AI+ 软件研发数字峰会
AI+ software Development Digital summit

AI驱动软件研发
全面进入数字化时代

中国·深圳 11.24-25

AI+
software
Development
Digital
summit

助力基于AI生态的软件开发:
深度学习模型训练过程的可视化解释与调试

林云 上海交通大学计算机系

科技生态圈峰会 + 深度研习

——1000 + 技术团队的共同选择

KEYLINK ing

K+峰会

K+峰会 深圳站

K+ 全球软件研发行业创新峰会

会议时间：2024.05.24-25

K+峰会 上海站

K+ 全球软件研发行业创新峰会

会议时间：2024.09.20-21

AiDD 峰会

AiDD峰会 深圳站

AI+ 软件研发数字峰会

会议时间：2023.11.24-25

AiDD峰会 北京站

AI+ 软件研发数字峰会

会议时间：2024.07.19-20

AiDD峰会 深圳站

AI+ 软件研发数字峰会

会议时间：2024.11.15-16

演讲嘉宾

林云

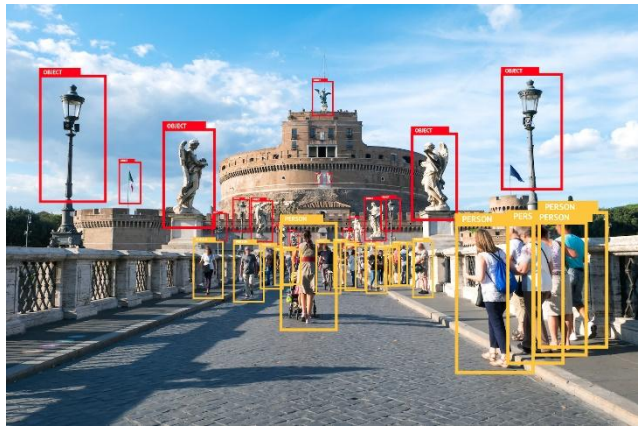
上海交通大学计算机系副教授，博士生导师

林云，上海交通大学计算机系副教授，博士生导师，原新加坡国立大学助理教授（研究岗），入选2021年国家海外高层次青年人才计划。主要研究领域为软件工程，侧重代码、网页和AI模型的自动分析技术。在ICSE、FSE、USENIX Security、ISSTA、ASE、 NeurIPS、AAAI、IJCAI、KDD、TSE、TDSC等领域相关的国际顶级会议和期刊发表论文40余篇，国内外专利受理2项。担任PRDC2023国际会议程序委员会联合主席，以及FSE、USENIX Security, ISSTA、ICML、NeurIPS、AAAI等重要国际会议的程序委员会委员、IEEE TSE/ACM TOSEM/IEEE TDSC等顶级期刊的审稿人。主持国家基金委优青项目（海外）。获得过ICSE2018最佳论文奖。

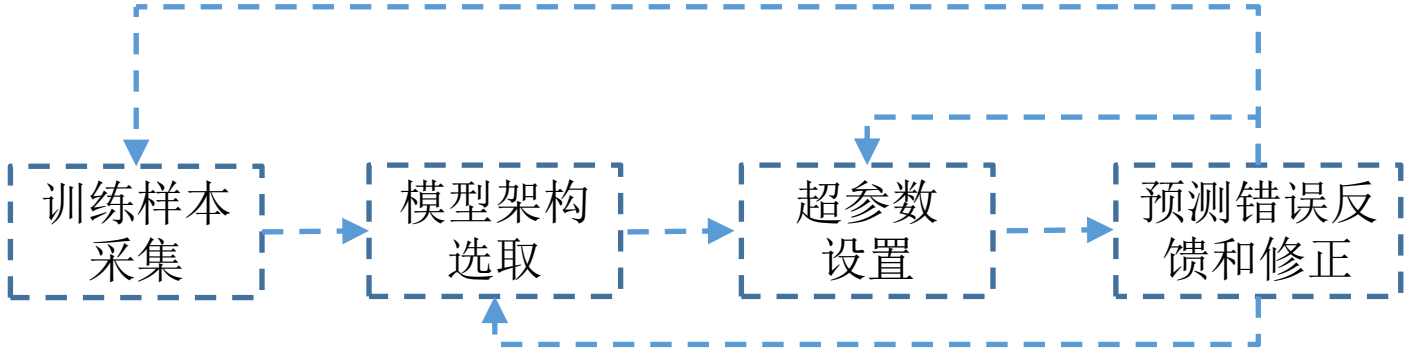
软件开发新形态

自动驾驶

图像识别

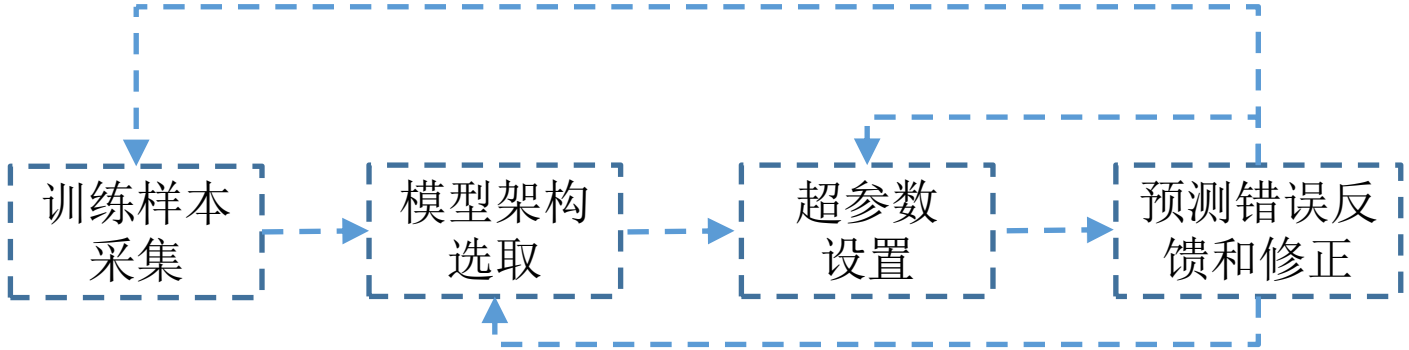


大语言模型插件
商店 （~900）



软件工程技术

- 代码调试
- 软件测试
- 变更影响分析

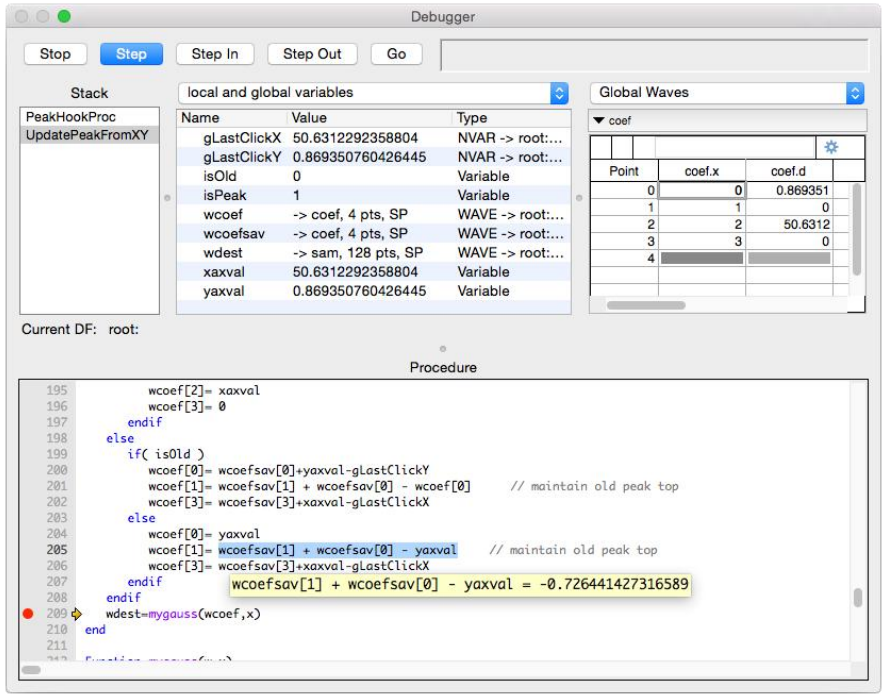


面向AI生态的软件工程技术

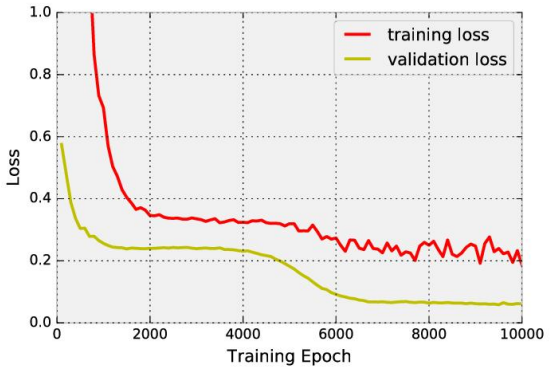
- 模型调试
- 模型测试
- 样本影响函数

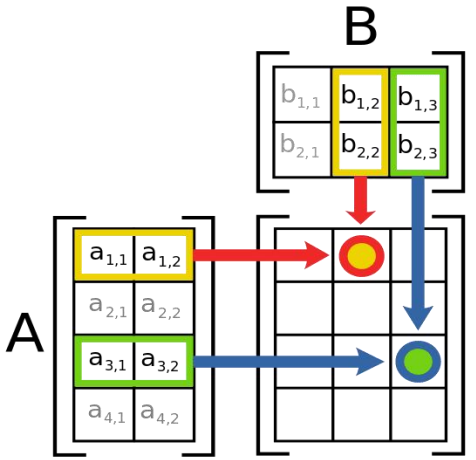
基本问题：模型的预测结果是如何一步一步产生的?

