JUNE 18-22, 2023
CVPR
VANCOUVER, CANADA

WinCLIP: Zero-/Few-Shot Anomaly Classification and Segmentation

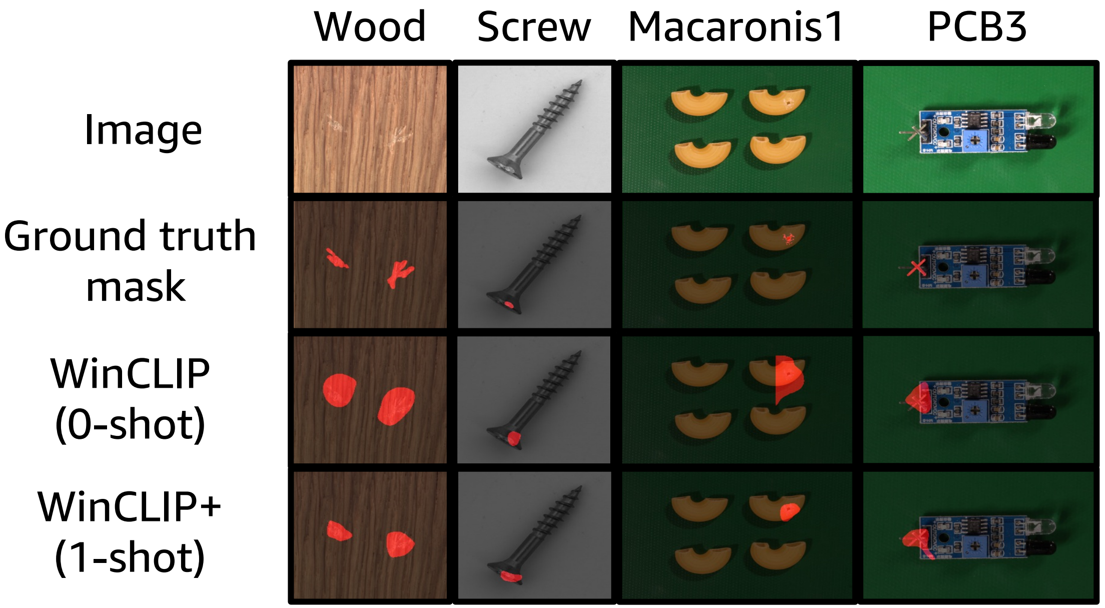
Jongheon Jeong[*❄], Yang Zou[❄], Taewan Kim, Dongqing Zhang, Avinash Ravichandran[☆], Onkar Dabeer

Poster: THU-AM-297

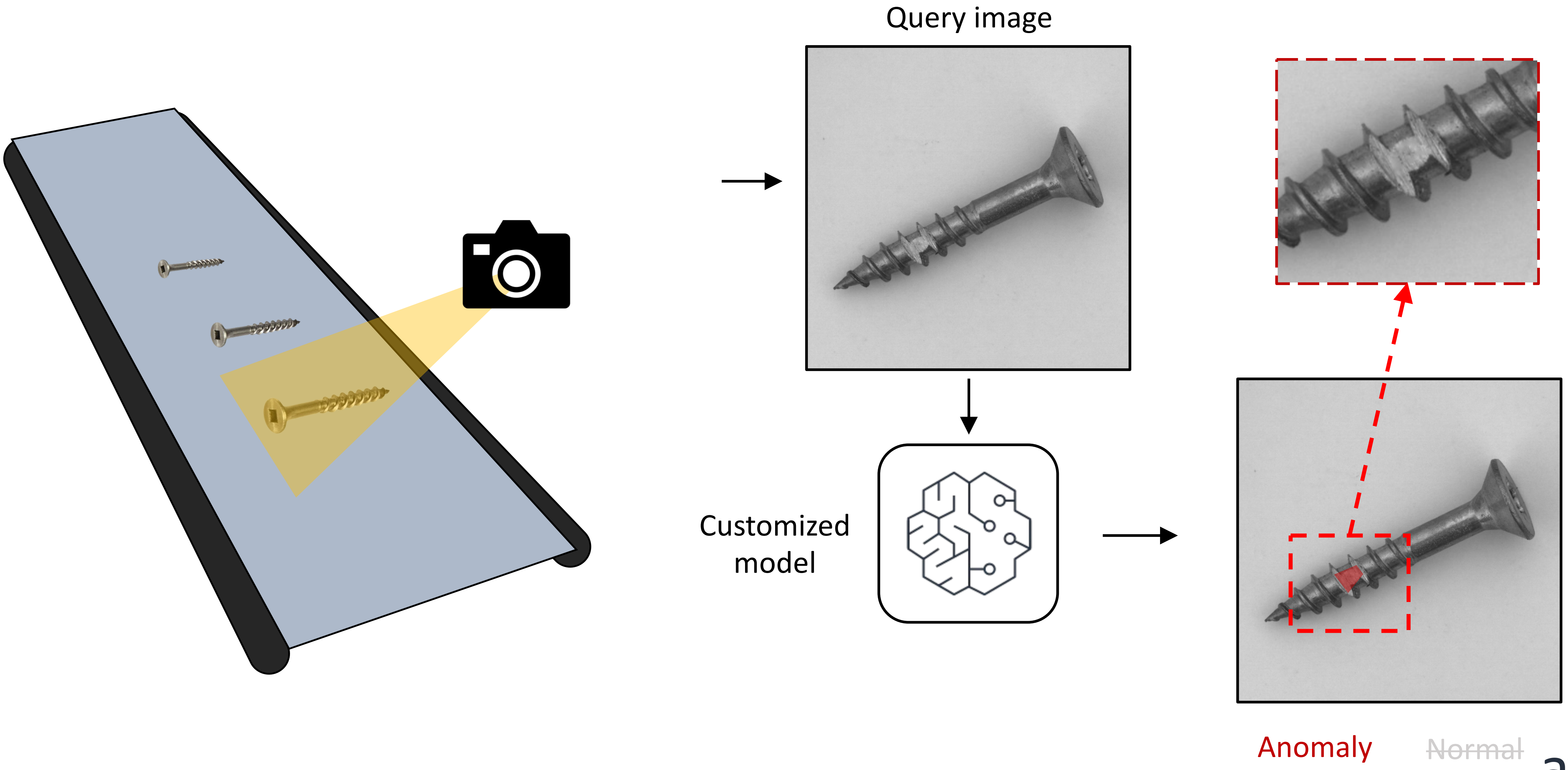
[*] PhD student at KAIST, work done during internship at AWS AI Labs
[❄] Equal contribution
[☆] Work done in AWS AI Labs

