PyramidFlow: High-Resolution Defect Contrastive Localization using Pyramid Normalizing Flow

CVPR 2023

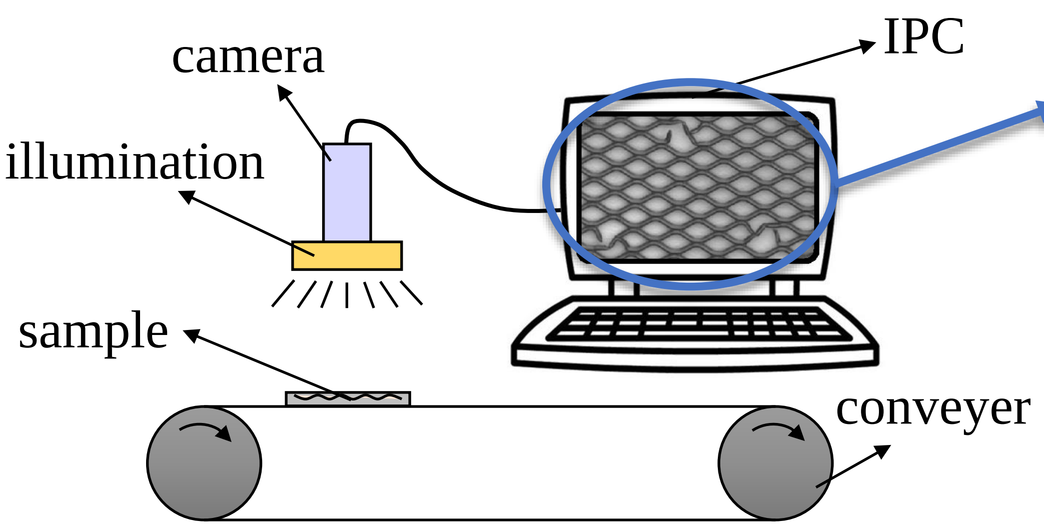
Jiarui Lei, Xiaobo Hu, Yue Wang, Dong Liu
Zhejiang University

WED-PM-171

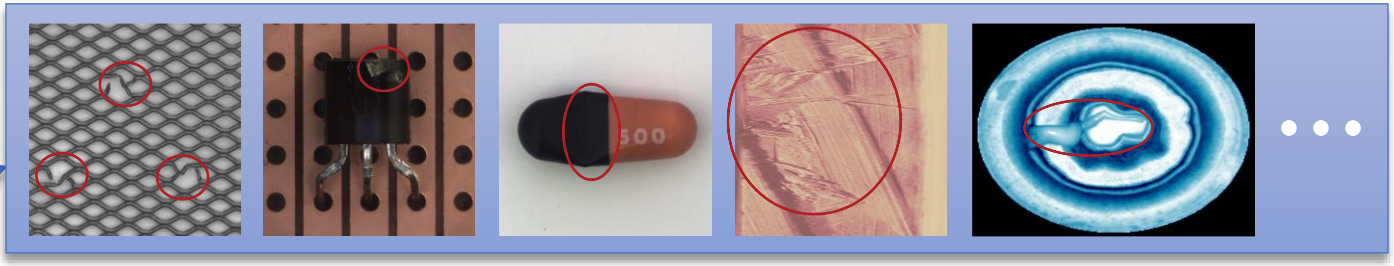
Project Page

0 Motivation

Visual Anomaly Detection via Machine Visions



Machine vision system



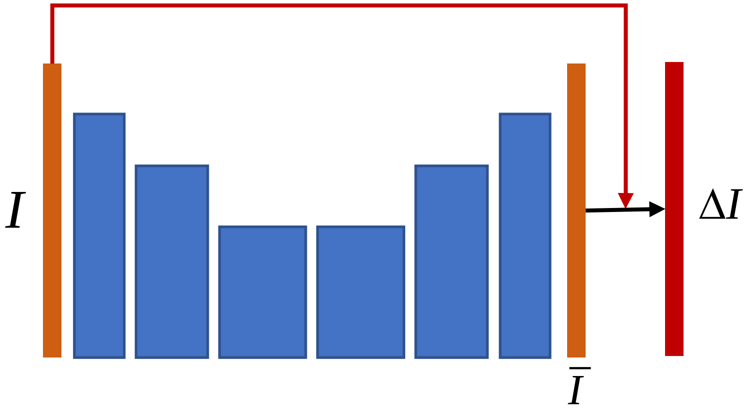
various defects

- **Defects as Open Set Objects: Unpredictable**
- **More Practical Scenarios: High Resolution**

Task: detect and localize any defects automatically

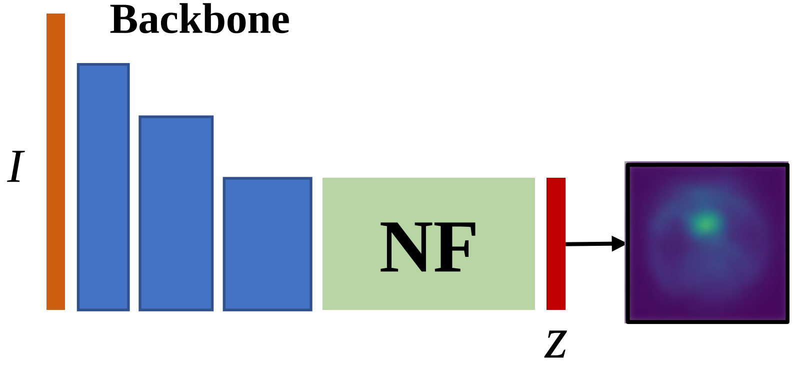
How to overcome these challenges ?

