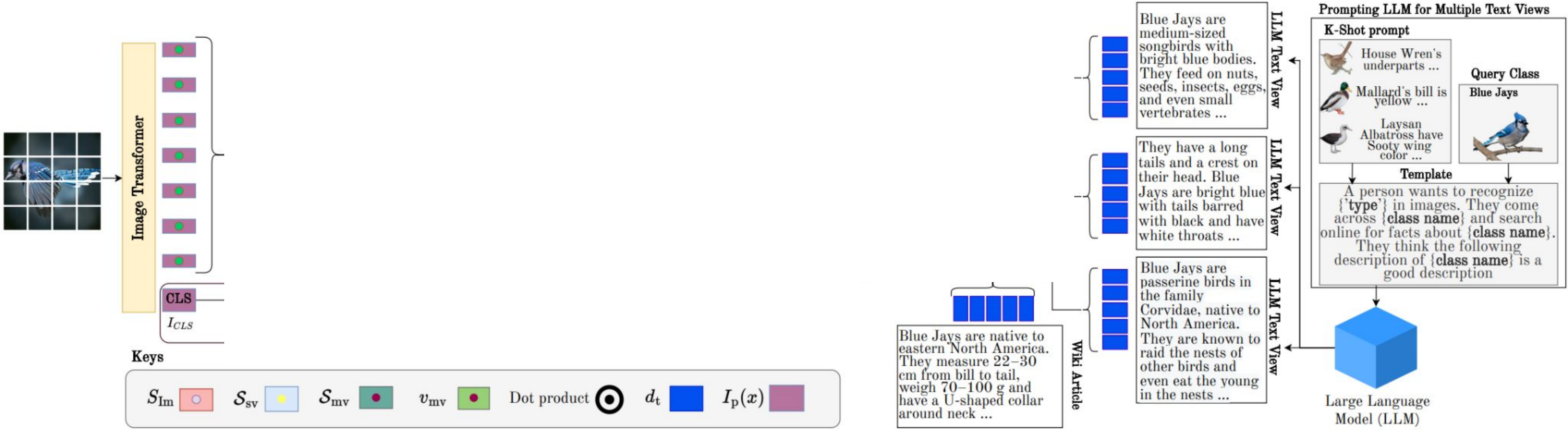
Views of class Blue Jay

Blue Jays have white chest and underparts. They can be found in most of the eastern and central United States.

Blue Jays feed on seeds, soft fruits and small vertebrates. Their bill, eyes and legs are all blacks.

Blue Jays are noisy bold and aggressive but are slow fliers. They have blue hue and crown of feathers on head.

# Method



Blue Jays are medium-sized songbirds with bright blue bodies. They feed on nuts, seeds, insects, eggs, and even small vertebrates ...

They have a long tails and a crest on their head. Blue Jays are bright blue with tails barred with black and have white throats ...

Blue Jays are passerine birds in the family Corvidae, native to North America. They are known to raid the nests of other birds and even eat the young in the nests ...

Blue Jays are native to eastern North America. They measure 22–30 cm from bill to tail, weigh 70–100 g and have a U-shaped collar around neck ...

