46th International Conference on Software Engineering

COCA: Improving and Explaining GNN-Based Vulnerability Detection

Sicong Cao[1], Xiaobing Sun[1], Xiaoxue Wu[1], David Lo[2], Lili Bo[1], Bin Li[1], and Wei Liu[1]

[1] Yangzhou University
[2] Singapore Management University

Vulnerability Detection Advancement

Phase 1

Manual

Code Review, Reverse Engineering, Expertise

1960s

Vulnerability Detection Advancement

Phase 1

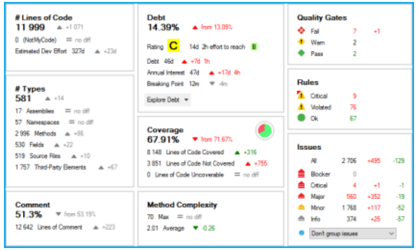
Manual

Code Review, Reverse Engineering, Expertise

Phase 2

Rule

Static/Taint Analysis, Model Checking

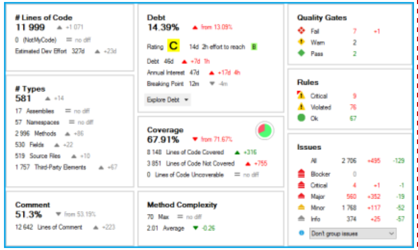


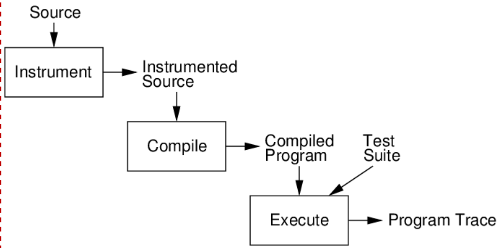
1960s

1970s

Vulnerability Detection Advancement

Phase 1	Phase 2	Phase 3
Manual	**Rule**	**Dynamic**
Code Review, Reverse Engineering, Expertise	Static/Taint Analysis, Model Checking	Fuzzing, Symbolic Execution

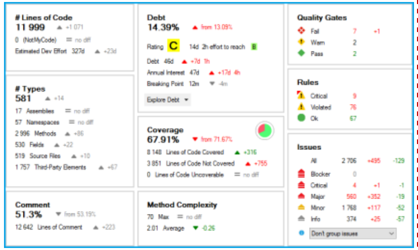


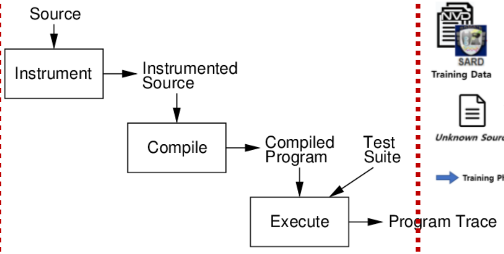


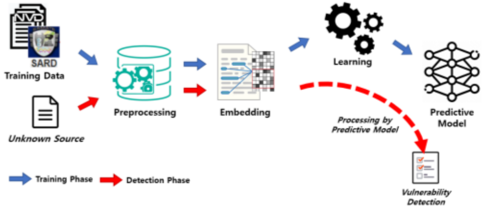
1960s → 1970s → 1990s

Vulnerability Detection Advancement

Phase 1	Phase 2	Phase 3	Phase 4
Manual	**Rule**	**Dynamic**	**Intelligent**
Code Review, Reverse Engineering, Expertise	Static/Taint Analysis, Model Checking	Fuzzing, Symbolic Execution	Machine/Deep Learning-Assisted