0 Overview



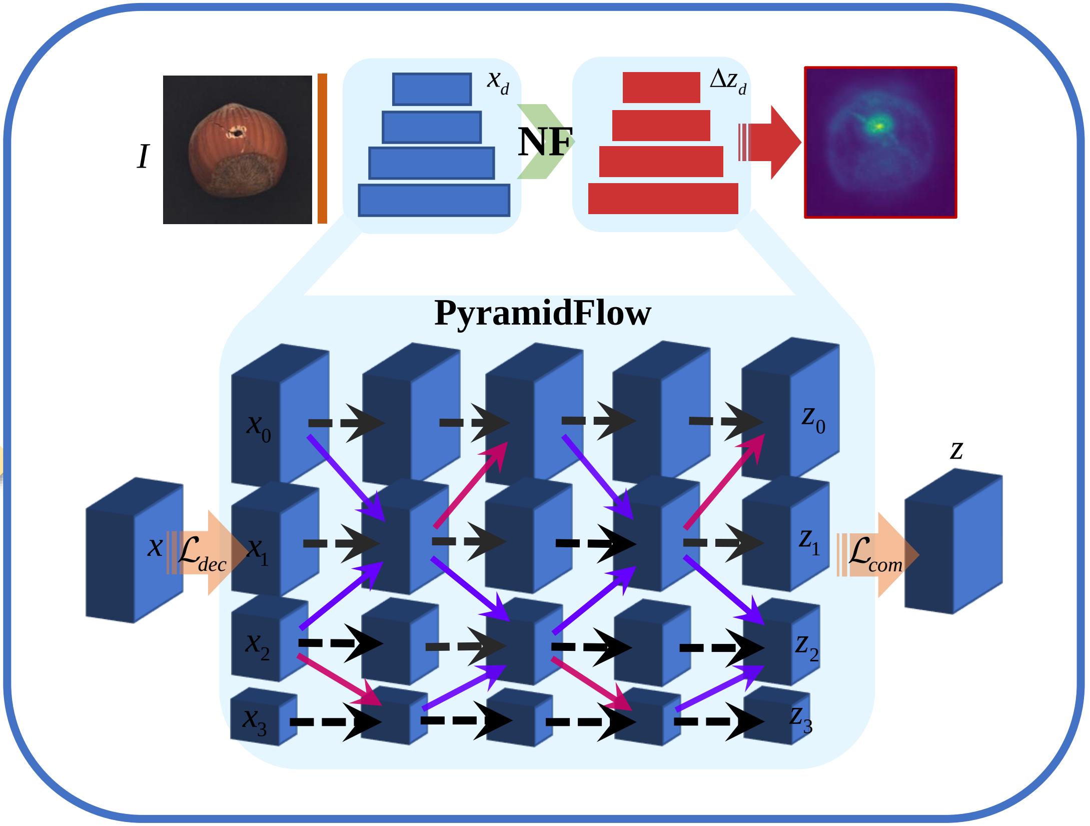
(a) reconstruction-based method

High Resolution
But
Slower Convergence



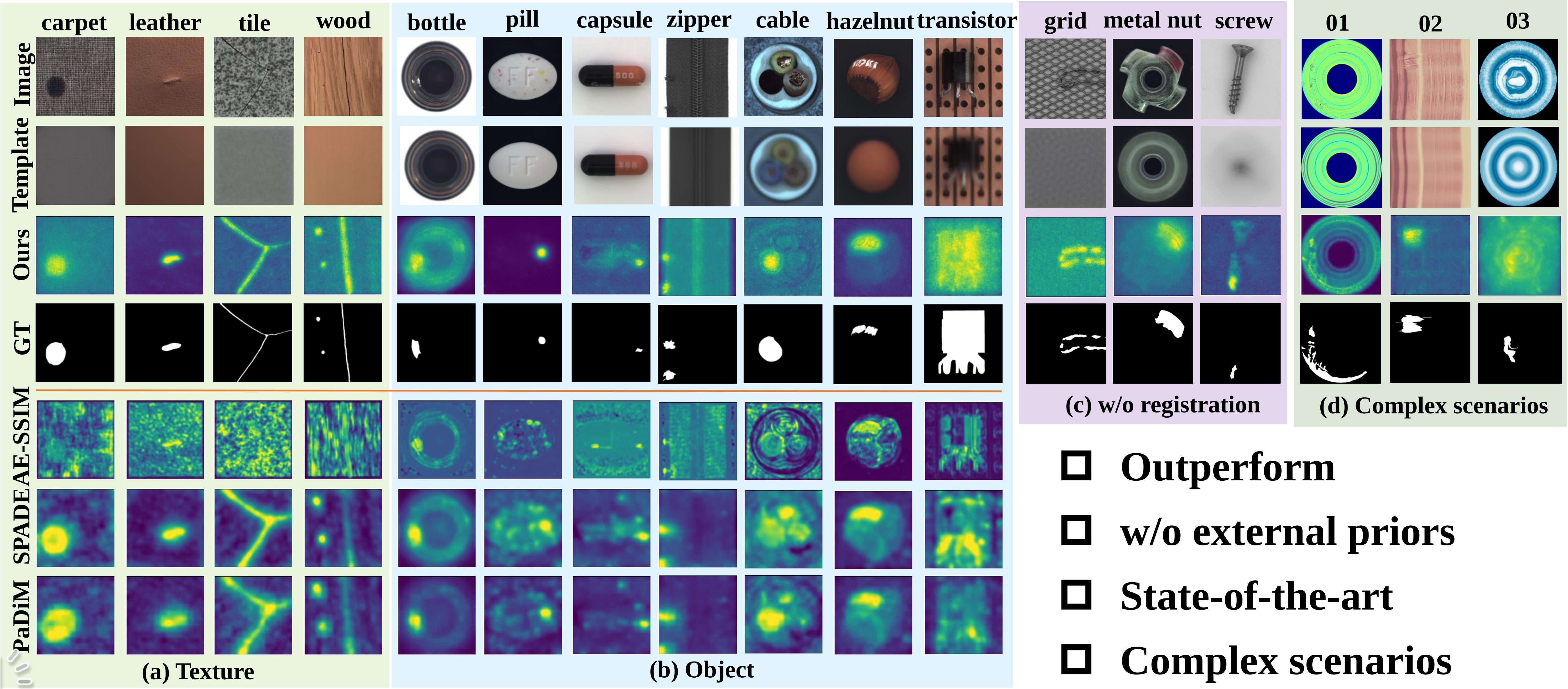
(b) anomaly-based method

High Performance
But
Low Resolution



The proposed latent template-based defect contrastive localization paradigm.

