

Paper Review

AMI-Net: Adaptive Mask Inpainting Network for Industrial Anomaly Detection and Localization

YeongHyeon Park

Department of Electrical and Computer Engineering

SungKyunKwan University

IEEE TRANSACTIONS ON AUTOMATION SCIENCE AND ENGINEERING



AMI-Net: Adaptive Mask Inpainting Network for Industrial Anomaly Detection and Localization

Wei Luo^{ID}, *Student Member, IEEE*, Haiming Yao^{ID}, *Graduate Student Member, IEEE*,
Wenyong Yu^{ID}, *Senior Member, IEEE*, and Zhengyong Li^{ID}

Manuscript received 15 December 2023; accepted 17 February 2024

Browse Journals & Magazines > IEEE Transactions on Automatio... [?](#)

IEEE Transactions on Automation Science and Engineering



Submit
Manuscript



Add Title
To My Alerts



Add to
My Favorites



Home

Popular

Early Access

Current Issue

All Issues

About Journal

5.6

Impact
Factor

0.01123

Eigenfactor

1.253

Article
Influence
Score

10.9

CiteScore
Powered by Scopus®



AMI-Net

Adaptive Mask Inpainting Network

Summaries

Motivation and solution

- **Motivation:** Developing adaptive masking network that only masks defective regions
- **Solution:** Learnable clustering token and outlier token masking network (clustering concept)

Contributions

- Investigating and organizing various methods in the field of anomaly detection
 - Reconstruction-based method (w/ various masking methods)
 - Feature embedding-based method
- Proposing clustering-based token masking method
 - Same concept under the names Padim and PatchCore, but...
- ~~The highest performance on MVTec AD dataset~~

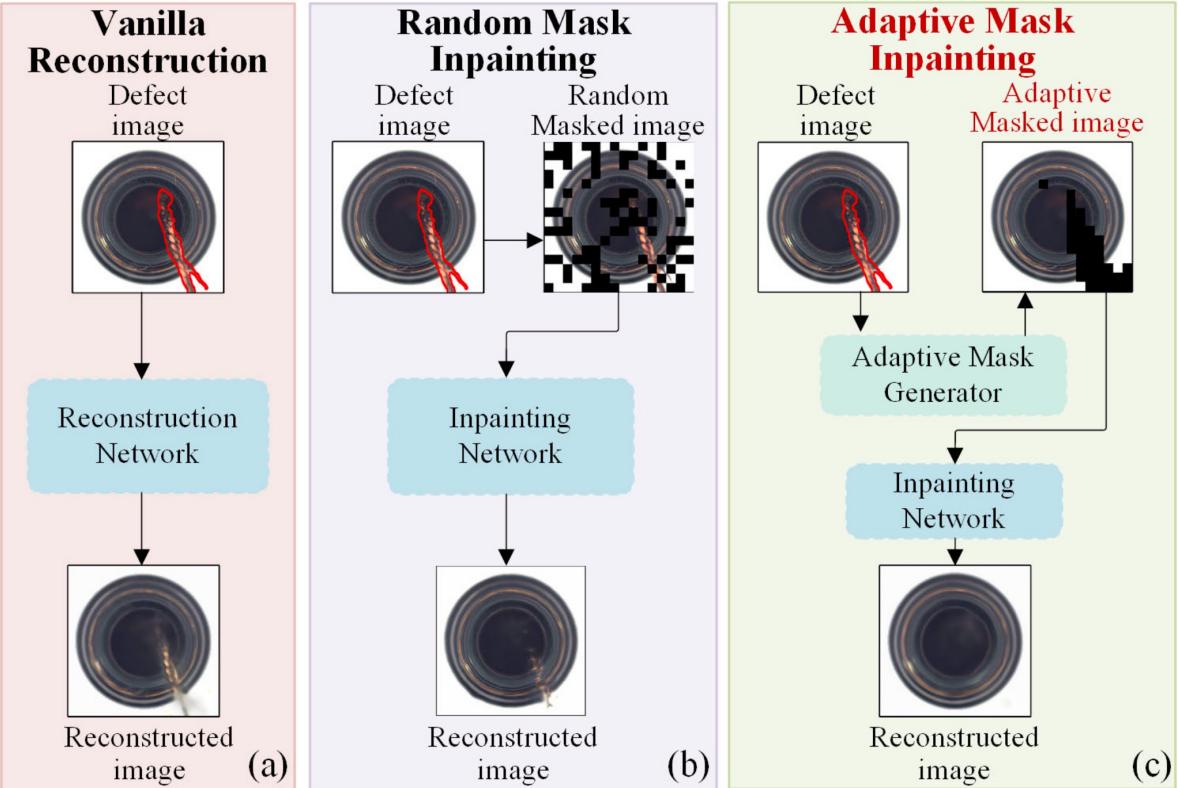
Paper Strengths

- Shows extensive experiments
 - MVTec AD, BTAD datasets
 - AD on single task setting, AD on unified setting
 - Few-shot AD
 - Hyperparameter tuning and analysis (written as ablation study)

Paper Weaknesses

- The structure of the manuscript is somewhat complicated
 - No symbols in the schematic diagrams (hard to follow the explanation)
 - Too many and confusing symbol expressions

Motivation



Vanilla Reconstruction Model

- Well reconstruction of abnormal regions by strong generalization capabilities
- To address this issue mask-based approaches have been proposed

Random Mask Inpainting Model

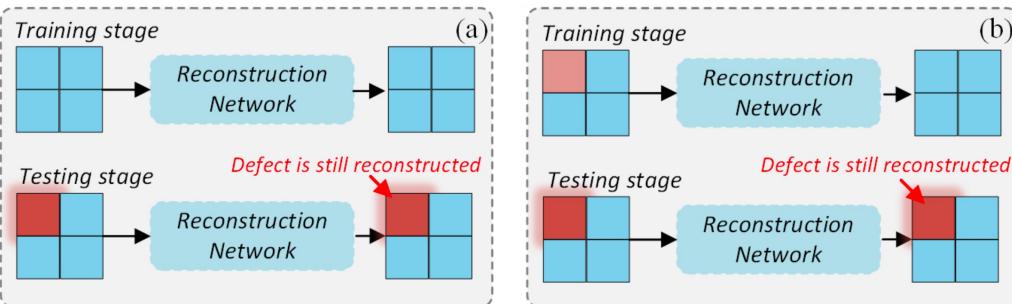
- Disparate detection outcomes by random masking
- Multiple complementary masks cause decrease in the inference speed

Adaptive Mask Inpainting Model

- Avoid randomness during the test process
- Make favorable balance between detection performance and inference speed

Masking Strategies

Normal feature Anomalous feature Artificial anomalous feature Masked feature



(a) Vanilla reconstruction

- Unintended strong generalization

(b) Artificial defect-based method

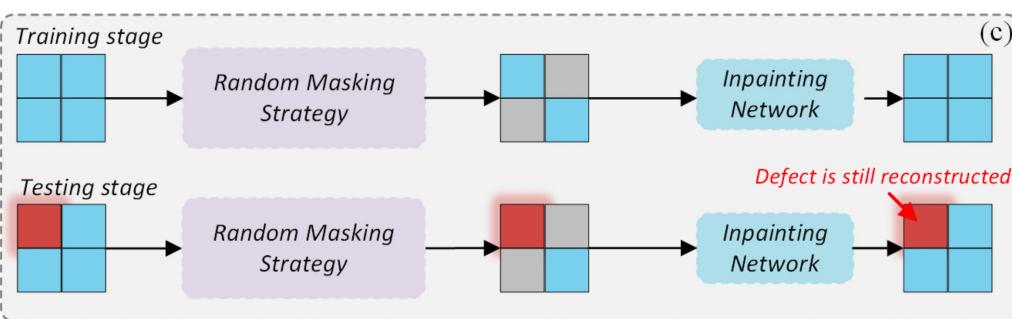
- lack of authenticity in artificial defects

(c) Random masking-based method

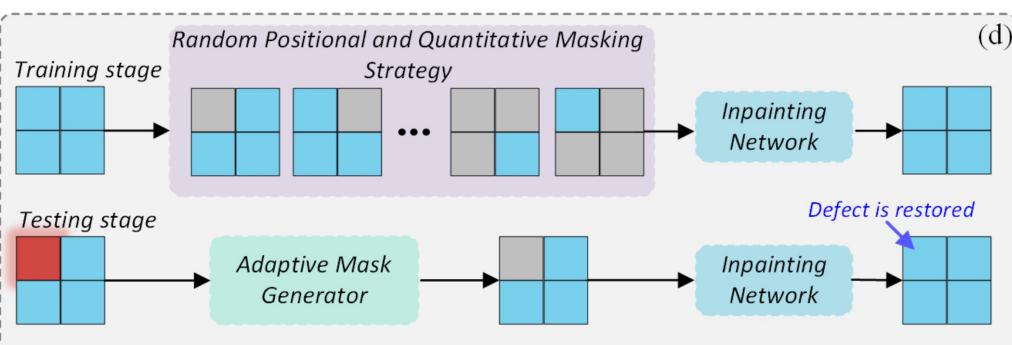
- Disparate detection outcomes

(d) Proposed method (adaptive mask inpainting)

- Effectively concealing all defect regions



(d)