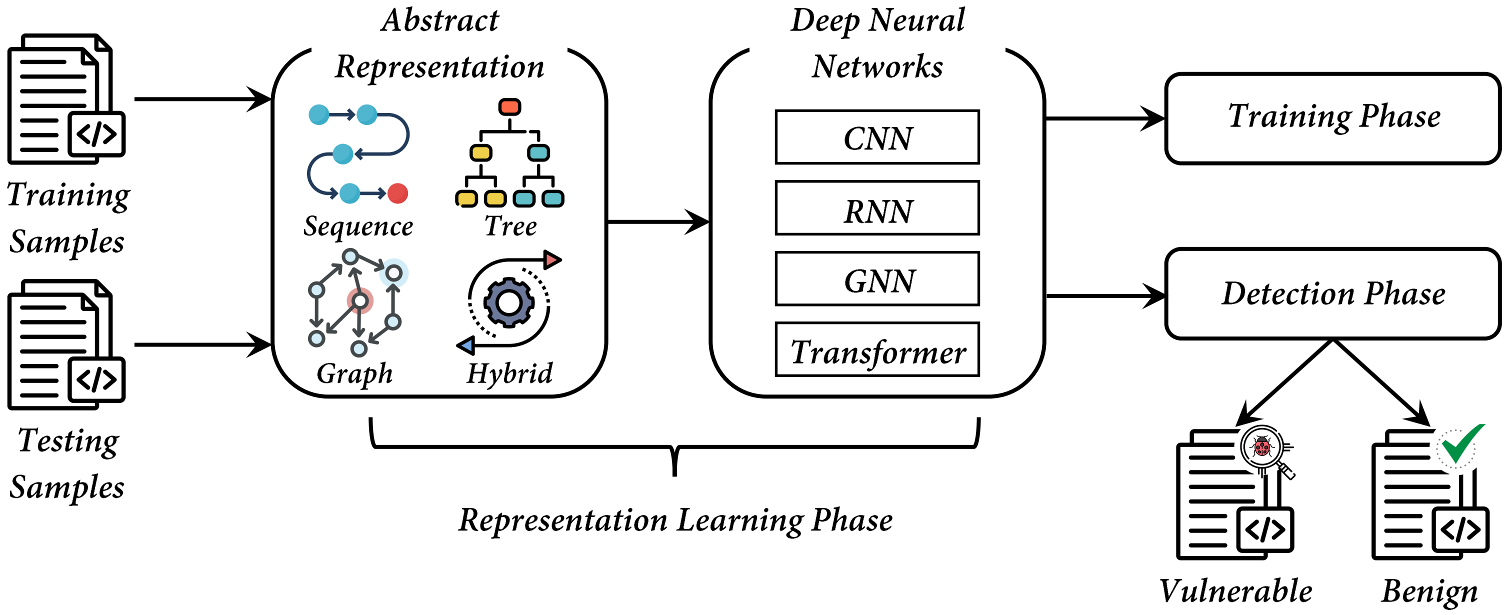


1960s 1970s 1990s 2013s

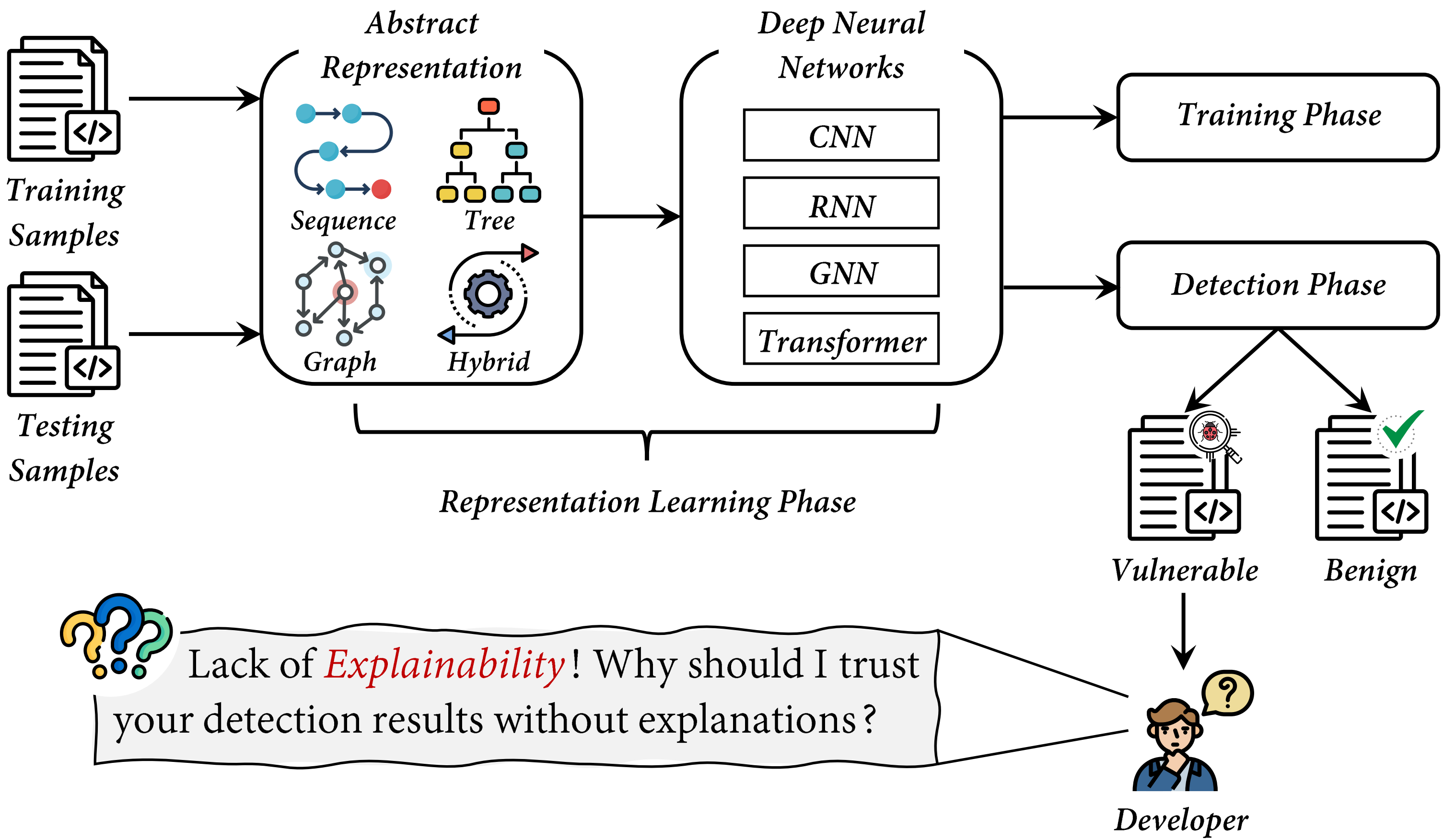
knowledge-Intensive, High FP, Poor scalability

