One-for-All: Proposal Masked Cross-Class Anomaly Detection

AAAI2023 Main Track

Xincheng Yao
Shanghai Jiao Tong University

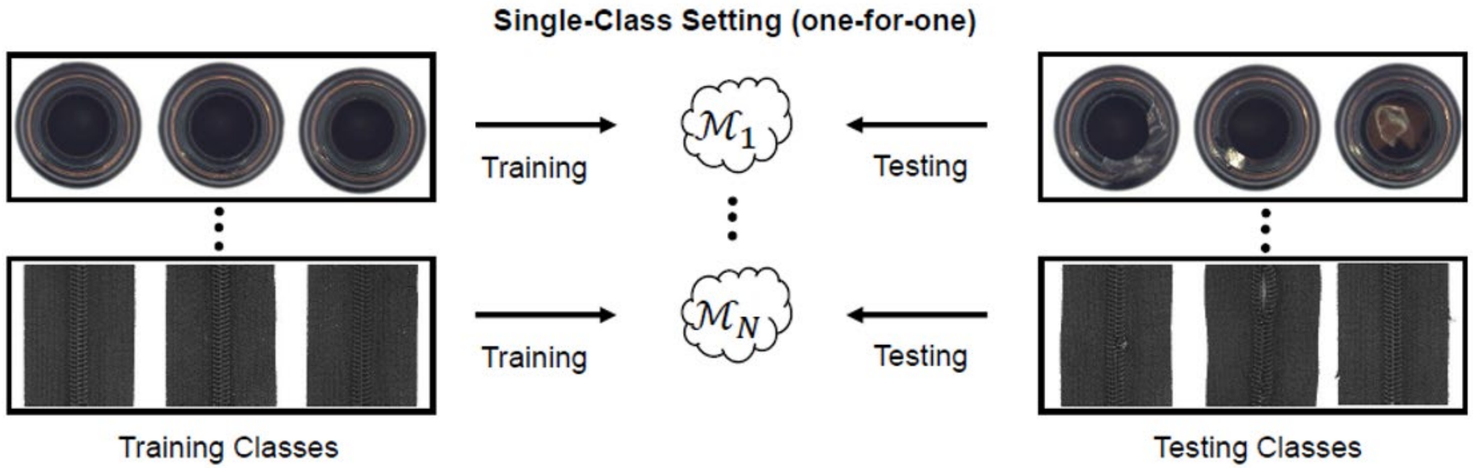
Coauthors:
Chongyang Zhang, Ruoqi Li, Jun Sun, Zhenyu Liu

Outline

1. Motivation
2. Background & Related Works
3. Our Approach: PMAD
4. Experiments
5. Ablations
6. Conclusions and Limitations

Motivation

Single-Class Setting (One-for-One)

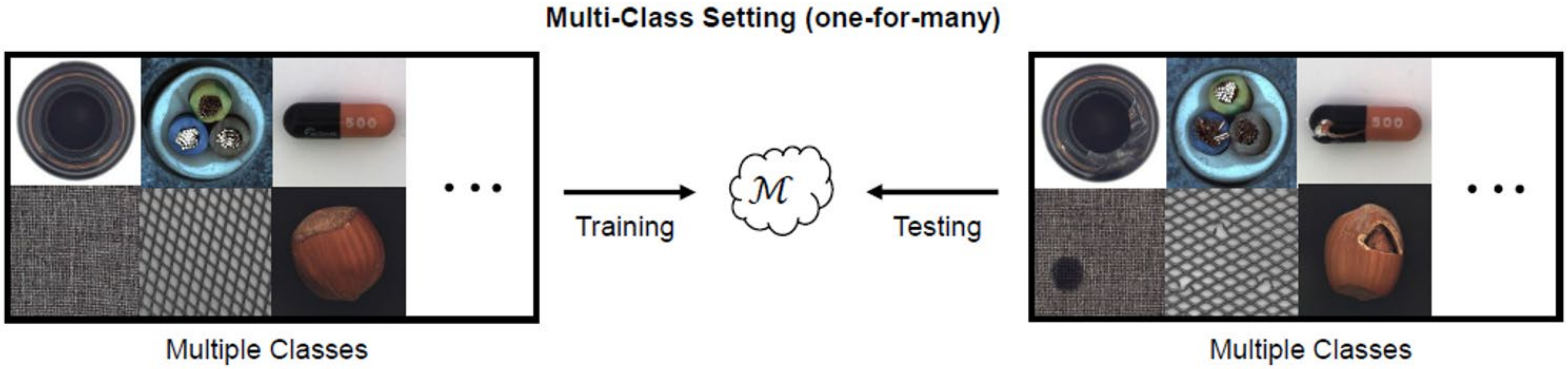


Previous methods often need to train a specific model for each object class.

- The one-for-one paradigm would require more computational and memory overhead.
- More resources are required to store different model weights in real-world applications.
- E.g., MVTecAD dataset has 15 classes, previous methods need to train 15 models.
- The trained models cannot generalize directly to new classes, which may cause the system to fail in new scenarios.