传统软件的调试过程



深度神经网络训练过程





- 可观测性问题
- 意图检测问题

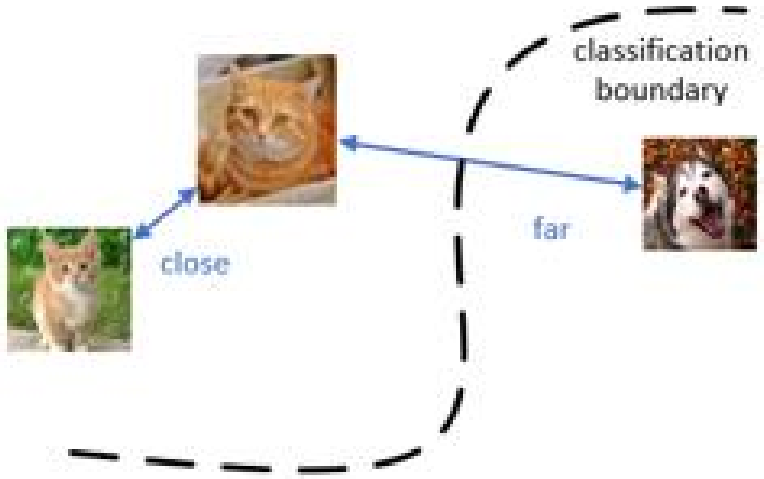
可观测性问题

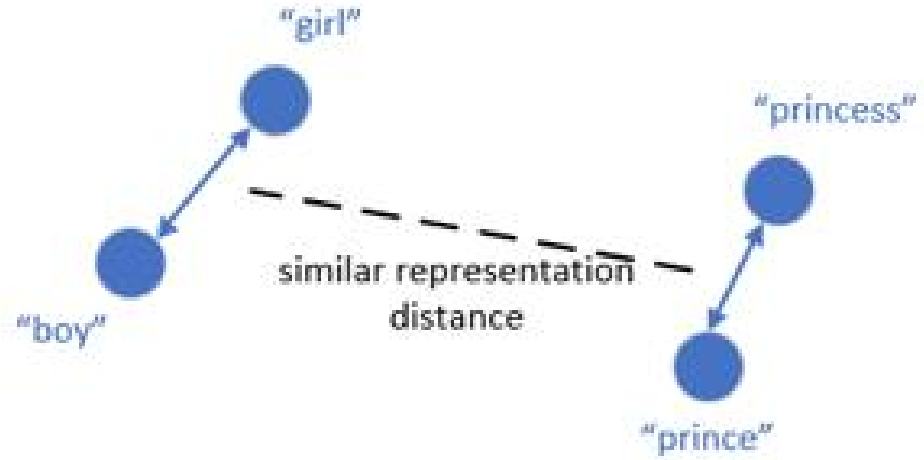
- 什么样的训练信息需要被观察?
- 这些信息如何获得和提炼?

意图检测问题

- 模型开发人员有哪些意图?
- 如果用轻量级地方式来侦测和推断这些意图?