Concept for Adaptive Masking

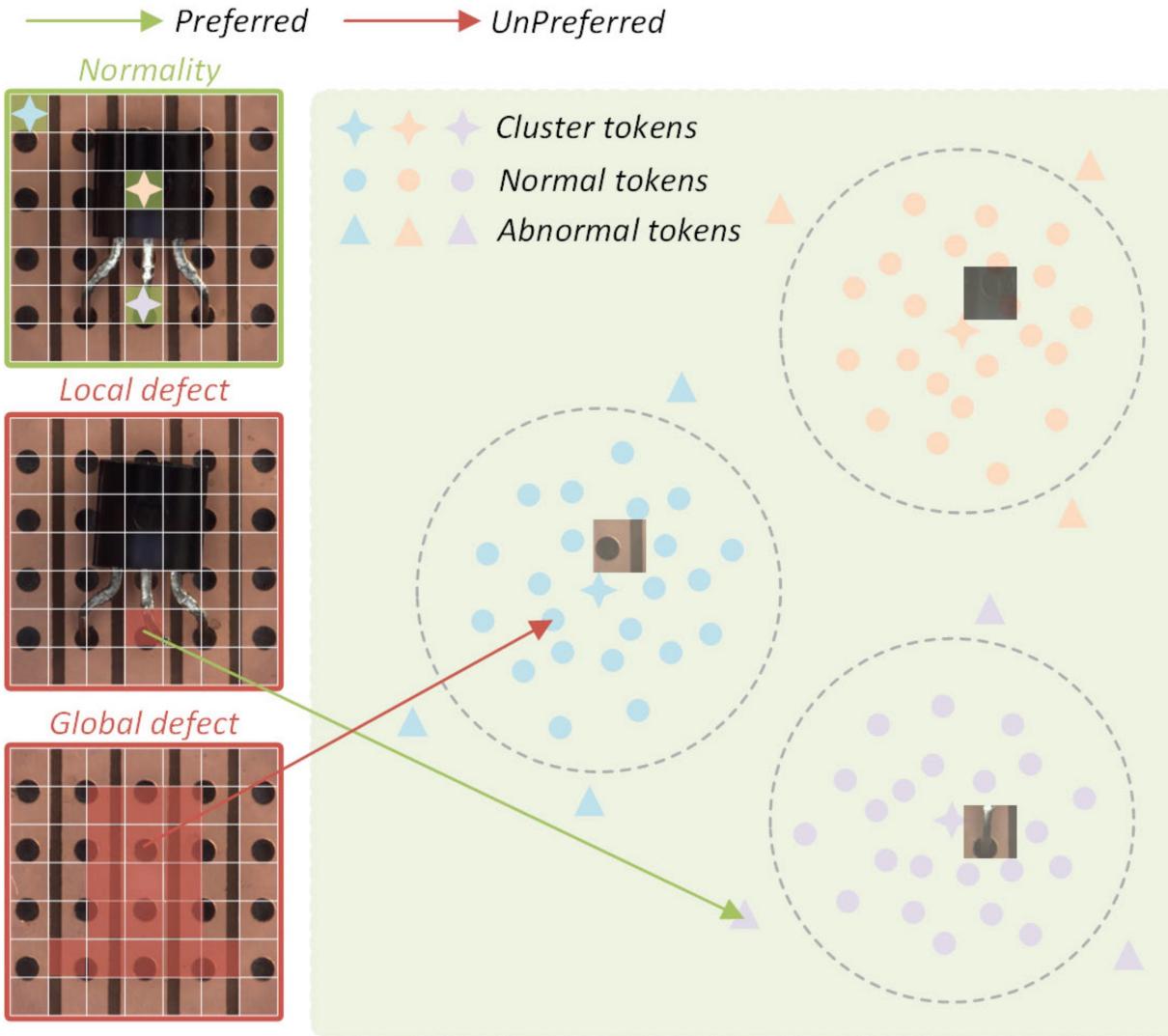
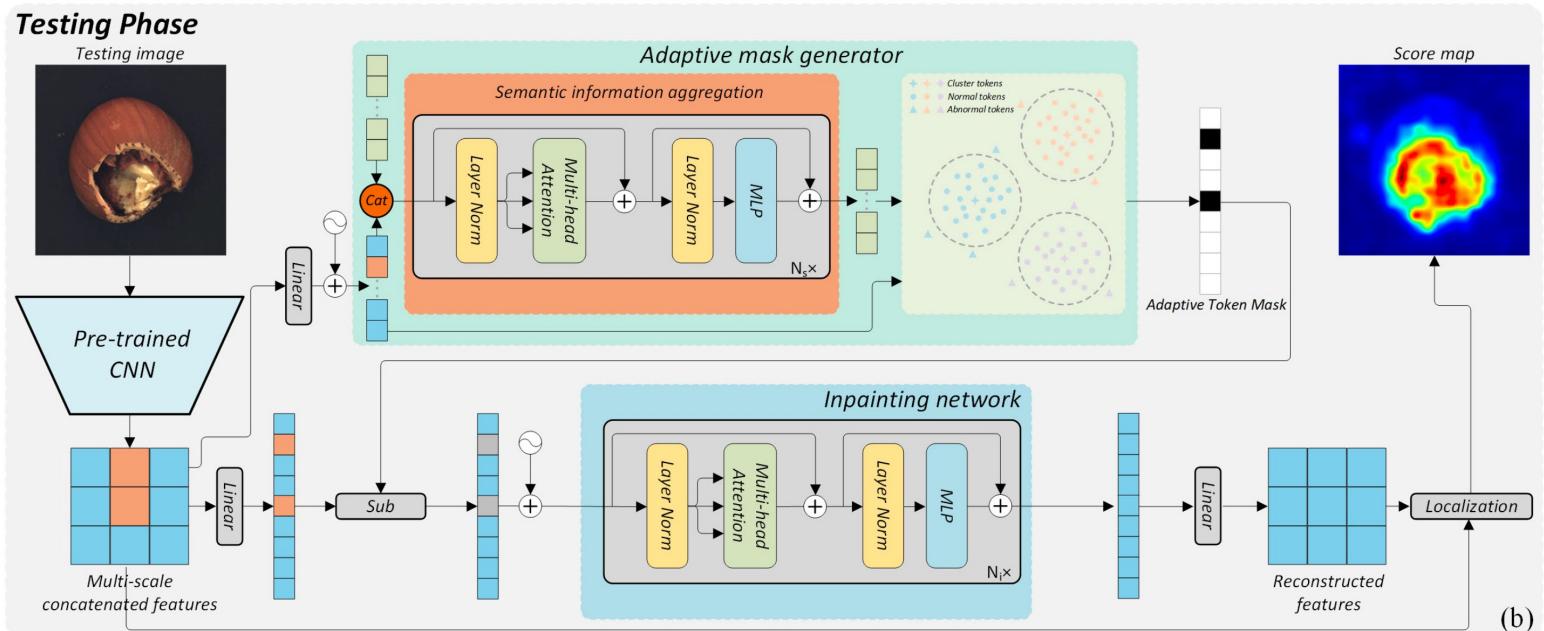
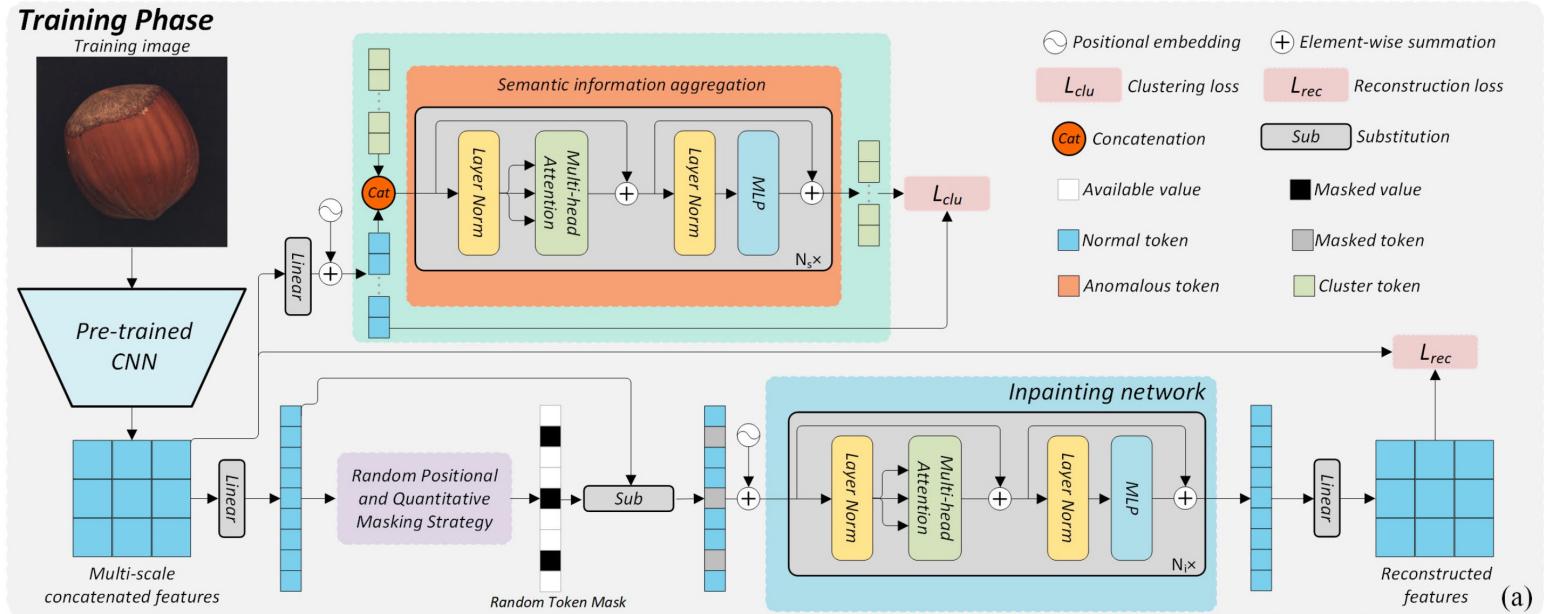


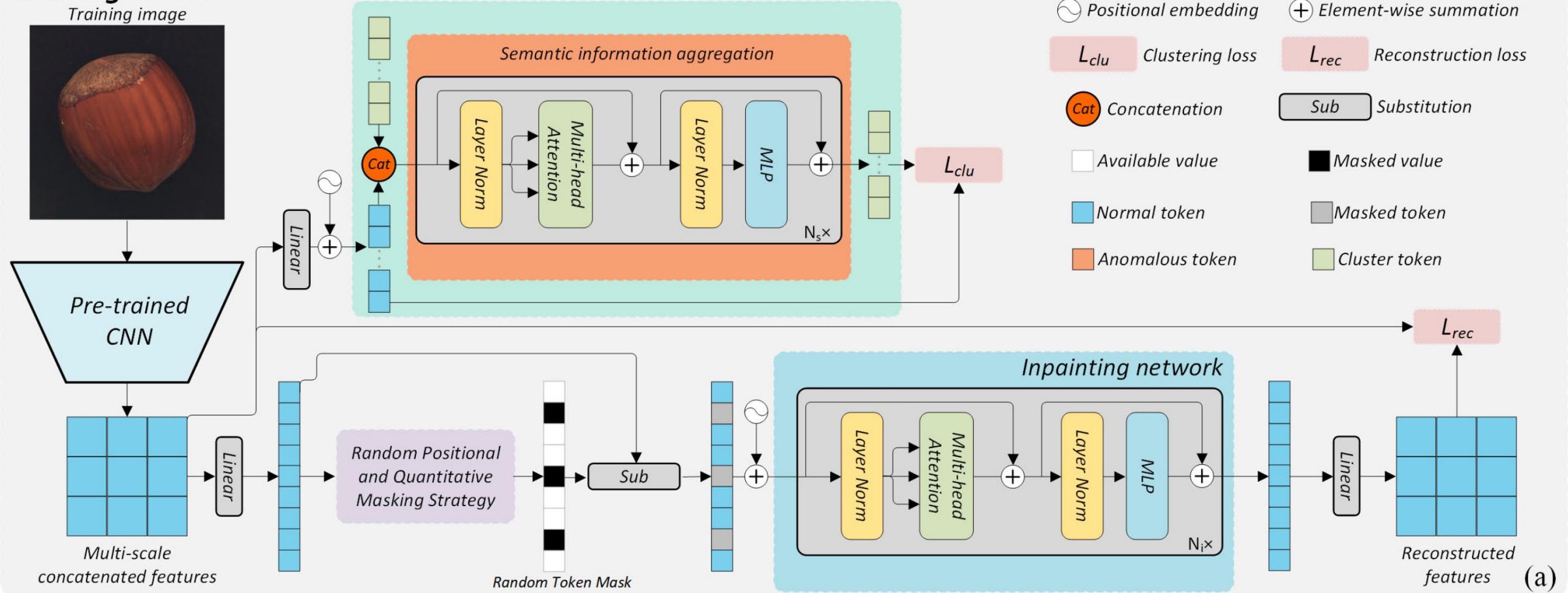
Fig. 5. Issues arising from clustering methods that do not take positional information into account.