High FN、 Low Coverage

DL-based VD Workflow



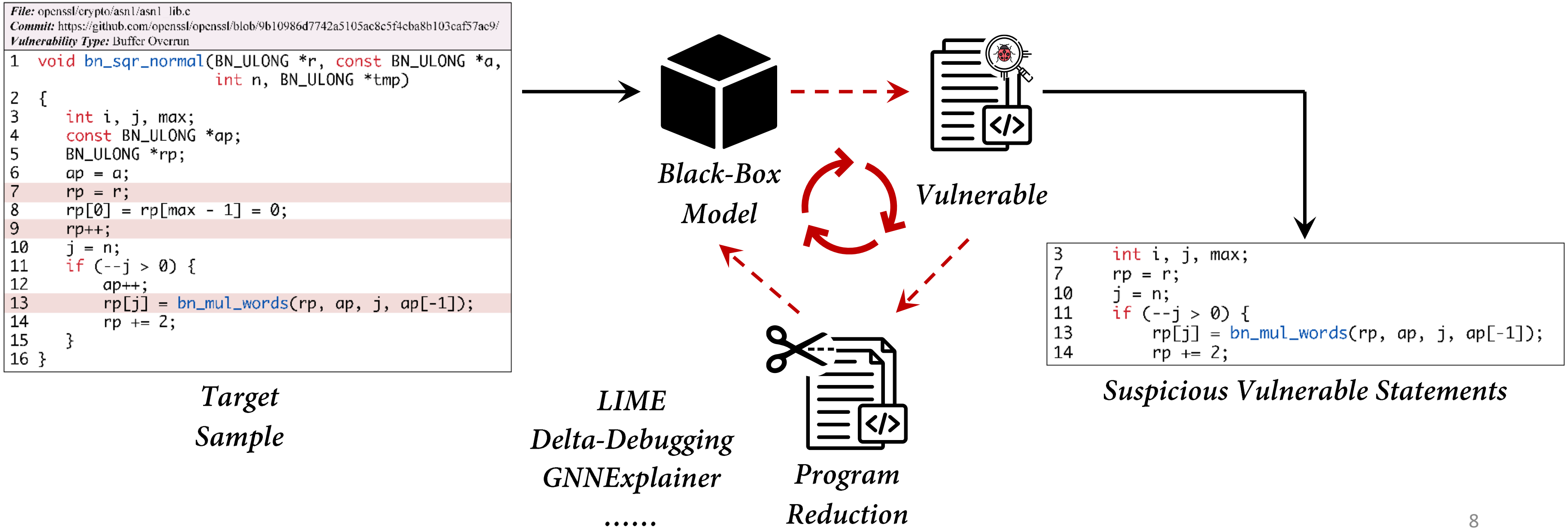
DL-based VD Workflow



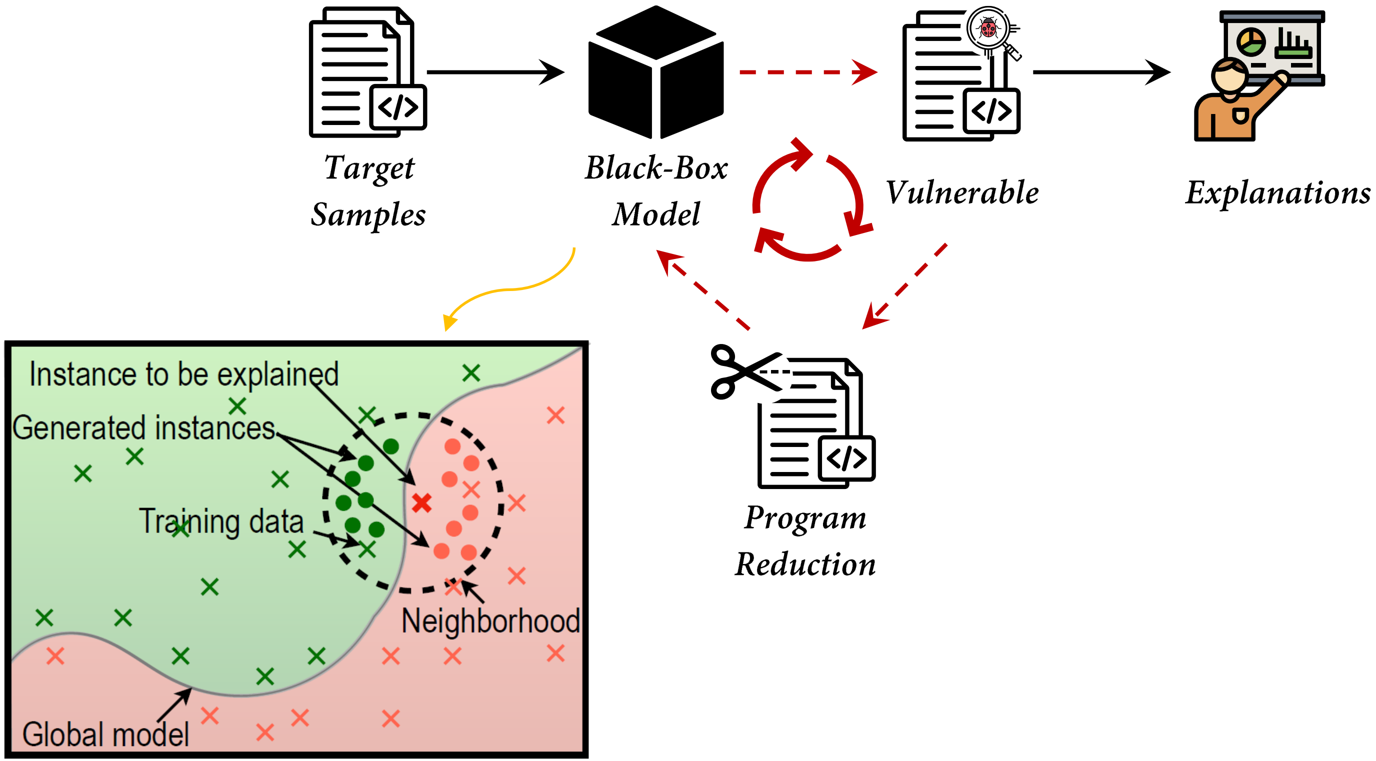
Explainable VD Workflow

Definition 1

Given an input program $P = \{s_1, \cdots, s_m\}$ which is detected as vulnerable, the explanation is a set of crucial statements $\{s_i, \cdots, s_j\}$ that are most relevant to the decision of the model.