可观测性思路：深度学习即表征学习

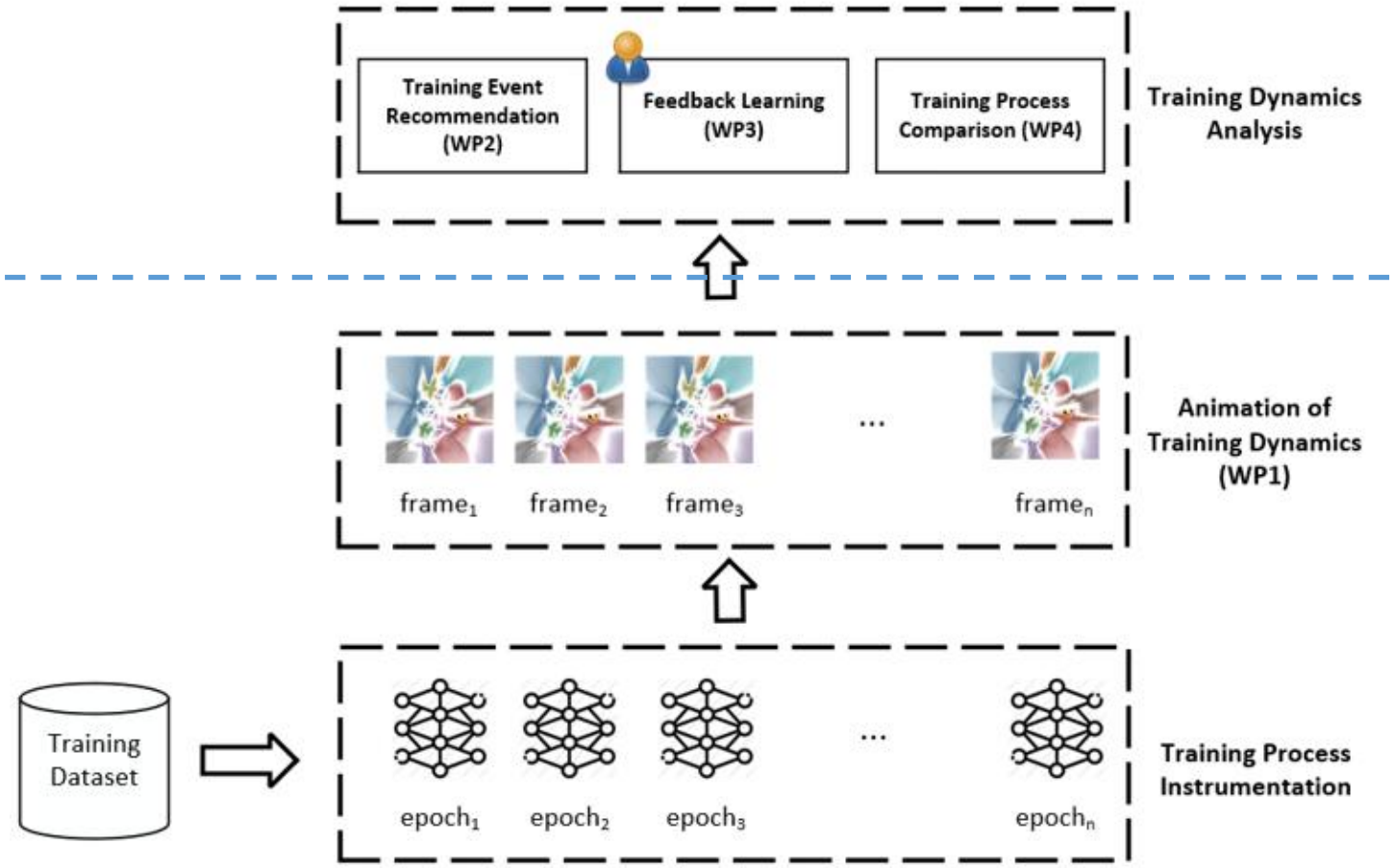




可视化模型训练调试框架示意图

意图检测

可观测性



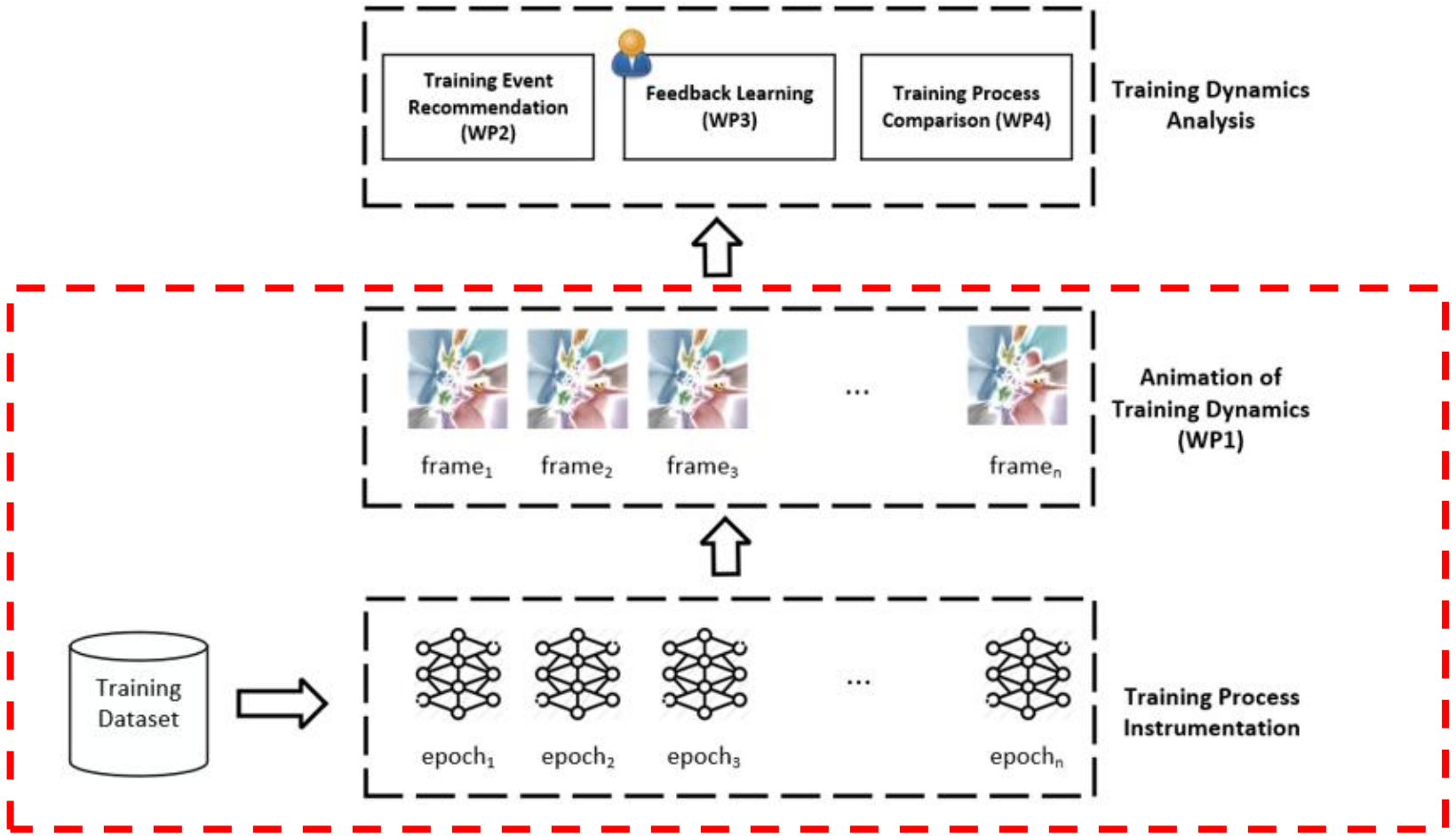
FSE’23
NeurIPS’22

AAAI’22,
IJCAI’22
ChinaSoft’22
（原型竞赛一等奖）

可视化模型训练调试框架示意图

意图检测

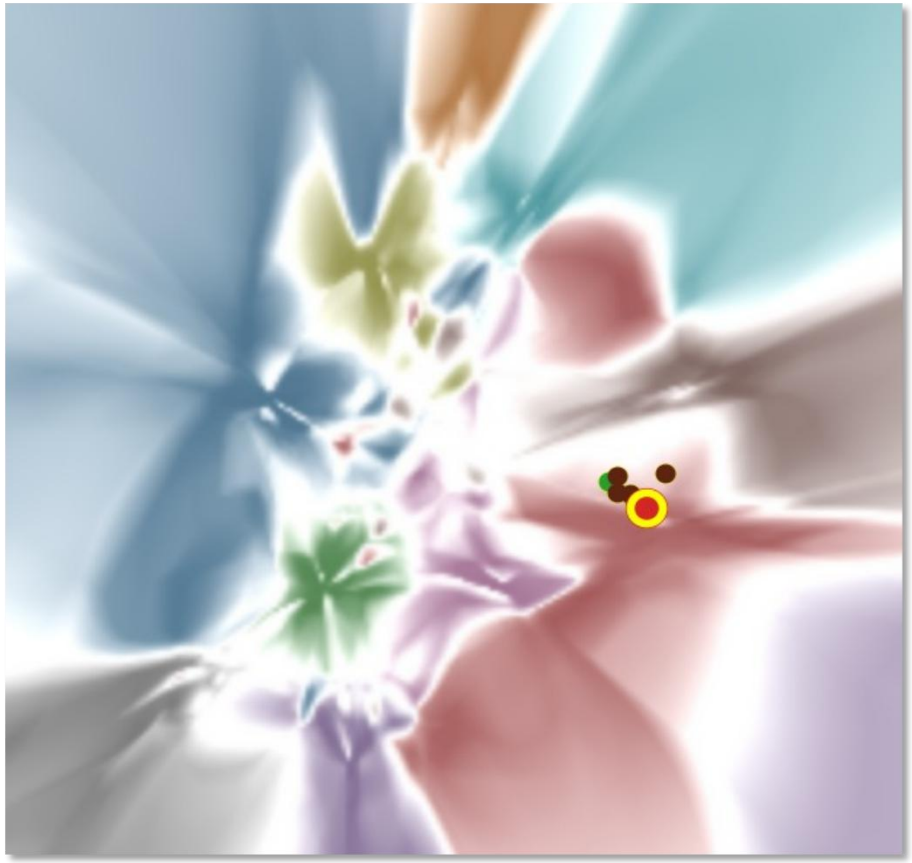
可观测性



FSE’23
NeurIPS’22

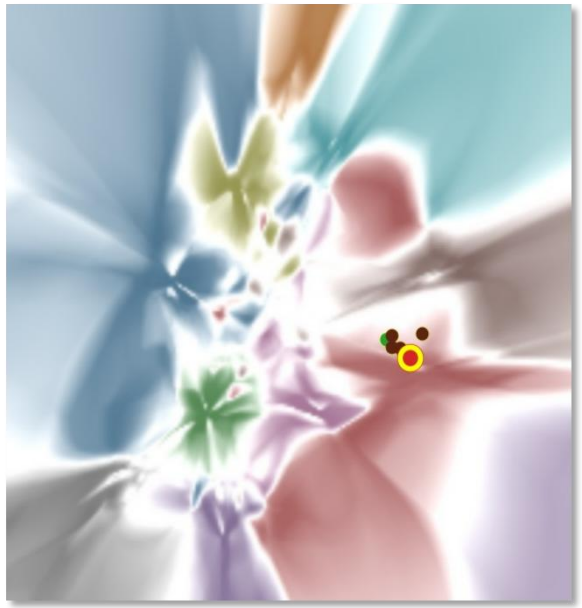
AAAI’22,
IJCAI’22
ChinaSoft’22
（原型竞赛一等奖）

DeepVisualInsight (DVI): Time-travelling Visualization for Deep Classifier Training

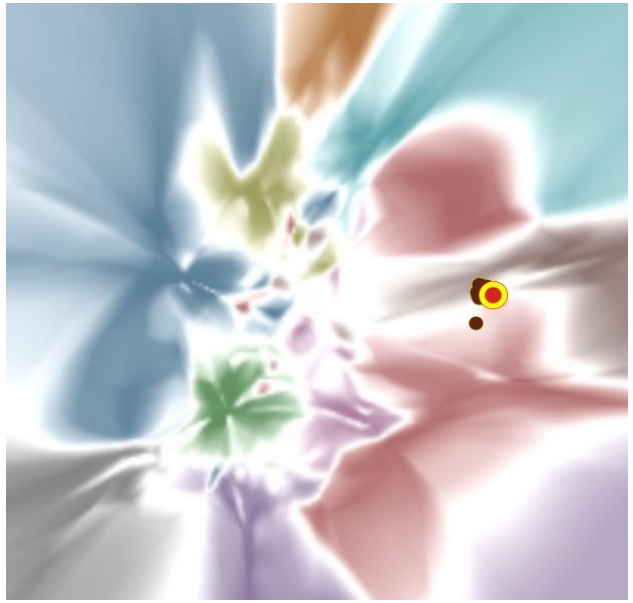


Time-travelling Visualization for Deep Classifier Training

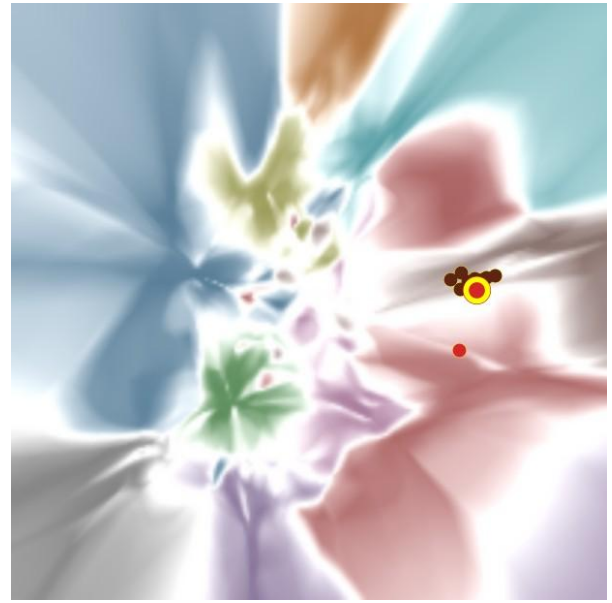
Epoch 1



Epoch 2



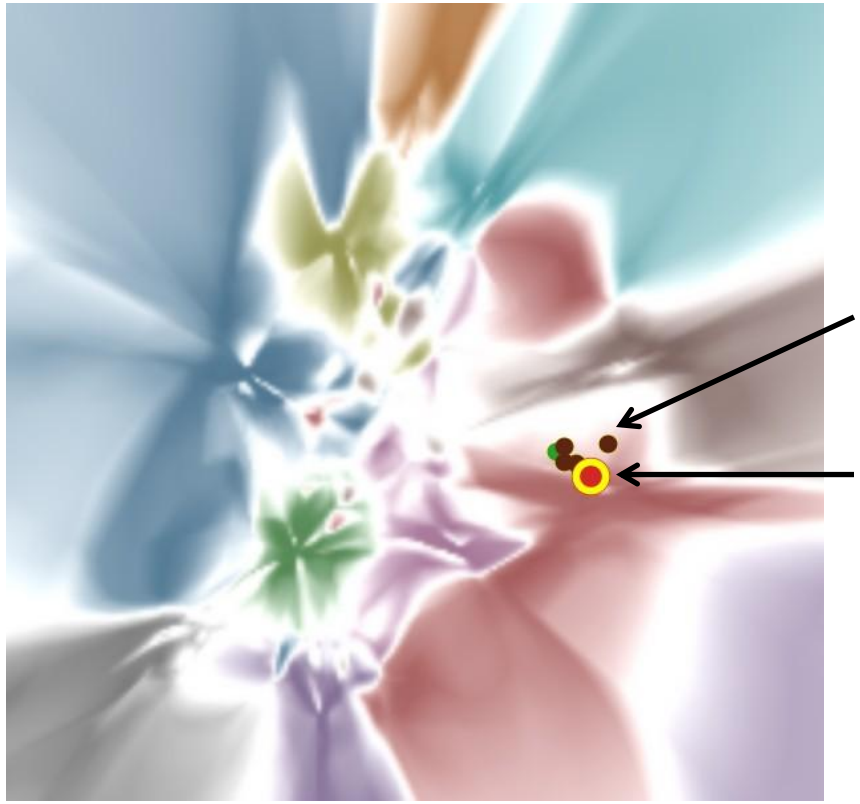
Epoch 3



Motivating example: how adversarial training cause the performance degeneration?

	Accuracy
Adversarial Samples	51.3%
Testing Samples	92.3%

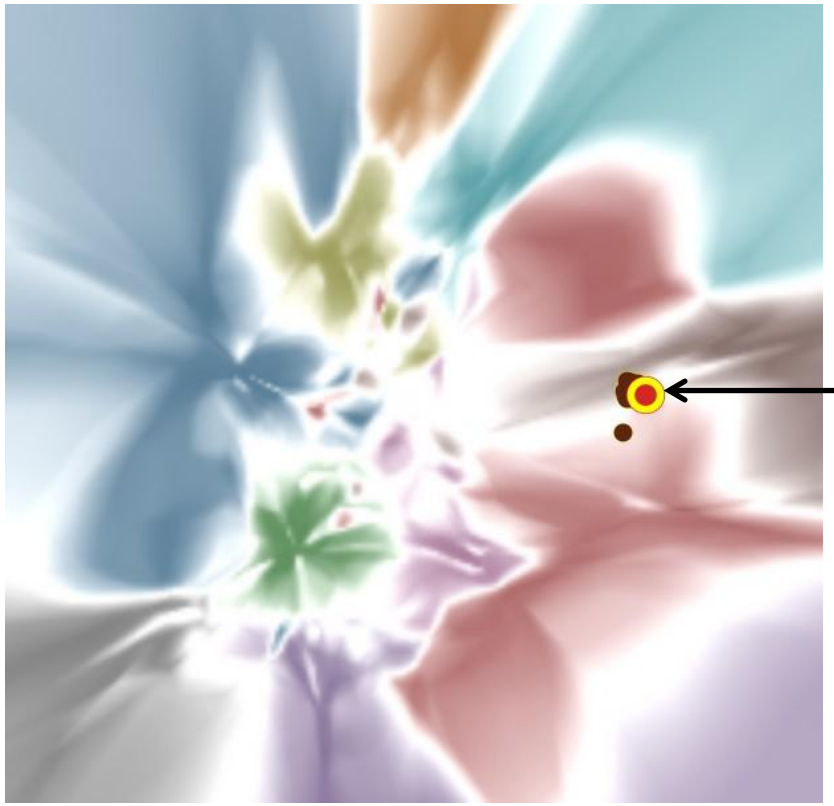
Epoch 1



Motivating example: how adversarial training cause the performance degeneration?