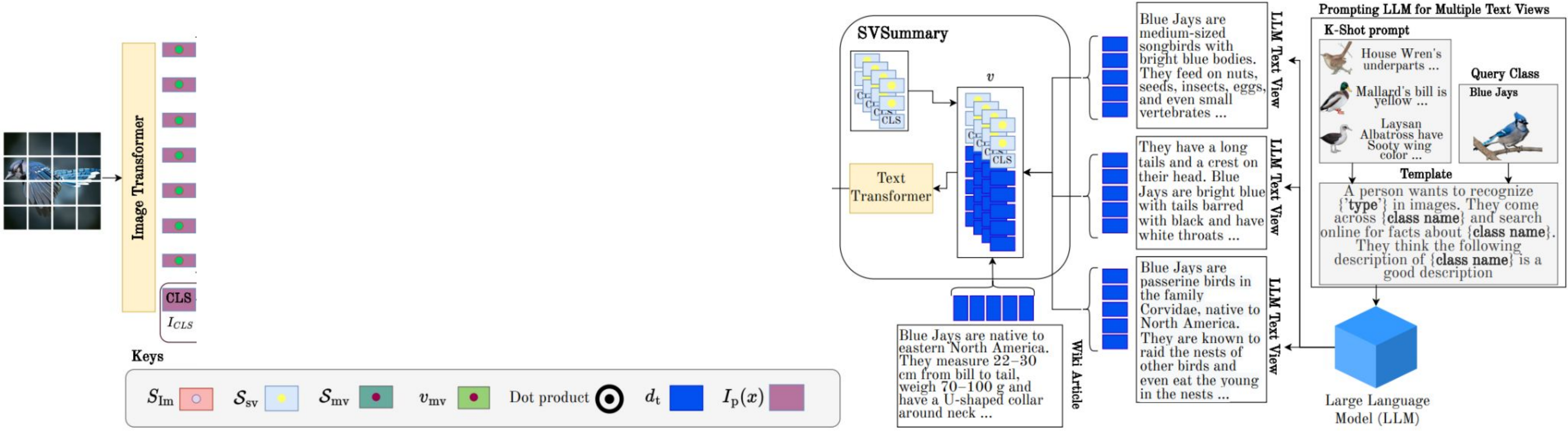
House Wren's underparts ...

Mallard's bill is yellow ...

Laysan Albatross have Sooty wing color ...

A person wants to recognize {'type'} in images. They come across {**class name**} and search online for facts about {**class name**}. They think the following description of {**class name**} is a good description

# Method



Blue Jays are native to eastern North America. They measure 22–30 cm from bill to tail, weigh 70–100 g and have a U-shaped collar around neck ...

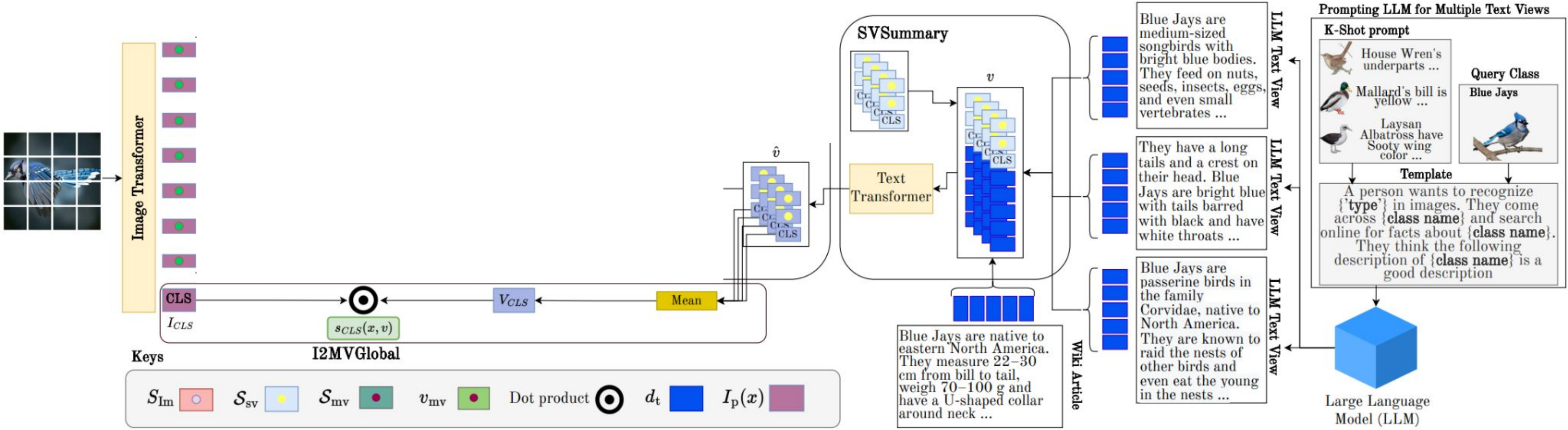
Blue Jays are medium-sized songbirds with bright blue bodies. They feed on nuts, seeds, insects, eggs, and even small vertebrates ...

They have a long tails and a crest on their head. Blue Jays are bright blue with tails barred with black and have white throats ...