Motivation

Multi-Class Setting (One-for-Many)



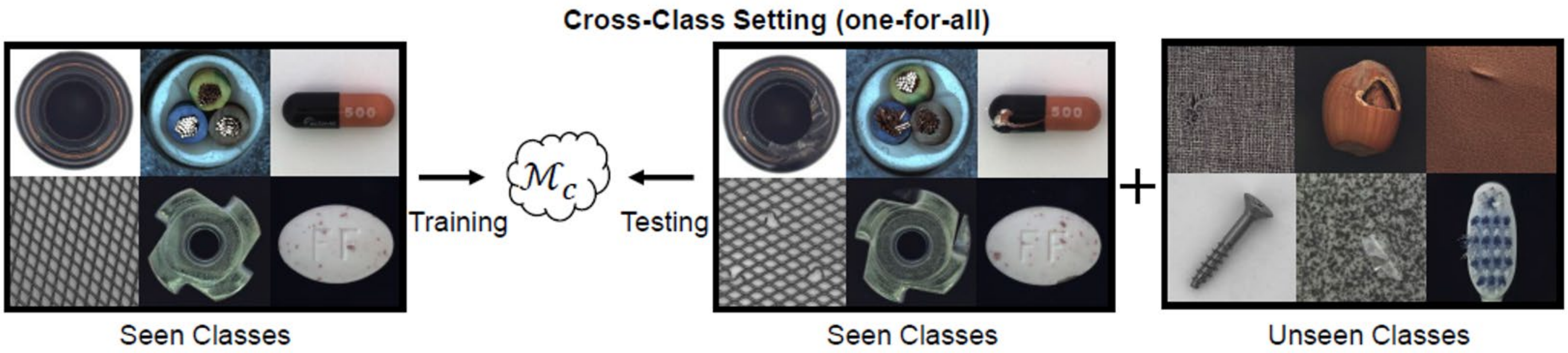
One unified model is trained and then used for multiple known classes.

- One unified model is more attractive to real-world applications.

Model needs to be **class agnostic!**

Motivation

Cross-Class Setting (One-for-All)



One unified model is trained with normal data from seen classes, and aims to detect anomalies from both seen and unseen classes.

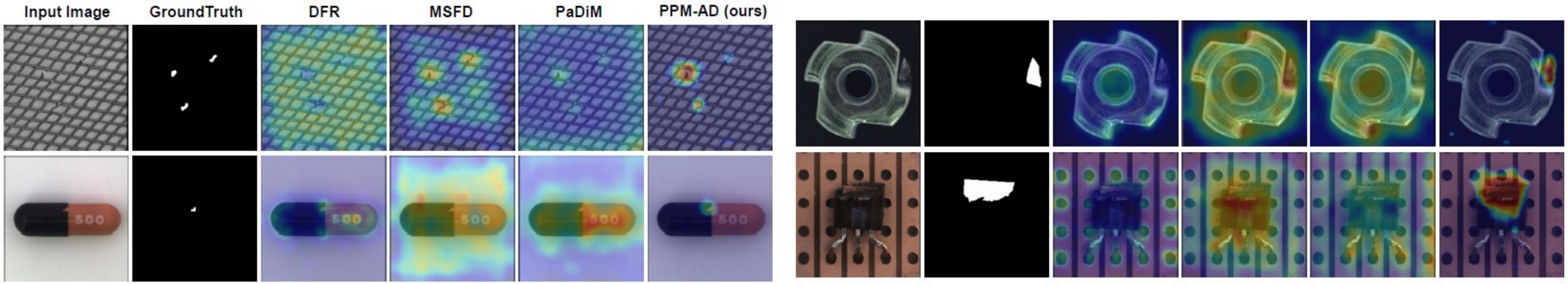
- This is the final goal for anomaly detection: one unified and generalizable model.

Model needs to be **class agnostic** and **class adaptive!**

Background & Related Works

- **Anomaly Detection:**

  Anomaly detection aims to distinguish an instance containing anomalous patterns from those normal samples, and further localize those anomalous regions.