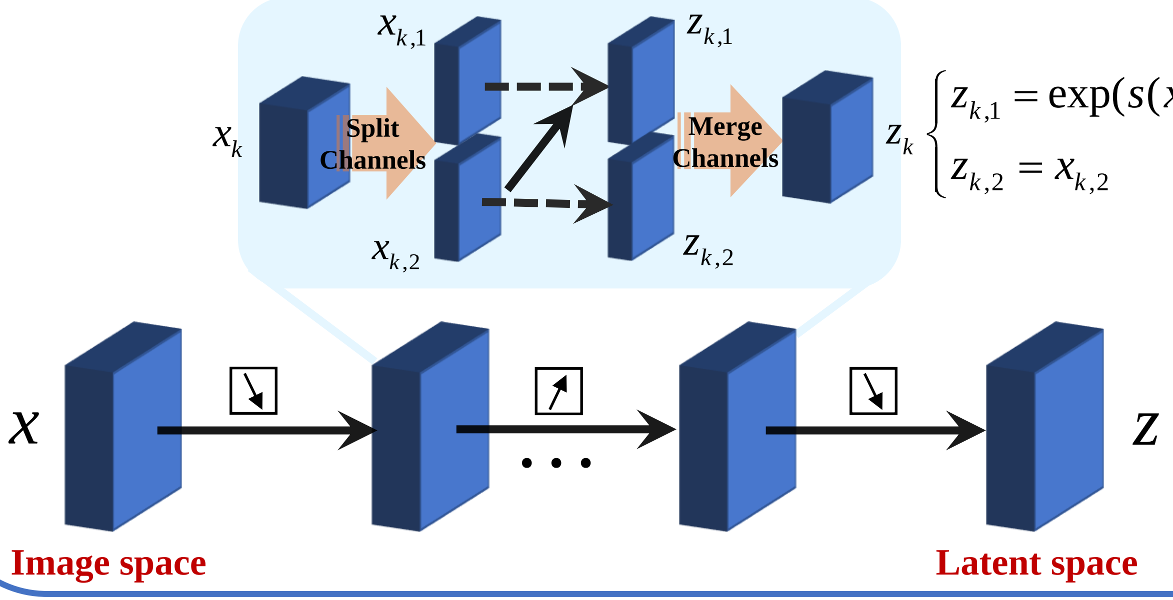
0 Results



- Outperform
- w/o external priors
- State-of-the-art
- Complex scenarios

Implementation Details

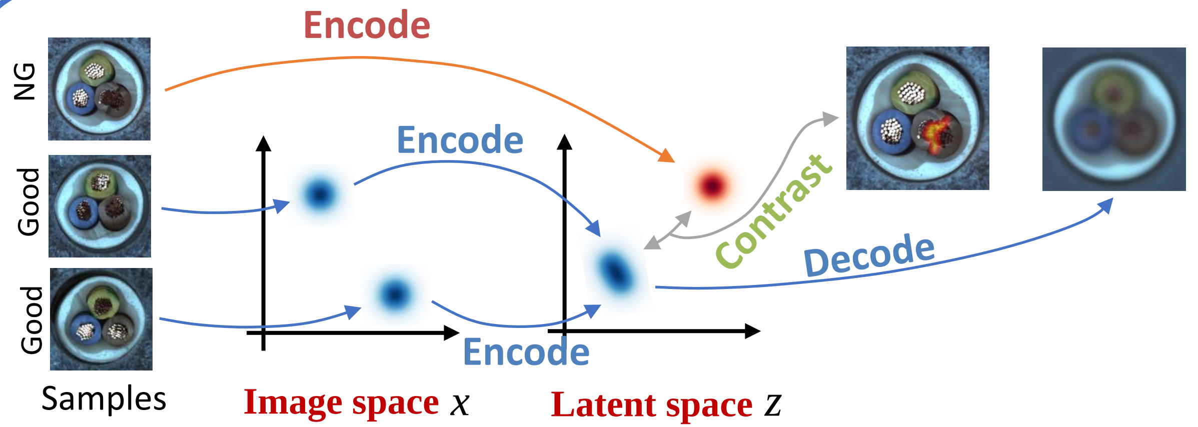
1 Basic Framework



$$\begin{cases} z_{k,1} = \exp(s(x_{k,2})) \odot x_{k,1} + t(x_{k,2}) \\ z_{k,2} = x_{k,2} \end{cases} \xrightarrow{\text{reversible}} \begin{cases} x_{k,1} = \exp(-s(x_{k,2})) \odot (z_{k,1} - t(x_{k,2})) \\ x_{k,2} = z_{k,2} \end{cases}$$

Log-probability transformation:

$$\log P(x) = \underbrace{\log P(z)}_{\text{basic probability model}} + \underbrace{\sum_k s(x_{k,2})}_{\text{subcomponent}}$$



Contrastive Learning:

- Reduce intra-class variance
- ~~Enlarge inter-class variance (w/o negative)~~

$$\boxed{\mathcal{L} = -\log P(x)}$$

1 Volume Normalization

Probability transformation:

$$P(x) = P(z)\left|\frac{\partial f(x)}{\partial x}\right|$$

Keep $\left|\frac{\partial f(x)}{\partial x}\right| = 1 \Leftrightarrow \sum_k s(x_{k,2}) = 0 \Leftarrow s(x_{k,2}) = 0$

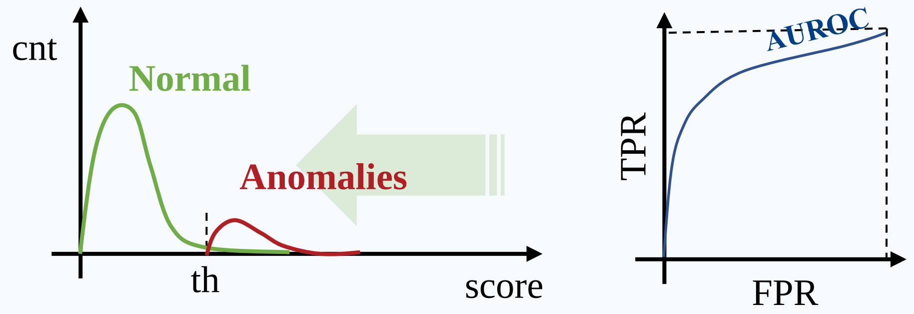
Zero Mean Normalization

Algorithm 2 Volume Normalization. (Pytorch-like Pseudocode)

```
Input: input x, momentum β
Output: output y
  def VolumeNorm2d(x, β = 0.1):
    if training:
      x̄ = mean(x, dim=1) % CVN: zero-mean normalization along
  channel dimensions
      y = x − x̄
      x̄_running = (1−β) × x̄_running + β × x̄ % update running mean
    else:
      y = x − x̄_running
    return y
```

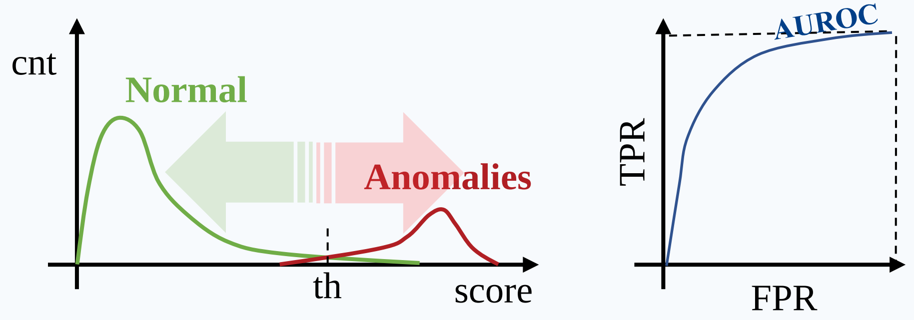
Case 1: maximize $\left|\frac{\partial f(x)}{\partial x}\right|$

Image likelihood maximization



Case 2: $\left|\frac{\partial f(x)}{\partial x}\right| = 1$

Intra-image probability redistribution



1 Latent Template

Probability transformation:

$$P(x) = P(z)\left|\frac{\partial f(x)}{\partial x}\right| = P(z)$$

Loss function $\mathcal{L} = -\log P(x) = -\log P(z)$

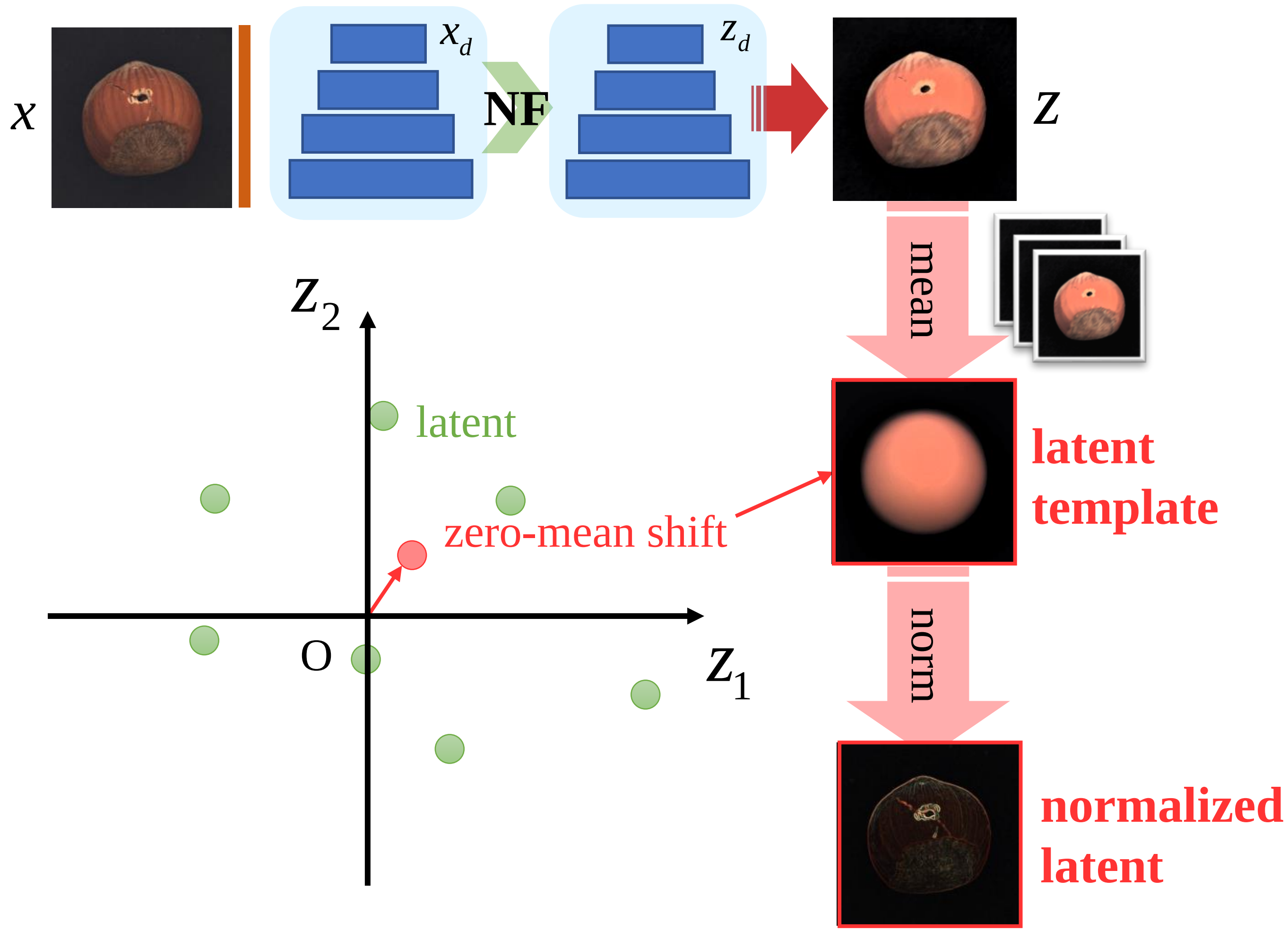
Modeling with Base Distribution

High-dimensional case (e.g. PaDiM)

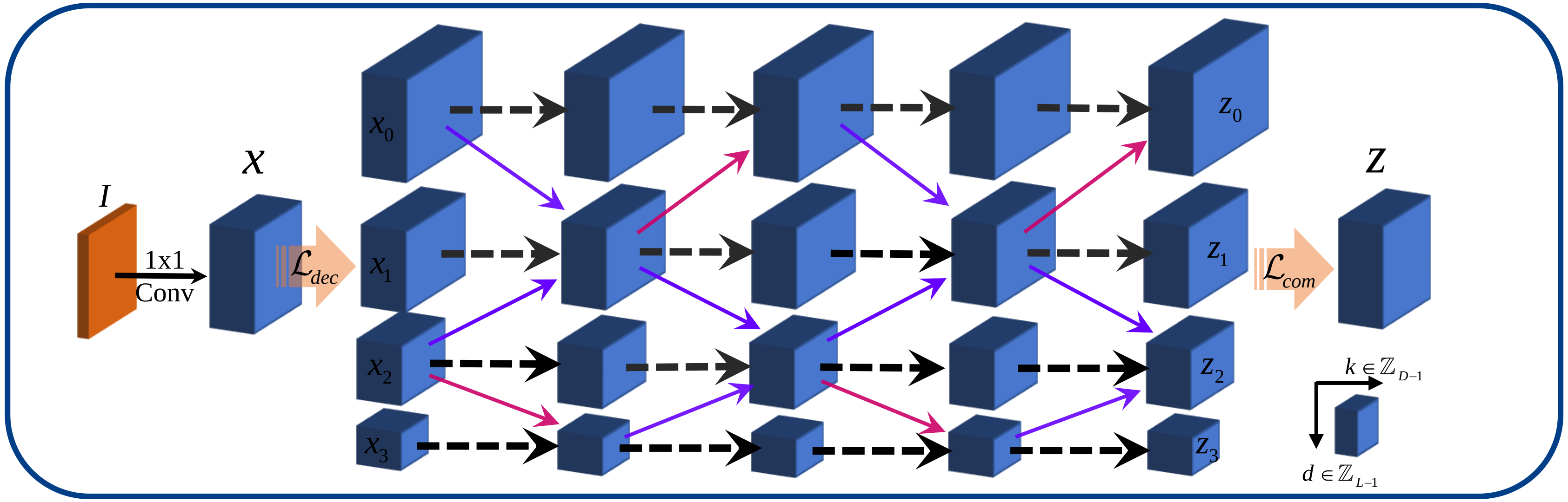
- zero center
- using standard Gaussian distribution

Low-dimensional case (e.g. PyramidFlow)

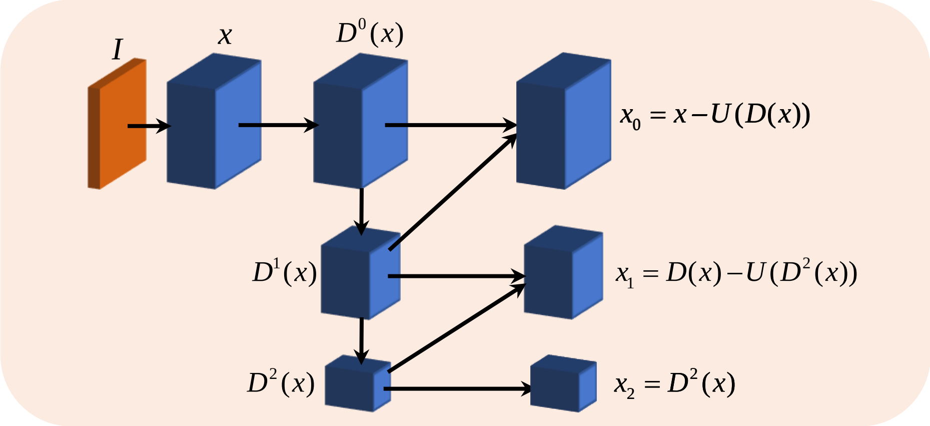
- zero-mean shift
- using zero-mean normalization