Blue Jays are passerine birds in the family Corvidae, native to North America. They are known to raid the nests of other birds and even eat the young in the nests ...

A person wants to recognize {'type'} in images. They come across {**class name**} and search online for facts about {**class name**}. They think the following description of {**class name**} is a good description

# Method



A person wants to recognize {'type'} in images. They come across {class name} and search online for facts about {class name}. They think the following description of {class name} is a good description

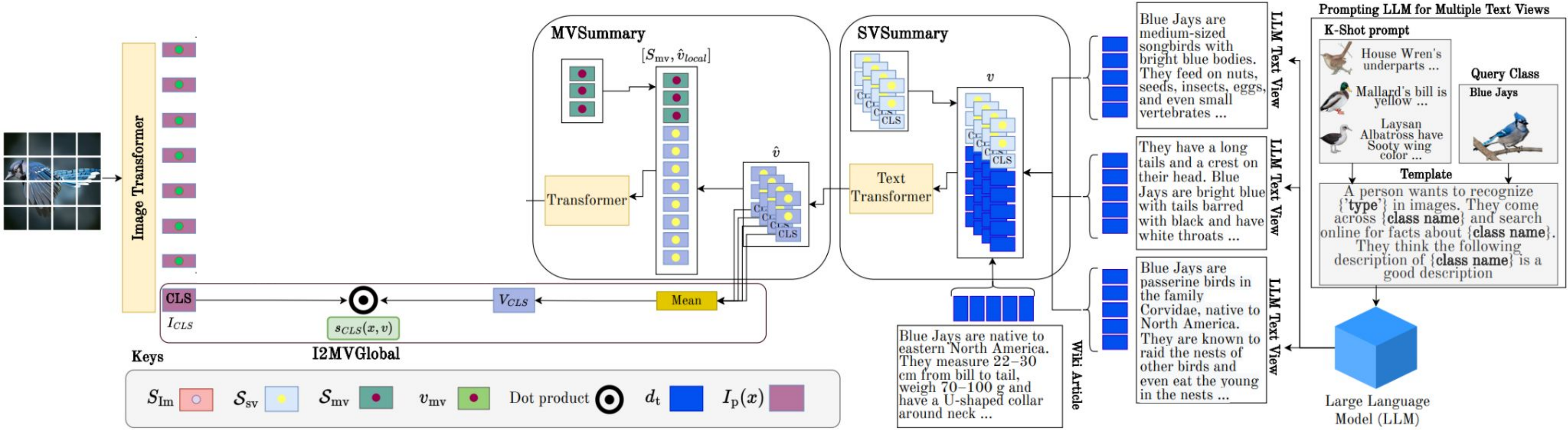
Blue Jays are medium-sized songbirds with bright blue bodies. They feed on nuts, seeds, insects, eggs, and even small vertebrates ...

They have a long tails and a crest on their head. Blue Jays are bright blue with tails barred with black and have white throats ...

Blue Jays are passerine birds in the family Corvidae, native to North America. They are known to raid the nests of other birds and even eat the young in the nests ...

Blue Jays are native to eastern North America. They measure 22–30 cm from bill to tail, weigh 70–100 g and have a U-shaped collar around neck ...

# Method



Blue Jays are native to eastern North America. They measure 22–30 cm from bill to tail, weigh 70–100 g and have a U-shaped collar around neck ...

Blue Jays are medium-sized songbirds with bright blue bodies. They feed on nuts, seeds, insects, eggs, and even small vertebrates ...