	Accuracy	
Adversarial Samples	67.8%	↑ 16.5%
Testing Samples	90.3%	↓ 2.0%

Epoch 2

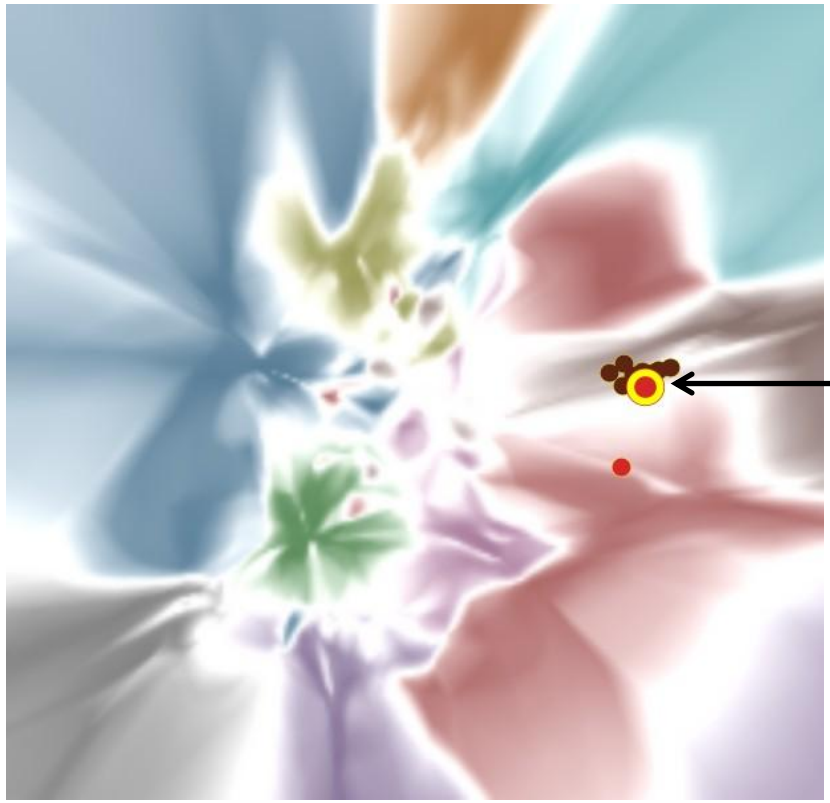


testing sample

Motivating example: how adversarial training cause the performance degeneration?

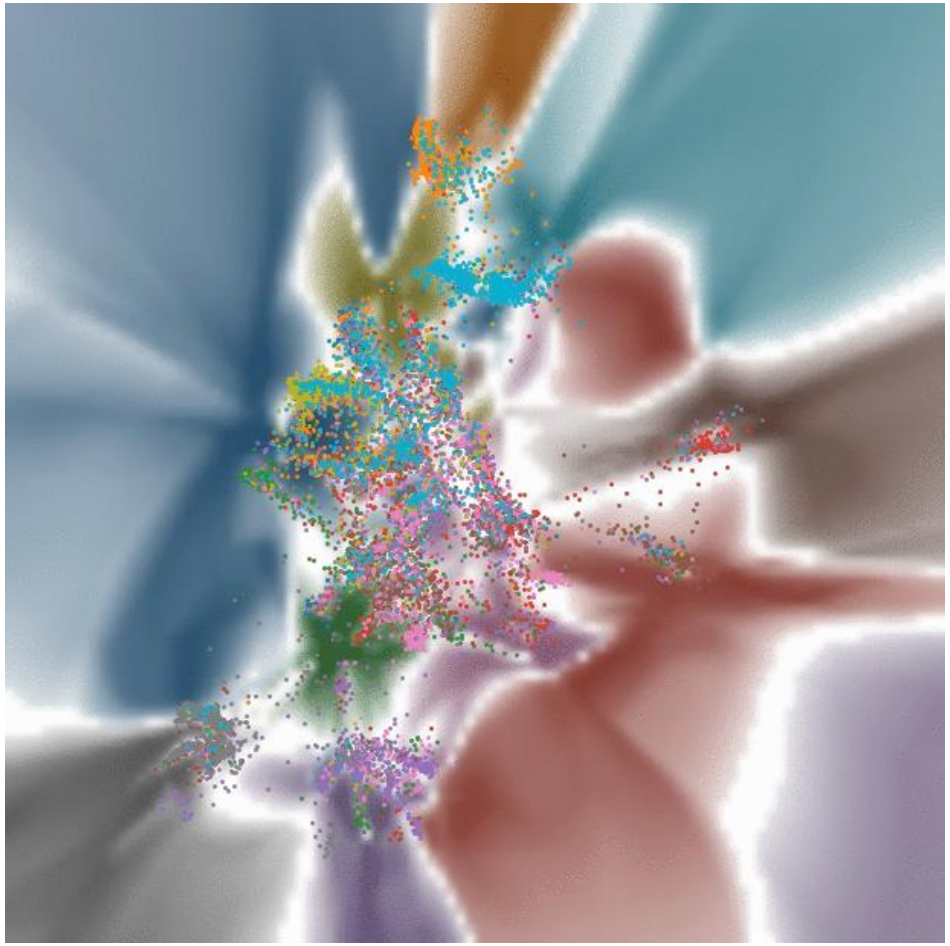
	Accuracy	
Adversarial Samples	68.8%	↑ 1.0%
Testing Samples	88.2%	↓ 1.1%