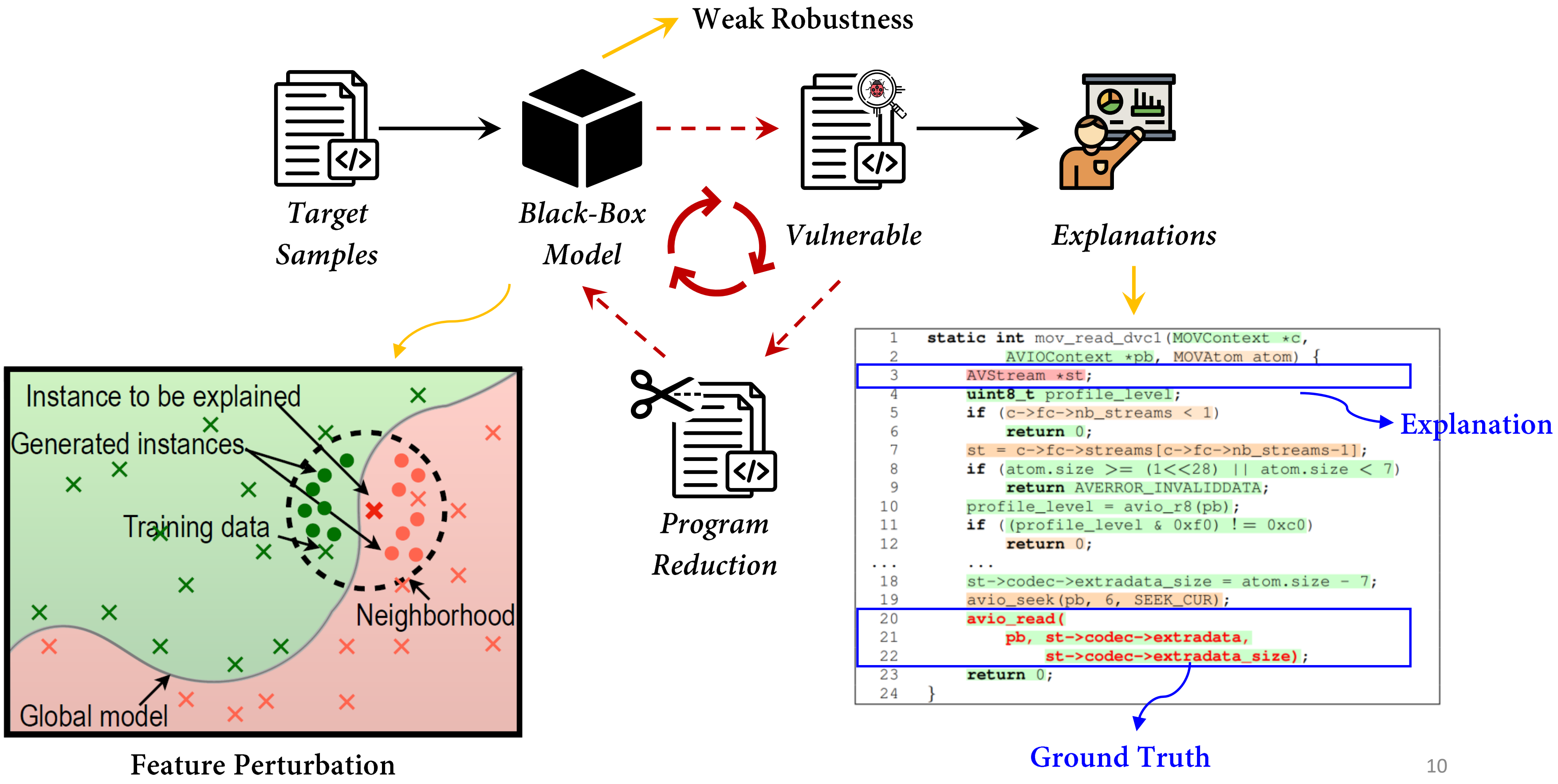


Challenge of Explainable VD



Feature Perturbation

Challenge of Explainable VD

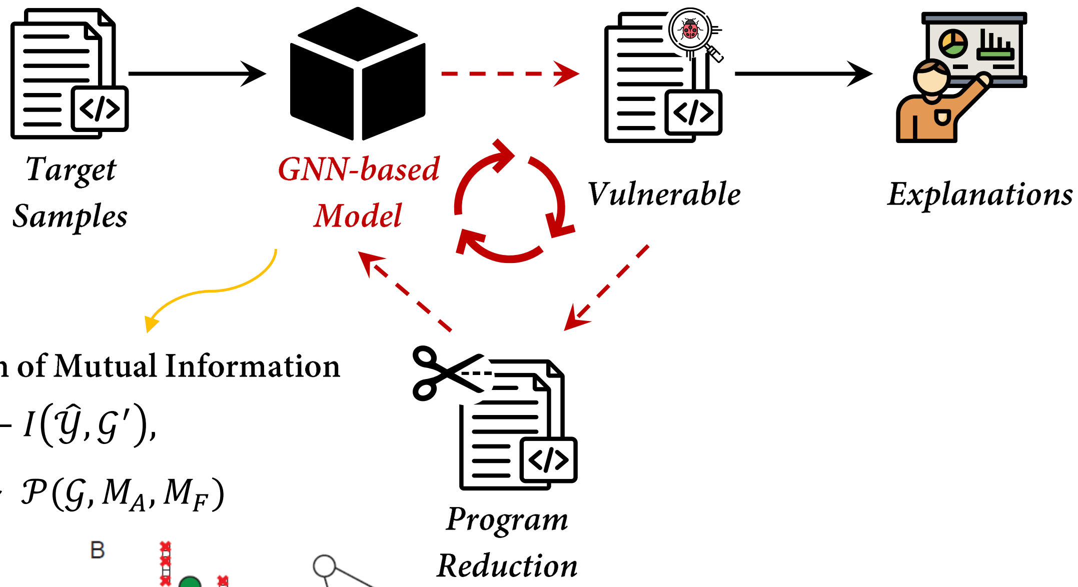


Challenge of Explainable VD

Challenge 1

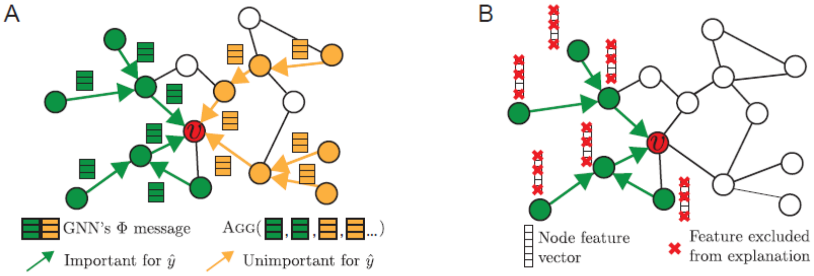
Due to the weak robustness of existing DL-based vulnerability detectors, their explanations are easy to be altered due to small perturbations, or even random noise. As a result, explanations built on top of the detection results from such weakly-robust models just reveal spurious correlations, which are hard to be tolerated by security applications.

Challenge of Explainable VD

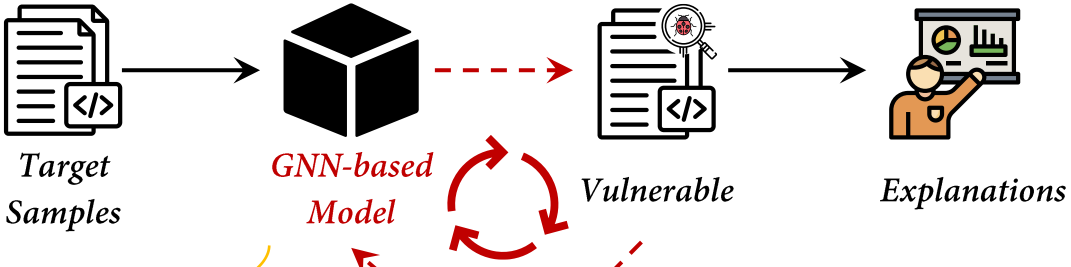


Maximization of Mutual Information

$$\min_{\mathcal{G}'} -I(\hat{\mathcal{Y}}, \mathcal{G}'),$$
$$s.t. \quad \mathcal{G}' \sim \mathcal{P}(\mathcal{G}, M_A, M_F)$$