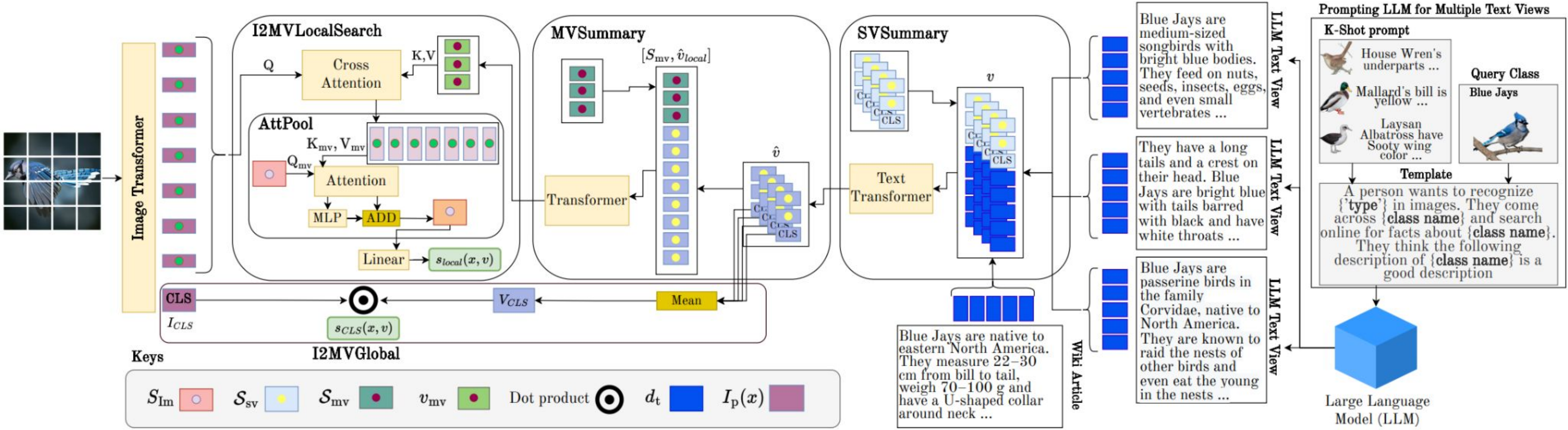
They have a long tails and a crest on their head. Blue Jays are bright blue with tails barred with black and have white throats ...

Blue Jays are passerine birds in the family Corvidae, native to North America. They are known to raid the nests of other birds and even eat the young in the nests ...

A person wants to recognize {'**type**'} in images. They come across {**class name**} and search online for facts about {**class name**}. They think the following description of {**class name**} is a good description

ETH zürich Google CVPR JUNE 18-22, 2023 VANCOUVER, CANADA

# Method



Blue Jays are native to eastern North America. They measure 22–30 cm from bill to tail, weigh 70–100 g and have a U-shaped collar around neck ...

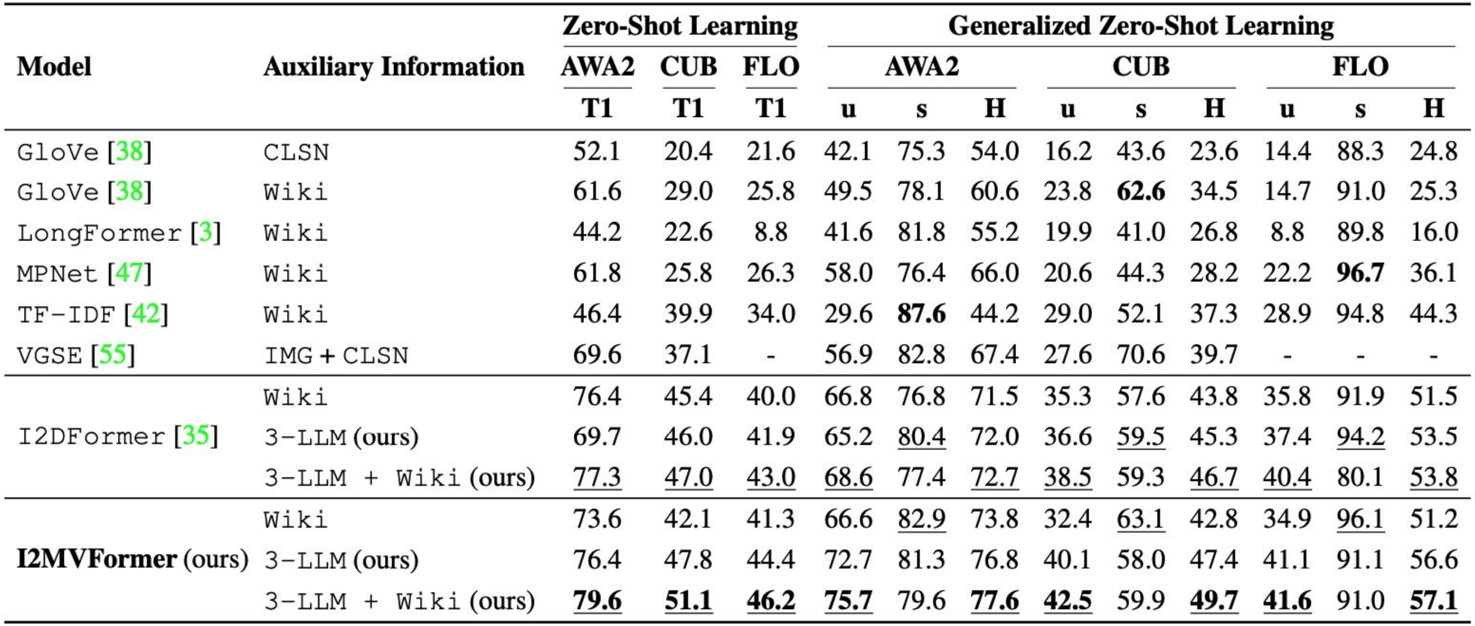
Blue Jays are medium-sized songbirds with bright blue bodies. They feed on nuts, seeds, insects, eggs, and even small vertebrates ...