- **Reconstruction-based Anomaly Detection:**

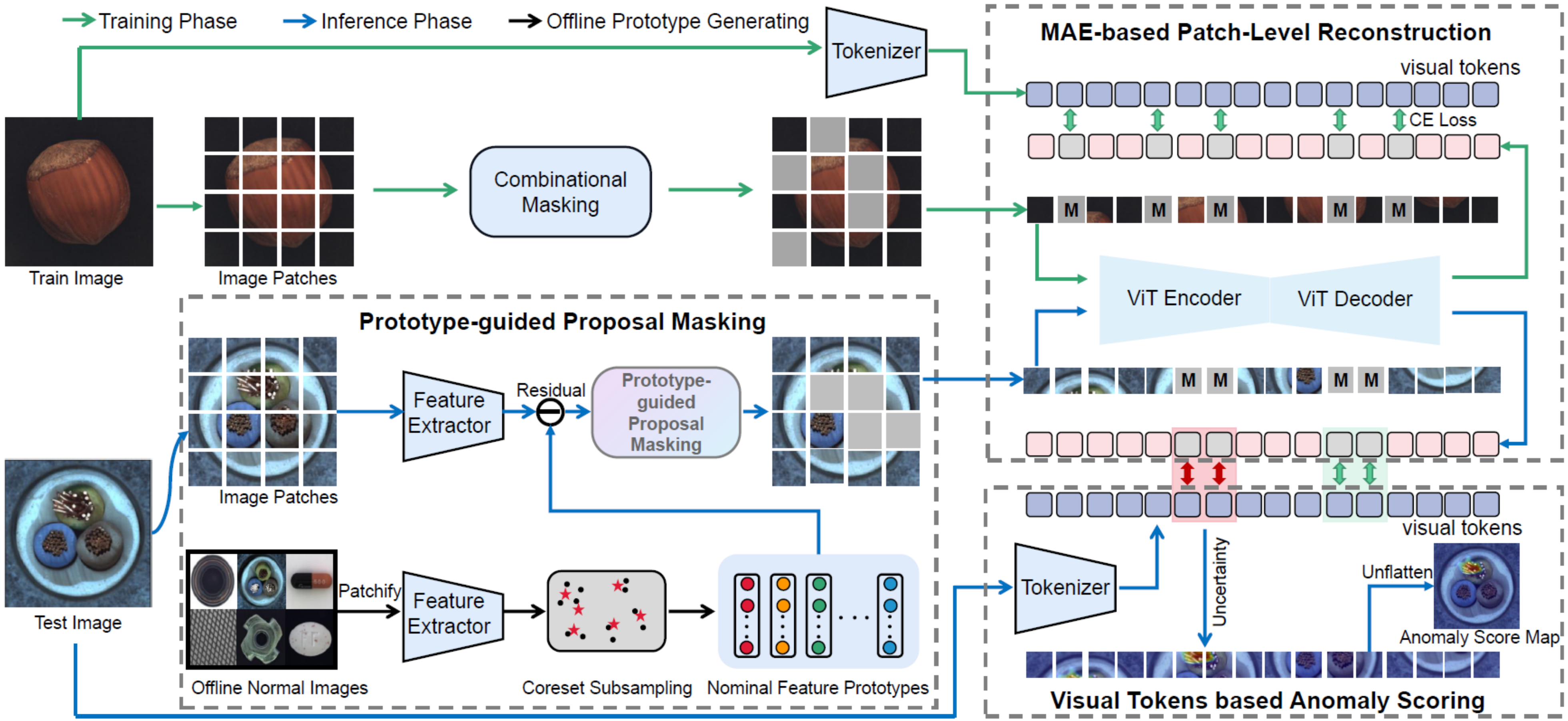
  These methods are based on the assumption that reconstruction models trained by normal samples only can reconstruct normal regions, but fail in abnormal regions.

- **Masked Image Modeling:**

  MIM-based patch-level reconstruction models are more adaptive and generalizable for unseen classes than conventional image-level reconstruction models.

Our Approach: PMAD

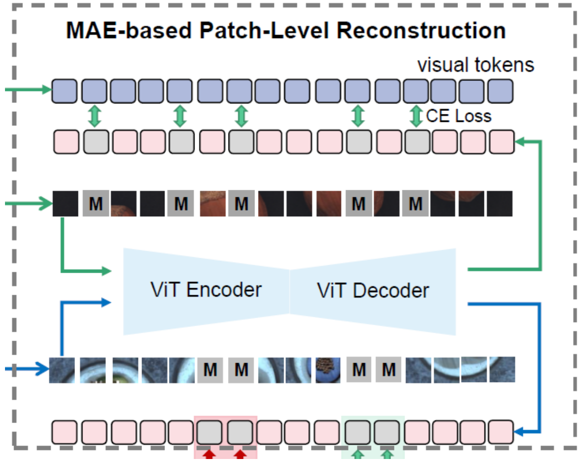
- **Proposal Masked Anomaly Detection, Model Overview:**



Three parts: MAE-based patch-level reconstruction, prototype-guided proposal masking, visual tokens based anomaly scoring.