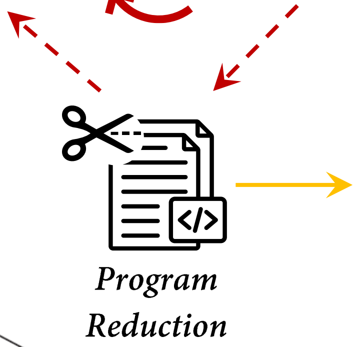
Challenge of Explainable VD

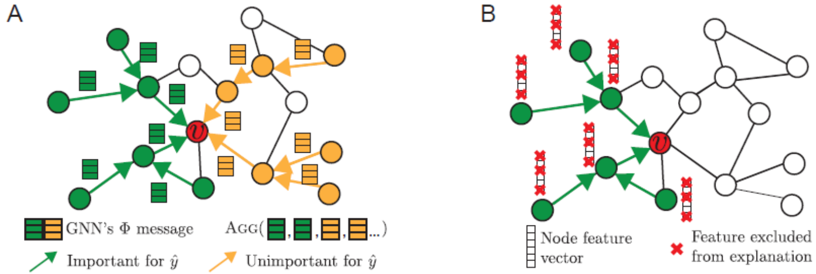


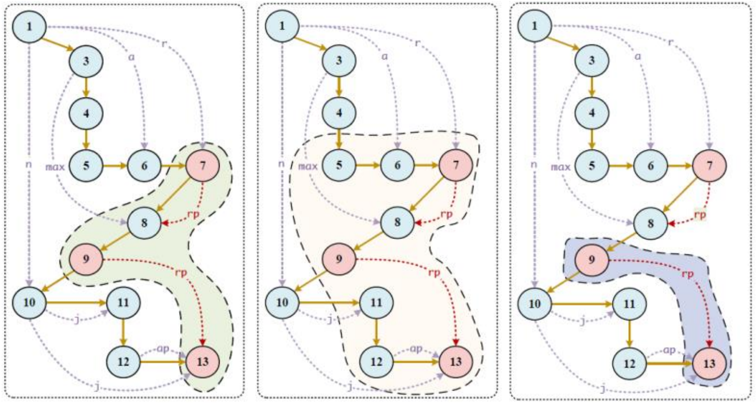
Maximization of Mutual Information

$$\min_{\mathcal{G}'} -I(\hat{\mathcal{Y}}, \mathcal{G}'),$$

$$s.t. \quad \mathcal{G}' \sim \mathcal{P}(\mathcal{G}, M_A, M_F)$$

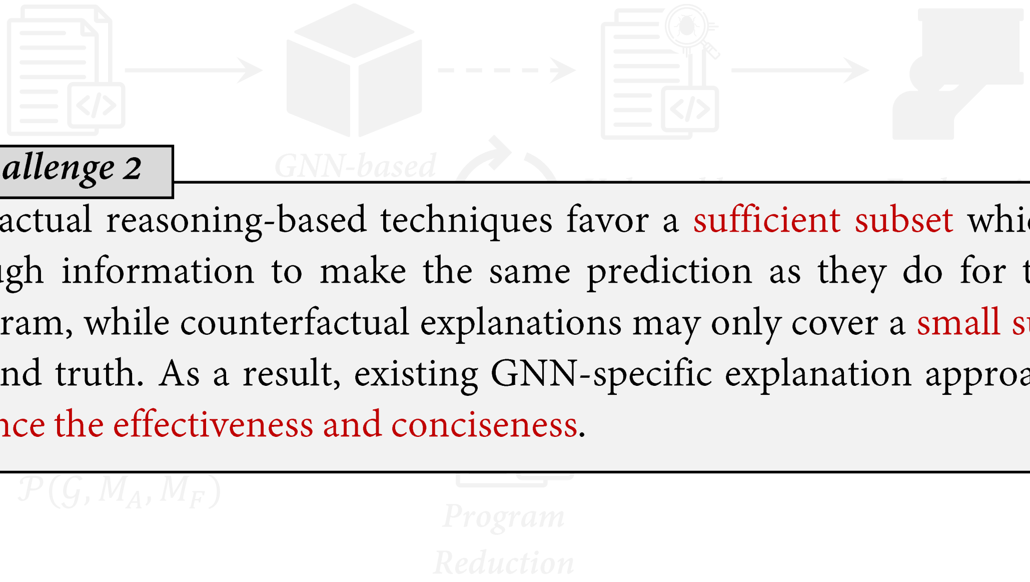






(a) CocA_{Exp} (b) GNNExplainer (c) CF-GNNExplainer

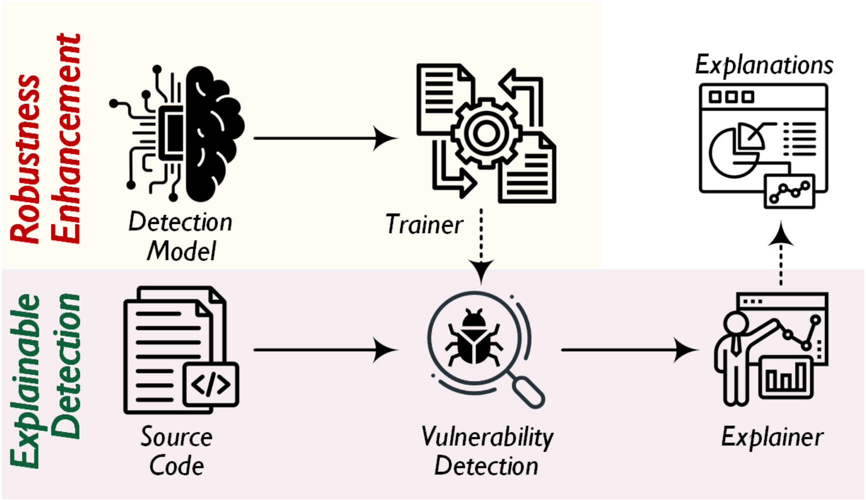
Challenge of Explainable VD



Challenge 2

Factual reasoning-based techniques favor a sufficient subset which contains enough information to make the same prediction as they do for the original program, while counterfactual explanations may only cover a small subset of the ground truth. As a result, existing GNN-specific explanation approaches fail to balance the effectiveness and conciseness.

Our approach: COCA



Workflow of COCA

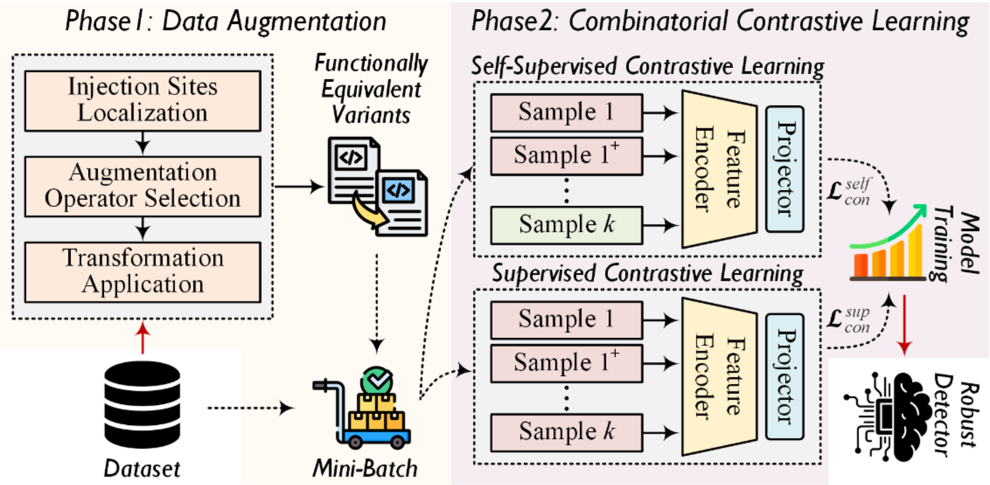
A General Optimization Framework for GNN-based Explainable Vulnerability Detection

- *Combinatorial Contrastive Learning-based Robustness Enhancement*
- *Vulnerability Explanation via Dual-View Causal Inference*

Our approach: CoCa

*A **General Optimization Framework** for GNN-based **Explainable** Vulnerability Detection*

- ◼ *Combinatorial Contrastive Learning-based Robustness Enhancement*
- ◻ *Vulnerability Explanation via Dual-View Causal Inference*



*How to **enhance the robustness** of Classifier against **random perturbations**?*

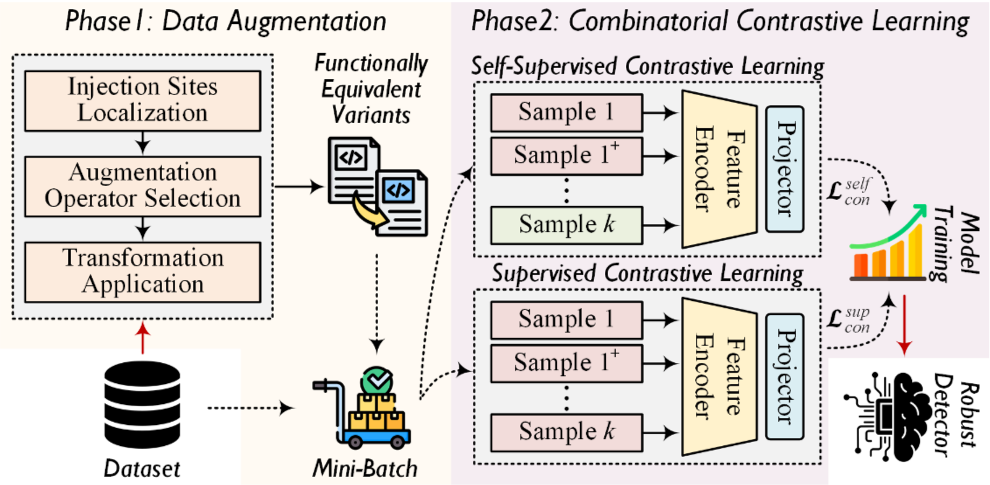
- Perform data augmentation to construct functionally-equivalent variants.