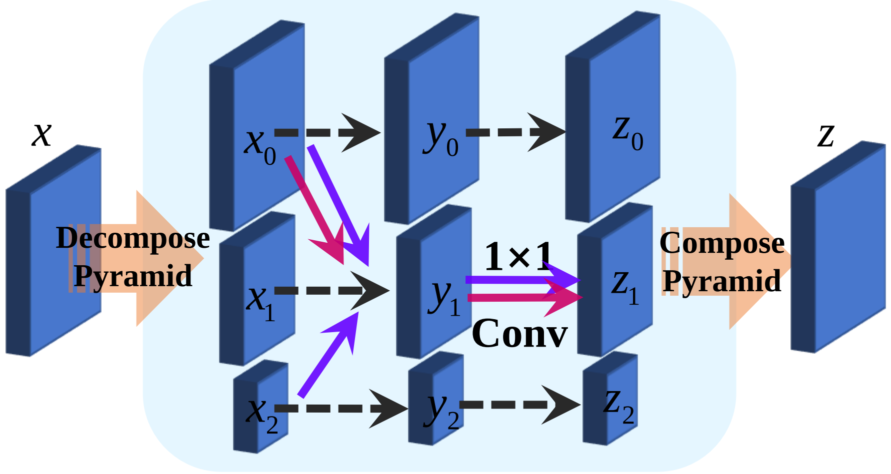
1 Model Architecture



PyramidFlow Architecture

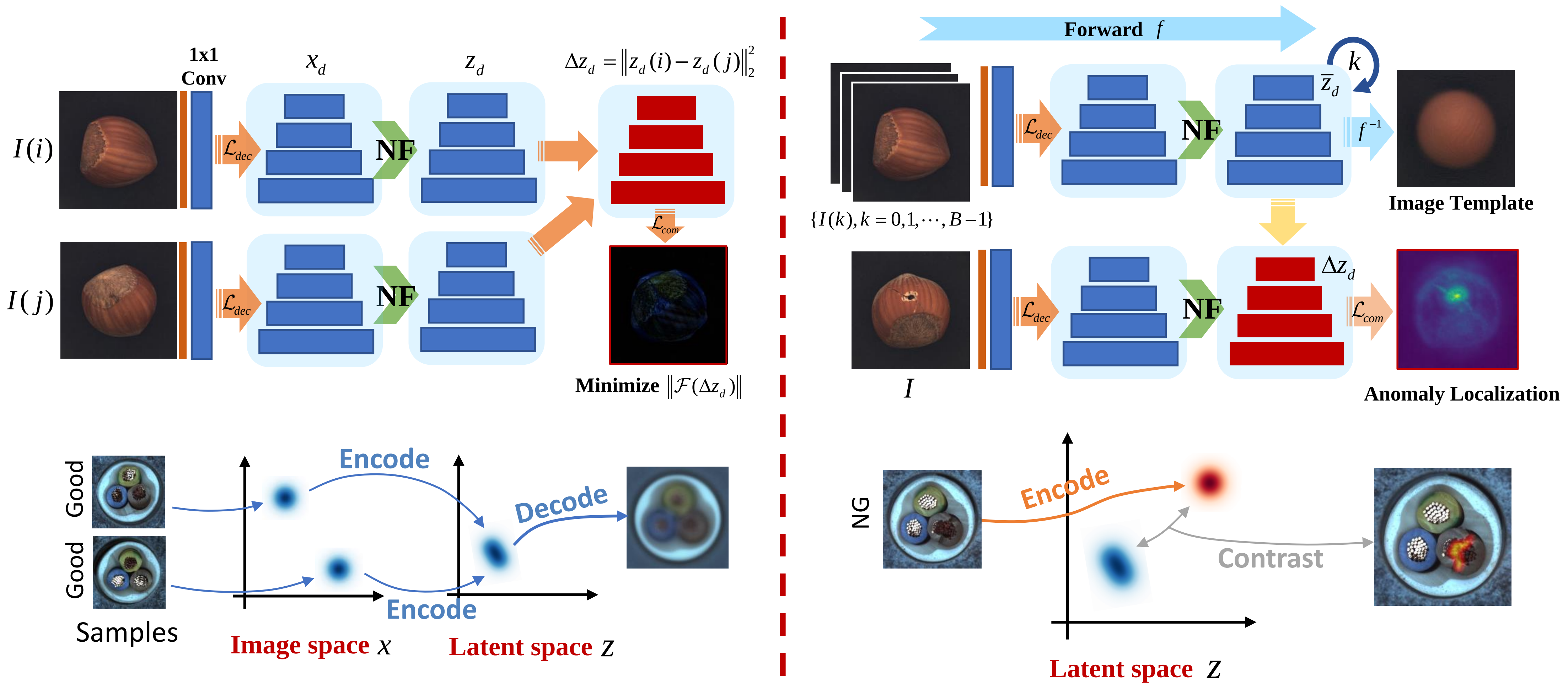


Invertible Pyramid Decomposition $\mathcal{L}_{dec}$



Pyramid Coupling Block

1 Training & Evaluation



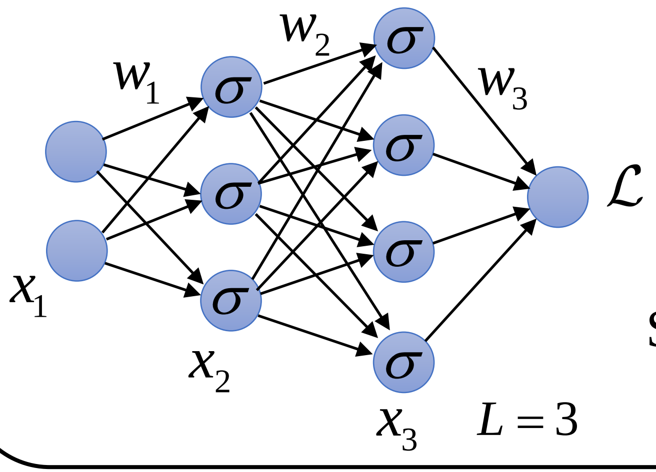
(a) Training

(b) Evaluation

Comprehensive Studies

2 Complexity Analysis

Net

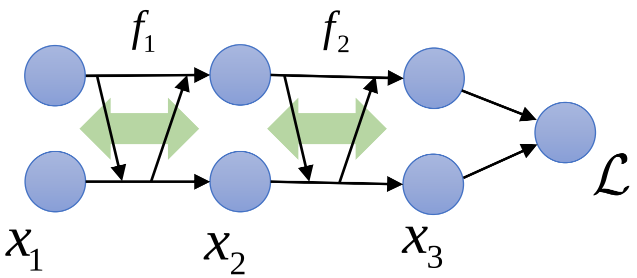


$L = 3$

$$\frac{\partial\mathcal{L}}{\partial w_2}=\frac{\partial\mathcal{L}}{\partial x_3}\frac{\partial x_3}{\partial w_2}=\frac{\partial\mathcal{L}}{\partial x_3}\cdot\sigma'(\boxed{w_2x_2+b_2})\cdot\boxed{x_2}$$