Overview



Training [1/7]

Training Phase



Training [2/7]

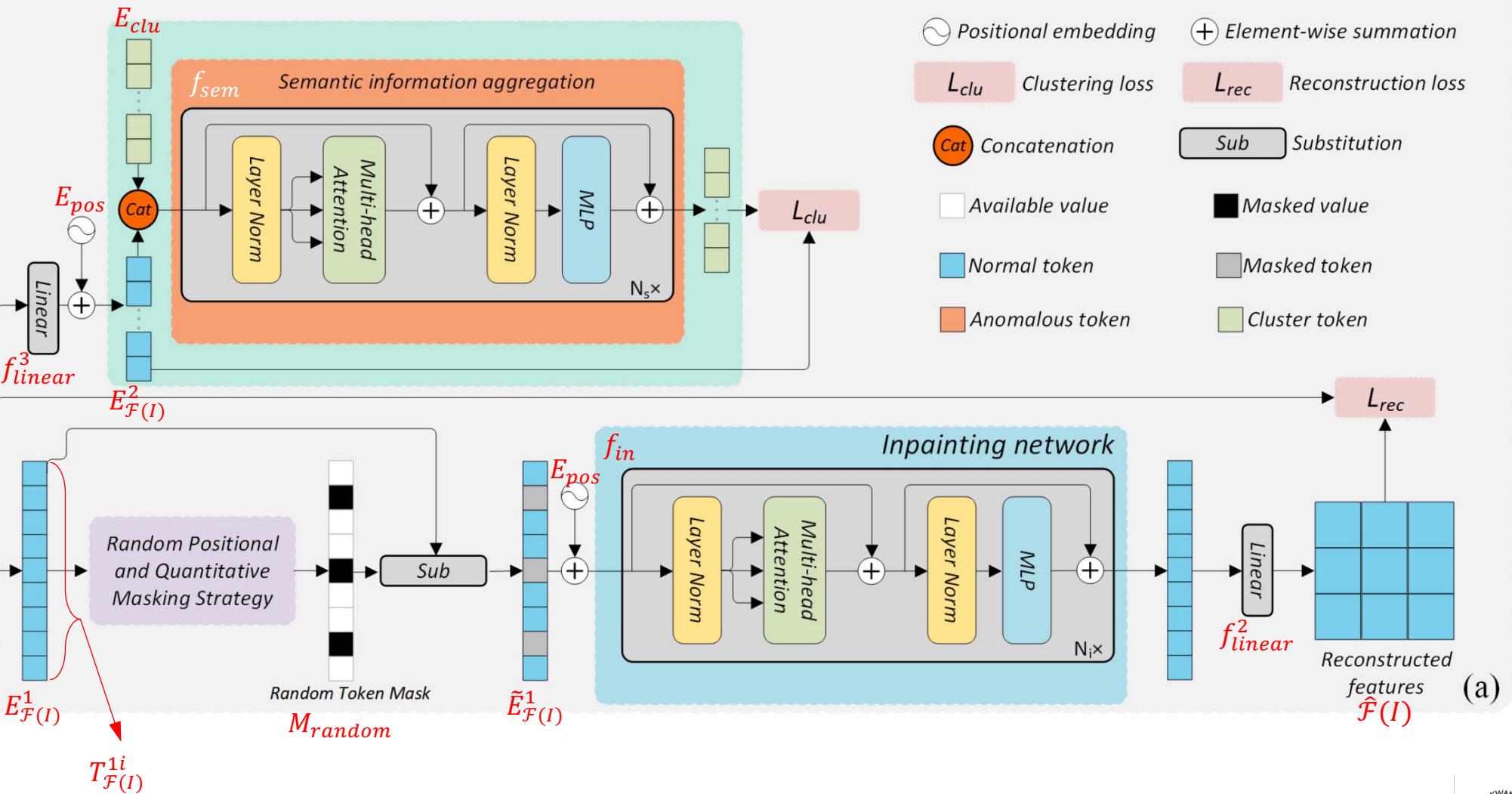
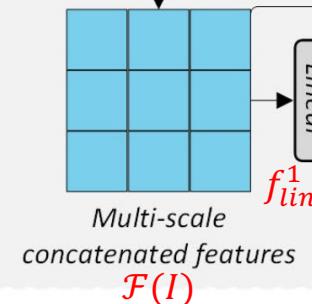
w/ Symbols

Training Phase

Training image

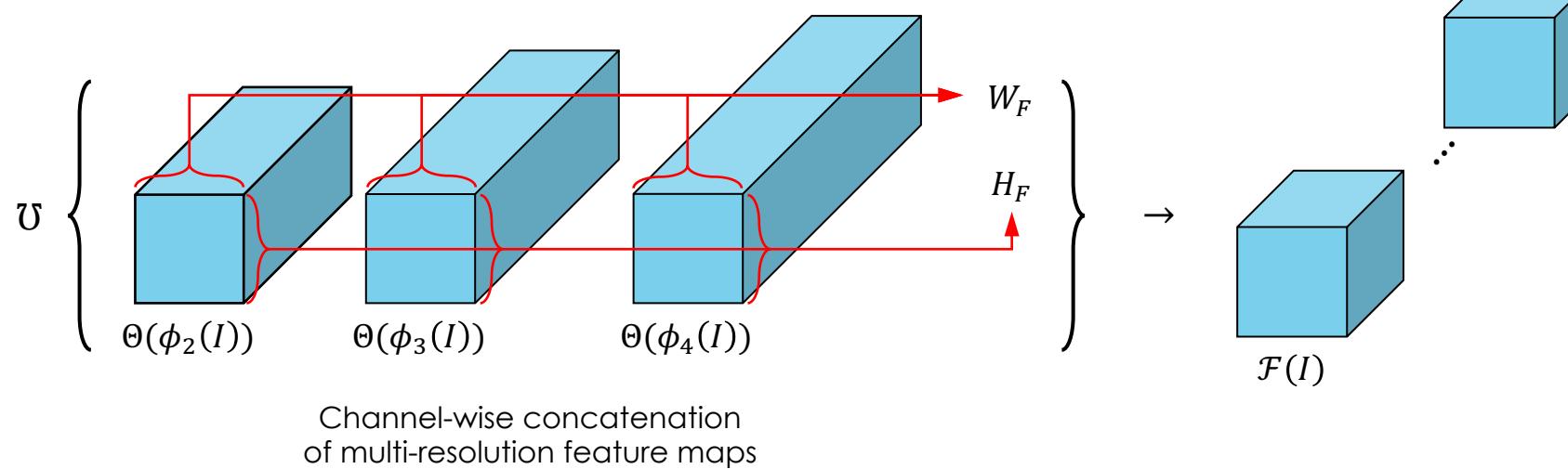
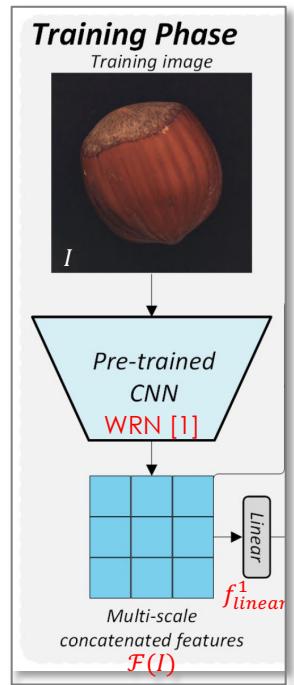
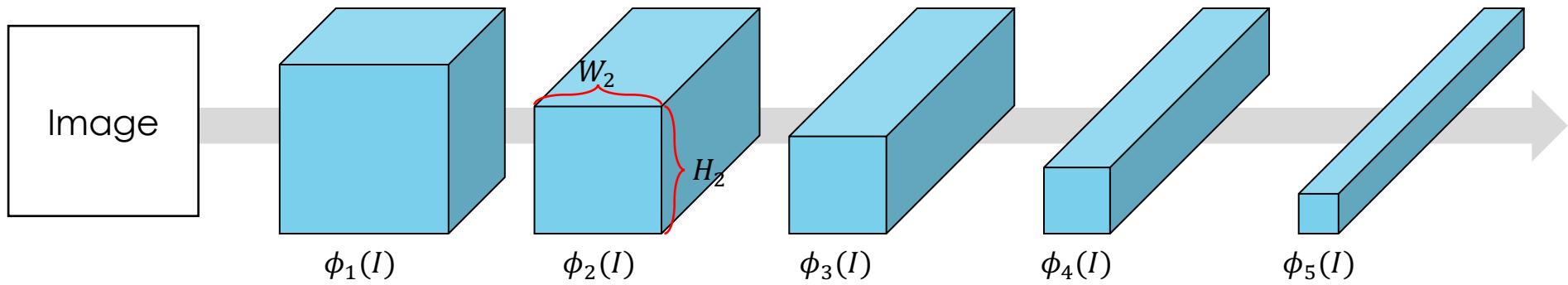


Pre-trained
CNN
WRN [1]



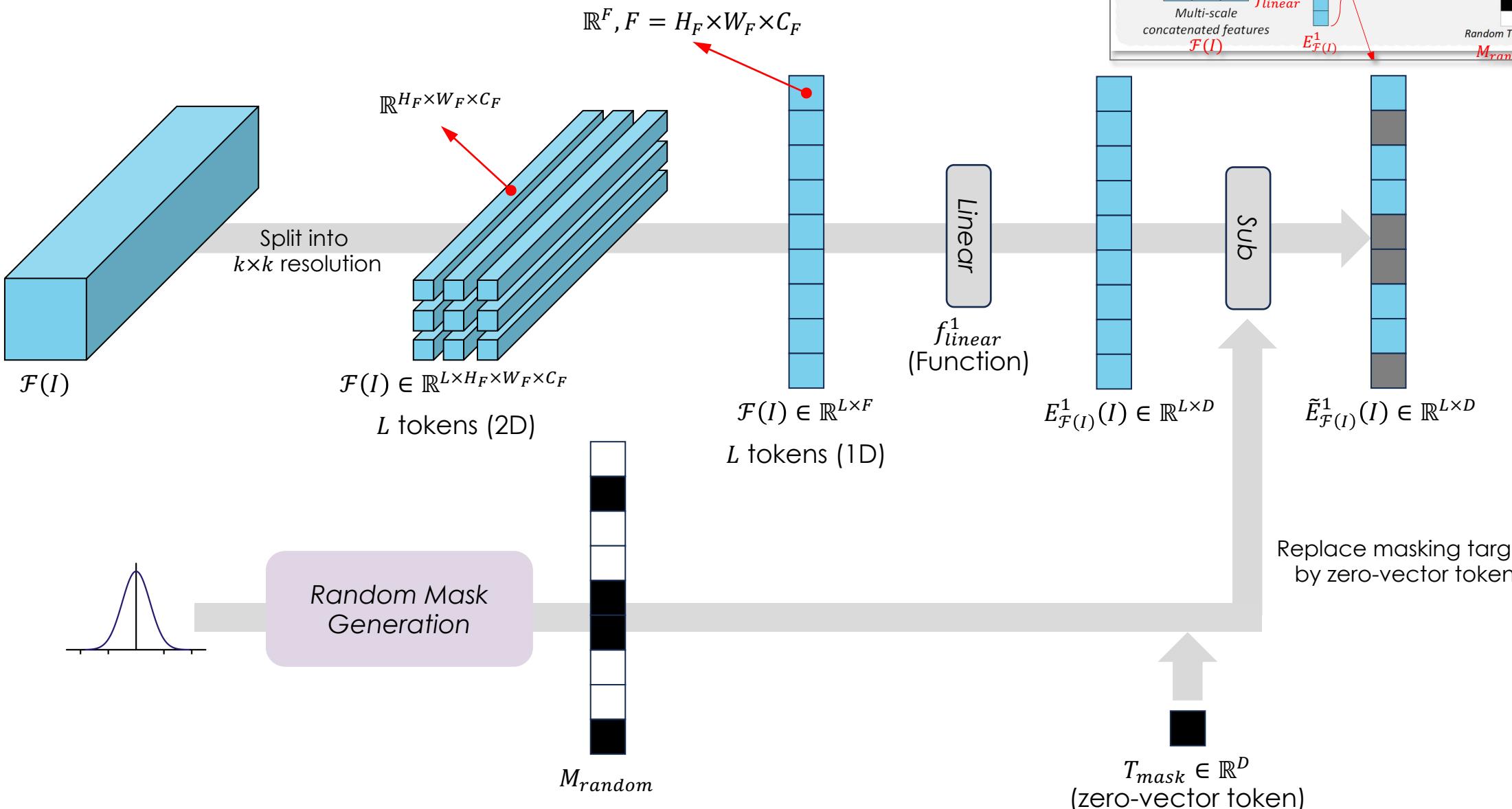
Training [3/7]

