Toward Generalist Anomaly Detection via In-context Residual Learning with Few-shot Sample Prompts *[CVPR 24]*

Jiawen Zhu, Guansong Pang
School of Computing and Information Systems, Singapore Management University

Sunghyun Ahn
skd@yonsei.ac.kr

1 Generalist Anomaly Detection

Generalist Anomaly Detection (GAD)

- 추가 학습 없이 여러 분야의 다양한 데이터에서 Anomaly Detection을 할 수 있는 하나의 모델을 만드는 것

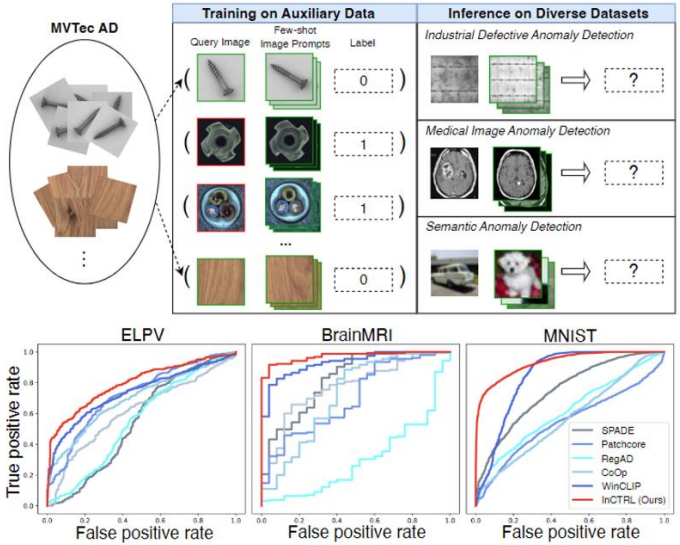
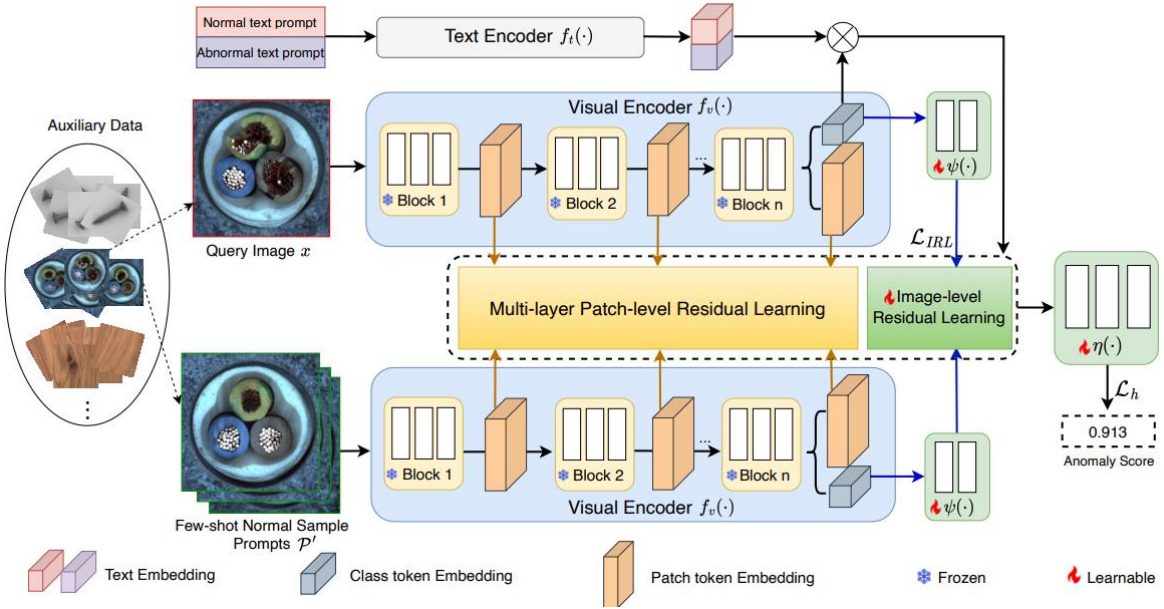


Figure 1. **Top**: An illustration of InCTRL: a one-for-all model using few-shot normal images as sample prompts. **Bottom**: AUROC curves of InCTRL and competing few-shot methods on three different application datasets without any training on the target data.