Save $\mathcal{O}(L)$ hidden nodes for back propagation

NF

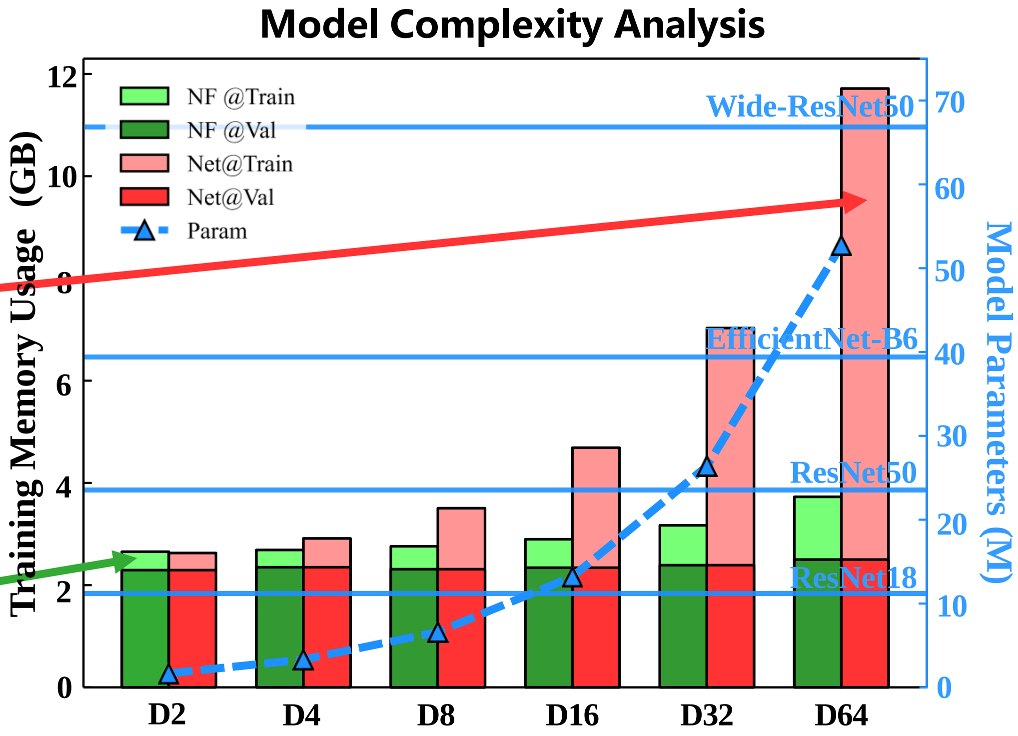


$L = 3$

$$\frac{\partial\mathcal{L}}{\partial w_2}=\frac{\partial\mathcal{L}}{\partial x_3}\frac{\partial x_3(x_2,w_2)}{\partial w_2}=\frac{\partial\mathcal{L}}{\partial x_3}\cdot x'_{3,2}(\boxed{x_3},\boxed{f_2^{-1}(x_3)})$$

Save $\mathcal{O}(1)$ hidden nodes

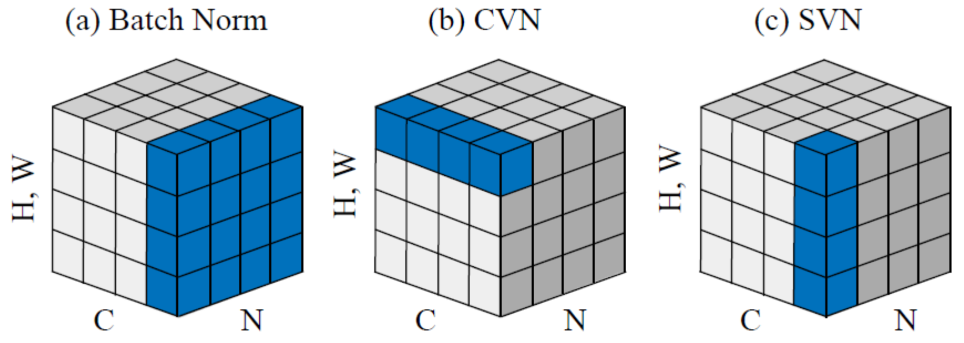
(**All nodes** can be obtained from the last one.)



We propose **autoFlow**, a easy to use normalizing flow framework with

- ☐ memory saving
- ☐ automatic Jacobian tracking
- ☐ object-oriented programming

2 Ablation Study



Volume Normalization

Table 1. Quantitative results of CVN and SVN on different categories. For each case in the table, the first column is Pixel-AUROC% and the second is AUPRO%, while the values within parentheses represent the relative improvement.

Classes	CVN AUROC	CVN AUPRO	SVN AUROC	SVN AUPRO
capsule	**96.1**(+2.6)	**93.1**(+5.1)	93.5(+0.0)	88.0(+0.0)
pill	**96.2**(+1.8)	**96.3**(+1.4)	94.4(+0.0)	94.9(+0.0)
toothbrush	**98.9**(+2.5)	**97.9**(+4.3)	96.4(+0.0)	93.6(+0.0)
carpet	88.9(+0.0)	88.3(+0.0)	**90.8**(+1.9)	**91.0**(+2.7)
grid	86.2(+0.0)	84.5(+0.0)	**94.2**(+8.0)	**92.7**(+8.2)
zipper	92.2(+0.0)	91.9(+0.0)	**95.4**(+3.2)	**95.1**(+3.2)

Others Ablation

Table 2. The ablation study on full MVTecAD. For each cell in the table, the first row is Pixel-AUROC% and the second is AUPRO%. The number within parentheses means the change relative to baseline, the larger absolute value with larger importance.

Method	Classes: Texture	Classes: Object	MEAN
Ours(baseline)	**95.2**(+0.0) **95.1**(+0.0)	**95.7**(+0.0) **93.5**(+0.0)	**95.5**(+0.0) **94.0**(+0.0)
I. w/o Volume Normalization	89.4(-5.8) 87.5(-7.6)	85.2(-10.5) 83.6(-9.9)	86.6(-8.9) 84.9(-9.1)
II. w/o Latent Template	93.1(-2.1) 91.9(-3.2)	90.7(-5.0) 84.7(-8.8)	91.5(-4.0) 87.1(-6.9)
III. w/o Pyramid Difference	87.8(-7.4) 87.8(-7.3)	93.1(-2.6) 89.4(-4.1)	91.3(-4.2) 88.9(-5.1)
IV. w/o Fourier Loss	92.0(-3.2) 92.8(-2.3)	93.3(-2.4) 91.9(-1.6)	92.9(-2.6) 92.2(-1.8)