WinCLIP - Preview

- WinCLIP: The first language-guided zero-shot anomaly recognition model
  - Use pre-trained CLIP model with *compositional prompt ensemble*
  - Aggregate multi-scale spatial features aligned with language
- WinCLIP+: The first language-guided few-shot anomaly recognition model
  - WinCLIP + vision-based reference association
- WinCLIP (zero-shot) even outperforms SOTA few-shot anomaly classification methods



Anomaly Classification & Segmentation for Visual Inspection

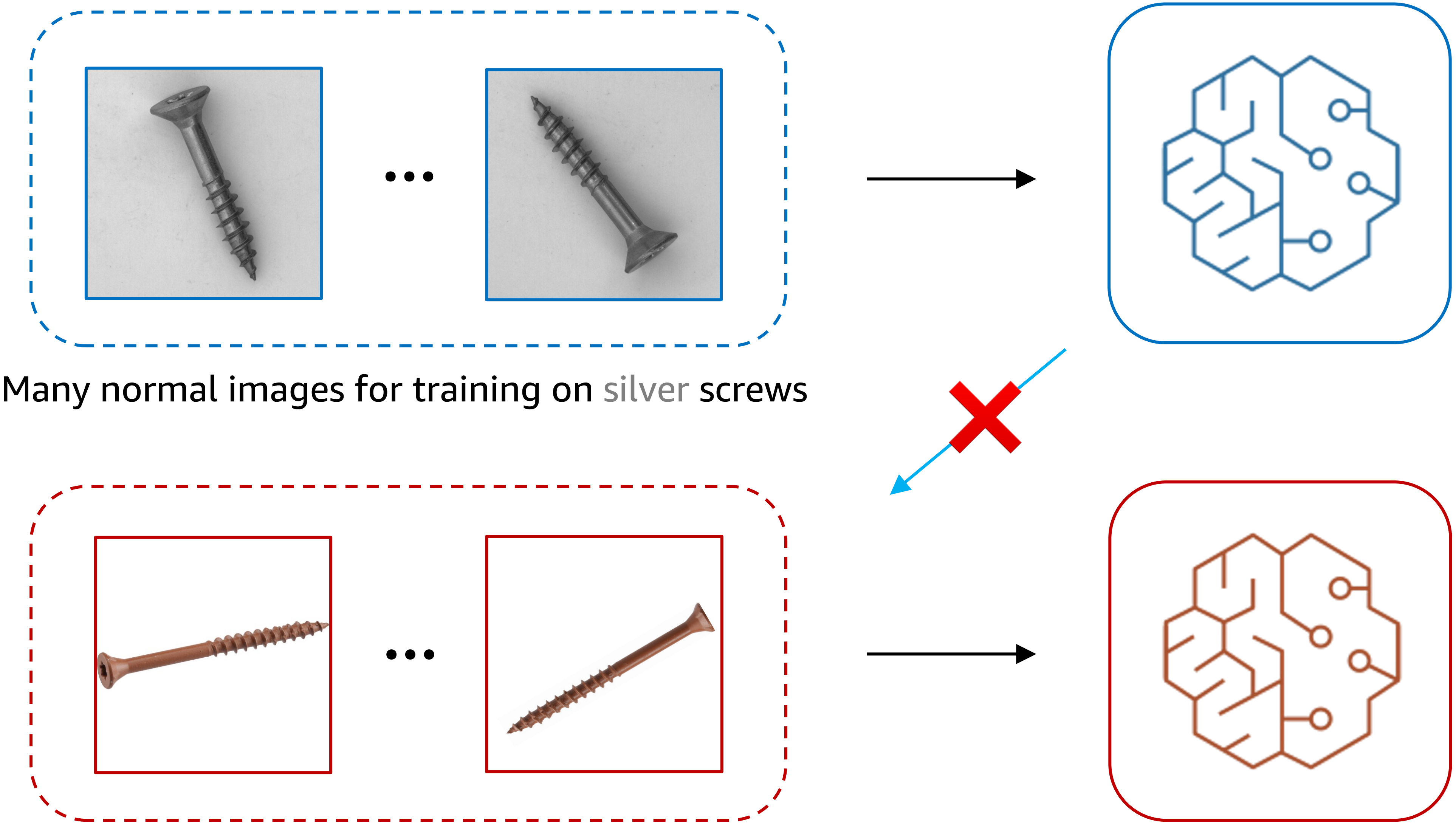


Query image

Customized model

Anomaly ~~Normal~~

Limited Generality Hinders Inspection at Scale

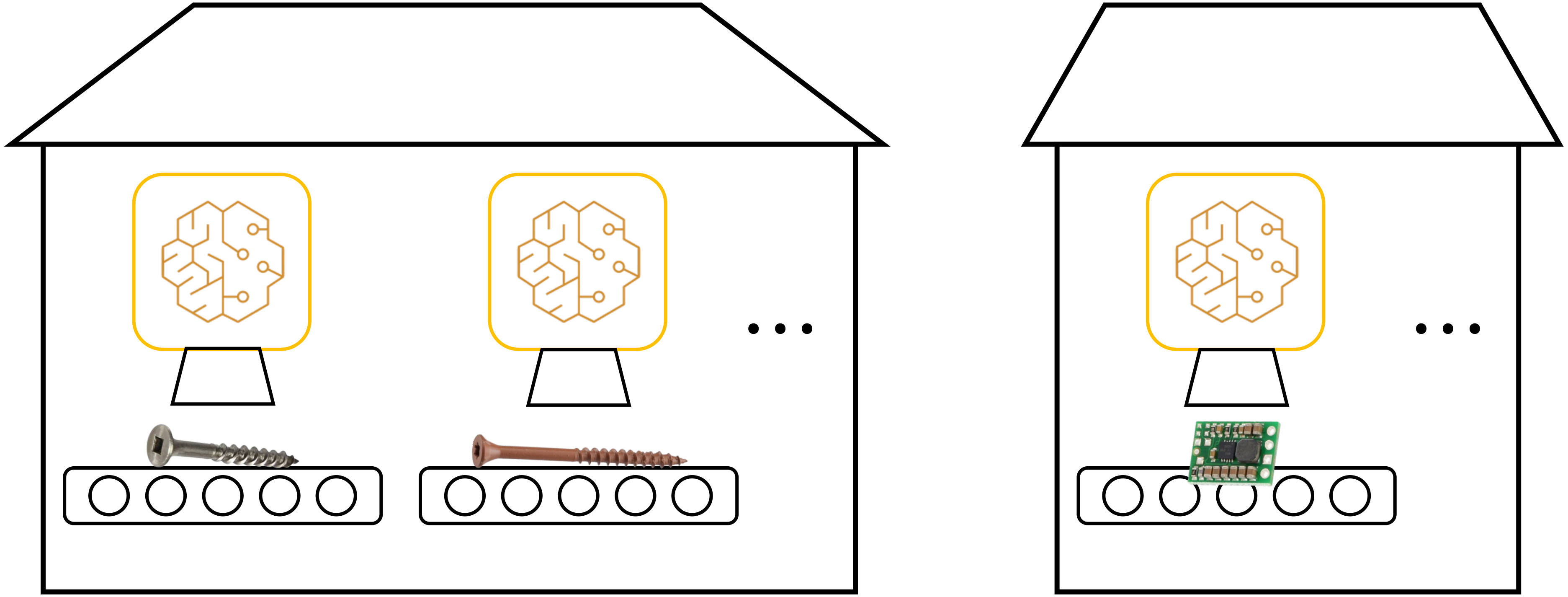


Many normal images for training on silver screws