1 Generalist Anomaly Detection

In-context residual learning model for GAD (InCTRL)

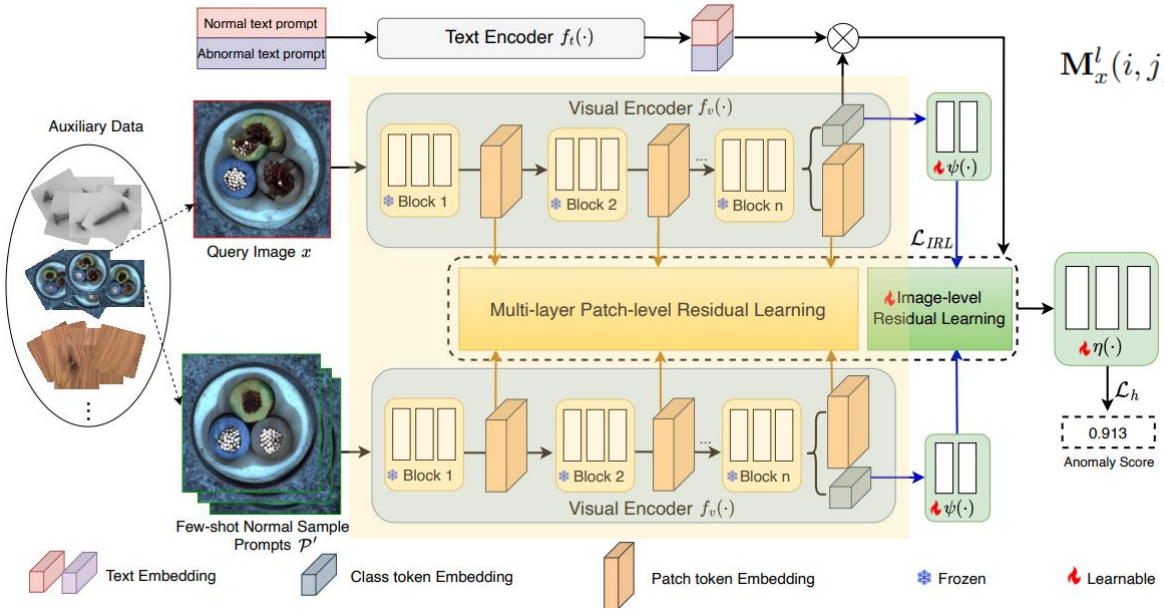
- 정상 샘플을 기반으로 입력 데이터의 잔차(residual)을 감지하여 이상을 식별하는 InCTRL 모델 제안
- 비정상일수록 정상 데이터(few-normal shots)와 residual이 클 것이라 가정함
- Image 및 patch level에서 residual learning을 수행하여 9개의 AD 데이터셋에서 기존 방법들보다 뛰어난 성능을 보임



1 Generalist Anomaly Detection

Multi-Layer Patch-Level Residual learning

- Fine-grained한 residual을 포착하고자 각 스테이지의 임베딩을 활용해 query 패치와 few-shot 패치 간의 cosine similarity를 계산
- 이를 통해 local한 이상 점수 맵(M)을 생성하고, 각 레이어에서 얻은 M을 평균내어 최종 M을 제작함