Epoch 3

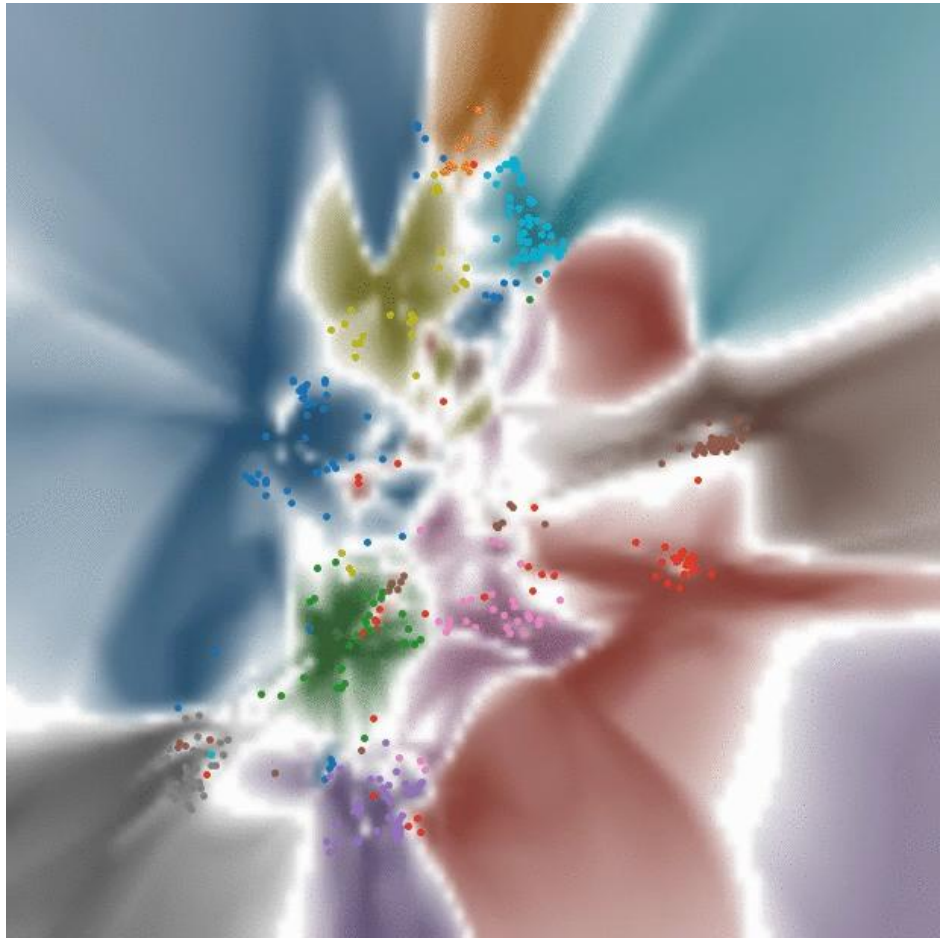


testing sample

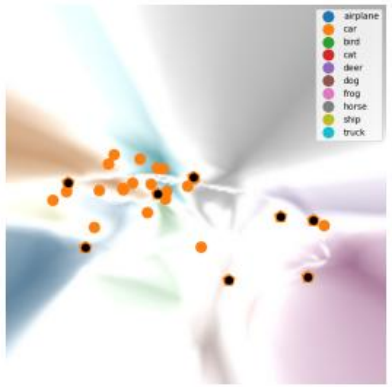
Adversarial Samples Dynamics



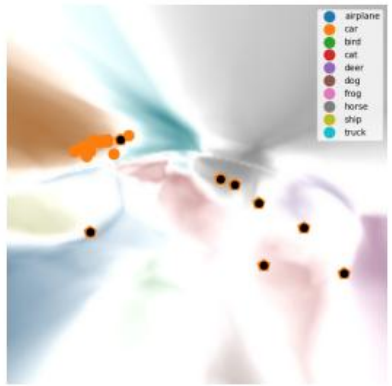
Testing Samples Dynamics



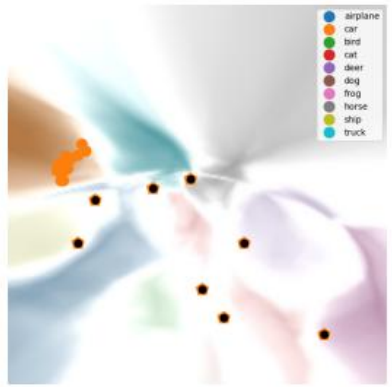
Training with Noisy Data



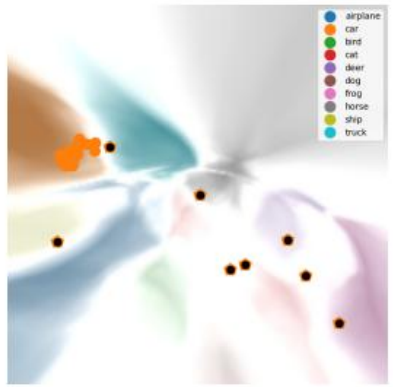
e1



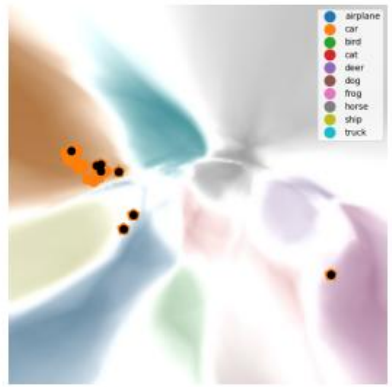
e2



e3

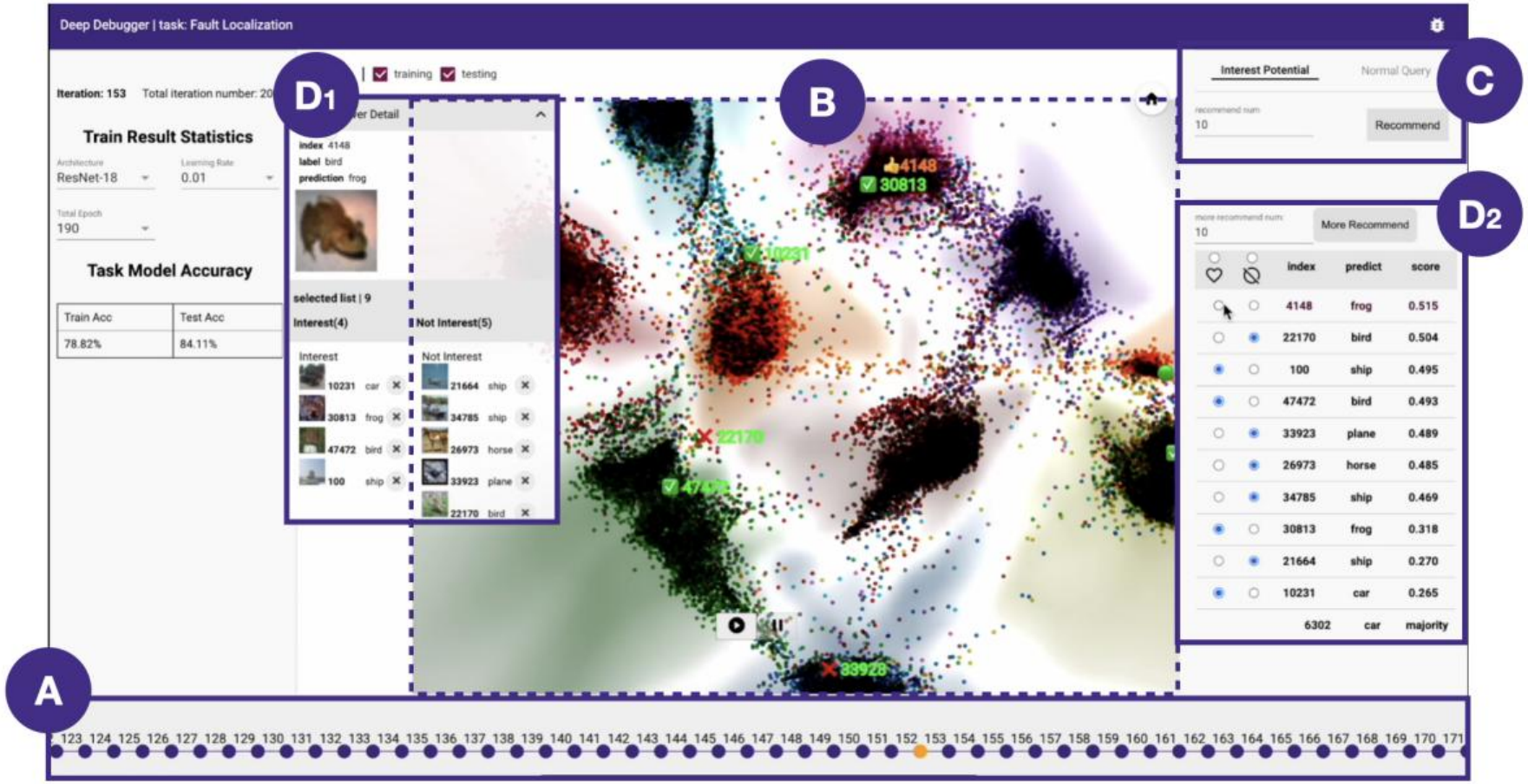


e4

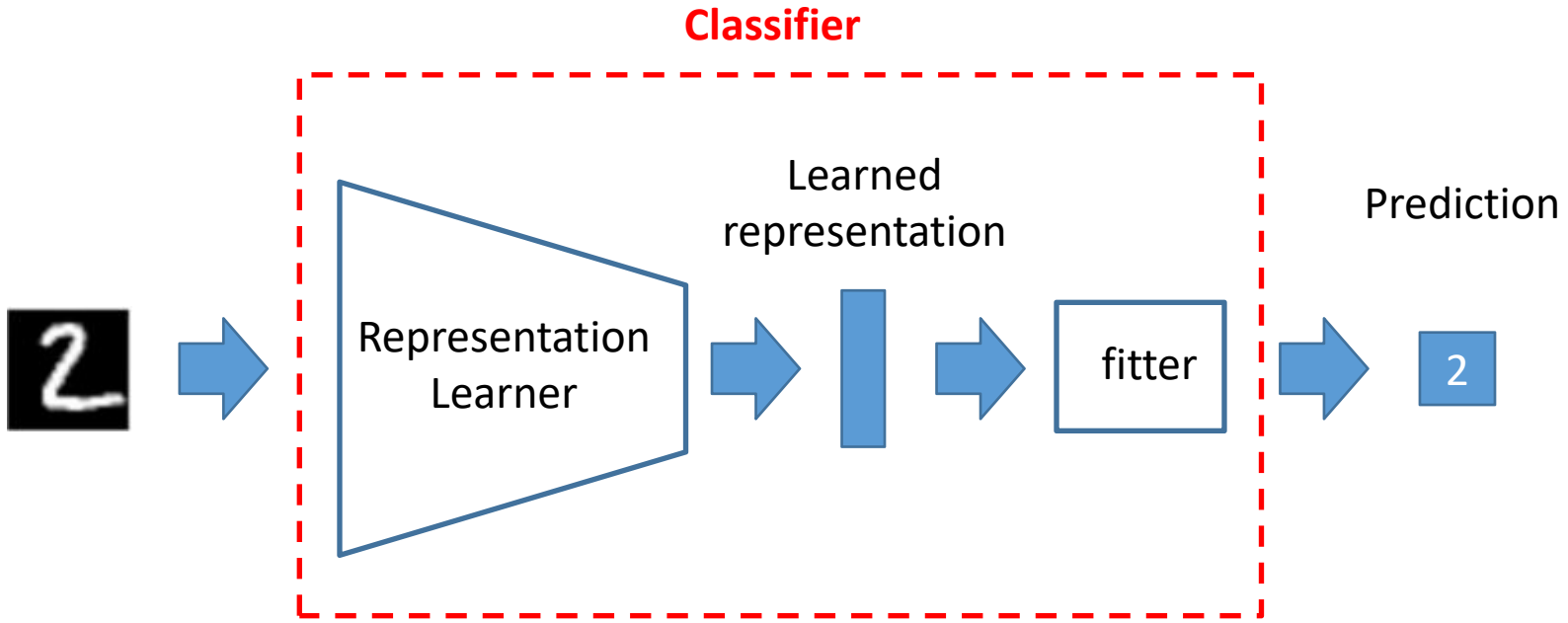


e5

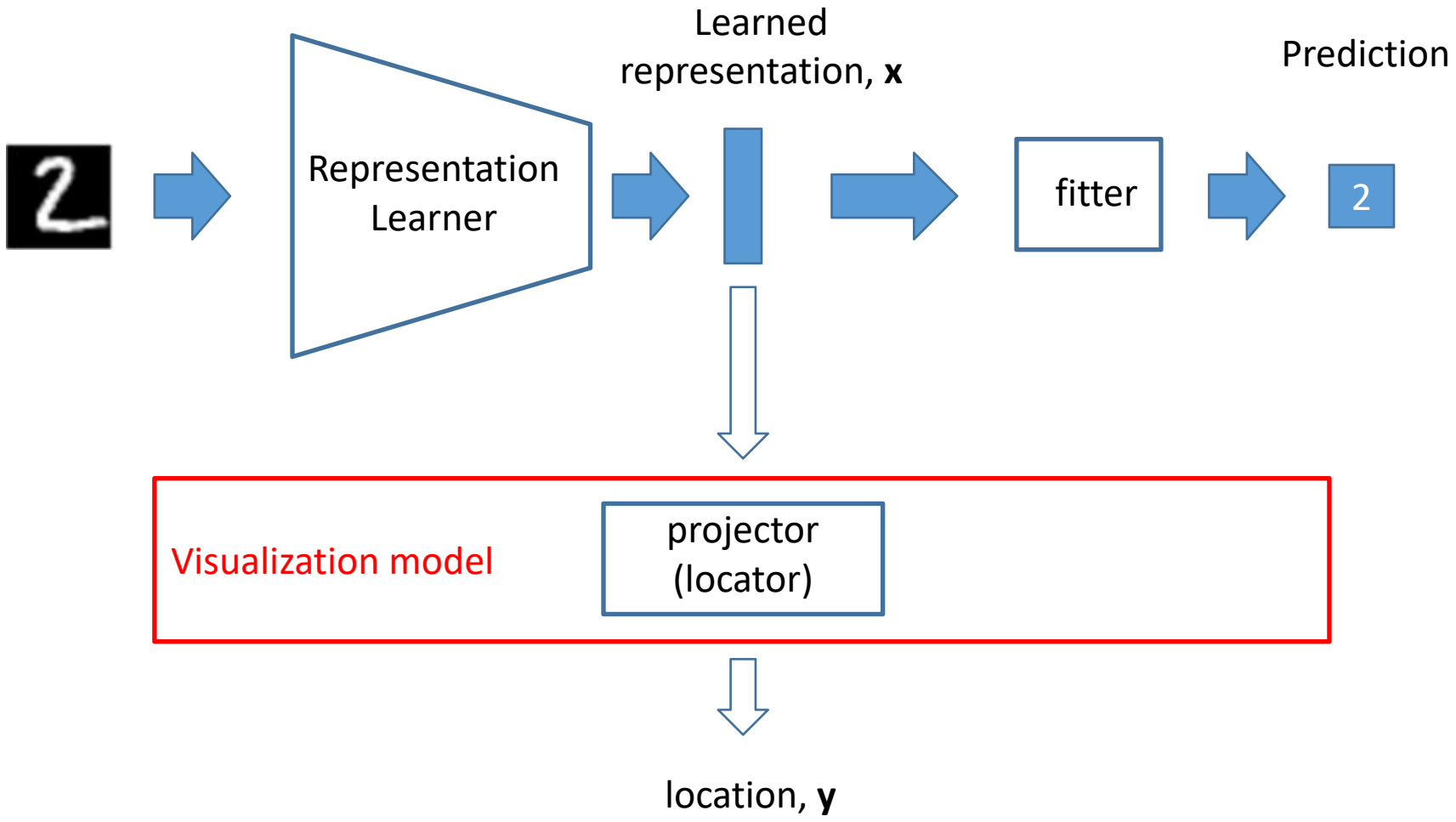
▶▶调试技术工具化 (TensorBoard插件 =》MindSpore插件)