Many normal images for training on red screws

Scalable Visual Inspection with An Unified Model

An unified model with **zero-/few-normal-shot anomaly** recognition ability, requiring no tuning for each task



Car, fabric, …

Hardware manufacturer

Electronics manufacturer

Principle 1: Language for Generalizable Anomaly Detection

- Language defines normality and anomaly that vary case by case

Normality: flawless/undamaged
Anomaly: crack/scratch/…

Normality: fresh/uncontaminated
Anomaly: mouldy/rotten/bitten/…

- We confirm this hypothesis with the CLIP

Principle 2: Multi-Scale Inspection for Comprehensive View

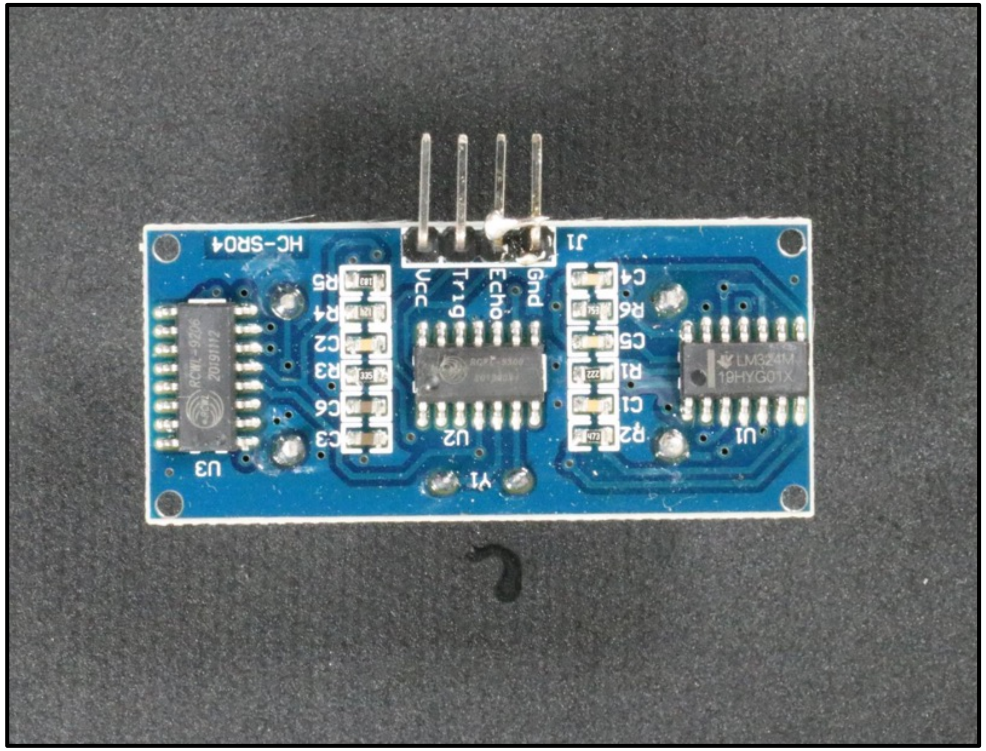
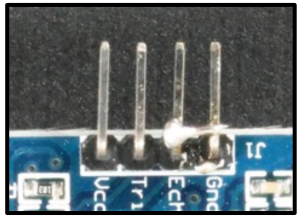
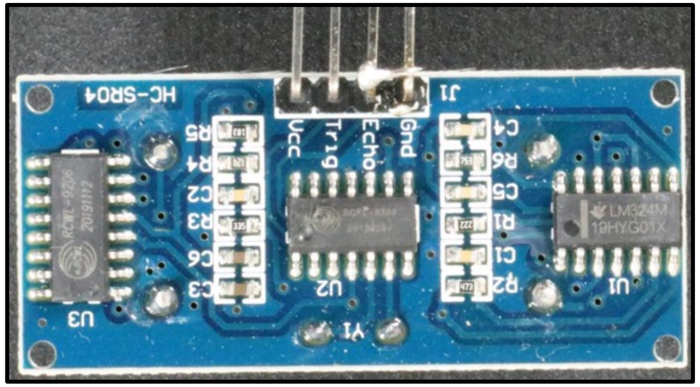