They have a long tails and a crest on their head. Blue Jays are bright blue with tails barred with black and have white throats ...

Blue Jays are passerine birds in the family Corvidae, native to North America. They are known to raid the nests of other birds and even eat the young in the nests ...

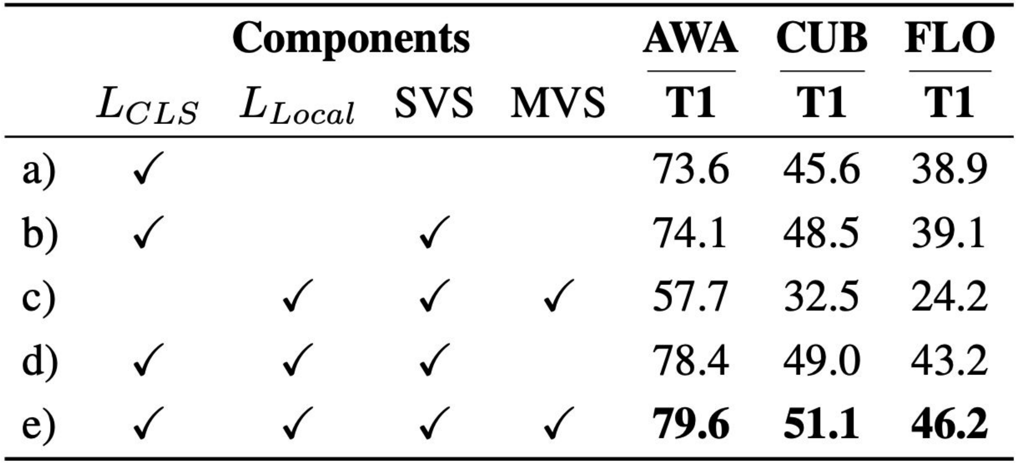
A person wants to recognize {'type'} in images. They come across {class name} and search online for facts about {class name}. They think the following description of {class name} is a good description