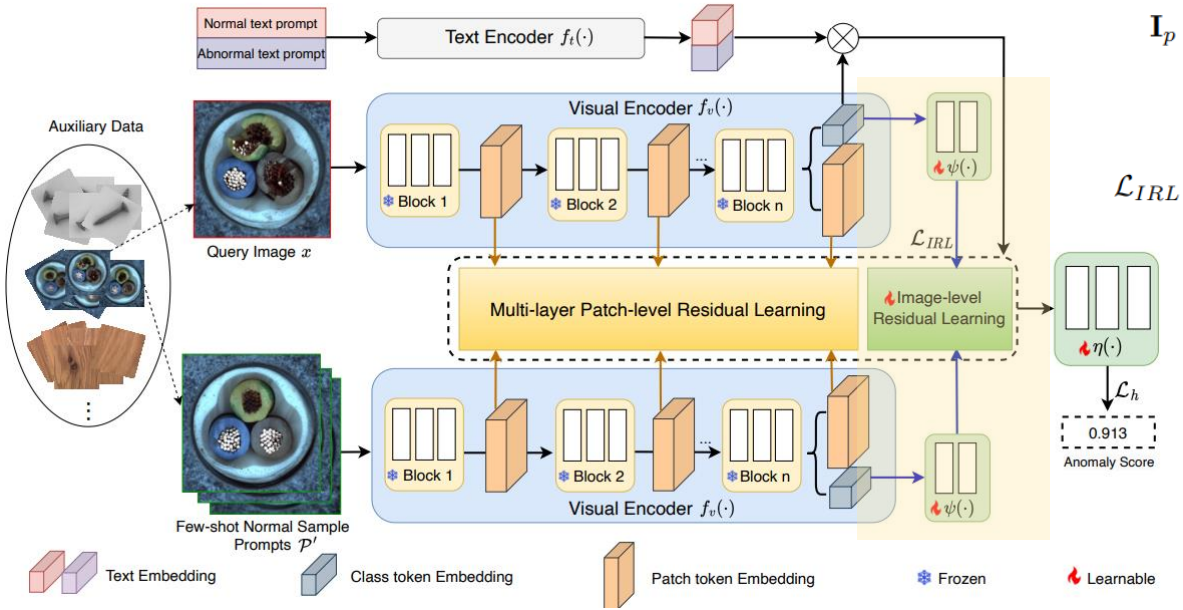
$$\mathbf{M}_x^l(i,j) = 1 - \langle T_x^l(i,j), h(T_x^l(i,j)|\mathcal{P}')\rangle, \quad (1)$$

$$\mathbf{M}_x = \frac{1}{n}\sum_{l=1}^{n}\mathbf{M}_x^l. \quad (2)$$

1 Generalist Anomaly Detection

Image-level Residual Learning

- 이미지의 global 정보를 활용하기 위해 visual Encoder의 class token을 사용해 few-shot class token들과 학습을 진행함
- 학습 방식은 few-shot sample의 평균 feature와 query feature의 차이가 y가 되도록 함으로써, 차이가 클수록 1이 되는 것을 기대함



$$\mathbf{I}_p = \frac{1}{K} \sum_{x'_k \in \mathcal{P}'} \psi(f_v(x'_k); \Theta_\psi), \tag{3}$$

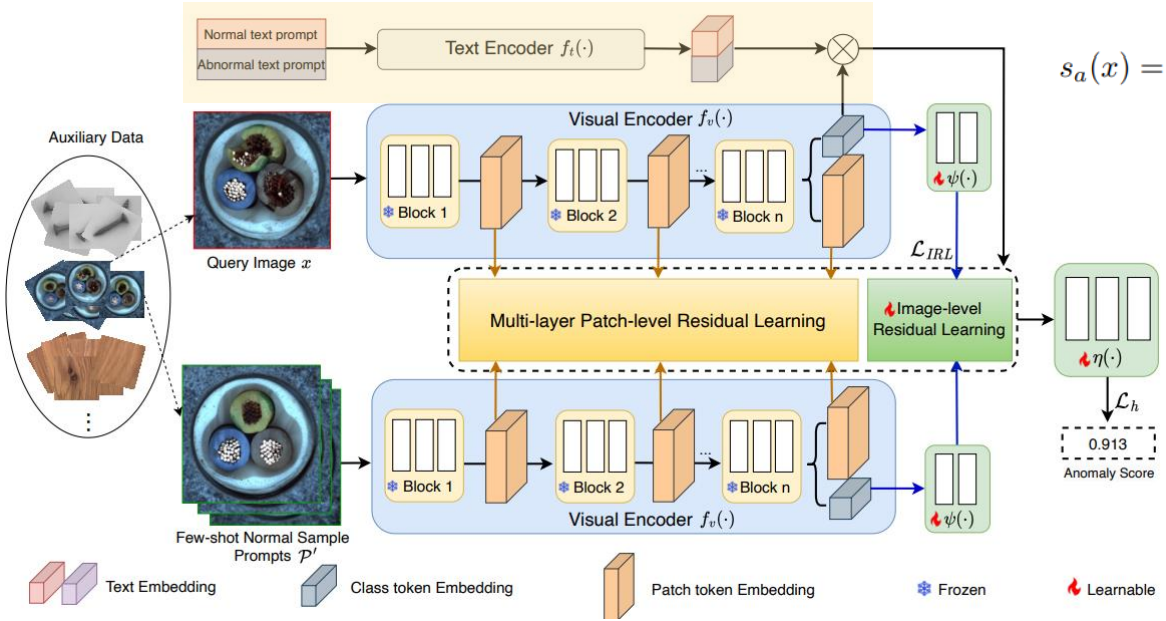
$$\mathbf{F}_x = \mathbf{I}_x \ominus \mathbf{I}_p, \tag{4}$$