Image-level

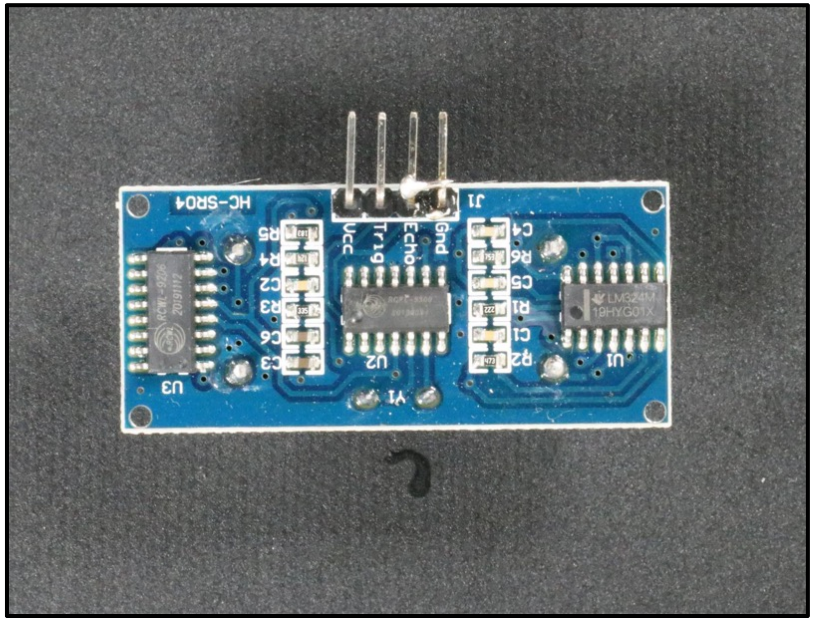


Window at small-scale

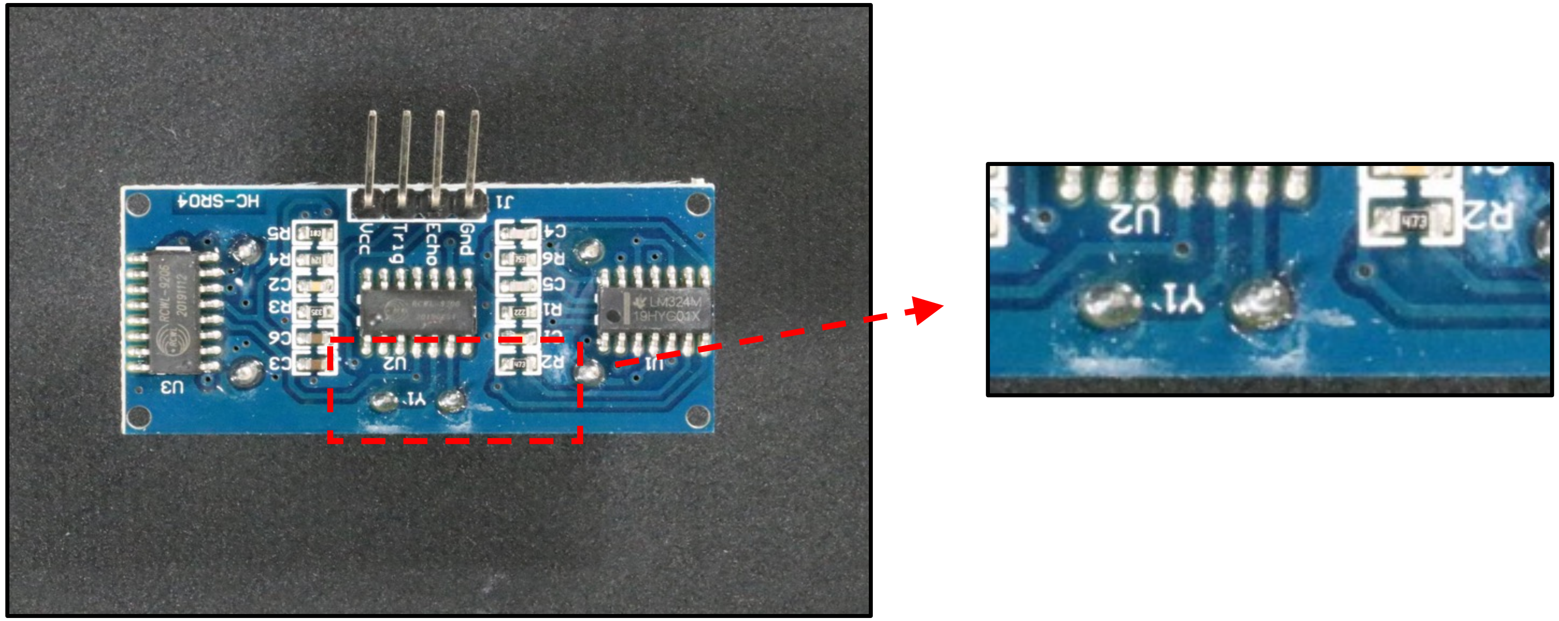


Window at mid-scale

Principle 3: Normal Image Clarifies Deviation from Normality



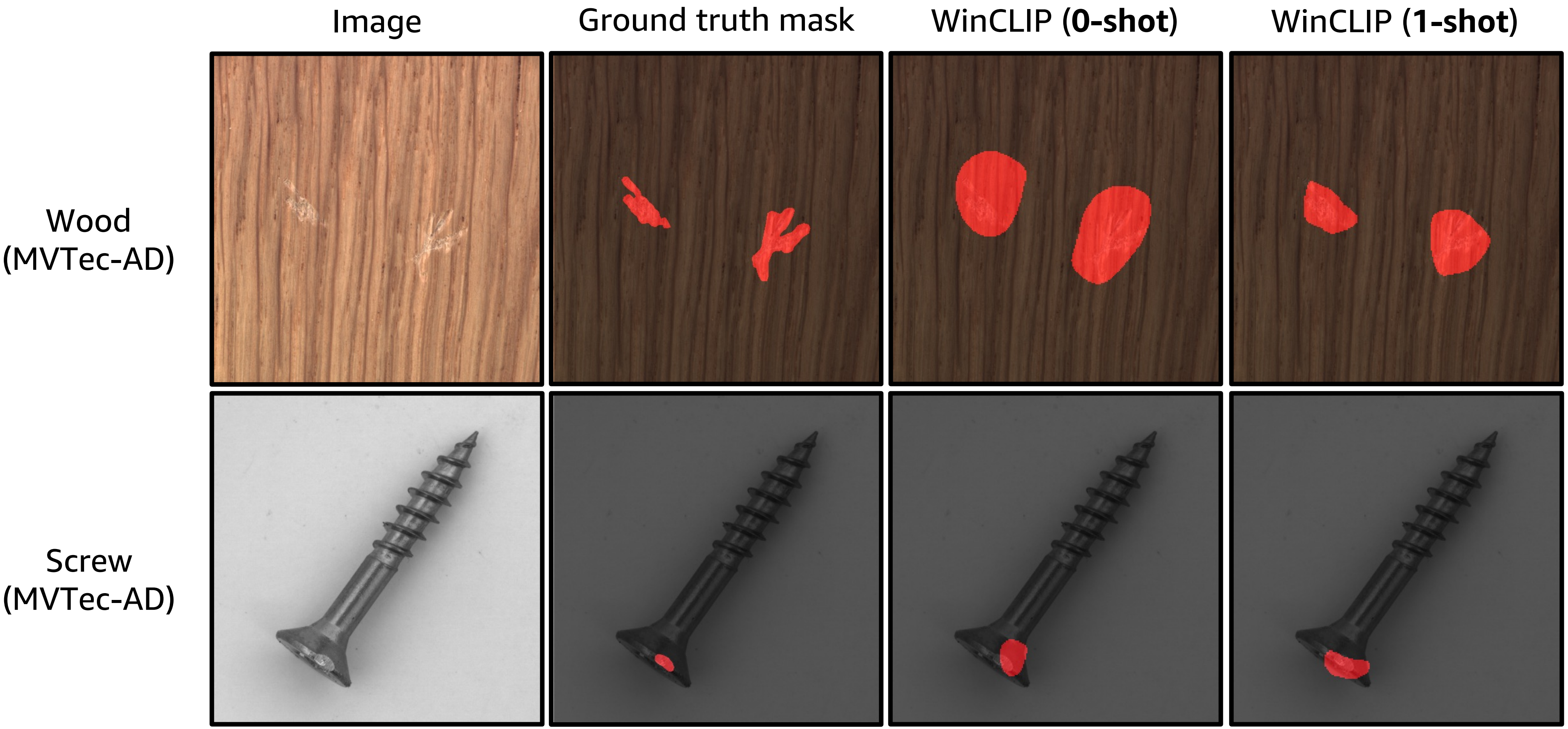