# Baseline Comparison



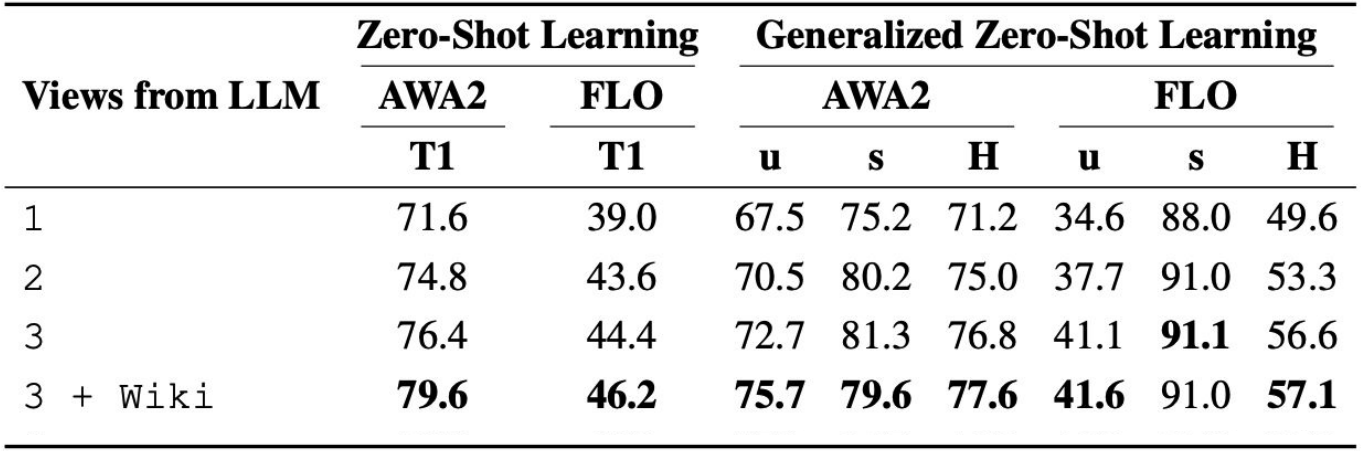
| **Model** | **Auxiliary Information** | **Zero-Shot Learning** | | | **Generalized Zero-Shot Learning** | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | **AWA2** | **CUB** | **FLO** | **AWA2** | | | **CUB** | | | **FLO** | | |
| | | **T1** | **T1** | **T1** | **u** | **s** | **H** | **u** | **s** | **H** | **u** | **s** | **H** |
| GloVe [38] | CLSN | 52.1 | 20.4 | 21.6 | 42.1 | 75.3 | 54.0 | 16.2 | 43.6 | 23.6 | 14.4 | 88.3 | 24.8 |
| GloVe [38] | Wiki | 61.6 | 29.0 | 25.8 | 49.5 | 78.1 | 60.6 | 23.8 | **62.6** | 34.5 | 14.7 | 91.0 | 25.3 |
| LongFormer [3] | Wiki | 44.2 | 22.6 | 8.8 | 41.6 | 81.8 | 55.2 | 19.9 | 41.0 | 26.8 | 8.8 | 89.8 | 16.0 |
| MPNet [47] | Wiki | 61.8 | 25.8 | 26.3 | 58.0 | 76.4 | 66.0 | 20.6 | 44.3 | 28.2 | 22.2 | **96.7** | 36.1 |
| TF-IDF [42] | Wiki | 46.4 | 39.9 | 34.0 | 29.6 | **87.6** | 44.2 | 29.0 | 52.1 | 37.3 | 28.9 | 94.8 | 44.3 |
| VGSE [55] | IMG + CLSN | 69.6 | 37.1 | - | 56.9 | 82.8 | 67.4 | 27.6 | 70.6 | 39.7 | - | - | - |
| I2DFormer [35] | Wiki | 76.4 | 45.4 | 40.0 | 66.8 | 76.8 | 71.5 | 35.3 | 57.6 | 43.8 | 35.8 | 91.9 | 51.5 |
| | 3-LLM (ours) | 69.7 | 46.0 | 41.9 | 65.2 | 80.4 | 72.0 | 36.6 | 59.5 | 45.3 | 37.4 | 94.2 | 53.5 |
| | 3-LLM + Wiki (ours) | 77.3 | 47.0 | 43.0 | 68.6 | 77.4 | 72.7 | 38.5 | 59.3 | 46.7 | 40.4 | 80.1 | 53.8 |
| **I2MVFormer** (ours) | Wiki | 73.6 | 42.1 | 41.3 | 66.6 | 82.9 | 73.8 | 32.4 | 63.1 | 42.8 | 34.9 | 96.1 | 51.2 |
| | 3-LLM (ours) | 76.4 | 47.8 | 44.4 | 72.7 | 81.3 | 76.8 | 40.1 | 58.0 | 47.4 | 41.1 | 91.1 | 56.6 |
| | 3-LLM + Wiki (ours) | **79.6** | **51.1** | **46.2** | **75.7** | 79.6 | **77.6** | **42.5** | 59.9 | **49.7** | **41.6** | 91.0 | **57.1** |