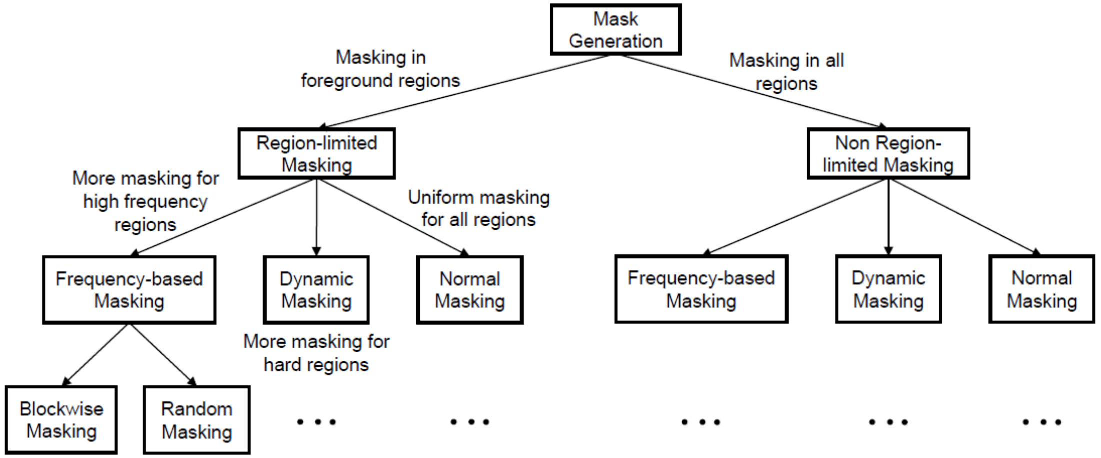
Our Approach: PMAD

- **MAE-based Patch-level Reconstruction:**



Network Architecture

- Standard ViT structure as both the encoder and decoder.
  - In the AD task, the decoder matters.



Combinational Masking

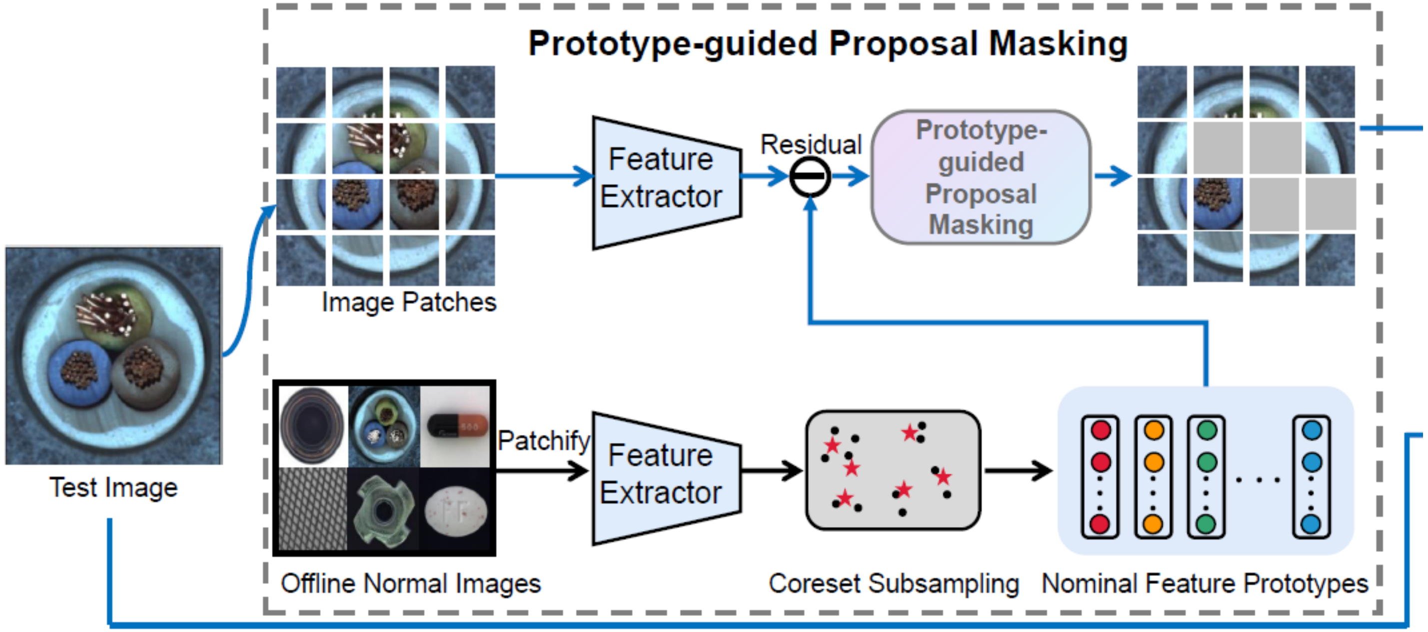
- Random Masking
- Blockwise Masking (continuous regions)
- Dynamic Masking...

Why patch-level reconstruction?

- The model learns how to utilize the contextual relationship to infer the features of masked patches.
- Even in unseen classes, the masked patches can be reconstructed well by employing the non-masked patches.