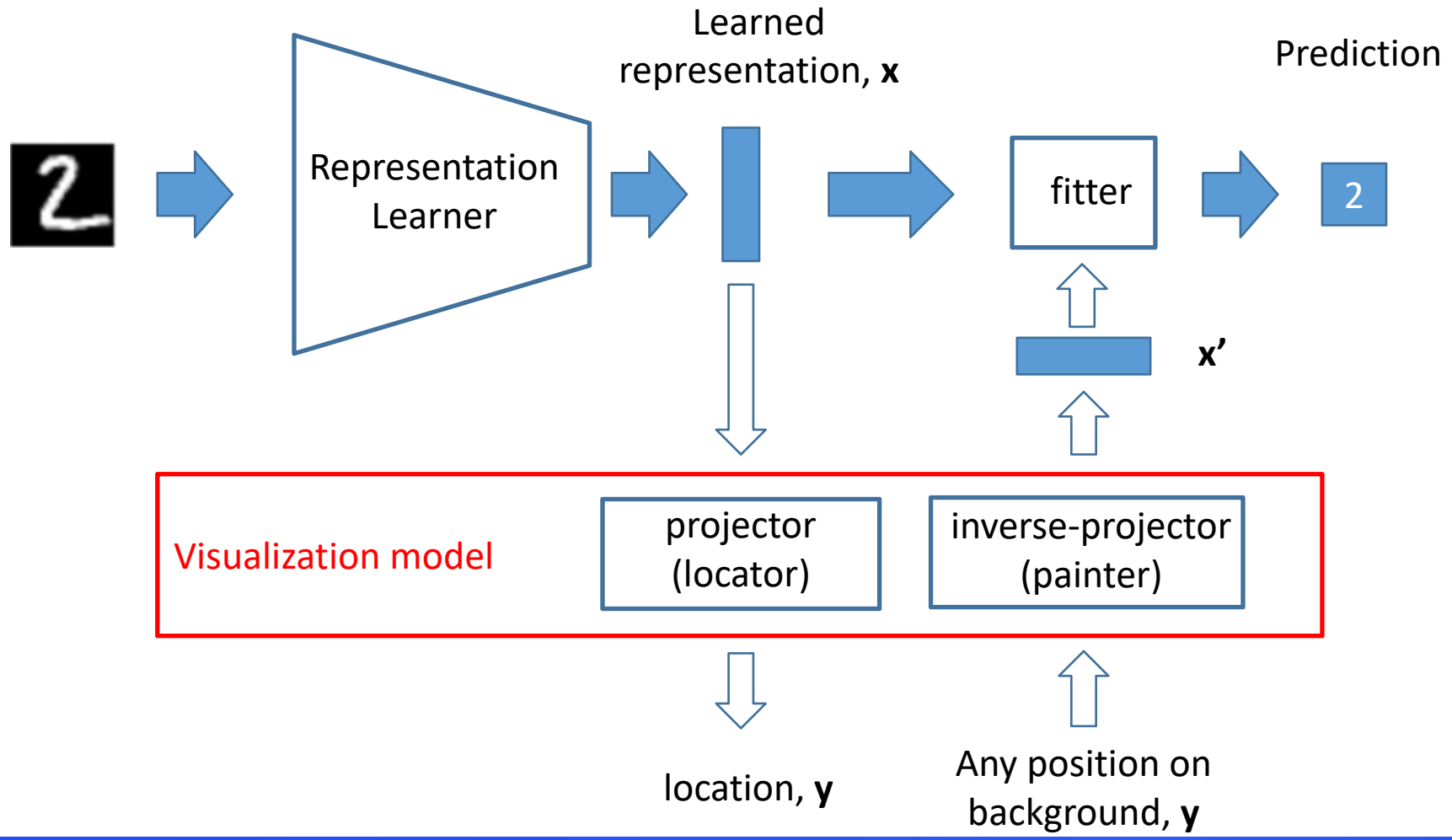
Technical Assumption



Overview

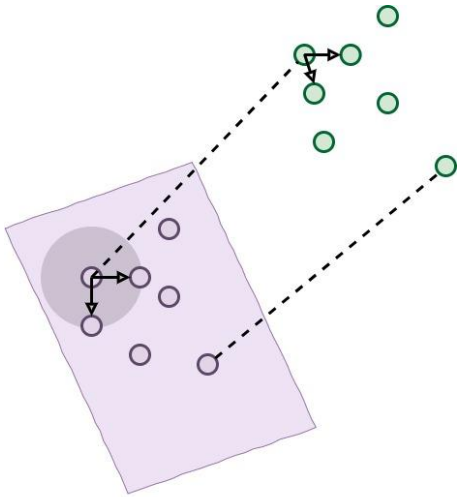

Learned
representation, x
Prediction
Representation
Learner
fitter
2
Visualization model
projector
(locator)
location, y

Overview


Learned
representation, x
Prediction
Representation
Learner
fitter
2
x'
Visualization model
projector
(locator)
inverse-projector
(painter)
location, y
Any position on
background, y

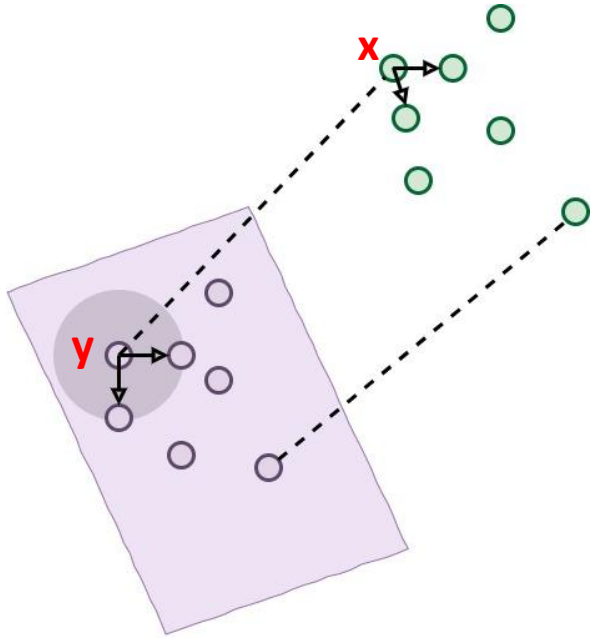
Time-travelling visualization for deep classifier training

- Spatial and temporal properties *any* time-travelling visualization shall abide:
  - Neighbor Preserving
  - Boundary Distance Preserving
  - Inverse Projection Preserving
  - Temporal Continuity



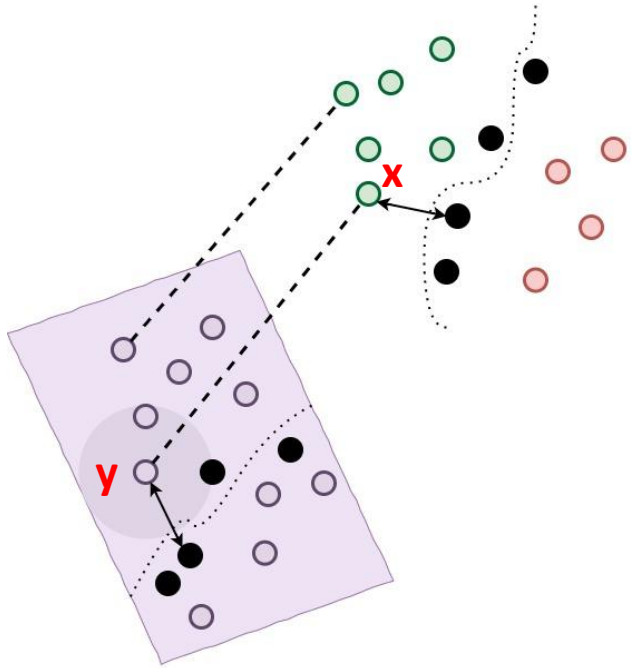
Neighbor Preserving Property

- Given a high-dimensional point **x**, its neighbours should be preserved after projection into the visible low-dimensional space.



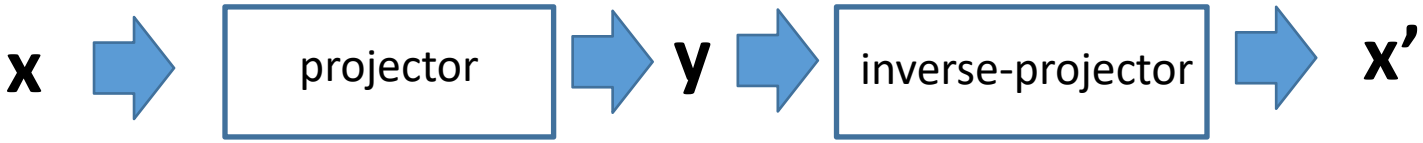
Boundary Distance Preserving Property

- Given a high-dimensional point **x**, its neighbouring boundary points should be preserved after projection into the visible low-dimensional space.



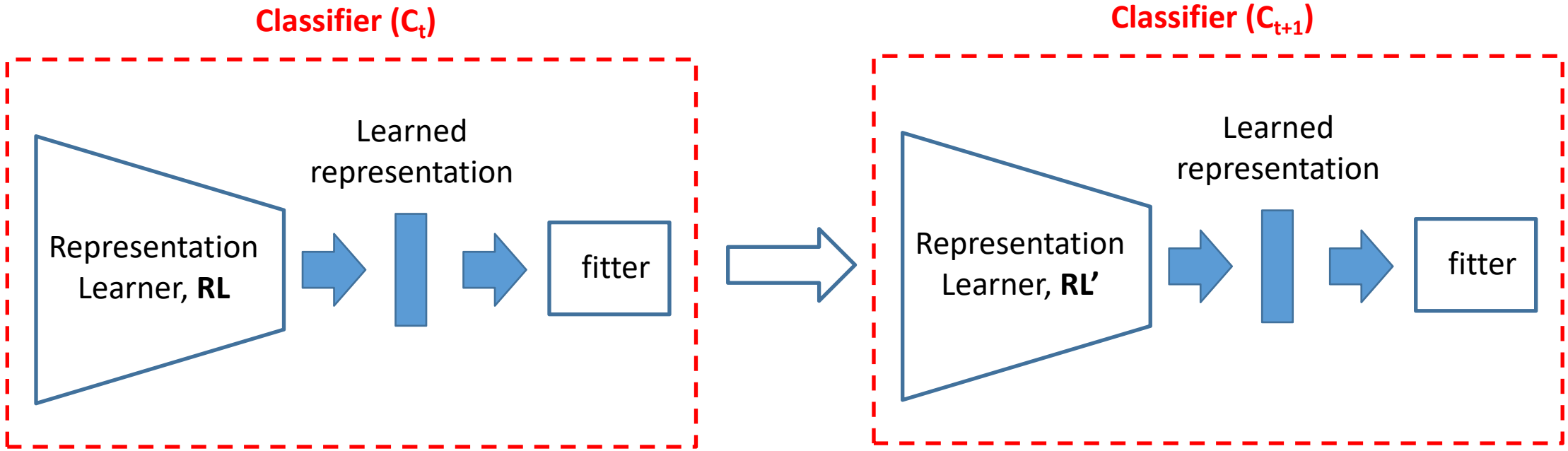
Inverse Projection Preserving Property

- Given a high-dimensional point **x**, after projection into a visible low-dimensional point **y**, we shall be able to inverse-project it back to the high-dimensional space **x'**, **x'** ~ **x**.



x ~ x'

