Zero-Shot Anomaly Detection with Dual-Branch Prompt Selection

Zihan Wang[1,2], Samira Ebrahimi Kahou[2,3,4], Narges Armanfard[1,2]

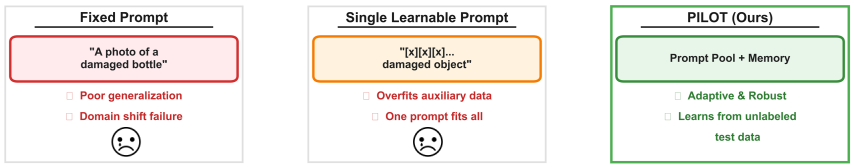
[1]Department of Electrical & Computer Engineering, McGill University, Montreal, QC, Canada
[2]Mila – Quebec AI Institute, Montreal, QC, Canada
[3]Department of Electrical and Software Engineering, University of Calgary Calgary, AB, Canada
[4]CIFAR AI Chair Canada

Scan Here!

Introduction

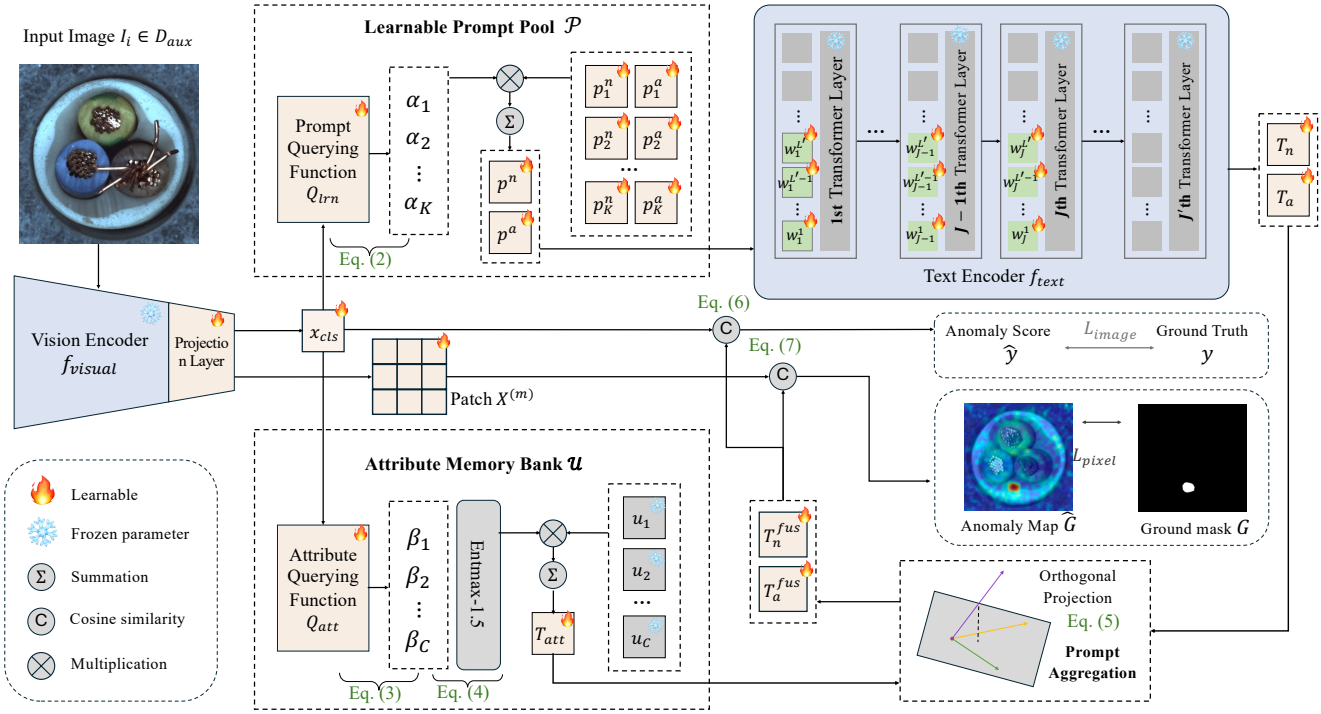
Anomaly detection is essential in safety-critical applications like industrial inspection and medical diagnostics. However, existing zero-shot approaches using vision–language models such as CLIP struggle under domain shifts due to limited prompt diversity and poor adaptability. We propose **PILOT**, a dual-branch framework that integrates a learnable prompt pool with a semantic attribute memory bank to dynamically select relevant anomaly cues. Additionally, a **label-free test-time adaptation** strategy refines prompts using high-confidence pseudo-labels from unlabeled test data, achieving state-of-the-art performance in both anomaly detection and localization across diverse domains.

Motivation



Existing zero-shot anomaly detectors fail under domain shifts because fixed or single prompts cannot capture diverse anomaly patterns.

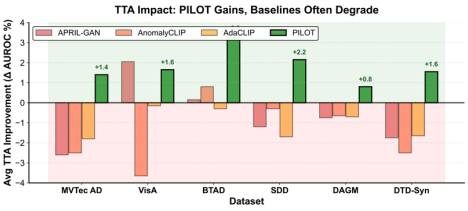
Overview of PILOT framework



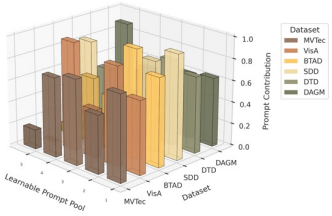
Test Time Adaption Phase

$$A = \mathrm{Top}_{\lceil \rho B \rceil}\{\tilde{y}_t\}_{t=1}^{B}, \quad N = \mathrm{Bot}_{\lceil \rho B \rceil}\{\tilde{y}_t\}_{t=1}^{B}$$
$$\mathcal{L}_{\text{image}} = \mathrm{BCE}(\hat{y}_i, \tilde{y}_i),\ i \in (A \cup N)$$

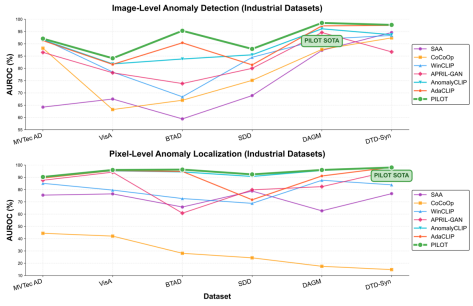
Unlike prior TTA methods that overfit or destabilize localization, **PILOT** performs selective prompt-weighted updates guided by high-confidence samples, preserving visual–semantic consistency and yielding stable gains in both image-level detection and pixel-level accuracy across domains.



Conclusion



The diagram reveal that PILOT adaptively combines diverse prompts, each capturing distinct semantic cues. By leveraging this diversity, the model adjusts its focus according to domain-specific features, achieving stronger generalization under distribution shifts.



Acknowledgments

Digital Research Alliance of Canada
Calcul Québec
Fonds de recherche Québec
NSERC CRSNG