Multi-Scale Feature Extraction



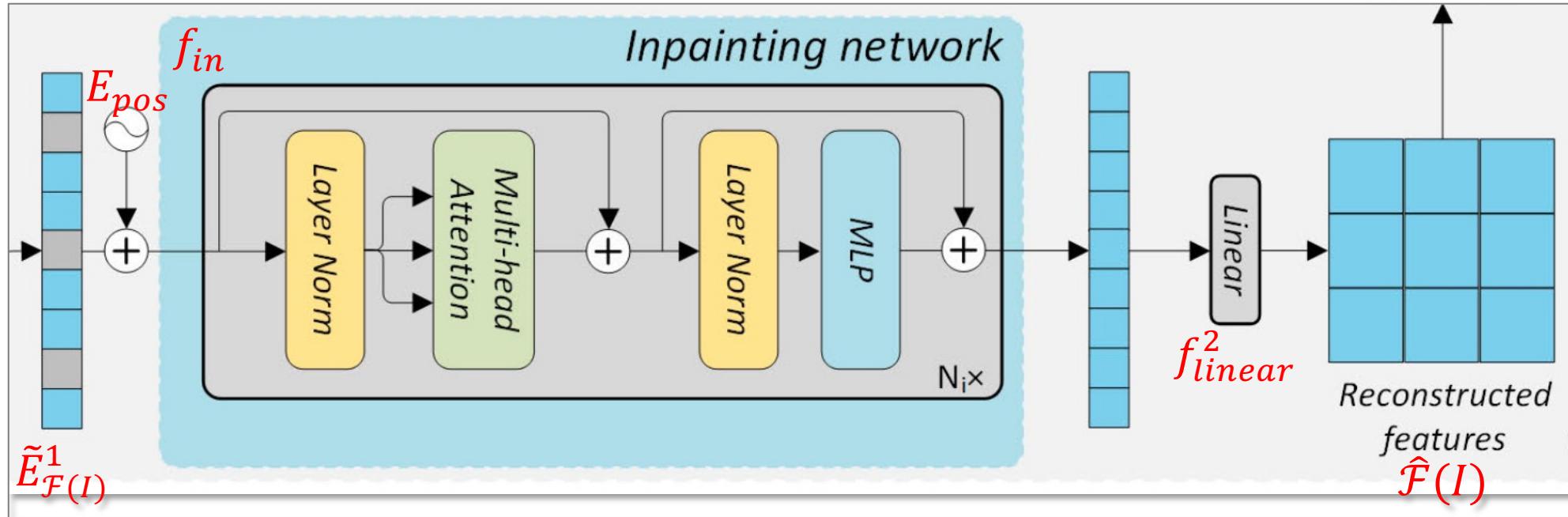
Training [4/7]

Random Positional & Quantitative Masking



Training [5/7]

Feature Inpainting



$$L_{mse} = \mathbb{E}_{P_I} \left[\|\hat{\mathcal{F}}(I) - \mathcal{F}(I)\|^2 \right] \quad (7)$$

$$L_{cos} = \mathbb{E}_{P_I} \left[1 - \frac{\hat{\mathcal{F}}(I) \cdot \mathcal{F}(I)}{\|\hat{\mathcal{F}}(I)\| \times \|\mathcal{F}(I)\|} \right] \quad (8)$$

$$L_{rec} = w_1 L_{mse} + w_2 L_{cos} \quad (9)$$

Training [6/7]

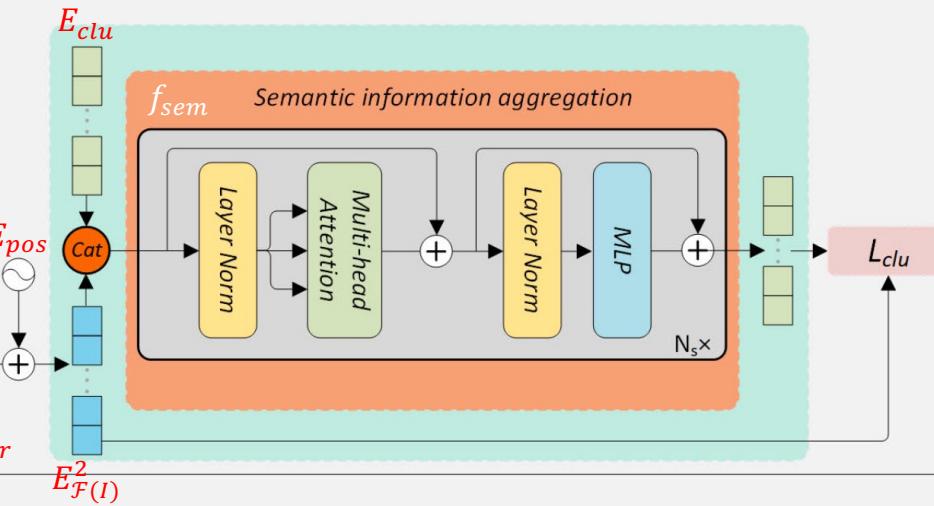
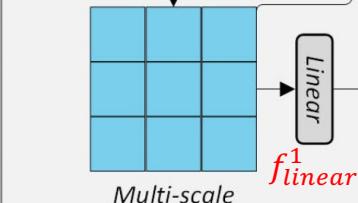
Adaptive Mask Generator

Training Phase

Training image



Pre-trained
CNN
WRN [1]



- Minimizing intra-normal class (each specific local info.) distance trainable cluster token (E_{clu}) and input feature token ($E_{\mathcal{F}(I)}^2$)
- Maximizing inter-normal class (specific local info. and the others) distance

$$L_{clu} = \mathbb{E}_{P_I} \left[w_3 \underbrace{\sum_i d_i}_{intra-class} - w_4 \underbrace{\sum_i \sum_j R(T_{clu}^i, T_{clu}^j)}_{inter-class} \right] \quad (15)$$

$$d_{ij} = \min_{i \in \{1, \dots, P\}} R(T_{clu}^i, T_{\mathcal{F}(I)}^{2j}) \quad (13)$$

$$d_i = \sum_j d_{ij} \quad (14)$$

$$R(T_{clu}^i, T_{\mathcal{F}(I)}^{2j}) = \underbrace{\|T_{clu}^i - T_{\mathcal{F}(I)}^{2j}\|^2}_{Euclidean\ distance} \times \underbrace{(1 - \frac{T_{clu}^i \cdot T_{\mathcal{F}(I)}^{2j}}{\|T_{clu}^i\| \times \|T_{\mathcal{F}(I)}^{2j}\|}}_{Cosine\ similarity} \quad (12)$$

Training [7/7]