No.	Name	Description
1	Identifier Renaming	Substitute the function/variable name with a random token.
2	Operand Swap	Swap the operands of binary logical operations.
3	Statement Permutation	Swap two lines of statements that have no dependency.
4	Loop Exchange	Replace for loops with while loops or vice versa.
5	Block Swap	Swap `then` block of a chosen `if` statement with its corresponding `else` block.
6	Switch to If	Replace a `switch` statement with its equivalent `if` statement.

Our approach: COCA

*A **General Optimization Framework** for GNN-based **Explainable** Vulnerability Detection*

- ☐ *Combinatorial Contrastive Learning-based Robustness Enhancement*
- ☐ *Vulnerability Explanation via Dual-View Causal Inference*



How to enhance the robustness of Classifier against random perturbations?

- ➢ Perform data augmentation to construct functionally-equivalent variants.
- ➢ Combine self-supervised contrastive learning with supervised contrastive learning to optimize the learned feature representations.

$$\mathcal{L}_{total} = (1-\lambda)\mathcal{L}_{con}^{self} + \lambda\mathcal{L}_{con}^{sup}$$

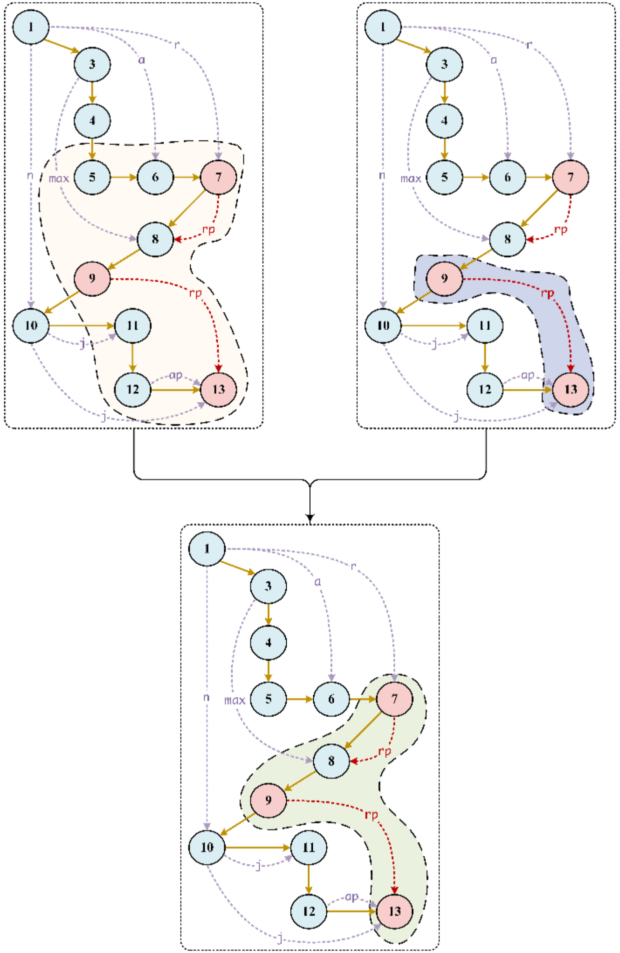
Our approach: COCA

A General Optimization Framework for GNN-based Explainable Vulnerability Detection

- ☐ *Combinatorial Contrastive Learning-based Robustness Enhancement*
- ☐ *Vulnerability Explanation via Dual-View Causal Inference*

How to make a trade-off between effectiveness and conciseness?

➢ Combine factual inference with counterfactual inference to search the explanation subgraph.



Our approach: COCA

*A **General Optimization Framework** for GNN-based **Explainable** Vulnerability Detection*

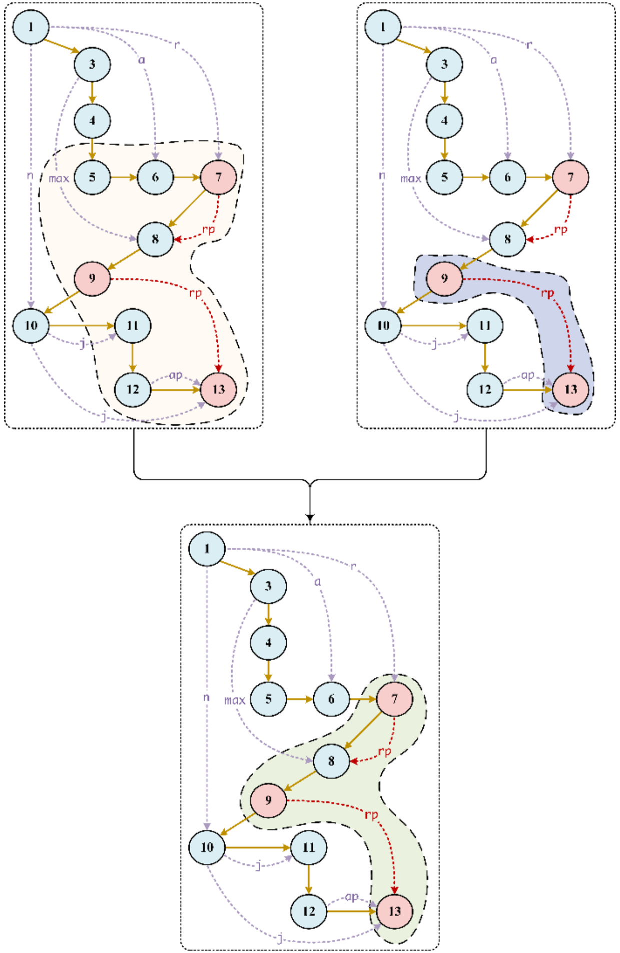
- ☐ *Combinatorial Contrastive Learning-based Robustness Enhancement*
- ☐ *Vulnerability Explanation via **Dual-View Causal Inference***

*How to make a trade-off between **effectiveness** and **conciseness**?*

$$
\begin{aligned}
&\text{minimize } C(M_k, F_k) \\
&\text{subject to} \\
&S_f(M_k, F_k) > P(\hat{y}_{k,s} | A_k \odot M_k, X_k \odot F_k) \\
&S_c(M_k, F_k) > -P(\hat{y}_{k,s} | A_k - A_k \odot M_k, X_k - X_k \odot F_k)
\end{aligned}
$$

Factual Inference

Counterfactual Inference



Performance of COCA

Dataset

Dataset	# Vul	# Non-vul	# Total	% Ratio
Devign	11,888	14,149	26,037	45.66
ReVeal	1,664	16,505	18,169	9.16
Big-Vul	11,823	253,096	264,919	4.46
CrossVul	6,884	127,242	134,126	5.13
CVEFixes	8,932	159,157	168,089	5.31
Merged	**29,844**	**305,827**	**335,671**	**8.89**

Detection Performance

Config	Loss	Approach	Acc	Pre	Rec	F1
Default	CE	Devign	**89.74**	32.59	31.40	31.98
		ReVeal	86.05	31.43	38.45	34.59
		DeepWuKong	87.21	28.55	26.04	27.24
COCA $_{Tra}$	Ours	Devign	88.15	34.68	37.12	35.86
		ReVeal	87.42	**35.96**	**40.61**	**38.14**
		DeepWuKong	88.30	30.07	34.79	32.26
	InfoNCE	Devign	86.33	28.38	30.11	29.22
		ReVeal	84.95	29.64	34.27	31.78
		DeepWuKong	86.20	25.99	24.83	25.40
	NCE	Devign	83.97	26.15	27.69	26.90
		ReVeal	81.52	26.73	31.76	29.03
		DeepWuKong	83.06	22.40	21.46	21.92