$$\mathcal{L}_{IRL} = \frac{1}{N} \sum_{x \in X_{train}} \mathcal{L}_b(\eta(\mathbf{F}_x; \Theta_\eta), y_x), \tag{5}$$

1 Generalist Anomaly Detection

Fusing Text Prompt-based Prior Knowledge

- Residual learning을 하는 것만으로는 해당 이미지가 정상에 해당하는지 비정상에 해당하는지에 대한 semantic 정보가 부족함
- Normal, abnormal text와 class token간의 유사도를 추가로 고려하여, 입력 query가 정상에 가까운지, 비정상에 가까운지 판단함



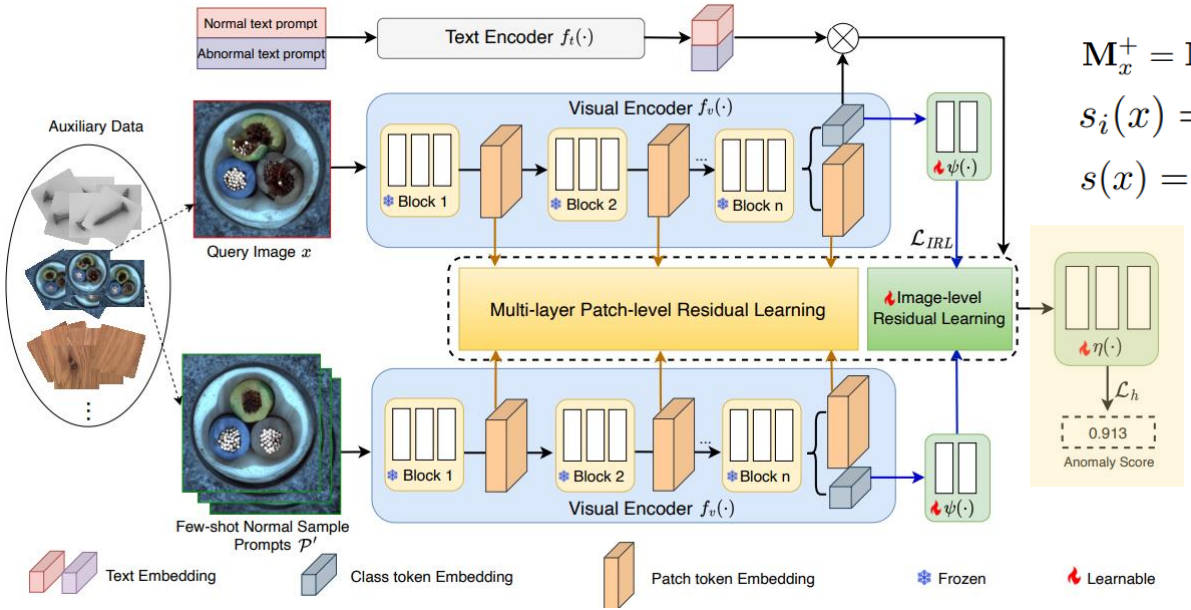
$$s_a(x) = \frac{\exp(\mathbf{F}_a^\intercal f_v(x))}{\exp(\mathbf{F}_n^\intercal f_v(x)) + \exp(\mathbf{F}_a^\intercal f_v(x))}, \quad (6)$$

$$\mathbf{F}_a = \frac{1}{|\mathcal{P}_t^a|} \sum_{p_j \in \mathcal{P}_t^a} f_t(p_j)$$

1 Generalist Anomaly Detection

In-Context Residual Learning

- 세 과정에서 계산된 M_x, F_x, s_a는 Query가 비정상일 때 모두 높은 값을 가진다는 특징이 있음 → 세 변수의 합을 Final Anomaly Score로 이용함
- Final Anomaly Score가 y가 되도록 학습하여 비정상 이미지에 대해 1(이상)을 출력하도록 함.