$$L_{total} = L_{rec} + L_{clu}$$

Concept for Adaptive Masking

Revisit

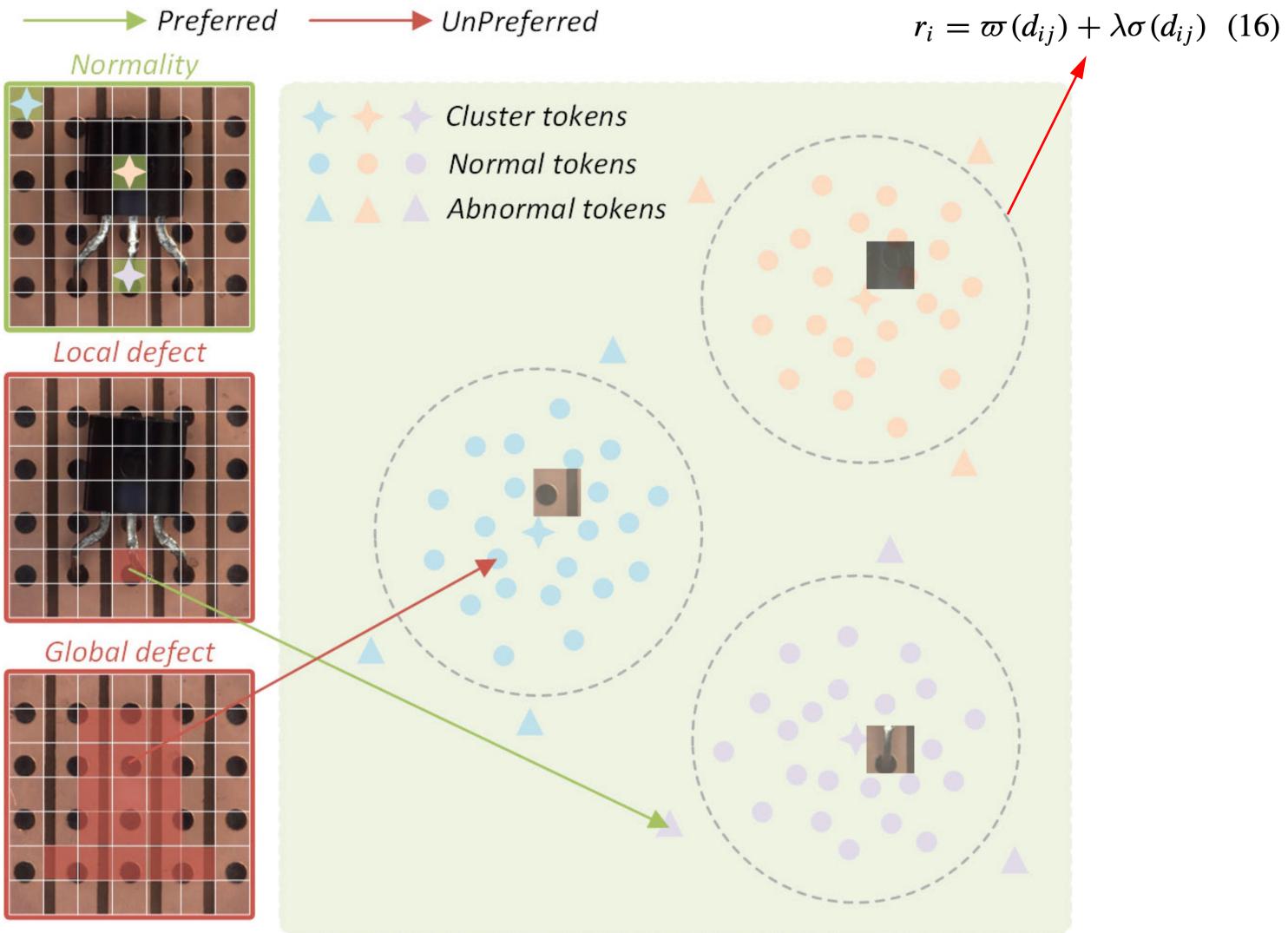
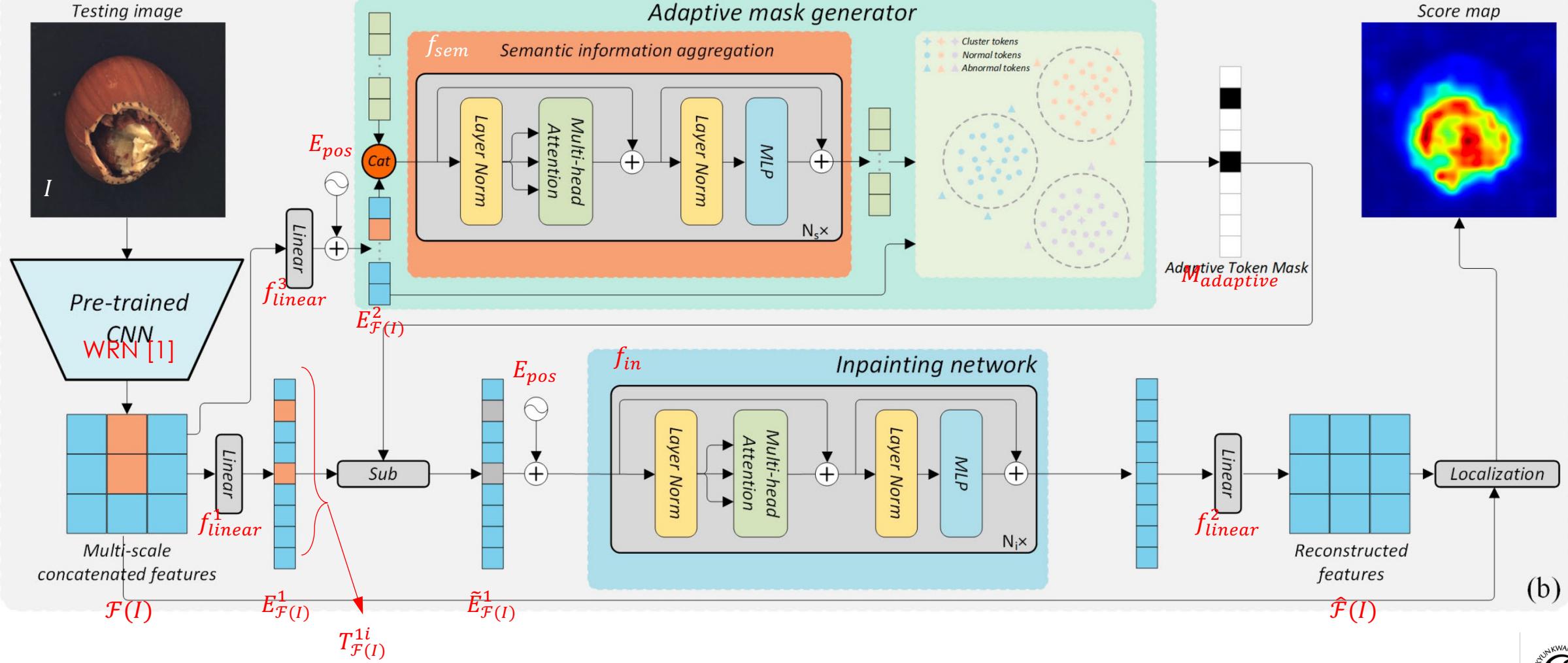


Fig. 5. Issues arising from clustering methods that do not take positional information into account.

Test

Is the input token in out-of-bound for every cluster tokens?
 → It will be masked for inpainting

Testing Phase



Experiments

Anomaly Scoring

Reconstruction loss for training
 $L_{rec} = w_1 L_{mse} + w_2 L_{cos}$

During the testing process, given a test image $I_t \in R^{H \times W \times C}$, AMI-Net employs a pre-trained CNN to extract multi-scale features $\mathcal{F}(I_t)$. Subsequently, the adaptive mask generator generates an adaptive mask for the features. Following this, the masked features are input into the inpainting network in order to obtain reconstruction features $\hat{\mathcal{F}}(I_t)$. Ultimately, the anomaly score map A_s can be generated through the utilization of both the reconstructed features and the input features.

$$A_s = \underbrace{\|\hat{\mathcal{F}}(I_t) - \mathcal{F}(I_t)\|}_{MSE} \times \underbrace{(1 - \frac{\hat{\mathcal{F}}(I_t) \cdot \mathcal{F}(I_t)}{\|\hat{\mathcal{F}}(I_t)\| \times \|\mathcal{F}(I_t)\|})}_{cosine-distance} \quad (18)$$

where $A_s \in R^{H_F \times W_F}$. Then, we employ a bilinear upsampling operation to scale A_s to the size of $H \times W$.

Quantitative Comparisons [1/3]

TABLE I

ANOMALY DETECTION AND LOCALIZATION RESULTS IN TERMS OF IMAGE/PIXEL LEVEL AUROC ON THE MVTEC AD DATASET [13]. [‡] MEANS
OUR TRAINING WITH THE FEATURE JITTERING STRATEGY PROPOSED BY REFERENCE [40]

Almost same

| | Category | AE_SSIM | TrustMAE | RIAD | DFR | DRAEM | PatchCore | MKD | MBPFM | Ours | Ours [‡] |
|-----------------|------------|-----------|------------------|------------------|-----------|------------------|------------------|------------------|------------------|------------------|-------------------|
| Texture | Carpet | 67.0/87.0 | 97.4/98.5 | 84.2/94.2 | -/97.0 | 97.0/95.5 | 98.7/99.0 | 79.3/95.6 | 100/99.2 | <u>99.8/99.2</u> | <u>99.8/99.2</u> |
| | Grid | 69.0/94.0 | 99.1/97.5 | 99.6/96.3 | -/98.0 | 99.9/99.7 | 98.2/98.7 | 78.0/91.8 | 98.0/98.8 | <u>100/98.8</u> | <u>99.9/98.9</u> |
| | Leather | 46.0/78.0 | <u>95.1/98.1</u> | 100/99.4 | -/98.0 | 100/98.6 | 100/99.3 | <u>95.1/98.1</u> | 100/99.4 | <u>100/99.3</u> | <u>100/99.4</u> |
| | Tile | 52.0/59.0 | 97.3/82.5 | 98.7/89.1 | -/87.0 | 99.6/99.2 | 98.7/95.6 | 91.6/82.8 | <u>99.6/96.2</u> | <u>99.9/95.9</u> | <u>100/96.0</u> |
| | Wood | 83.0/73.0 | 99.8/92.6 | 93.0/85.8 | -/94.0 | 99.1/96.4 | 99.2/95.0 | 94.3/84.8 | <u>99.5/95.6</u> | 99.4/94.8 | 99.3/95.3 |
| Average Texture | | 63.4/78.2 | 97.7/93.8 | 95.1/93.9 | -/94.8 | 99.1/97.9 | 99.0/97.5 | 87.7/90.6 | <u>99.4/97.8</u> | 99.8/97.6 | 99.8/97.8 |
| Object | Bottle | 88.0/93.0 | 97.0/93.4 | <u>99.9/98.4</u> | -/97.0 | 99.2/99.1 | 100/98.6 | 99.4/96.3 | 100/98.4 | <u>100/98.7</u> | <u>100/98.8</u> |
| | Cable | 61.0/82.0 | 85.1/92.9 | <u>81.9/84.2</u> | -/92.0 | 91.8/94.7 | 99.5/98.4 | 89.2/82.4 | 98.8/96.7 | <u>99.1/98.1</u> | <u>99.5/98.6</u> |
| | Capsule | 61.0/94.0 | 78.8/87.4 | 88.4/92.8 | -/99.0 | 98.5/94.3 | 98.1/98.8 | 80.5/95.9 | 94.5/98.3 | 95.7/98.6 | <u>98.4/98.9</u> |
| | Hazelnut | 54.0/97.0 | 98.5/98.5 | 83.3/96.1 | -/99.0 | 100/99.7 | 100/98.7 | 98.4/94.6 | 100/99.1 | <u>99.9/98.3</u> | <u>100/98.6</u> |
| | Metal nut | 54.0/89.0 | 76.1/91.8 | 88.5/92.5 | -/93.0 | 98.7/99.5 | 100/98.4 | 73.6/86.4 | 100/97.2 | 99.2/95.3 | <u>99.8/96.5</u> |
| | Pill | 60.0/91.0 | 83.3/89.9 | 83.8/95.7 | -/97.0 | 98.9/97.6 | 96.6/97.4 | 82.7/89.6 | 96.5/97.2 | 95.9/97.7 | <u>96.0/98.4</u> |
| | Screw | 51.0/96.0 | 83.4/97.6 | 84.5/98.8 | -/99.0 | 93.9/97.6 | 98.1/99.4 | 83.3/96.0 | 91.8/98.7 | 97.1/99.0 | <u>97.9/99.4</u> |
| | Toothbrush | 74.0/92.0 | <u>96.9/98.1</u> | 100/98.9 | -/98.1 | 100/98.1 | 100/98.7 | 92.2/96.1 | 88.6/98.6 | 93.6/98.8 | <u>96.1/98.9</u> |
| | Transistor | 52.0/90.0 | <u>87.5/92.7</u> | 90.9/87.7 | -/80.0 | 93.1/90.9 | 100/96.3 | 85.6/76.5 | <u>97.8/87.8</u> | 100/96.7 | <u>100/98.2</u> |
| | Zipper | 80.0/88.0 | 87.5/97.8 | 98.1/97.8 | -/96.0 | 100/98.8 | 99.4/98.5 | 93.2/93.9 | 97.4/98.2 | 97.8/98.2 | <u>98.5/98.5</u> |
| Average Object | | 63.5/91.2 | 87.4/94.0 | 89.9/94.3 | -/95.0 | 97.4/97.0 | 99.2/98.3 | 87.8/90.8 | 96.5/97.0 | 97.8/97.9 | <u>98.6/98.5</u> |
| Average All | | 63.5/87.0 | 90.9/94.0 | 91.7/94.2 | 93.8/95.5 | 98.0/97.3 | 99.1/98.1 | 87.7/90.7 | 97.5/97.3 | 98.5/97.8 | <u>99.0/98.2</u> |

¹ The best result is in **bold**, and the second best is underlined.