Baselines

- Devign (NeurIPS'19)
- ReVeal (TSE'21)
- DeepWuKong (TOSEM'21)

Performance of CoCA

Dataset

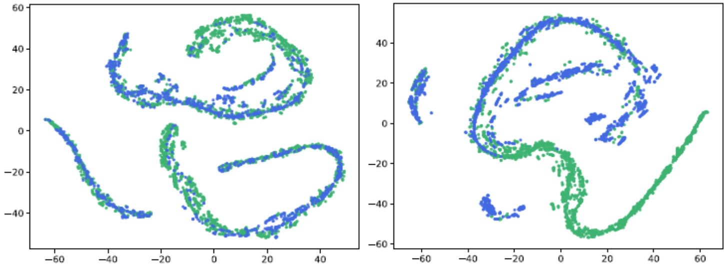
Dataset	# Vul	# Non-vul	# Total	% Ratio
Devign	11,888	14,149	26,037	45.66
ReVeal	1,664	16,505	18,169	9.16
Big-Vul	11,823	253,096	264,919	4.46
CrossVul	6,884	127,242	134,126	5.13
CVEFixes	8,932	159,157	168,089	5.31
Merged	**29,844**	**305,827**	**335,671**	**8.89**

Baselines

- Devign (NeurIPS'19)
- ReVeal (TSE'21)
- DeepWuKong (TOSEM'21)

Detection Performance

Config	Loss	Approach	Acc	Pre	Rec	F1
Default	CE	Devign	**89.74**	32.59	31.40	31.98
		ReVeal	86.05	31.43	38.45	34.59
		DeepWuKong	87.21	28.55	26.04	27.24
CoCA_{Tra}	Ours	Devign	88.15	34.68	37.12	35.86
		ReVeal	87.42	**35.96**	**40.61**	**38.14**
		DeepWuKong	88.30	30.07	34.79	32.26
	InfoNCE	Devign	86.33	28.38	30.11	29.22
		ReVeal	84.95	29.64	34.27	31.78
		DeepWuKong	86.20	25.99	24.83	25.40
	NCE	Devign	83.97	26.15	27.69	26.90
		ReVeal	81.52	26.73	31.76	29.03
		DeepWuKong	83.06	22.40	21.46	21.92



(a) DeepWuKong (Default) (b) DeepWuKong (CoCA_{Tra})

Performance of CoCA

Dataset

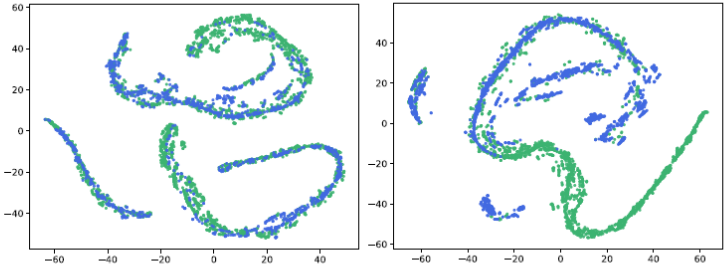
Dataset	# Vul	# Non-vul	# Total	% Ratio
Devign	11,888	14,149	26,037	45.66
ReVeal	1,664	16,505	18,169	9.16
Big-Vul	11,823	253,096	264,919	4.46
CrossVul	6,884	127,242	134,126	5.13
CVEFixes	8,932	159,157	168,089	5.31
Merged	**29,844**	**305,827**	**335,671**	**8.89**

Baselines

- Devign (NeurIPS'19)
- ReVeal (TSE'21)
- DeepWuKong (TOSEM'21)

Detection Performance

Config	Loss	Approach	Acc	Pre	Rec	F1
Default	CE	Devign	**89.74**	32.59	31.40	31.98
		ReVeal	86.05	31.43	38.45	34.59
		DeepWuKong	87.21	28.55	26.04	27.24
CoCA_{Tra}	Ours	Devign	88.15	34.68	37.12	35.86
		ReVeal	87.42	**35.96**	**40.61**	**38.14**
		DeepWuKong	88.30	30.07	34.79	32.26
	InfoNCE	Devign	86.33	28.38	30.11	29.22
		ReVeal	84.95	29.64	34.27	31.78
		DeepWuKong	86.20	25.99	24.83	25.40
	NCE	Devign	83.97	26.15	27.69	26.90
		ReVeal	81.52	26.73	31.76	29.03
		DeepWuKong	83.06	22.40	21.46	21.92



(a) DeepWuKong (Default) (b) DeepWuKong (CoCA_{Tra})

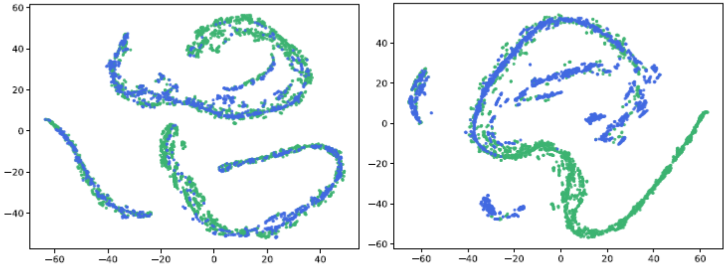
Performance of CoCA

Dataset

Dataset	# Vul	# Non-vul	# Total	% Ratio
Devign	11,888	14,149	26,037	45.66
ReVeal	1,664	16,505	18,169	9.16
Big-Vul	11,823	253,096	264,919	4.46
CrossVul	6,884	127,242	134,126	5.13
CVEFixes	8,932	159,157	168,089	5.31
Merged	**29,844**	**305,827**	**335,671**	**8.89**

Detection Performance

Config	Loss	Approach	Acc	Pre	Rec	F1
Default	CE	Devign	**89.74**	32.59	31.40	31.98
		ReVeal	86.05	31.43	38.45	34.59
		DeepWuKong	87.21	28.55	26.04	27.24
CoCA_{Tra}	Ours	Devign	88.15	34.68	37.12	35.86
		ReVeal	87.42	**35.96**	**40.61**	**38.14**
		DeepWuKong	88.30	30.07	34.79	32.26
	InfoNCE	Devign	86.33	28.38	30.11	29.22
		ReVeal	84.95	29.64	34.27	31.78
		DeepWuKong	86.20	25.99	24.83	25.40
	NCE	Devign	83.97	26.15	27.69	26.90
		ReVeal	81.52	26.73	31.76	29.03
		DeepWuKong	83.06	22.40	21.46	21.92



(a) DeepWuKong (Default)

(b) DeepWuKong (CoCA_{Tra})

Baselines

- Devign (NeurIPS'19)
- ReVeal (TSE'21)
- DeepWuKong (TOSEM'21)