# Ablation over components of Model



|  | Components |  |  |  | AWA | CUB | FLO |
|---|---|---|---|---|---|---|---|
|  | $L_{CLS}$ | $L_{Local}$ | SVS | MVS | T1 | T1 | T1 |
| a) | ✓ |  |  |  | 73.6 | 45.6 | 38.9 |
| b) | ✓ |  | ✓ |  | 74.1 | 48.5 | 39.1 |
| c) |  | ✓ | ✓ | ✓ | 57.7 | 32.5 | 24.2 |
| d) | ✓ | ✓ | ✓ |  | 78.4 | 49.0 | 43.2 |
| e) | ✓ | ✓ | ✓ | ✓ | **79.6** | **51.1** | **46.2** |

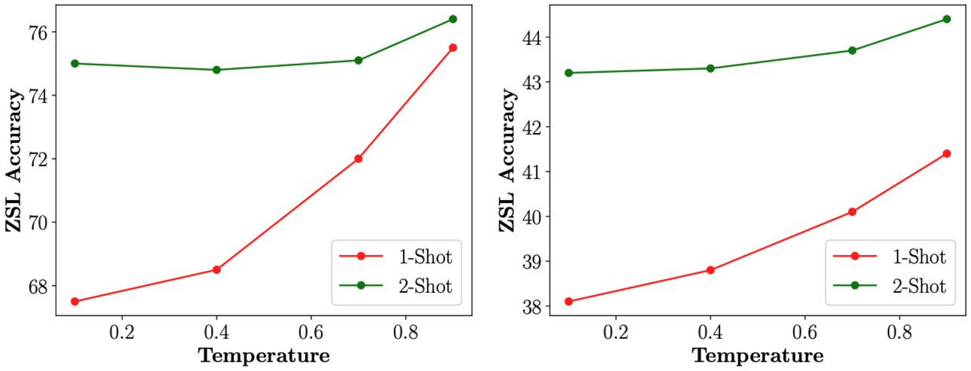
# Each view provides complimentary information



| Views from LLM | Zero-Shot Learning | | Generalized Zero-Shot Learning | | | | | |
|---|---|---|---|---|---|---|---|---|
| | AWA2 | FLO | AWA2 | | | FLO | | |
| | T1 | T1 | u | s | H | u | s | H |
| 1 | 71.6 | 39.0 | 67.5 | 75.2 | 71.2 | 34.6 | 88.0 | 49.6 |
| 2 | 74.8 | 43.6 | 70.5 | 80.2 | 75.0 | 37.7 | 91.0 | 53.3 |
| 3 | 76.4 | 44.4 | 72.7 | 81.3 | 76.8 | 41.1 | **91.1** | 56.6 |
| 3 + Wiki | **79.6** | **46.2** | **75.7** | **79.6** | **77.6** | **41.6** | 91.0 | **57.1** |

# How to prompt the LLM?

| Shots | Zero-Shot Learning | | Generalized Zero-Shot Learning | | | | | |
|---|---|---|---|---|---|---|---|---|
| | AWA2 | FLO | AWA2 | | | FLO | | |
| | T1 | T1 | u | s | H | u | s | H |
| 0 shot | 73.0 | 40.7 | 66.6 | 79.1 | 72.3 | 38.0 | 85.7 | 52.7 |
| 1 shot unique | 74.2 | 42.1 | 68.8 | **82.8** | 75.1 | 39.8 | 89.9 | 55.2 |
| 2 shots unique | **76.4** | **44.4** | **72.7** | 81.3 | **76.8** | **41.1** | **91.1** | **56.6** |



(a) AWA

(b) FLO

# Conclusion

- LLM can benefit a ZSL model by providing text supervision for both seen and unseen classes
- I2MVFormer uses multiple complementary class descriptions from the LLM to learn class embeddings
- I2MVFormer achieves SOTA performance across multiple ZSL benchmark datasets