Query image



Normal reference image

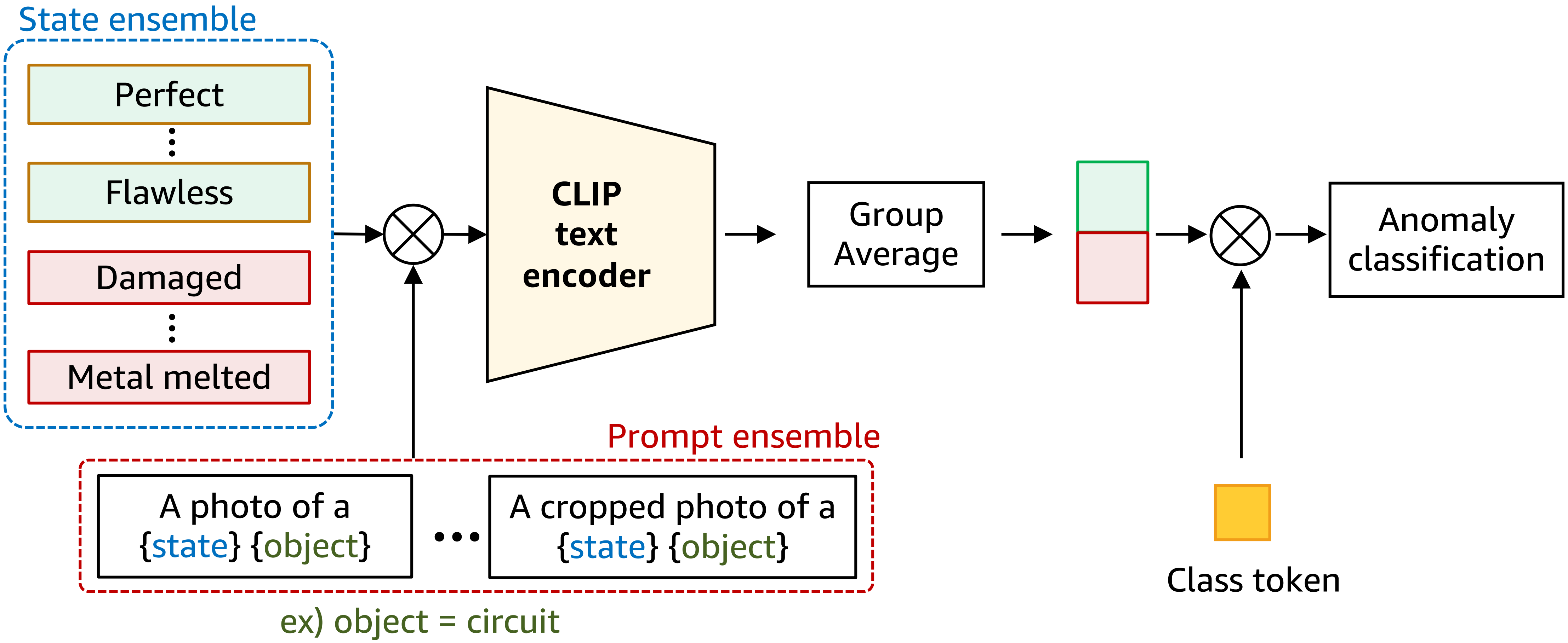
WinCLIP for Zero-/One-Shot Anomaly Segmentation

- **Window-based CLIP (WinCLIP)**



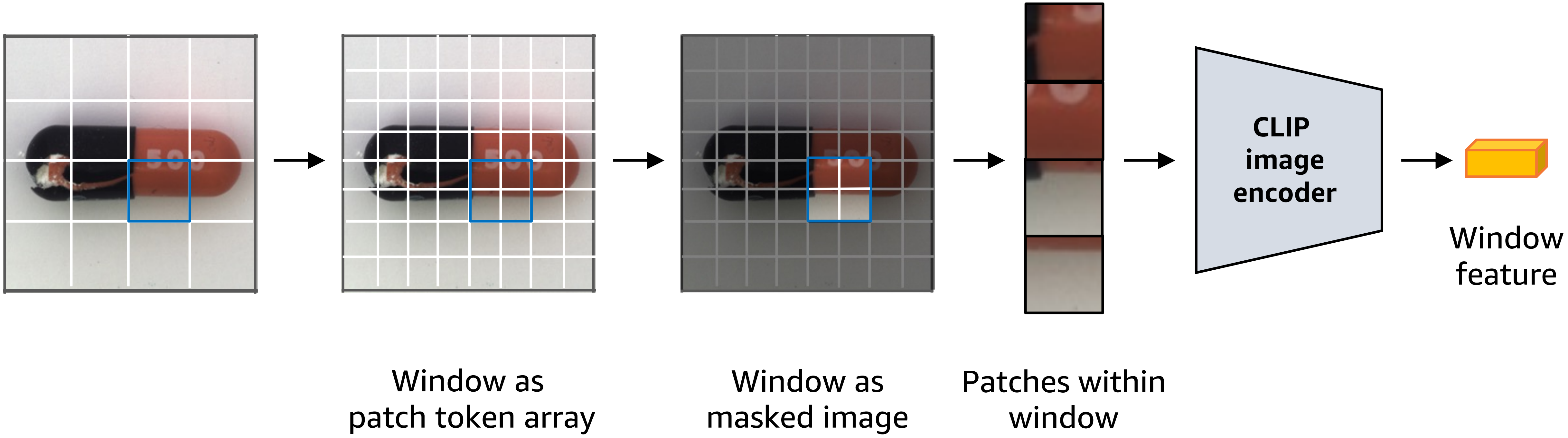
Language Driven Zero-Shot Anomaly Classification