Our Approach: PMAD

- **Prototype-guided Proposal Masking:**

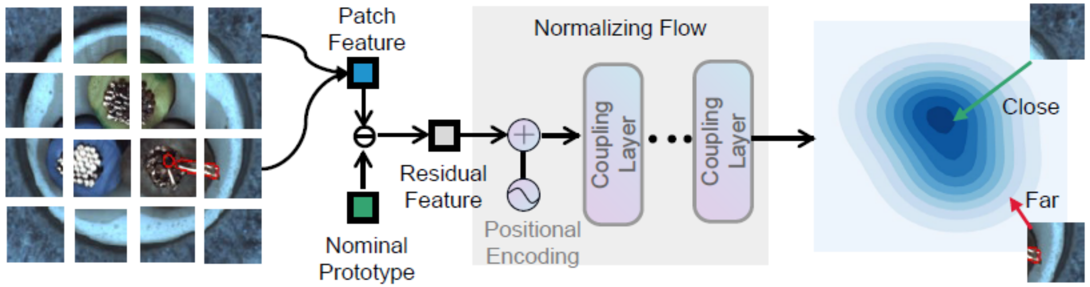


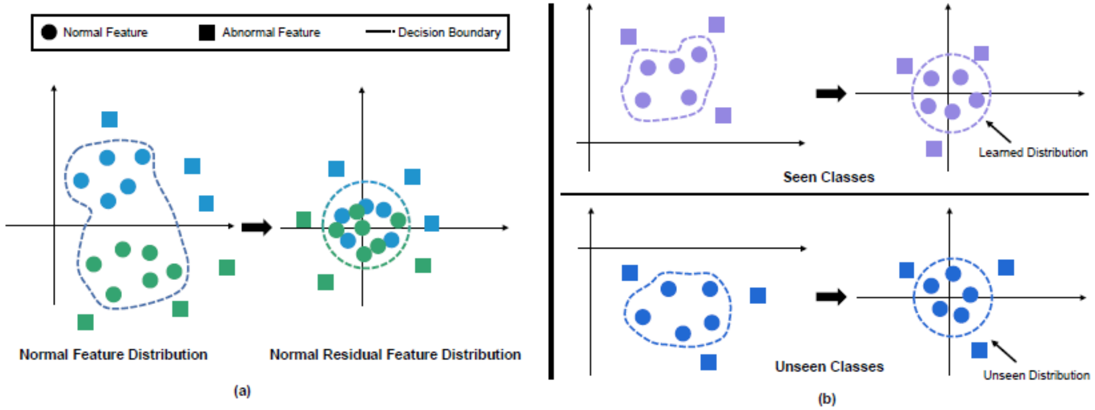
Why & Goal & How

- **Why:** Random and blockwise masking may leak a large amount of abnormal information.
- **Goal:** Masking suspicious anomaly proposals as much as possible.
- **How:** Forming an abnormality ranking of image patches, and selecting the top m percent of the image patches as masked patches.

Our Approach: PMAD

- **Prototype-guided Proposal Masking:**





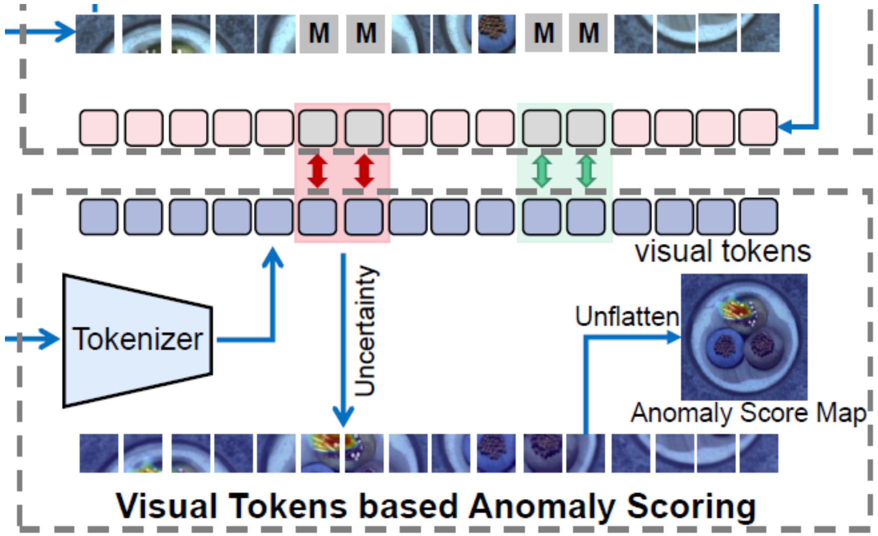
``Mis-masking`` Issue

- Normal patches are incorrectly masked in unseen classes.
- The normal patterns of unseen classes may be significantly different from the known patterns.

- Nominal prototypes guidance: The distribution of normal residual features would not be remarkably shifted from the learned distribution even in unseen classes.
- The distribution of normal residual features can be also significantly simplified.

Our Approach: PMAD

- **Visual Tokens based Anomaly Scoring:**



Anomaly Scoring

- We calculate cross-entropy to measure the uncertainty of each patch.
- The larger the uncertainty, the more likely the patch is to be abnormal.

$$s = -\sum_{i=1}^{|\mathcal{V}|} p_i \log(p_i)$$

Why & How

- Raw pixels as targets have a potential risk of overfitting to local statistics and high-frequency details.
- It would be affected by the image details.
- **Visual Tokens:** we follow DALL-E to compress an image with a dVAE codebook, each patch is encoded into a discrete visual token.

