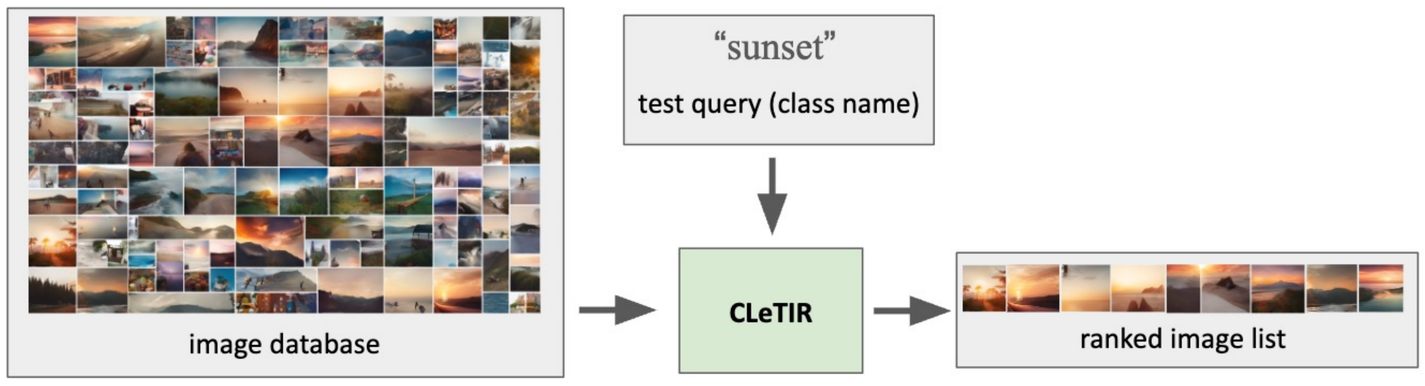
Category-level Text-to-Image Retrieval Improved: Bridging the Domain Gap with Diffusion Models and Vision Encoders

Faizan Farooq Khan[1], Vladan Stojnić[2], Zakaria Laskar[2], Mohamed Elhoseiny[1], Giorgos Tolias[2]
[1]KAUST & [2]VRG, CTU in Prague