Compositional Prompt Ensemble

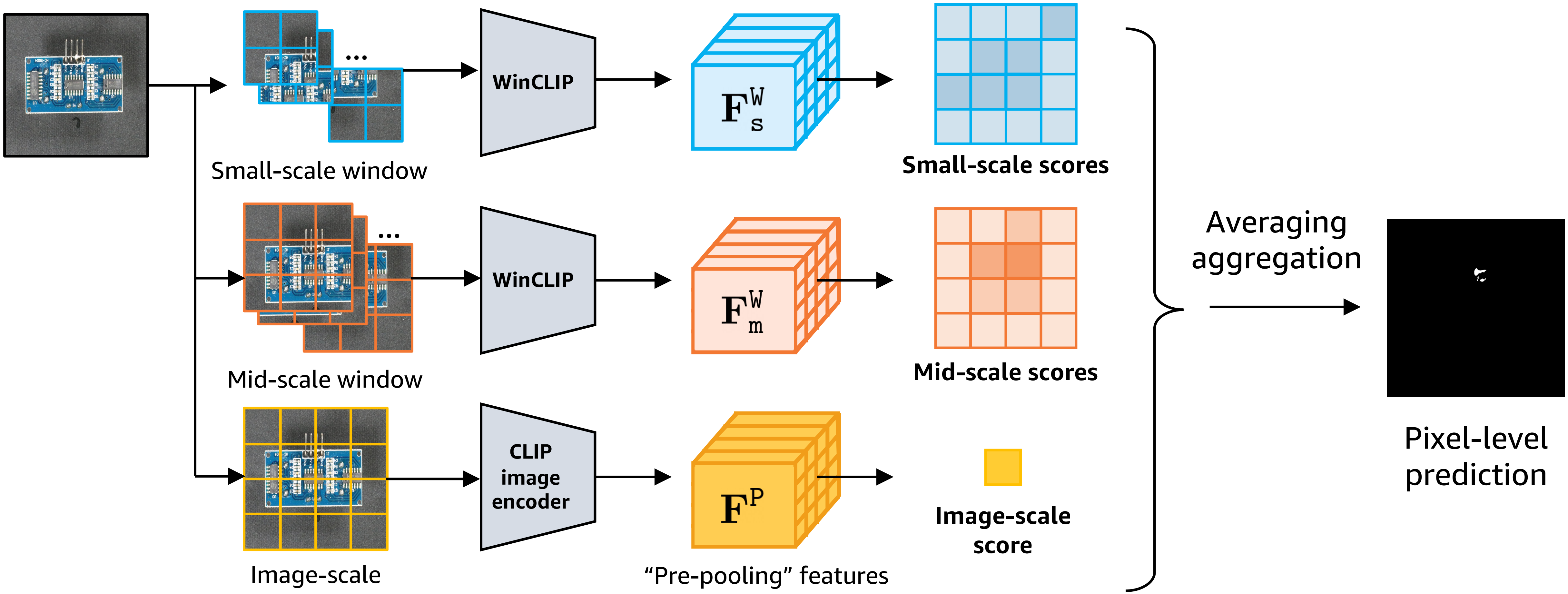


Efficient Window Feature Extraction via Maskable Inference

- Window based CLIP-ViT feature extraction

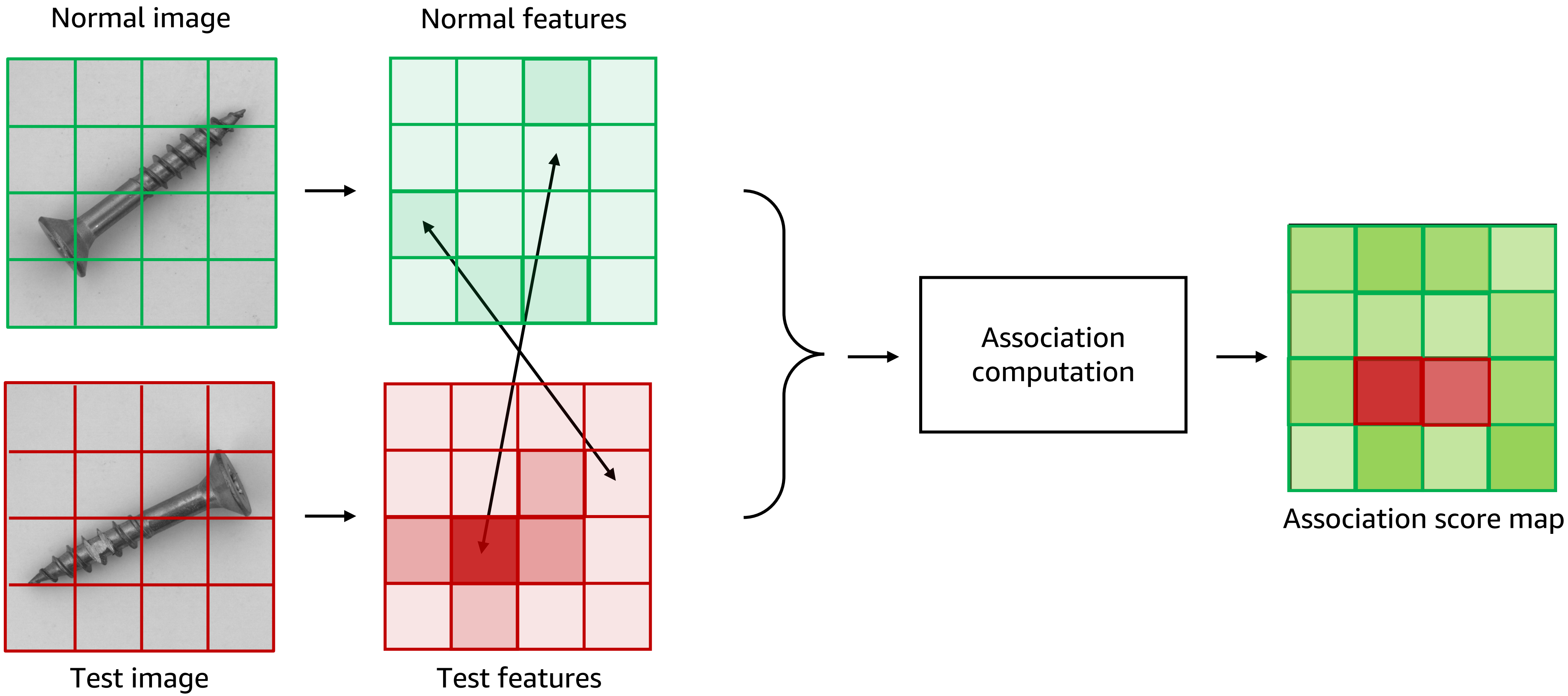


Multi-Scale Feature Extraction

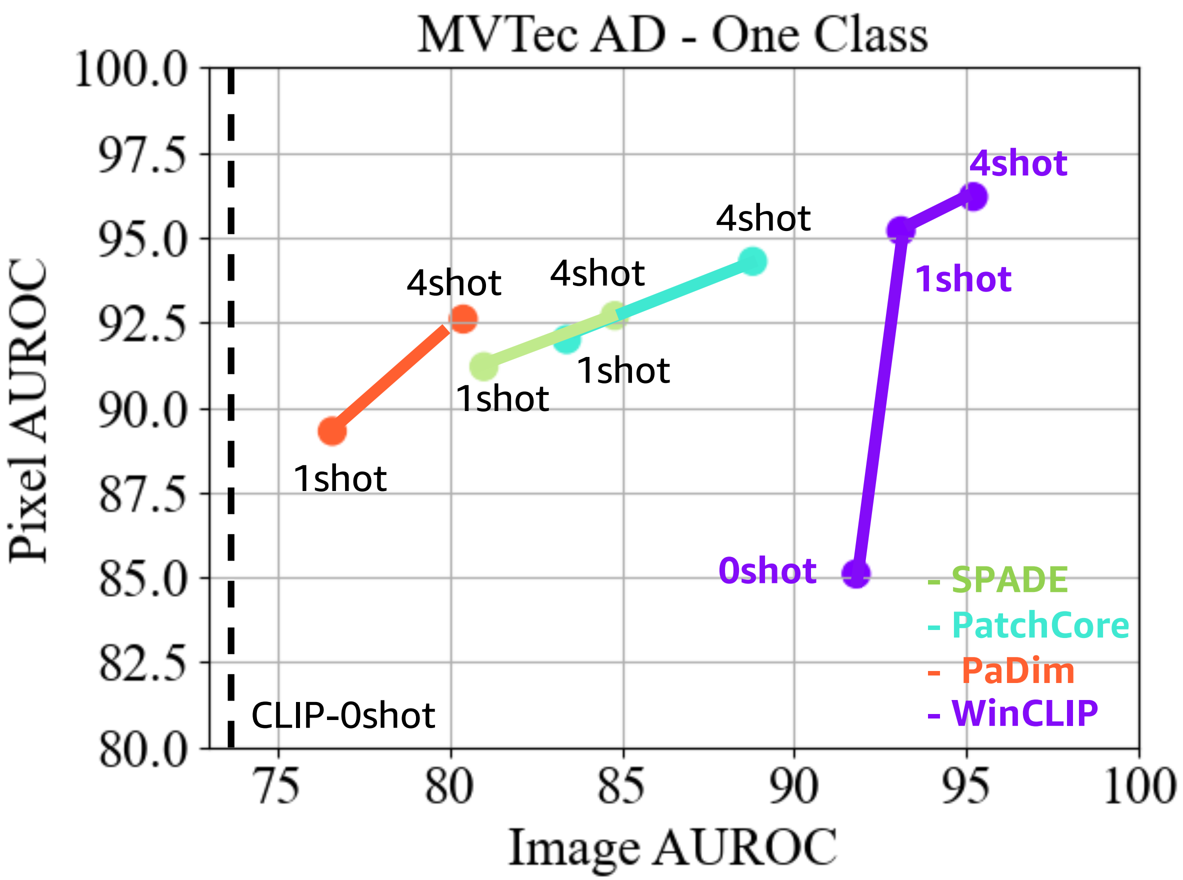


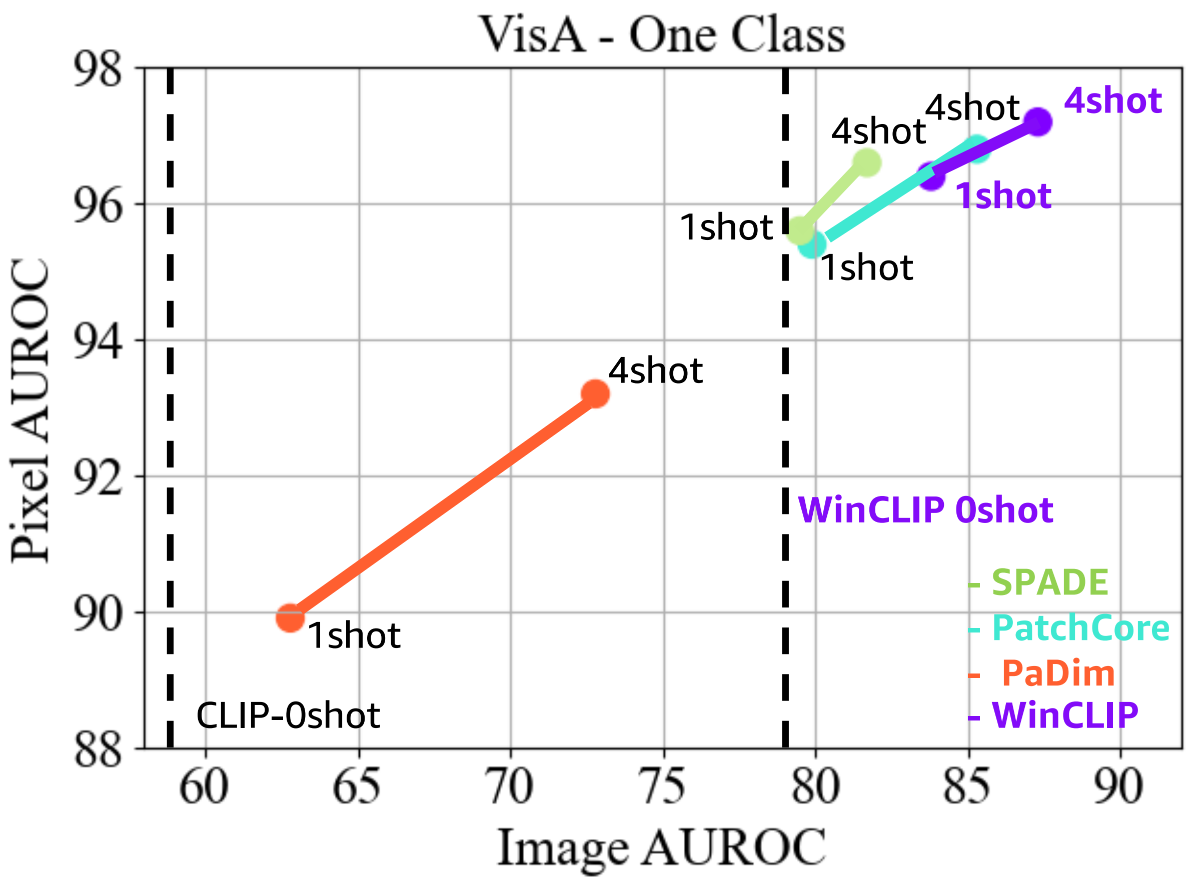
Reference Association for Visual Anomaly