2 Quantitative results

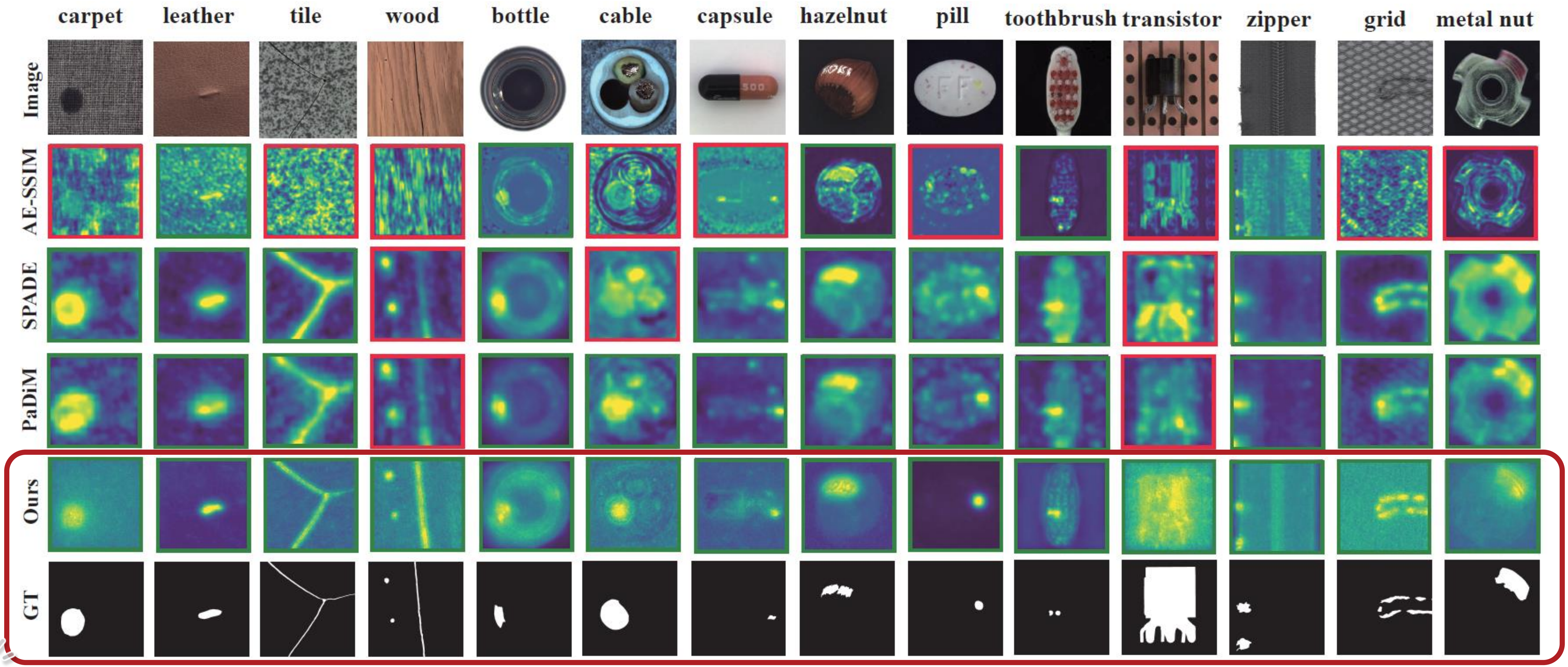
Experiments on MVTecAD

Table 3. Quantitative results of various challenging methods on MVTecAD. In the table, the fully normalized flow method is labeled as FNF, while the abbreviations Res18, WRes50, EffiB5, and DTD are denoted as ResNet18, Wide-ResNet50-2, EfficientNet-B5, and Describable Textures Dataset, respectively. For each case in the table, the first row is Pixel-AUROC% and the second is AUPRO% , where the best results are marked in bold.

External Prior	Methods	carpet	leather	tile	wood	bottle	cable	capsule	hazelnut	pill	toothbrush	transistor	zipper	MEAN
×	AnoGAN [23]	54.2	64.1	49.7	62.1	85.8	78.0	84.1	87.1	86.8	90.0	79.9	78.1	75.0
		20.4	37.8	17.7	38.6	62.0	38.3	30.6	69.8	77.6	74.9	54.9	46.7	47.4
	Vanilla VAE [15]	62.0	83.5	52.0	69.9	89.4	81.6	90.7	95.1	87.9	95.3	85.1	77.5	80.8
		61.9	64.9	24.2	57.8	70.5	77.9	77.9	77.0	79.3	85.4	61.0	60.8	66.6
	AE-SSIM [4]	87.0	78.0	59.0	73.0	93.0	82.0	94.0	97.0	91.0	92.0	80.0	88.0	84.5
		64.7	56.1	17.5	60.5	83.4	47.8	86.0	91.6	83.0	78.4	72.4	66.5	67.3
	Ours (FNF)	**90.8**	**99.6**	**97.9**	**93.8**	**95.9**	**92.1**	**96.1**	**98.0**	**96.2**	**98.9**	**97.4**	**95.4**	**96.0**
		91.0	**99.7**	**95.8**	**96.2**	**94.0**	**86.4**	**93.1**	**97.3**	**96.3**	**97.7**	**91.4**	**95.1**	**94.5**
Res18	S-T [3]	93.5	97.8	92.5	92.1	97.8	91.9	96.8	98.2	**96.5**	97.9	73.7	95.6	93.7
		87.9	94.5	94.6	91.1	93.1	81.8	96.8	96.5	**96.1**	93.3	66.6	95.1	90.6
WRes50	SPADE [6]	97.5	97.6	87.4	88.5	**98.4**	**97.2**	**99.0**	**99.1**	**96.5**	97.9	94.1	96.5	95.8
		94.7	97.2	75.9	87.4	95.5	90.9	93.7	95.4	94.6	93.5	**97.4**	92.6	92.4
WRes50	PaDiM [7]	**99.1**	99.2	94.1	94.9	98.3	96.7	98.5	98.2	95.7	**98.8**	**97.5**	**98.5**	**97.5**
		96.2	97.8	86.0	91.1	94.8	88.8	93.5	92.6	92.7	93.1	84.5	**95.9**	92.3
EffiB5	CS-Flow [22]	98.0	98.4	93.9	88.6	90.9	95.3	97.9	96.3	95.7	96.3	95.5	96.4	95.3
		98.0	98.5	94.5	92.9	88.7	**94.0**	96.1	95.1	91.1	89.9	96.9	95.4	94.2
DTD	DRÆM [29]	94.9	**96.6**	**99.6**	**97.3**	97.6	95.4	94.0	99.2	95.0	98.1	90.0	94.4	96.0
		96.1	97.9	**99.7**	**97.9**	**97.2**	90.4	96.5	**98.7**	93.7	97.1	92.9	94.7	96.1
Res18	Ours	97.4	98.7	97.1	97.0	97.8	91.8	98.6	98.1	96.1	98.5	96.9	96.6	97.1
		97.2	**99.2**	97.2	**97.9**	95.5	90.3	**98.3**	98.1	**96.1**	**97.9**	94.7	95.4	**96.5**

SOTA AUPRO!