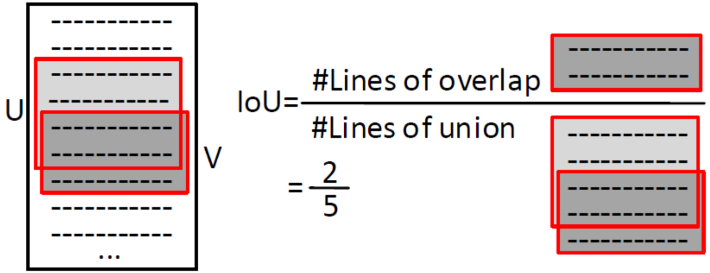
Performance of Coca

Baselines

- mVulPreter (TDSC'22)
- IVDetect (ESEC/FSE'21)
- P2IM (ESEC/FSE'21)

Evaluation Metrics

- Mean Statement Precision (MSP)
- Mean Statement Recall (MSR)
- Mean Intersection over Union (MIoU)



Explanation Performance

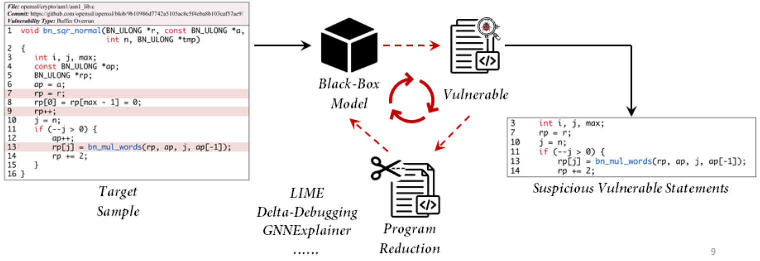
Config	Approach	MSP	MSR	MIoU
Default	mVulPreter	25.86	29.01	22.88
	IVDetect	32.54	23.79	17.06
	P2IM (Devign)	27.99	43.85	22.56
	P2IM (ReVeal)	31.04	46.10	28.94
	P2IM (DeepWuKong)	26.57	38.12	23.11
	Coca$_{Exp}$ (Devign)	33.84	44.06	30.89
	Coca$_{Exp}$ (ReVeal)	35.61	52.94	34.36
	Coca$_{Exp}$ (DeepWuKong)	29.77	40.16	25.83
Coca$_{Tra}$	IVDetect	39.81	31.64	25.19
	P2IM (Devign)	33.01	48.33	29.27
	P2IM (ReVeal)	40.62	55.73	36.29
	P2IM (DeepWuKong)	32.97	44.85	28.10
	Coca$_{Exp}$ (Devign)	43.61	52.98	39.64
	Coca$_{Exp}$ (ReVeal)	**49.52**	**58.39**	**44.97**
	Coca$_{Exp}$ (DeepWuKong)	40.33	47.61	34.22

Conclusion

Explainable VD Workflow

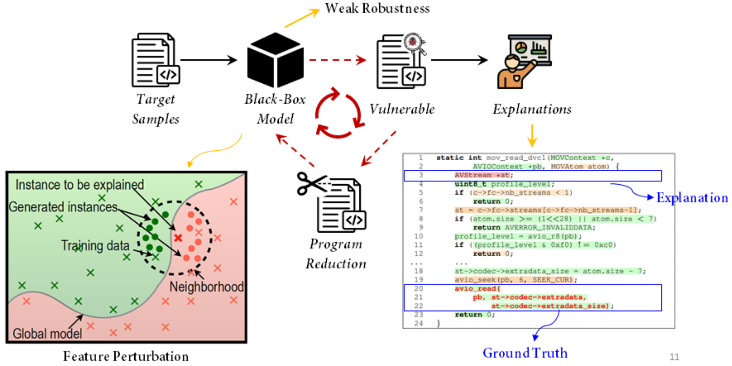
Definition 1

Given an input program $P = \{s_1, \cdots, s_m\}$ which is detected as vulnerable, the explanation is a set of crucial statements $\{s_i, \cdots, s_j\}$ that are most relevant to the decision of the model.



9

Challenge of Explainable VD



11

Our approach: CoCa

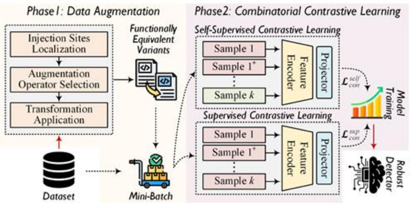
A General Optimization Framework for GNN-based Explainable Vulnerability Detection

- *Combinatorial Contrastive Learning-based Robustness Enhancement*
- *Vulnerability Explanation via Dual-View Causal Inference*

How to enhance the robustness of Classifier against random perturbations?

➢ Perform data augmentation to construct functionally-equivalent variants.

No.	Name	Description
1	Identifier Renaming	Substitute the function/variable name with a random token.
2	Operand Swap	Swap the operands of binary logical operations.
3	Statement Permutation	Swap two lines of statements that have no dependency.
4	Loop Exchange	Replace for loops with while loops or vice versa.
5	Block Swap	Swap `then` block of a chosen `if` statement with its corresponding `else` block.
6	Switch to If	Replace a `switch` statement with its equivalent `if` statement.



9

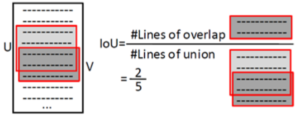
Performance of CoCa

Baselines

- mVulPreter (TDSC'22)
- IVDetect (ESEC/FSE'21)
- P2IM (ESEC/FSE'21)

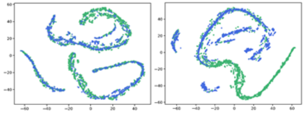
Evaluation Metrics

- Mean Statement Precision (MSP)
- Mean Statement Recall (MSR)
- Mean Intersection over Union (MIoU)



Explanation Performance

Config	Approach	MSP	MSR	MIoU
Default	mVulPreter	25.86	29.01	22.88
	IVDetect	32.54	23.79	17.06
	P2IM (Devign)	27.99	43.85	22.56
	P2IM (ReVeal)	31.04	46.10	28.94
	P2IM (DeepWuKong)	26.57	38.12	23.11
	CoCa_{Exp} (Devign)	33.84	44.06	30.89
	CoCa_{Exp} (ReVeal)	35.61	52.94	34.36
	CoCa_{Exp} (DeepWuKong)	29.77	40.16	25.83
CoCa_{Tra}	IVDetect	39.81	31.64	25.19
	P2IM (Devign)	33.01	48.33	29.27
	P2IM (ReVeal)	40.62	55.73	36.29
	P2IM (DeepWuKong)	32.97	44.85	28.10
	CoCa_{Exp} (Devign)	43.61	52.98	39.64
	CoCa_{Exp} (ReVeal)	**49.52**	**58.39**	**44.97**
	CoCa_{Exp} (DeepWuKong)	40.33	47.61	34.22



(a) DeepWuKong (Default) (b) DeepWuKong (CoCa_{Tra})

13

46th International Conference on Software Engineering

Thanks for listening !

sicongcao1996@gmail.com

https://github.com/CocaVul/Coca

Paper

Artifact

揚州大學
YANGZHOU UNIVERSITY

SINGAPORE
MANAGEMENT
UNIVERSITY