WRN50-based feat. ext.

Quantitative Comparisons [2/3]

TABLE III

ANOMALY DETECTION AND LOCALIZATION RESULTS IN TERMS OF IMAGE/PIXEL LEVEL AUROC ON THE MVTEC AD DATASET [13]. ALL METHODS ARE EVALUATED UNDER THE UNIFIED (ONE FOR ALL) CASE. IN THE UNIFIED CASE, THE LEARNED MODEL IS APPLIED TO DETECT ANOMALIES FOR ALL CATEGORIES WITHOUT FINE-TUNING. THE EXPERIMENTAL RESULTS OF OTHER SUPERIOR METHODS ARE SOURCED FROM REFERENCE [53]. [‡] MEANS OUR TRAINING WITH THE FEATURE JITTERING STRATEGY PROPOSED BY REFERENCE [40]

| | Category | US | Patch SVDD | PaDiM | MKD | DRAEM | SimpleNet | PatchCore | <u>UniAD</u> | Ours | Ours [‡] |
|-----------------|------------|-----------|------------------|------------------|-----------|------------------|------------------|------------------|------------------|------------------|-------------------|
| Texture | Carpet | 86.6/88.7 | 63.3/78.6 | 93.8/97.6 | 69.8/95.5 | <u>98.0/98.6</u> | 95.9/92.4 | 97.0/98.1 | 99.8/98.5 | 97.5/98.2 | 97.9/98.6 |
| | Grid | 69.2/64.5 | 66.0/70.8 | 73.9/71.0 | 83.8/82.3 | 99.3/98.7 | 49.8/46.7 | 91.4/98.4 | <u>98.2/96.5</u> | 95.1/94.1 | <u>99.2/98.5</u> |
| | Leather | 97.2/95.4 | 60.8/93.5 | <u>99.9/84.8</u> | 93.6/96.7 | 98.7/97.3 | 93.9/96.9 | 100/99.2 | 100/98.8 | 100/98.9 | <u>100/98.8</u> |
| | Tile | 93.7/82.7 | 88.3/92.1 | 93.3/80.5 | 89.5/85.3 | 99.8/98.0 | 93.7/93.1 | 96.0/90.3 | 99.3/91.8 | 98.7/91.9 | <u>99.5/94.3</u> |
| | Wood | 90.6/83.3 | 72.1/80.7 | 98.4/89.1 | 93.4/80.5 | <u>99.8/96.0</u> | 95.2/84.8 | 93.8/90.8 | 98.6/93.2 | 99.6/92.1 | 100/93.3 |
| Average Texture | | 87.5/82.9 | 70.1/83.1 | 91.9/84.6 | 86.0/88.1 | <u>99.1/97.7</u> | 85.7/82.8 | 95.6/95.4 | <u>99.2/95.8</u> | 98.2/95.0 | 99.3/96.7 |
| Object | Bottle | 84.0/67.9 | 85.5/86.7 | 97.9/96.1 | 98.7/91.8 | 97.5/87.6 | 97.7/91.2 | 100/97.4 | <u>99.7/98.1</u> | 100/97.8 | 100/98.4 |
| | Cable | 60.0/78.3 | 64.4/62.2 | 70.9/81.0 | 78.2/89.3 | 57.8/71.3 | 87.6/88.1 | 95.3/93.6 | <u>95.2/97.3</u> | <u>96.6/94.9</u> | 98.7/98.2 |
| | Capsule | 57.6/85.5 | 61.3/83.1 | 73.4/96.9 | 68.3/88.3 | 65.3/50.5 | 78.3/89.7 | 96.8/98.0 | <u>86.9/98.5</u> | 83.1/98.2 | 94.1/98.9 |
| | Hazelnut | 95.8/93.7 | 83.9/97.4 | 85.5/96.3 | 97.1/91.2 | 93.7/96.9 | 99.2/95.7 | 99.3/97.6 | 99.8/98.1 | <u>99.6/97.0</u> | 99.8/98.0 |
| | Metal nut | 62.7/76.6 | 80.9/96.0 | 88.0/84.8 | 64.9/64.2 | 72.8/62.2 | 85.1/90.9 | <u>99.1/96.3</u> | 99.2/94.8 | 98.6/92.3 | <u>98.5/96.0</u> |
| | Pill | 56.1/80.3 | <u>89.4/96.5</u> | 68.8/87.7 | 79.7/69.7 | 82.2/94.4 | 78.3/89.7 | 86.4/90.8 | 93.7/95.0 | 91.9/95.4 | <u>93.6/97.4</u> |
| | Screw | 66.9/90.8 | 80.9/74.3 | 56.9/94.1 | 75.6/92.1 | <u>92.0/95.5</u> | 45.5/93.7 | 94.2/98.9 | <u>87.5/98.3</u> | 80.0/97.1 | 83.7/98.9 |
| | Toothbrush | 57.8/86.9 | <u>99.4/98.0</u> | 95.3/95.6 | 75.3/88.9 | <u>90.6/97.7</u> | 94.7/97.5 | 100/98.8 | <u>94.2/98.4</u> | 97.2/97.7 | <u>95.3/98.7</u> |
| | Transistor | 61.0/68.3 | <u>77.5/78.5</u> | 86.6/92.3 | 73.4/71.7 | 74.8/64.5 | 82.0/86.0 | 98.9/92.3 | 99.8/97.9 | 98.6/90.5 | <u>99.5/95.5</u> |
| | Zipper | 78.6/84.2 | 77.8/95.1 | 79.7/94.8 | 87.4/86.1 | <u>98.8/98.3</u> | 99.1/97.0 | 97.1/95.7 | 95.8/96.8 | 97.7/97.8 | 97.5/98.3 |
| Average Object | | 68.1/81.3 | 80.1/86.8 | 80.3/92.0 | 79.9/83.3 | 82.6/81.9 | 84.8/92.0 | 96.7/95.9 | <u>95.2/97.3</u> | 94.3/95.8 | 96.1/97.8 |
| Average All | | 74.5/81.8 | 76.8/85.6 | 84.2/89.5 | 81.9/84.9 | 88.1/87.2 | 85.1/88.9 | <u>96.4/95.7</u> | <u>96.5/96.8</u> | 95.6/95.6 | 97.2/97.5 |

¹ The best result is in **bold**, and the second best is underlined.

Transformer backbones

WRN50-based feat. ext.

Quantitative Comparisons [3/3]

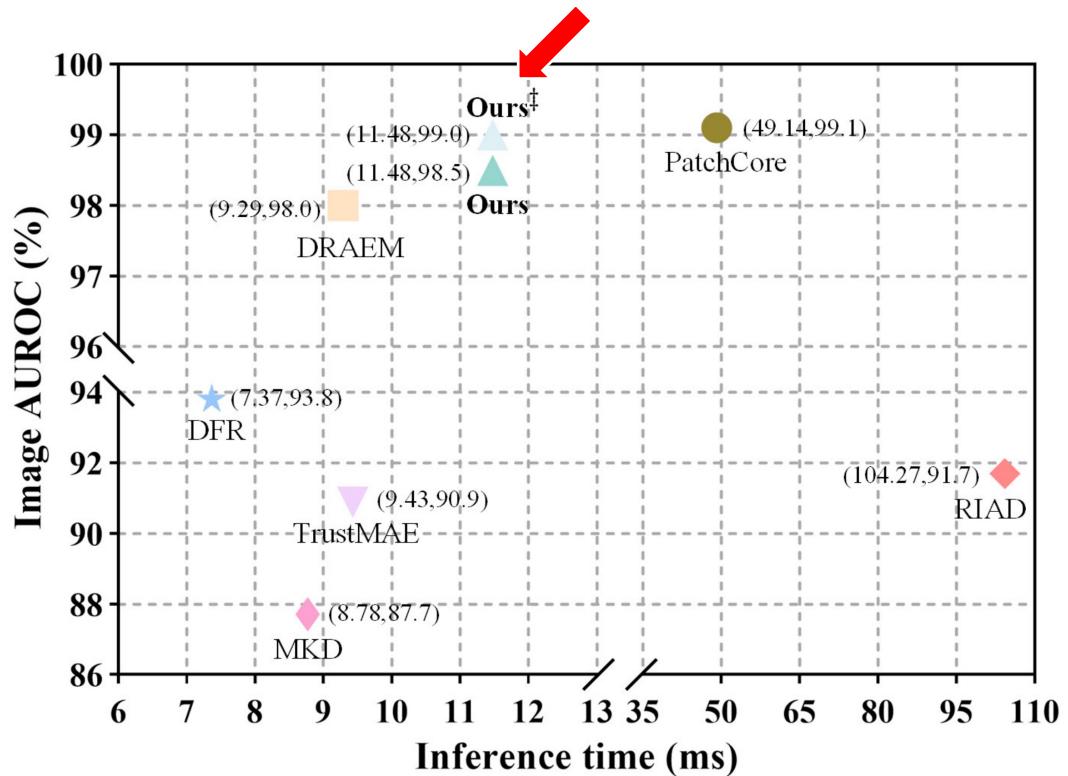


Fig. 8. Inference time versus Image level AUROC on MVTec AD dataset [13]. (·, ·) denotes (Inference time, Image AUROC).

TABLE II

ANOMALY DETECTION AND LOCALIZATION RESULTS IN TERMS OF IMAGE/PIXEL LEVEL AUROC ON THE BTAD DATASET [14]. THE EXPERIMENTAL RESULTS OF OTHER SUPERIOR METHODS ARE SOURCED FROM REFERENCES [41] AND [42]. [‡] MEANS OUR TRAINING WITH THE FEATURE JITTERING STRATEGY PROPOSED BY REFERENCE [40]

| Methods | Class 01 | Class 02 | Class 03 | Average |
|-------------------------|------------------|------------------|------------------|------------------|
| VT-ADL | 97.6/99.0 | 71.0/94.0 | 82.6/77.0 | 83.7/90.0 |
| Patch SVDD | 95.7/91.6 | 72.1/93.6 | 82.1/91.0 | 83.3/92.1 |
| SPADE | 91.4/97.3 | 71.4/94.4 | <u>99.9/99.1</u> | 87.6/96.9 |
| PatchCore | 90.9/95.5 | 79.3/94.7 | 99.8/99.3 | 90.0/96.5 |
| FastFlow | 99.4/97.1 | 82.4/93.6 | 91.1/98.3 | 90.1/96.3 |
| CFA | 98.1/95.9 | 85.5/96.0 | 99.0/98.6 | 94.2/96.8 |
| Ours | 99.9/96.7 | <u>85.4/95.2</u> | 100/99.5 | 95.1/97.1 |
| Ours[‡] | 99.9/96.8 | <u>85.4/96.0</u> | 100/99.6 | 95.1/97.5 |

¹ The best result is in **bold**, and the second best is underlined.