$$\mathbf{M}_x^+ = \mathbf{M}_x \oplus s_i(x) \oplus s_a(x), \quad (7)$$

$$s_i(x) = \eta(\mathbf{F}_x; \Theta_\eta)$$

$$s(x) = \phi(\mathbf{M}_x^+; \Theta_\phi) + \alpha s_p(x),$$

$$s_p(x) = \max(\mathbf{M}_x)$$

$$\mathcal{L}_h = \frac{1}{N} \sum_{x \in X_{train}} \mathcal{L}_b(s(x), y_x).$$

1 Generalist Anomaly Detection

Experiments

- Few-shot AD가 가능한 다른 모델들보다 우수한 성능을 보임 (Medical dataset에서도 유용함)
- 한계: 정상성이 다양한 경우에는 정상 query의 residual이 커져서 낮은 성능을 보임

Setup	Methods	Industrial Defects					Medical Anomalies		Semantic Anomalies			
									One-vs-all		Multi-class	
		ELPV	SDD	AITEX	VisA	MVTec AD	BrainMRI	HeadCT	MNIST	CIFAR-10	MNIST	CIFAR-10
Baseline (0-shot)		0.733±0.000	0.946±0.000	0.733±0.000	0.781±0.000	0.912±0.000	0.926±0.000	0.900±0.000	0.678±0.000	0.924±0.000	0.620±0.000	0.900±0.000
2-shot	SPADE	0.517±0.012	0.729±0.041	0.727±0.004	0.795±0.045	0.817±0.054	0.754±0.048	0.645±0.034	0.779±0.024	0.823±0.014	0.595±0.060	0.655±0.042
	PaDiM	0.594±0.083	0.721±0.015	**0.784±0.028**	0.680±0.042	0.785±0.025	0.657±0.122	0.595±0.036	-	-	-	-
	Patchcore	0.716±0.031	0.902±0.006	0.739±0.017	0.817±0.028	0.858±0.034	0.706±0.009	0.736±0.096	0.756±0.004	0.602±0.009	0.603±0.009	0.703±0.008
	RegAD	0.571±0.016	0.499±0.008	0.564±0.072	0.557±0.053	0.640±0.047	0.449±0.129	0.602±0.018	0.525±0.030	0.534±0.005	0.608±0.026	0.695±0.002
	CoOp	0.762±0.011	0.897±0.006	0.687±0.062	0.806±0.023	0.858±0.016	0.725±0.020	0.811±0.003	0.557±0.006	0.527±0.011	0.612±0.007	0.393±0.009
	WinCLIP	**0.726±0.020**	**0.942±0.006**	0.726±0.055	**0.842±0.024**	**0.931±0.019**	**0.934±0.012**	**0.915±0.015**	**0.810±0.008**	**0.925±0.001**	**0.632±0.000**	**0.914±0.005**
	Ours (InCTRL)	**0.839±0.003**	**0.972±0.011**	**0.761±0.029**	**0.858±0.022**	**0.940±0.015**	**0.973±0.027**	**0.929±0.025**	**0.892±0.009**	**0.935±0.002**	**0.635±0.010**	**0.924±0.005**
4-shot	SPADE	0.537±0.013	0.731±0.020	0.718±0.011	0.811±0.040	0.828±0.044	0.759±0.070	0.624±0.012	0.810±0.009	0.836±0.006	0.588±0.041	0.631±0.063
	PaDiM	0.612±0.080	0.742±0.014	**0.787±0.038**	0.735±0.031	0.805±0.018	0.792±0.048	0.622±0.013	-	-	-	-
	Patchcore	0.756±0.073	0.923±0.008	0.733±0.002	0.843±0.025	0.885±0.026	0.794±0.040	0.805±0.006	0.833±0.009	0.639±0.010	0.497±0.044	0.739±0.011
	RegAD	0.596±0.040	0.525±0.027	0.596±0.074	0.574±0.042	0.663±0.032	0.571±0.149	0.522±0.050	0.548±0.053	0.534±0.002	0.596±0.075	0.677±0.161
	CoOp	**0.781±0.002**	0.902±0.006	0.720±0.017	0.818±0.018	0.874±0.017	0.759±0.033	0.860±0.032	0.563±0.004	0.537±0.005	0.618±0.002	0.395±0.008
	WinCLIP	0.754±0.009	**0.943±0.004**	0.764±0.025	**0.858±0.025**	**0.940±0.021**	**0.941±0.002**	**0.912±0.003**	**0.851±0.010**	**0.927±0.001**	**0.632±0.004**	**0.915±0.003**
	Ours (InCTRL)	**0.846±0.011**	**0.975±0.006**	**0.790±0.018**	**0.877±0.019**	**0.945±0.018**	**0.975±0.016**	**0.933±0.013**	**0.902±0.016**	**0.940±0.010**	**0.643±0.007**	**0.928±0.009**
8-shot	SPADE	0.567±0.034	0.741±0.011	0.708±0.006	0.821±0.042	0.840±0.057	0.794±0.039	0.626±0.022	0.829±0.009	0.849±0.006	0.597±0.028	0.656±0.037
	PaDiM	0.724±0.017	0.769±0.037	0.792±0.025	0.768±0.032	0.820±0.016	0.758±0.025	0.661±0.039	-	-	-	-
	Patchcore	**0.837±0.016**	0.925±0.003	0.745±0.002	0.860±0.026	0.922±0.019	0.812±0.016	0.817±0.034	**0.876±0.004**	0.672±0.006	0.526±0.019	0.764±0.004
	RegAD	0.633±0.027	0.594±0.029	0.603±0.062	0.589±0.040	0.674±0.033	0.632±0.079	0.628±0.026	0.547±0.063	0.555±0.008	0.573±0.076	0.587±0.211
	CoOp	0.817±0.012	0.898±0.005	0.769±0.008	0.822±0.021	0.880±0.014	0.755±0.003	0.914±0.027	0.567±0.007	0.542±0.005	0.619±0.004	0.399±0.006
	WinCLIP	0.814±0.010	**0.941±0.001**	**0.796±0.015**	**0.868±0.020**	**0.947±0.025**	**0.944±0.001**	**0.915±0.008**	0.867±0.007	**0.928±0.001**	**0.641±0.004**	**0.916±0.003**
	Ours (InCTRL)	**0.872±0.013**	**0.978±0.006**	**0.806±0.036**	**0.887±0.021**	**0.953±0.013**	**0.983±0.012**	**0.936±0.008**	**0.920±0.003**	**0.945±0.002**	**0.646±0.003**	**0.934±0.008**