Experiments

SHANGHAI JIAO TONG UNIVERSITY

- **Datasets:**
  - MVTecAD: 5534 high-resolution images, 15 categories, 73 anomaly types, and 1900 abnormal regions.
  - BTAD: This dataset contains 2830 real-world images of 3 industrial products.
- **Metrics:**
  - Area under the curve of the receiver operating characteristic (AUROC), image-level and pixel-level.
- **Settings:**
  - Multi-Class Setting: train models with all classes from the dataset simultaneously.
  - Cross-Class Setting: select some classes as training classes and the remaining classes for testing.

Experiments

SHANGHAI JIAO TONG UNIVERSITY

- **Results under the Multi-Class Setting:**

Datasets	Multi-Class Setting						
	DFR	PaDiM	PatchSVDD	DRAEM	MSFD	CFLOW	PMAD (ours)
Carpet	0.975/0.982	0.997/0.985	0.399/0.698	0.967/0.979	0.992/0.988	0.988/0.975	0.990/0.979
Grid	0.923/0.953	0.845/0.870	0.685/0.696	0.995/0.990	1.000/0.970	0.959/0.941	0.962/0.956
Leather	0.972/0.991	1.000/0.988	0.727/0.732	0.993/0.985	1.000/0.972	1.000/0.981	1.000/0.992
Tile	0.885/0.807	0.956/0.923	0.886/0.738	1.000/0.989	0.999/0.950	0.979/0.922	0.998/0.945
Wood	0.991/0.935	0.989/0.916	0.946/0.774	1.000/0.954	0.983/0.930	0.990/0.927	0.996/0.890
Bottle	0.956/0.916	0.997/0.975	0.806/0.830	0.996/0.884	0.999/0.959	0.987/0.964	0.998/0.984
Cable	0.680/0.809	0.759/0.929	0.572/0.805	0.648/0.776	0.824/0.937	0.804/0.929	0.935/0.954
Capsule	0.832/0.969	0.789/0.979	0.706/0.878	0.739/0.603	0.649/0.943	0.755/0.977	0.805/0.970
Hazelnut	0.991/0.980	0.988/0.972	0.883/0.935	0.971/0.984	0.980/0.972	0.971/0.957	0.996/0.974
Metal nut	0.828/0.851	0.949/0.948	0.404/0.715	0.858/0.627	0.936/0.937	0.878/0.844	0.980/0.917
Pill	0.777/0.906	0.787/0.955	0.756/0.895	0.891/0.936	0.883/0.876	0.880/0.907	0.894/0.934
Screw	0.751/0.972	0.677/0.962	0.409/0.843	0.924/0.971	0.476/0.881	0.595/0.939	0.733/0.966
Toothbrush	0.831/0.959	0.878/0.977	0.781/0.844	0.975/0.983	0.728/0.983	0.780/0.957	0.958/0.982
Transistor	0.700/0.752	0.929/0.958	0.672/0.905	0.820/0.741	0.988/0.957	0.867/0.923	0.972/0.933
Zipper	0.919/0.951	0.865/0.969	0.727/0.681	0.998/0.984	0.888/0.880	0.922/0.957	0.960/0.961
MVTecAD Mean	0.867/0.916	0.894/0.954	0.691/0.798	0.918/0.891	0.888/0.944	0.890/0.940	0.945/0.956
Product 1	0.998/0.967	0.984/0.955	0.896/0.601	0.969/0.890	0.968/0.942	0.980/0.953	0.979/0.965
Product 2	0.866/0.960	0.839/0.948	0.725/0.812	0.772/0.928	0.796/0.955	0.825/0.952	0.834/0.957
Product 3	0.980/0.981	0.992/0.995	0.783/0.827	0.996/0.939	0.926/0.988	0.986/0.994	1.000/0.996
BTAD Mean	0.948/0.969	0.938/0.966	0.801/0.746	0.912/0.919	0.897/0.962	0.930/0.966	0.938/0.973

- Baseline methods drop dramatically under the multi-class setting.
- We beat the best competitor (DRAEM) under the multi-class setting by a large margin (2.7%).