TABLE IV

IMAGE/PIXEL LEVEL AUROC OF K-SHOT ANOMALY DETECTION ON THE MVTec AD DATASET [13]. THE RESULTS ARE AVERAGED OVER ALL CATEGORIES. [‡] MEANS OUR TRAINING WITH THE FEATURE JITTERING STRATEGY PROPOSED BY REFERENCE [40]

| Dataset | k | TDG | DiffNet | RegAD | Ours | Ours[‡] |
|---------|---|--------|---------|------------------|-------------|-------------------------|
| MVTec | 2 | 71.2/- | 80.6/- | 85.7/94.6 | 84.2/94.4 | <u>84.6/94.8</u> |
| | 4 | 72.7/- | 81.3/- | 88.2/95.8 | 86.7/95.5 | <u>86.9/95.8</u> |
| | 8 | 75.2/- | 82.3/- | 91.2/96.7 | 90.9/96.6 | <u>90.9/96.8</u> |

¹ The best result is in **bold**, and the second best is underlined.

Qualitative Evaluation [1/2]

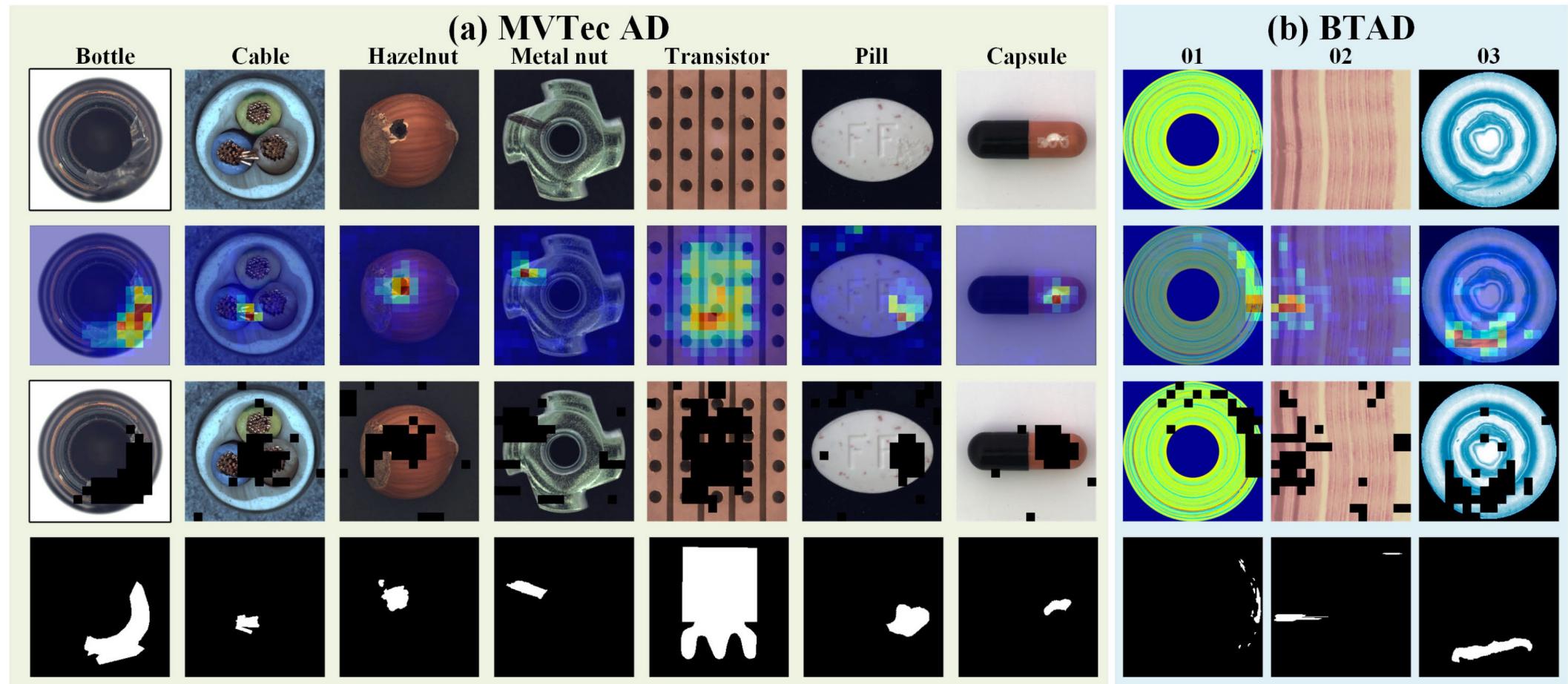


Fig. 4. Examples of the effectiveness of the adaptive mask generator. First Row: the defective image. Second Row: the distance map formed by the distance between latent feature and their corresponding cluster centers. Third Row: the adaptive mask. Final Row: the corresponding label.

Qualitative Evaluation [2/2]

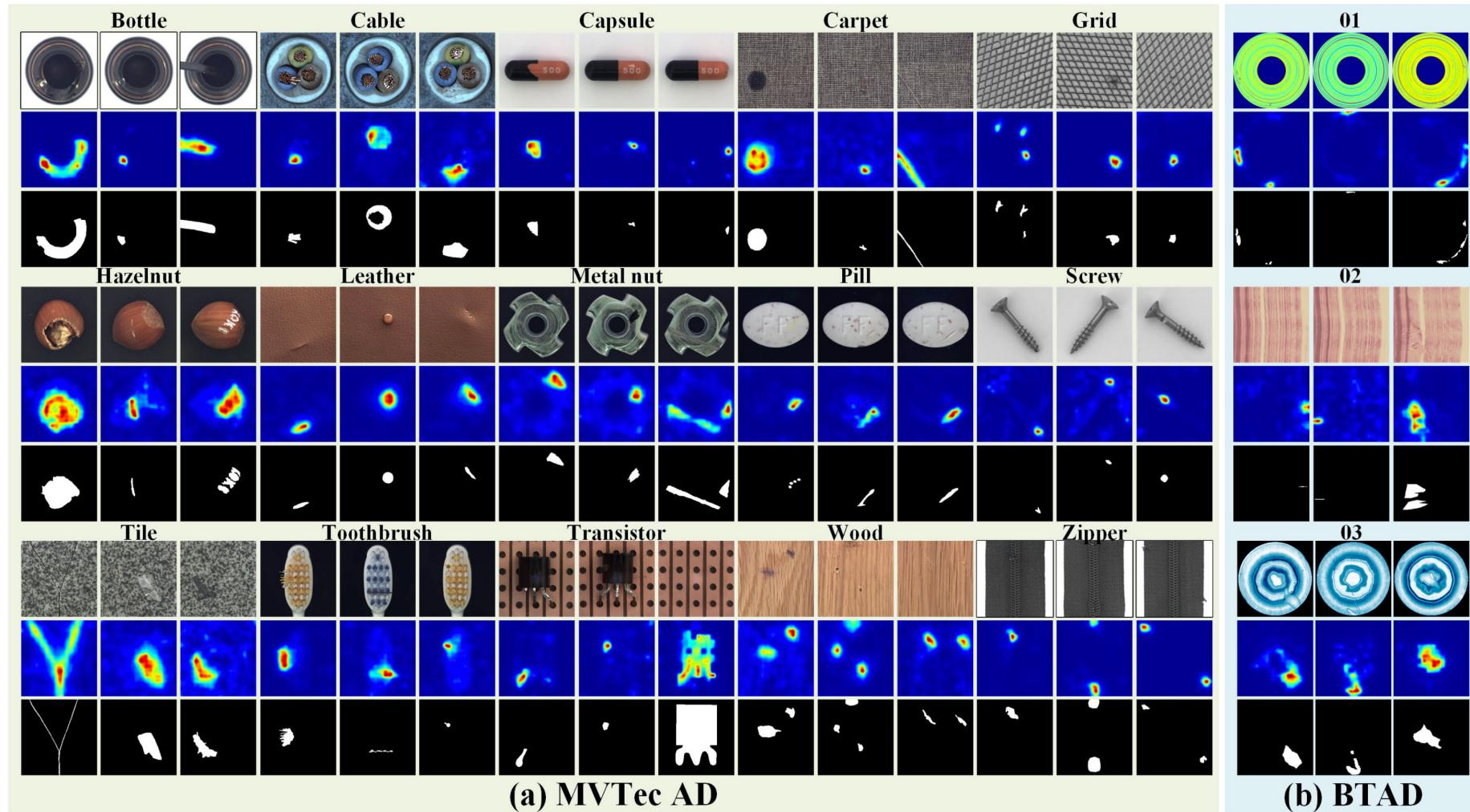


Fig. 7. Localization Results of AMI-Net on MVTec AD [13] and BTAD [14]. For each set, from top to bottom, there are the defect image, detection heat map, and corresponding label. (a) Competitive detection results across 15 categories in MVTec AD. (b) Competitive detection results across 3 categories in BTAD.