Table 1. AUROC results(mean±std) on nine real-world AD datasets under various few-shot AD settings. Best results and the second-best results are respectively highlighted in **red** and **blue**. 'Baseline' is a WinCLIP-based zero-shot AD model.

1 Generalist Anomaly Detection

Experiments

- 2-shot prompts에서 query의 정상성을 반영하지 못하여 residual이 커진 예시 (False Positive)

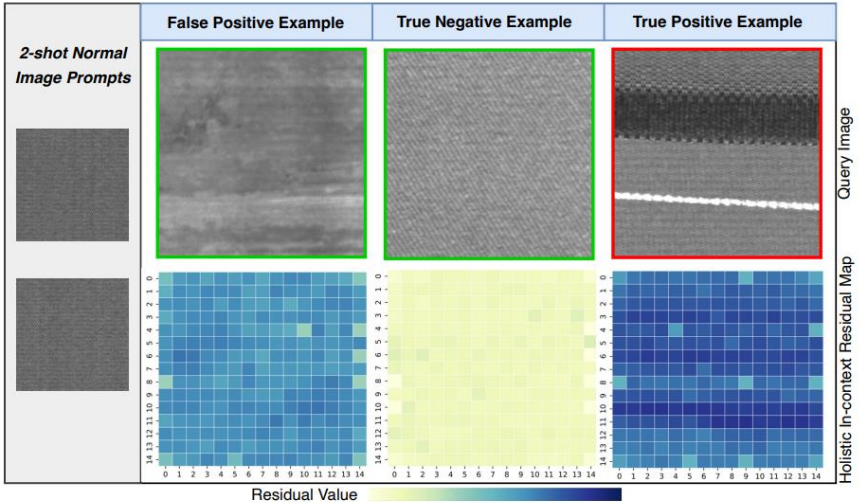


Figure 3. Visualization of query images x_t and their holistic in-context residual maps $\mathbf{M}^{+}_{x_t}$. Green and Red frames indicate normal and abnormal images respectively. Deeper colors in the residual maps represent larger residual values.

Thank You