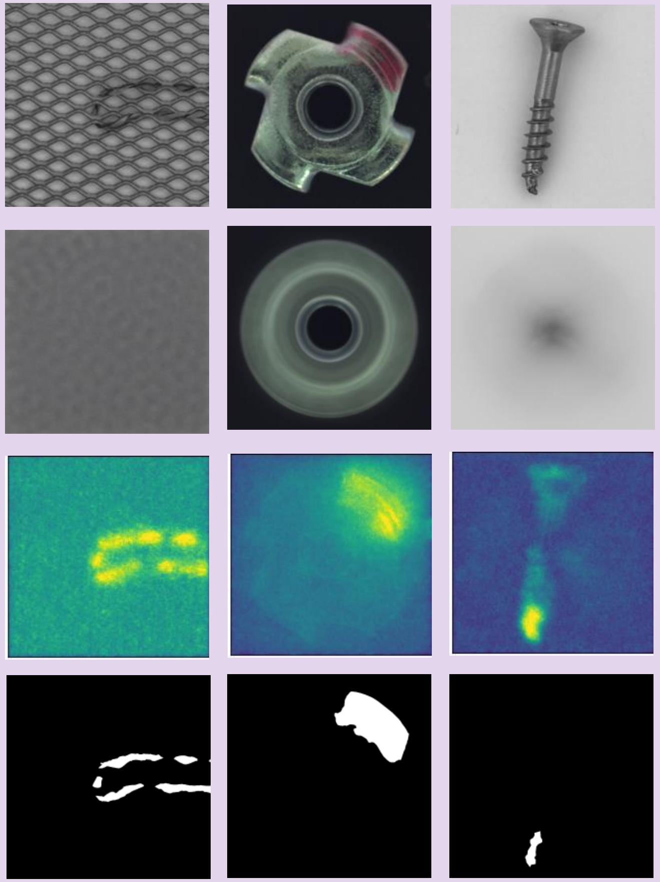
2 Visualization



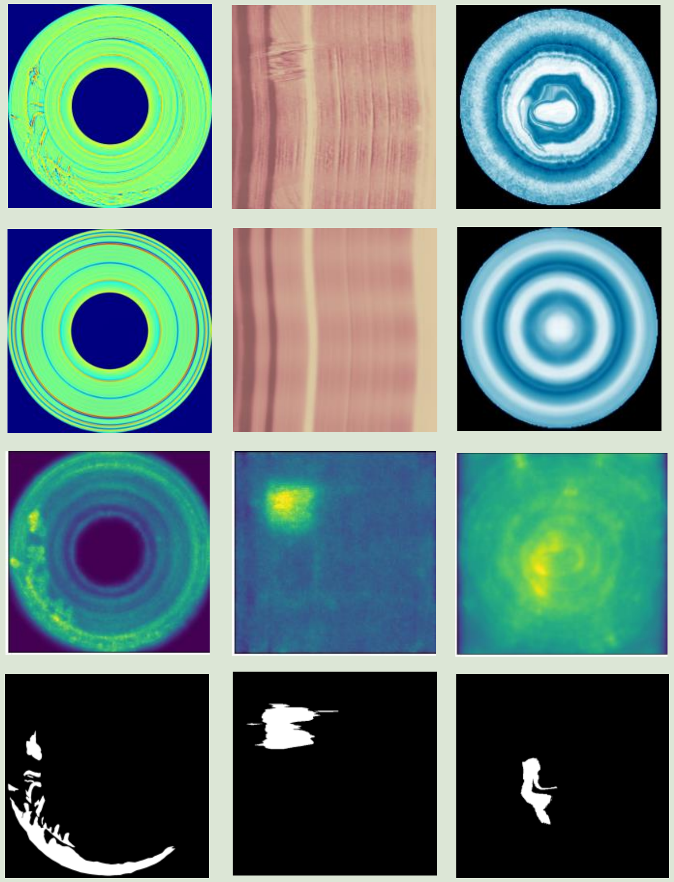
Visualization of competitive results on MVTecAD

2 Complex Scenarios

Experiments on BTAD



(a) w/o registration



(b) Complex texture

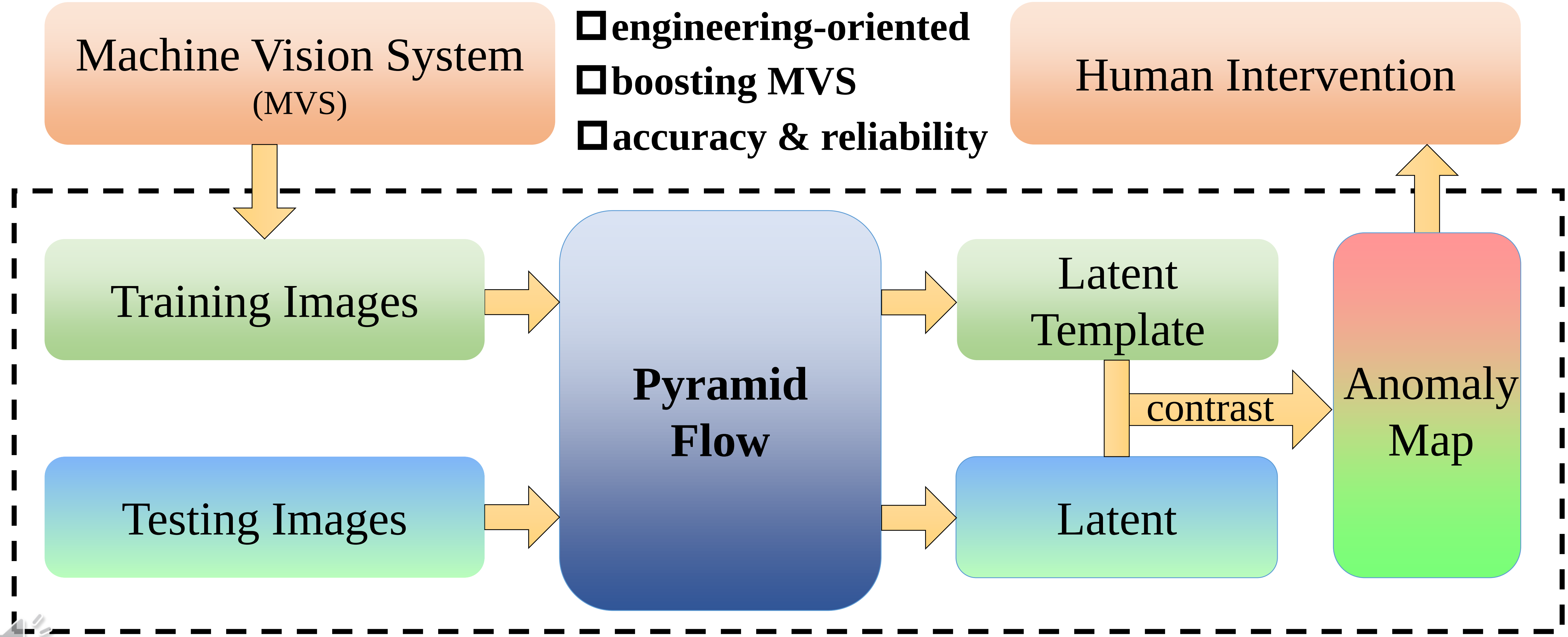
Table 4. Quantitative results of various challenging methods on BTAD. For each case in the table, the first row is Image-AUROC% and the second is Pixel-AUROC%, where the best results are marked in bold.

Experiments on BTAD

Methods	Classes 01	02	03	MEAM
VT-ADL [16]	97.6	71.0	82.6	83.7
	99.0	94.0	77.0	90.0
P-SVDD [26]	95.7	72.1	82.1	83.3
	91.6	93.6	91.0	92.1
SPADE [6]	91.4	71.4	**99.9**	87.6
	97.3	94.4	99.1	96.9
PatchCore [20]	90.9	79.3	99.8	90.0
	95.5	94.7	**99.3**	96.5
PaDiM [7]	99.8	82.0	99.4	93.7
	97.0	96.0	98.8	97.3
Ours (Res18)	**100.0**	**88.2**	99.3	**95.8**
	97.4	**97.6**	98.1	**97.7**

- **Handling complexity**
- **Complex textures**

3 Summarization



Thank you!