Ablation Study

a.k.a hyperparameter tuning

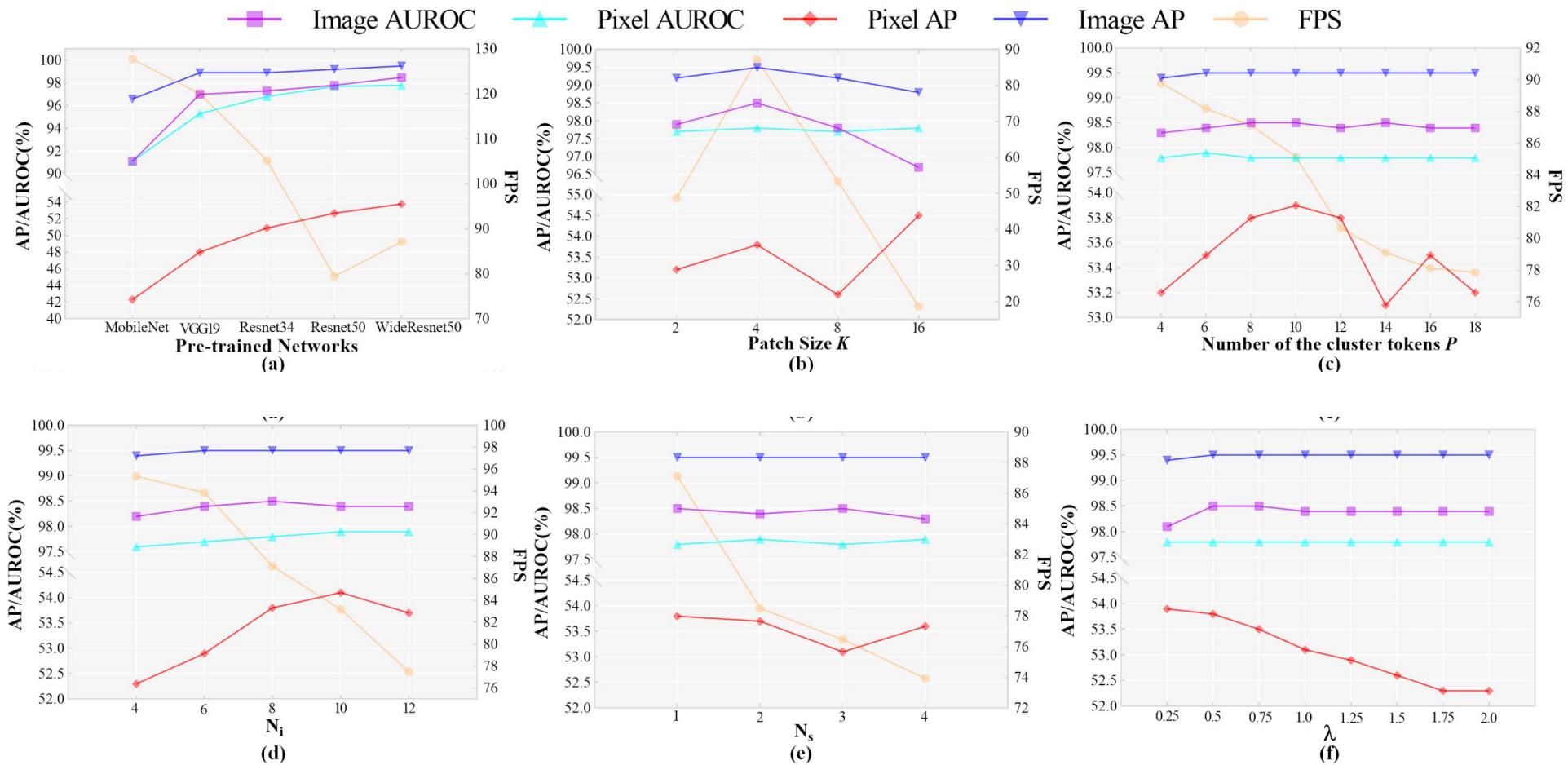


Fig. 9. The ablation experiment results. (a) Influence of pre-trained network. (b) Impact of patch size K . (c) Influence of the number of cluster tokens P . (d) Influence of the number of transformer block in inpainting network N_i . (e) Influence of the number of transformer block in semantic information aggregation network N_s . (f) Influence of the scaling factor λ .

Limitations

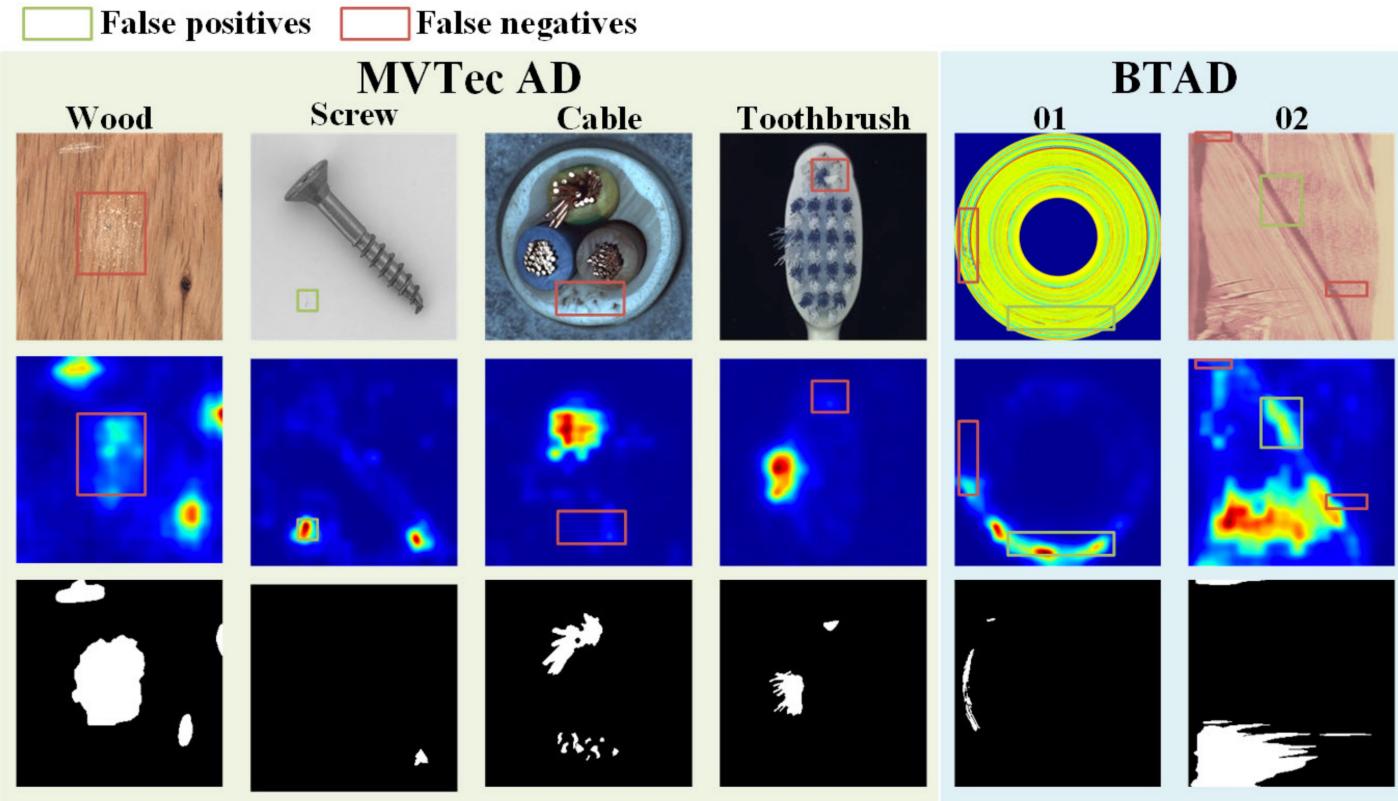


Fig. 11. Some instances of detection failures by AMI-Net. Top Row: the defective sample. Mid Row: The detection result. Bottom Row: The corresponding label.

AMI-Net exhibits reduced capabilities in detecting subtle defects and those with low contrast.

Appendix

WRN

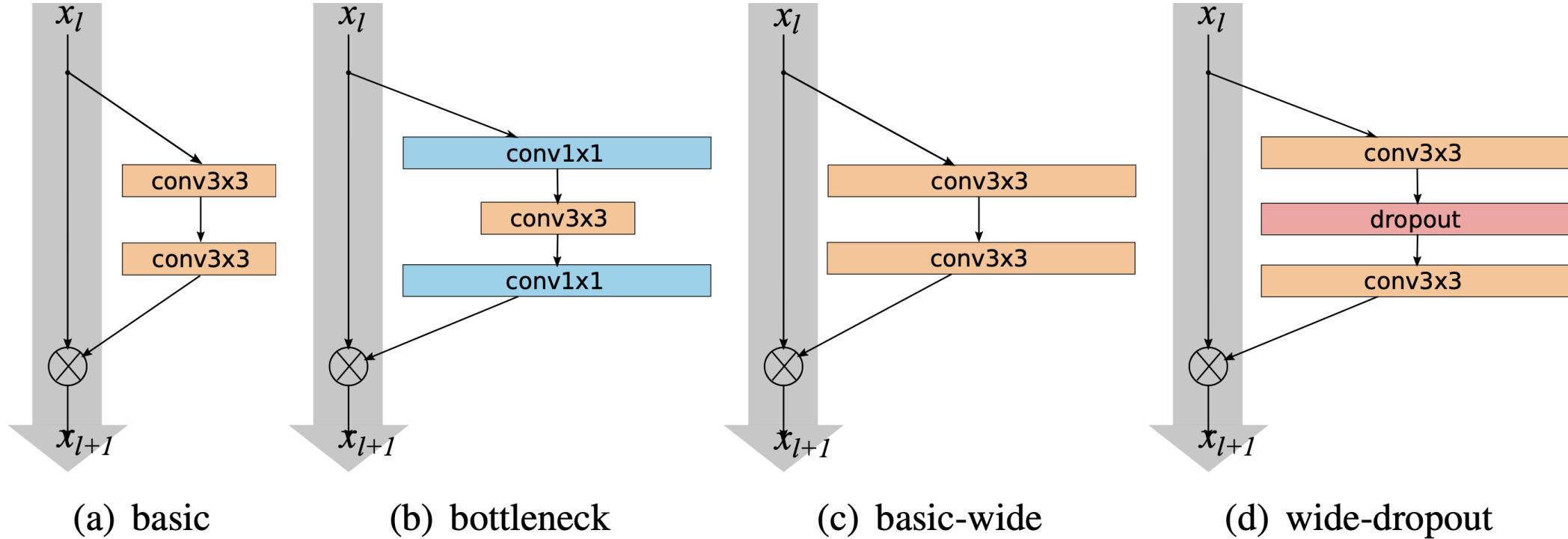


Figure 1: Various residual blocks used in the paper. Batch normalization and ReLU precede each convolution (omitted for clarity)

 **WideResNet_WRN-TF2** Public

master 1 Branch 0 Tags Go to file Add file Code

YeongHyeon Update neuralnet.py 8aca892 · 4 years ago 7 Commits

figures readme 4 years ago

source Update neuralnet.py 4 years ago

LICENSE license 4 years ago

README.md Update README.md 4 years ago

run.py source 4 years ago

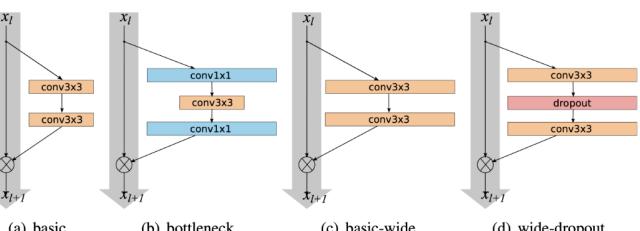
README MIT license

[TensorFlow 2] Wide Residual Networks

Related Repositories

[ResNet-TF2](#)
[ResNeXt-TF2](#)
[ResNet-with-LRWarmUp-TF2](#)
[ResNet-with-SGDR-TF2](#)

Concept



The diagram illustrates four types of residual blocks:

- (a) basic: A residual block where the input x_l is processed by two parallel paths. The first path has a residual connection that bypasses a convolutional layer (conv3x3). The second path has a residual connection that bypasses a residual connection (conv3x3).
- (b) bottleneck: A residual block where the input x_l is processed by two parallel paths. The first path has a residual connection that bypasses a residual connection (conv3x3). The second path has a residual connection that bypasses a residual connection (conv1x1).
- (c) basic-wide: A residual block where the input x_l is processed by two parallel paths. The first path has a residual connection that bypasses a residual connection (conv3x3). The second path has a residual connection that bypasses a residual connection (conv3x3).
- (d) wide-dropout: A residual block where the input x_l is processed by two parallel paths. The first path has a residual connection that bypasses a residual connection (conv3x3). The second path has a residual connection that bypasses a residual connection (conv3x3), which then passes through a dropout layer.

Figure 1: Various residual blocks used in the paper. Batch normalization and ReLU precede each convolution (omitted for clarity)

Concept of Wide ResNet [1].