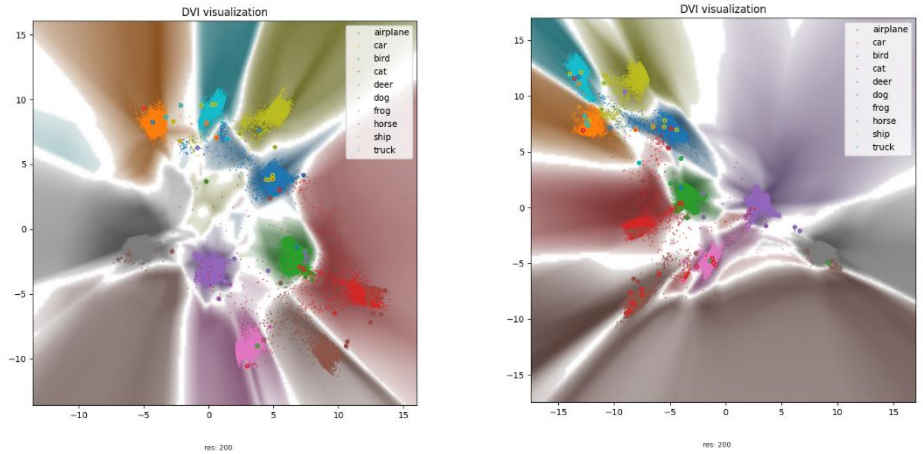
Temporal Continuity Property



RL ~ RL' → V ~ V'

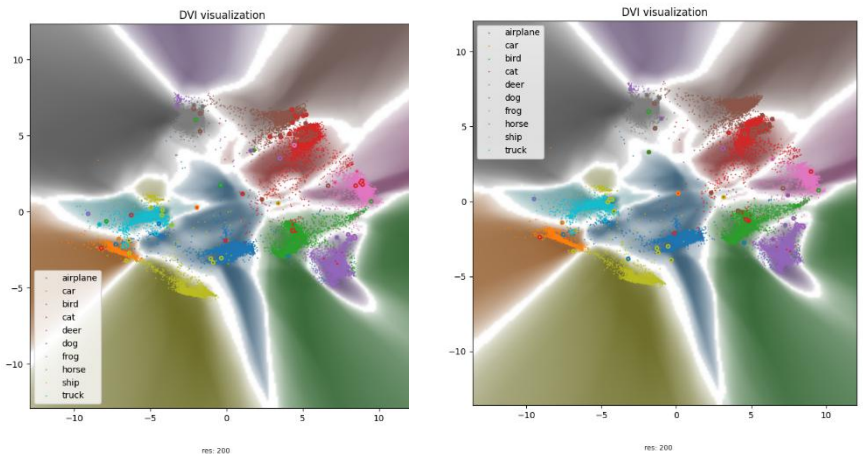
Temporal Continuity Property

E_t E_{t+1}



Temporal continuity NOT considered

E'_t E'_{t+1}



Temporal continuity considered

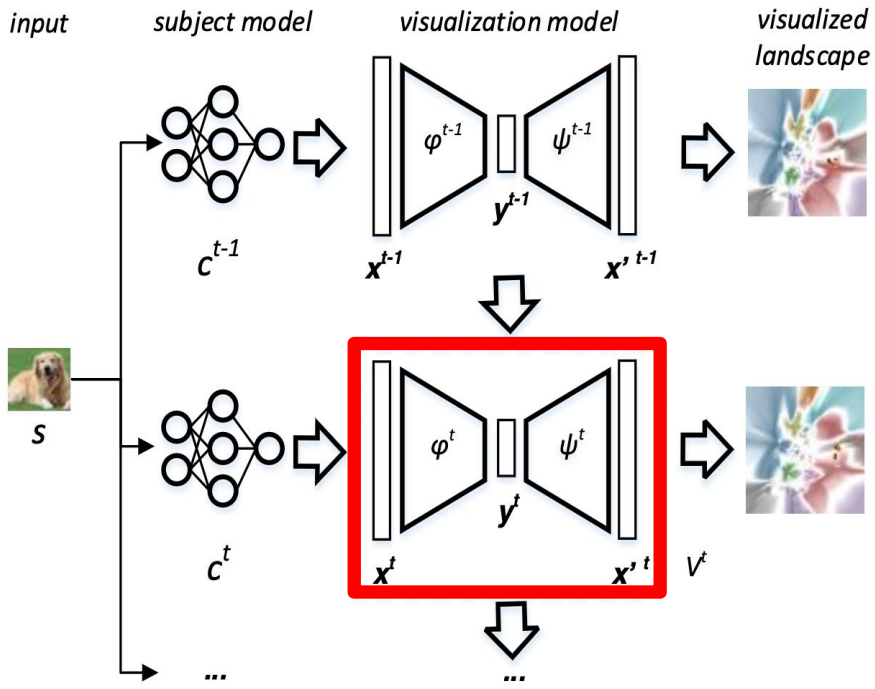
Approach

- Neighbor Preserving
- Boundary Distance Preserving
- Inverse Projection Preserving

Spatial Property

- Temporal Continuity

Temporal Property



Yang, X., and Lin, Y., et al. (2022). DeepVisualInsight: Time-Travelling Visualization for Spatio-Temporal Causality of Deep Classification Training. ***AAAI'22***

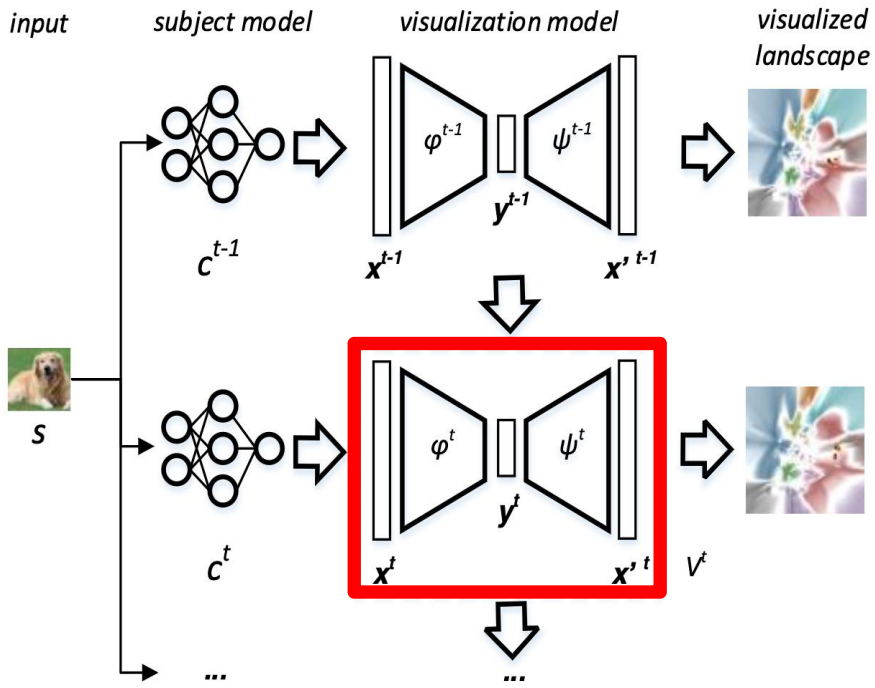
Yang, X., and Lin, Y., et al. (2022). Temporality Spatialization: A Scalable and Faithful Time-Travelling Visualization for Deep Classifier Training ***IJCAI'22***

Approach

- Neighbor Preserving
- Boundary Distance Preserving
- Inverse Projection Preserving
- Temporal Continuity

Spatial Property

Temporal Property

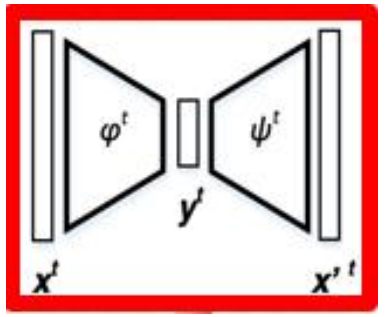


Yang, X., and Lin, Y., et al. (2022). DeepVisualInsight: Time-Travelling Visualization for Spatio-Temporal Causality of Deep Classification Training. ***AAAI'22***

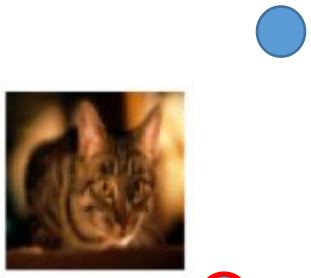
Yang, X., and Lin, Y., et al. (2022). Temporality Spatialization: A Scalable and Faithful Time-Travelling Visualization for Deep Classifier Training ***IJCAI'22***