Experiments

- **Results under the Cross-Class Setting:**

Datasets	Seen Classes (train)	Unseen Classes (test)	Cross-Class Setting DFR	PaDiM	PatchSVDD	DRAEM	MSFD	CFLOW	RegAD	PMAD (ours)
MVTecAD	Seen Textures	Grid	0.673/0.409	0.688/0.560	0.888/0.728	0.919/0.597	0.657/0.393	0.897/0.849	0.774/0745	0.950/0.907
		Tile	0.716/0.335	0.935/0.864	0.937/0.817	0.608/0.769	0.659/0.652	0.891/0.904	0.891/0.857	0.895/0.921
		Wood	0.986/0.764	0.987/0.895	0.934/0.775	0.711/0.661	0.845/0.896	0.964/0.913	0.956/0.913	0.989/0.860
		Mean	0.792/0.502	0.870/0.773	0.920/0.773	0.766/0.676	0.720/0.647	0.917/0.889	0.874/0.838	0.945/0.896
	Seen Objects	Hazelnut	0.943/0.958	0.795/0.917	0.882/0.945	0.472/0.858	0.422/0.840	0.725/0.889	0.832/0.908	0.984/0.953
		Metal nut	0.534/0.653	0.446/0.719	0.358/0.734	0.525/0.521	0.799/0.780	0.454/0.610	0.610/0.889	0.936/0.811
		Pill	0.533/0.758	0.470/0.786	0.784/0.905	0.669/0.677	0.683/0.805	0.405/0.708	0.526/0.901	0.870/0.901
		Toothbrush	0.572/0.892	0.361/0.875	0.789/0.865	0.606/0.926	0.642/0.949	0.519/0.887	0.647/0.934	0.794/0.959
		Zipper	0.394/0.733	0.291/0.837	0.793/0.786	0.474/0.498	0.914/0.948	0.772/0.925	0.720/0.925	0.889/0.933
		Mean	0.595/0.799	0.473/0.827	0.721/0.847	0.549/0.696	0.692/0.864	0.565/0.804	0.667/0.911	0.895/0.911
MVTecAD	Seen Textures	Carpet	0.475/0.229	0.979/0.986	0.836/0.805	0.789/0.605	0.970/0.981	0.970/0.981	0.887/0.891	0.995/0.984
		Leather	0.695/0.770	0.999/0.984	0.986/0.898	0.819/0.814	0.995/0.989	1.000/0.989	0.913/0.958	1.000/0.987
		Mean	0.585/0.499	0.989/0.985	0.911/0.852	0.804/0.709	0.982/0.985	0.985/0.985	0.900/0.924	0.997/0.986
	Seen Objects	Bottle	0.410/0.574	0.770/0.796	0.968/0.905	0.538/0.697	0.794/0.907	0.813/0.872	0.868/0.903	1.000/0.963
		Cable	0.534/0.699	0.530/0.686	0.771/0.829	0.398/0.337	0.608/0.747	0.446/0.718	0.594/0.825	0.889/0.944
		Capsule	0.244/0.895	0.401/0.905	0.711/0.908	0.259/0.801	0.686/0.855	0.519/0.891	0.623/0.969	0.794/0.959
		Screw	0.506/0.924	0.553/0.919	0.463/0.795	0.879/0.892	0.519/0.887	0.439/0.907	0.658/0.949	0.617/0.954
		Transistor	0.354/0.571	0.445/0.453	0.570/0.802	0.492/0.358	0.430/0.676	0.411/0.585	0.595/0.938	0.918/0.841
		Mean	0.409/0.733	0.536/0.752	0.697/0.848	0.513/0.617	0.607/0.814	0.525/0.795	0.668/0.917	0.844/0.932
BTAD	Product 1	Product 2	0.832/0.810	0.731/0.748	0.725/0.809	0.766/0.503	0.763/0.898	0.701/0.858	0.660/0.794	0.858/0.956
		Product 3	0.911/0.747	0.565/0.809	0.788/0.861	0.385/0.580	0.679/0.916	0.903/0.925	0.698/0.764	1.000/0.996
		Mean	0.872/0.778	0.648/0.778	0.756/0.835	0.576/0.542	0.721/0.907	0.802/0.892	0.679/0.779	0.929/0.976
	Product 2	Product 1	0.508/0.550	0.343/0.735	0.892/0.475	0.868/0.560	0.527/0.333	0.589/0.706	0.731/0.845	0.978/0.964
		Product 3	0.714/0.549	0.718/0.711	0.764/0.845	0.551/0.698	0.696/0.402	0.621/0.801	0.598/0.741	1.000/0.996
		Mean	0.611/0.550	0.531/0.723	0.828/0.660	0.709/0.629	0.611/0.368	0.605/0.753	0.665/0.793	0.989/0.980
	Product 3	Product 1	0.624/0.671	0.422/0.771	0.939/0.551	0.746/0.630	0.607/0.661	0.912/0.872	0.723/0.785	0.977/0.964
		Product 2	0.719/0.721	0.691/0.701	0.706/0.801	0.542/0.475	0.765/0.873	0.760/0.821	0.608/0.761	0.820/0.957
		Mean	0.672/0.696	0.556/0.736	0.823/0.676	0.644/0.553	0.686/0.767	0.836/0.846	0.666/0.773	0.898/0.960