1. Construct a local feature bank $\mathbf{R}$ by collecting those extracted from normal samples
2. The local anomaly score is defined as distance to the feature bank (distance to nearest neighbor)

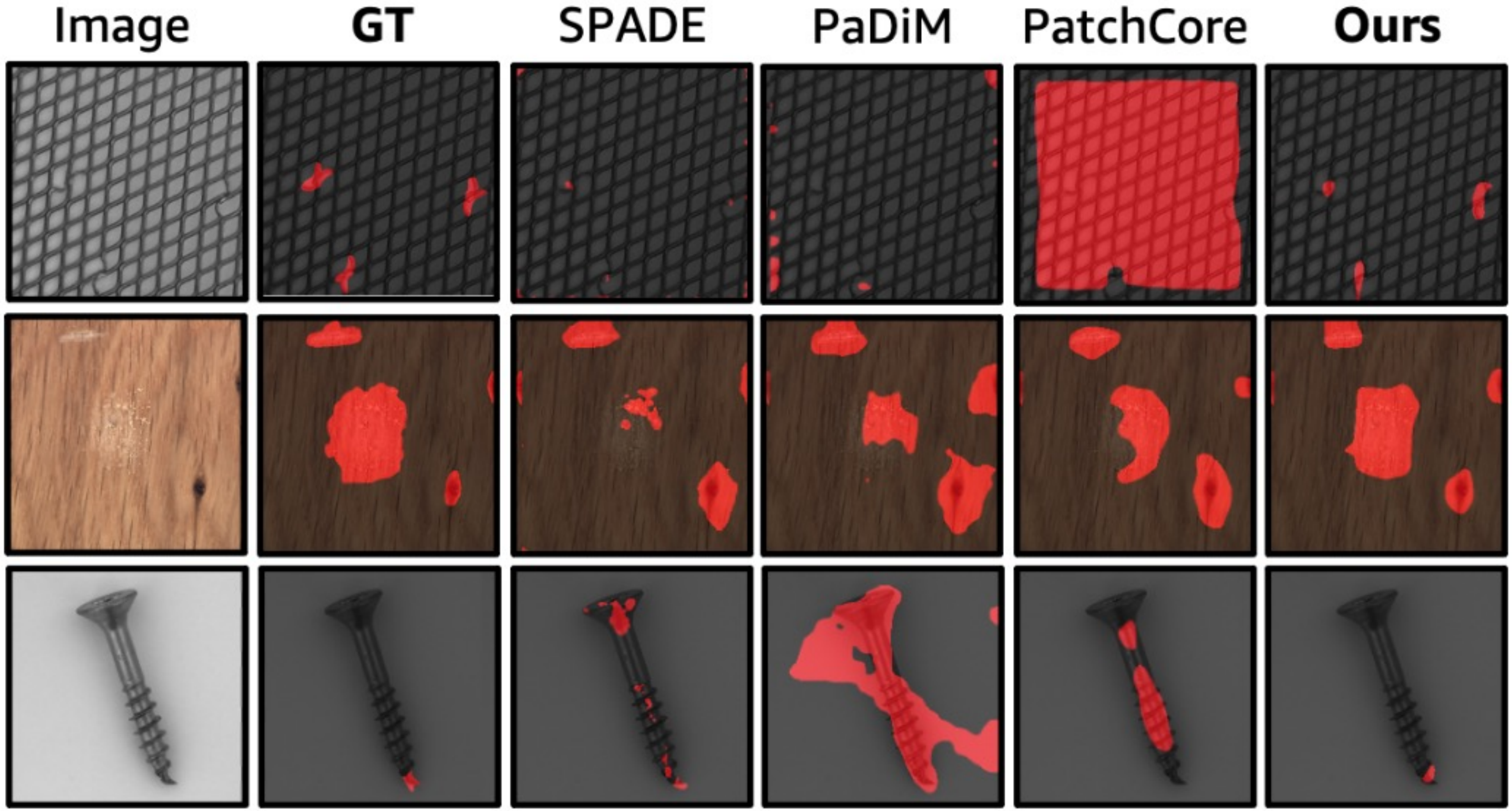


Quantitative Results





Qualitative Results: One-shot Anomaly Segmentation



Conclusion

- WinCLIP/WinCLIP+: a novel framework to define normality and anomaly with
  - Fine-grained text descriptions
  - Normal reference images
- CLIP pre-trained on large-scale web data provides a powerful representation
  - Alignment between texts and images for anomaly recognition
- Two-class design for zero-/few-shot anomaly recognition
  - Values complementary to standard one-class methods