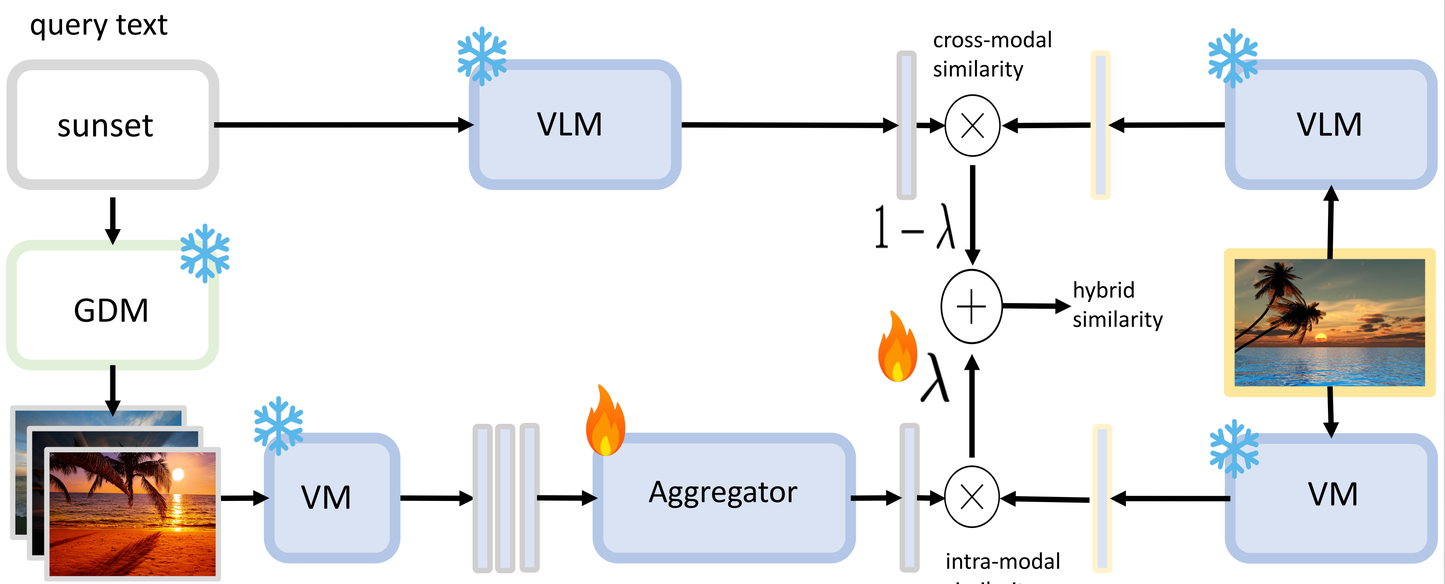
Category Level Text to Image Retrieval



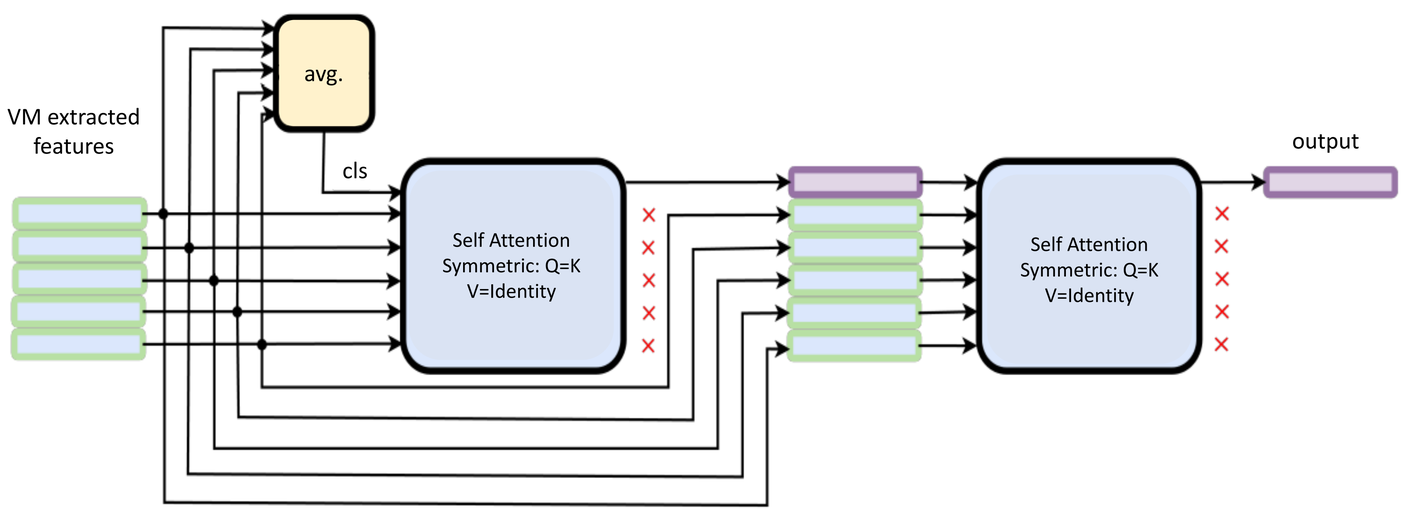
Contribution:

- **Bridging the modality gap by mapping text queries to visual space.**
- **Fusing queries from two modalities.**
- **Consistent improvements over the text-only baseline across 15 benchmarks.**

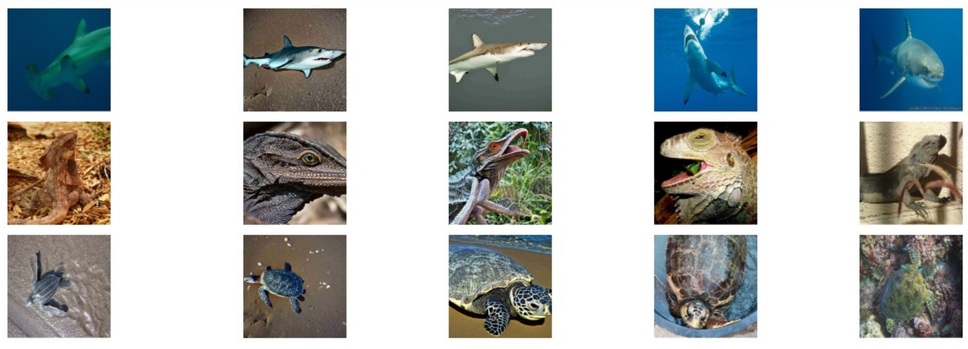
Overall Pipeline



Aggregator Architecture



Qualitative Samples

Class Name	Ground Truth	Generated	Top Negative
Hammerhead Shark			
Frilled-necked Lizard			
Leatherback Sea Turtle			
Pig			
Maltese			
Prairie Grouse			

Results

Method	Avg.
Query: text class	
Image Only(VM)	53.7
Text-only(VLM)	62.7
Image-only(VLM)	41.4
Text(VLM)+image(VLM)	56.4
Text(VLM)+image(VM)	61.0
Tip-Adapter	65.9
Ours	**67.9**
Query: Description Only	
Text-only(VLM)	37.2
Ours	**43.6**
Query: Description + Class	
Text-only(VLM)	63.6
Ours	**67.2**

Ablations

Method	Avg.
Average aggregation	61.0
Text-baseline	62.7
Ours-Full	67.9
w/o λ tuning	56.4
w/o dual generator	61.0
w/o dynamic negative mining	65.9
w/o repeating input	67.9

Multiple Generators

k	3	6	9
SD query images only			
mAP	67.8	68.0	68.0
SD+SD-Turbo+Flux query images			
mAP	68.4	68.8	69.0

Training Pipeline

Training data:

- Query classes: from OpenImages after removing test set classes
- Database: images from multiple text2image generators using query classes as prompts

- Aggregator and λ trained using a contrastive loss with hard negative mining

Description vs Class Name

Class Name	Description based text	Ground Truth	Description Based	Class Name
Ball Moss	the flowers of air plants are typically small and come in a variety of colors, such as white, pink, or purple.			
Dotted	material with a series of small dots printed on it is commonly seen.			
Lacelike	the delicate fabric is known for its thinness and intricate patterns, often featuring small holes or gaps.			
Moon Orchid	the white flower with yellow and orange stripes has a unique appearance.			
Silverbush	the plant's foliage has a silvery hue and its blooms are petite and white.			
Wild Cat	this feline creature has a compact and agile body, adorned with sharp ears, giving it a fierce and untamed appearance.			

Code, and data: https://github.com/faixan-khan/cletir