Datasets	Seen Classes (train)	Unseen Classes (test)	Cross-Class Setting DFR	PaDiM	PatchSVDD	DRAEM	MSFD	CFLOW	RegAD	PMAD (ours)
MVTecAD	Seen Objects	Carpet	0.456/0.202	0.944/0.978	0.532/0.656	0.514/0.506	0.863/0.945	0.948/0.976	0.880/0.899	0.985/0.971
		Grid	0.665/0.284	0.674/0.461	0.886/0.635	0.360/0.480	0.962/0.937	0.744/0.831	0.683/0.688	0.912/0.920
		Leather	0.617/0.611	0.982/0.977	0.754/0.670	0.429/0.482	0.812/0.970	0.914/0.986	0.899/0.967	1.000/0.984
		Tile	0.598/0.310	0.915/0.775	0.837/0.796	0.823/0.525	0.721/0.746	0.773/0.856	0.838/0.856	0.987/0.936
		Wood	0.984/0.846	0.976/0.853	0.905/0.775	0.907/0.612	0.806/0.890	0.835/0.899	0.854/0.869	0.983/0.885
		Mean	0.664/0.451	0.898/0.809	0.783/0.707	0.607/0.521	0.833/0.898	0.843/0.909	0.831/0.856	0.973/0.939
	Seen Textures	Bottle	0.362/0.474	0.817/0.775	0.811/0.859	0.629/0.272	0.891/0.879	0.887/0.832	0.863/0.920	0.953/0.935
		Cable	0.530/0.715	0.512/0.722	0.569/0.810	0.466/0.336	0.615/0.779	0.607/0.742	0.565/0.795	0.918/0.932
		Capsule	0.271/0.865	0.469/0.909	0.651/0.919	0.724/0.727	0.510/0.815	0.619/0.849	0.641/0.970	0.659/0.928
		Hazelnut	0.906/0.932	0.831/0.928	0.864/0.945	0.537/0.928	0.621/0.884	0.739/0.866	0.858/0.933	0.891/0.925
		Metal nut	0.607/0.634	0.430/0.725	0.471/0.797	0.464/0.536	0.731/0.769	0.653/0.600	0.600/0.881	0.753/0.662
		Pill	0.560/0.701	0.532/0.736	0.755/0.900	0.635/0.690	0.648/0.620	0.527/0.637	0.573/0.900	0.770/0.860
		Screw	0.636/0.911	0.547/0.915	0.227/0.835	0.974/0.587	0.947/0.894	0.447/0.885	0.667/0.948	0.575/0.885
		Toothbrush	0.492/0.843	0.392/0.864	0.797/0.861	0.631/0.862	0.525/0.815	0.533/0.845	0.664/0.942	0.892/0.920
		Transistor	0.382/0.584	0.291/0.837	0.686/0.919	0.198/0.557	0.648/0.660	0.439/0.638	0.539/0.910	0.867/0.805
		Zipper	0.576/0.674	0.331/0.826	0.730/0.637	0.522/0.542	0.881/0.801	0.736/0.923	0.728/0.927	0.871/0.917
		Mean	0.532/0.733	0.535/0.800	0.656/0.848	0.578/0.604	0.702/0.792	0.619/0.782	0.670/0.913	0.815/0.877

- Our method can outperform these SOTA methods significantly.
- For texture classes, outperform by (2.5%/0.7% and 0.8%/0.1%).
- For object classes, outperform by (17.4%/4.7% and 14.7%/8.4%).

Ablations

- **Ablation study results:**

Ablations		Multi-Class Setting MVTecAD
Network Structure	Asymmetric Architecture (MAE)	0.918/0.937
	ViT structure	**0.945/0.956**
Training Masking Strategy	Random Masking	0.929/0.945
	Blockwise Masking	0.939/0.950
	Combinational Masking	**0.945/0.956**
Reconstruction Objective	Raw Pixels	0.773/0.712
	Deep Features	0.844/0.867
	Visual Tokens	**0.945/0.956**
Inference Masking Strategy	Random Masking	0.730/0.667
	Blockwise Masking	0.749/0.700
	Proposal Masking	**0.945/0.956**

Network Structure

- The ViT architecture can achieve much better detection results than the asymmetric architecture.

Training Masking Strategy

- Our combinational masking strategy can enable the network to learn better reconstruction capabilities, thus achieving better detection results.

Reconstruction Objective

- Raw pixels will result in much worse performance.
- Visual tokens can achieve better results than deep features.

Inference Masking Strategy

- The proposal masking strategy can achieve a significant performance gain, because the suspicious abnormal patches will be masked as much as possible.

Conclusions and Limitations

- **Conclusions:**
  - Class adaptability is a critical but still not well-studied issue in the AD community.
  - We propose a novel PMAD approach based on two key designs: MAE-based patch-level reconstruction and prototype-guided proposal masking.
  - Our model illustrates better class adaptability than SOTA methods under multi- and cross-class settings.
  - Masked AutoEncoder is suitable for multi- and cross-class anomaly detection, and should be exploited more.
- **Limitations:**
  - Our model can only reconstruct 16x16 image patches, but cannot reconstruct more fine-grained image patches.
  - Thus, the anomaly localization ability of our model is limited.
  - Future Work: employ hierarchical transformers and design a multi-scale masking strategy.

SHANGHAI JIAO TONG UNIVERSITY

Thanks!

Contact Us:
sunny_zhang@sjtu.edu.cn