(Boundary) Neighbour Preserving Property

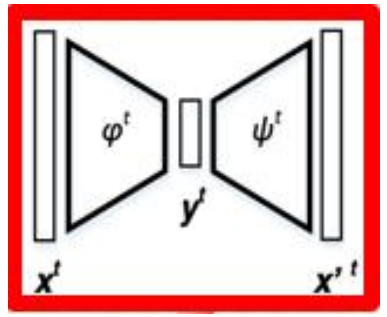
$V^t =$



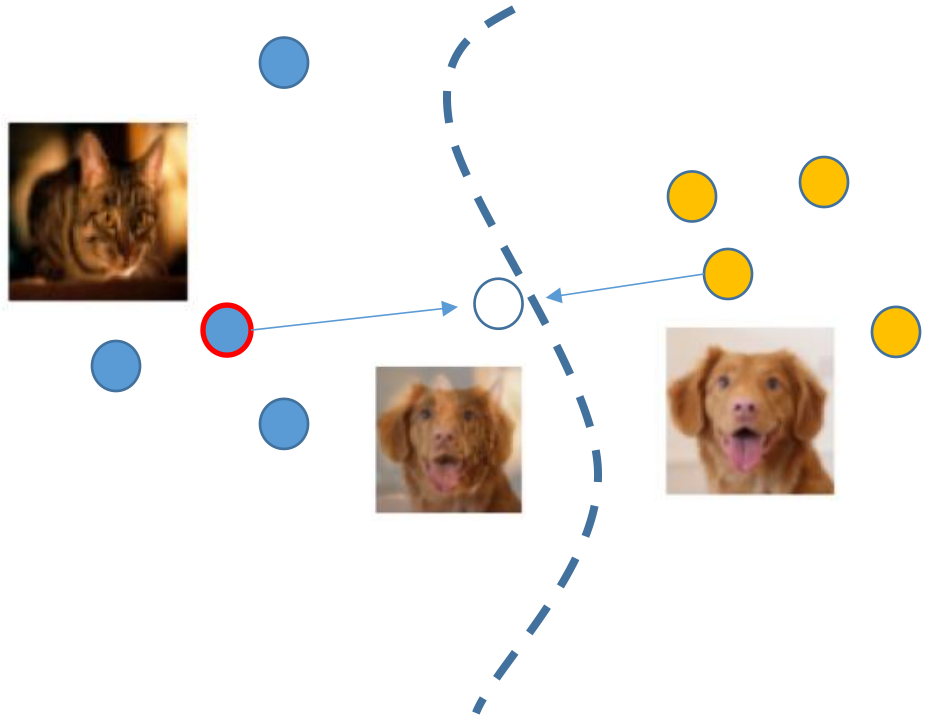
high-dimensional space



(Boundary) Neighbour Preserving Property

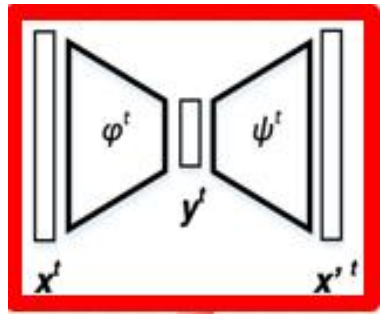
$V^t =$ 

high-dimensional space

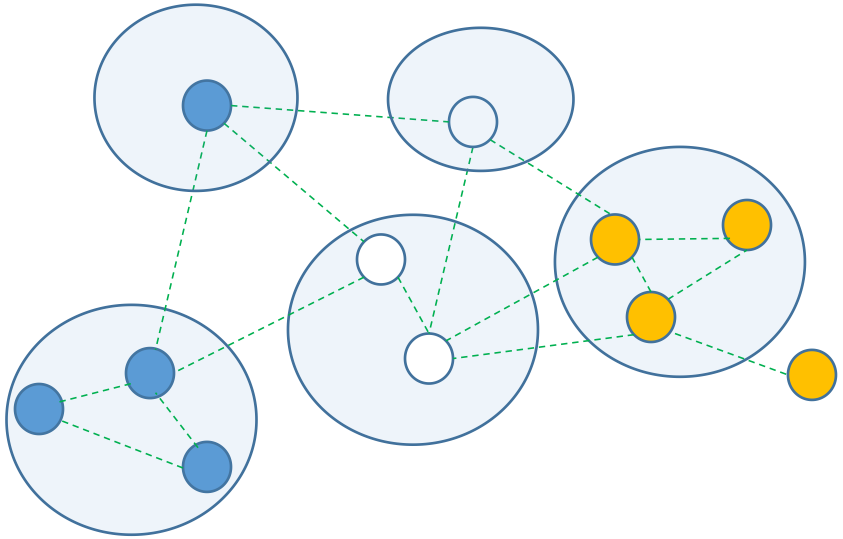


(Boundary) Neighbour Preserving Property

$V^t =$



high-dimensional space



The final selected pairs: $P^{+}_{x\text{-}x}$, $P^{-}_{x\text{-}x}$, $P^{+}_{x\text{-}b}$, $P^{-}_{x\text{-}b}$, $P^{+}_{b\text{-}b}$, $P^{-}_{b\text{-}b}$

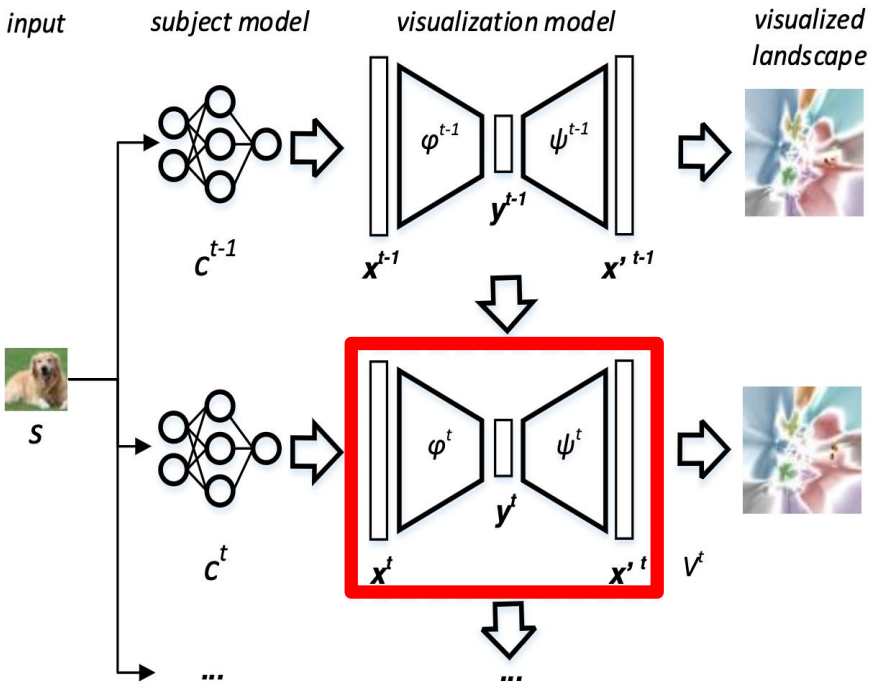
$$C_{umap} := \sum_i \sum_j \left[p_{ij} \cdot \log\left(\frac{p_{ij}}{q_{ij}}\right) + (1 - p_{ij}) \cdot \log\left(\frac{1 - p_{ij}}{1 - q_{ij}}\right) \right]$$

Approach

- Neighbor Preserving
- Boundary Distance Preserving
- Inverse Projection Preserving
- Temporal Continuity

Spatial Property

Temporal Property



$$\mathcal{L}_{total} = \lambda_1 \cdot \mathcal{L}_{umap} + \lambda_2 \cdot \mathcal{L}_{rec} + \lambda_3 \cdot \mathbb{1}(t \geq 2) \cdot \mathcal{L}_t \quad (12)$$

Evaluation

- Subject models: Resnet18 with 512 dimensions of representation vectors
- Dataset: MNIST, FMNIST, CIFAR-10
- Baselines: PCA, t-SNE, UMAP

Result: Better preserved neighboring boundaries

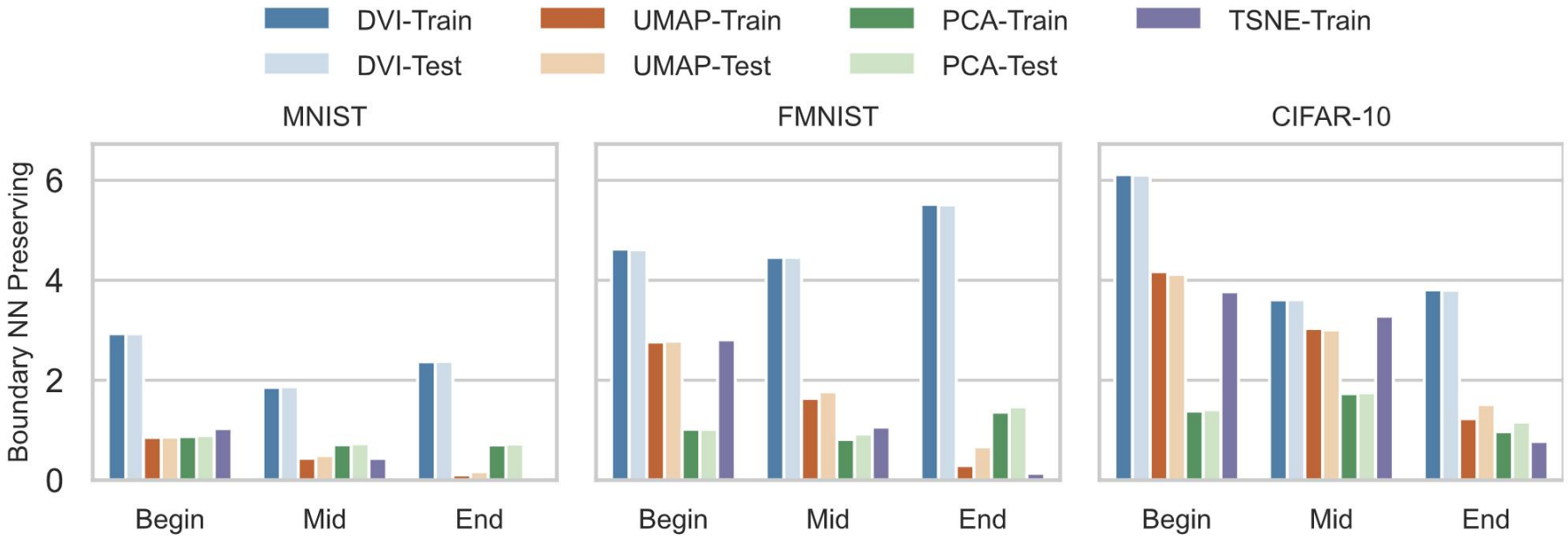


Figure: k Boundary Neighbour Preserving (k=15)

Results：Well Preserved Inverse-projection (~UMAP)

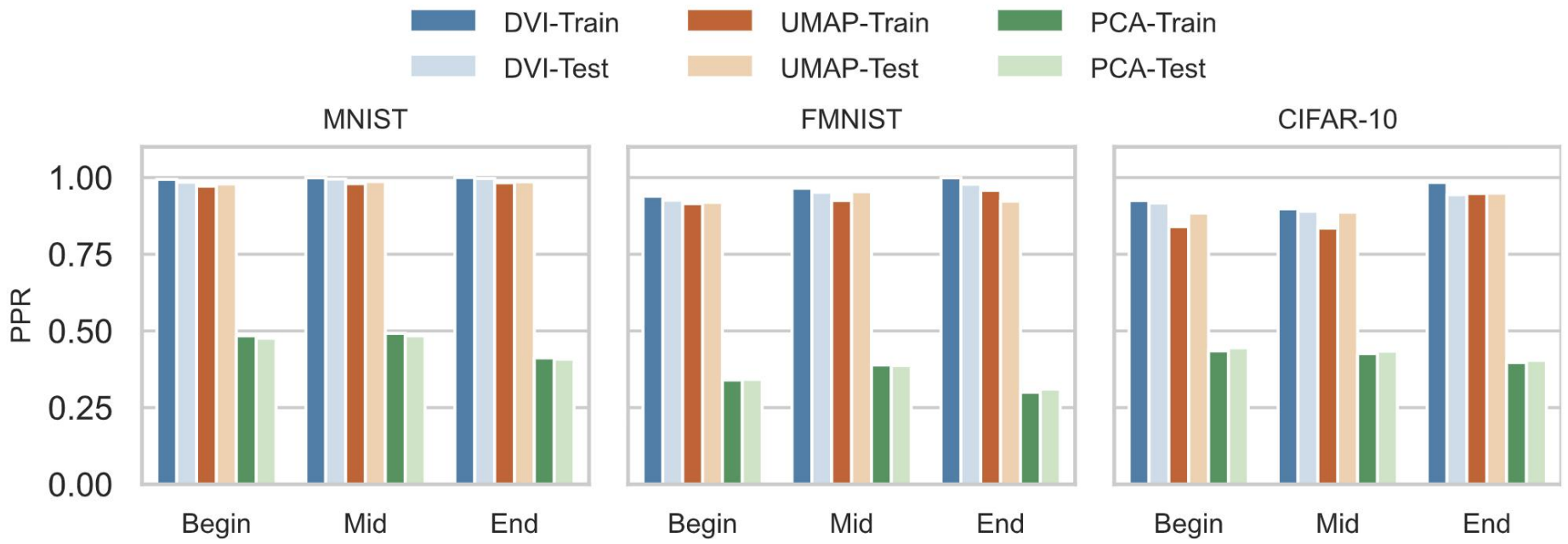


Figure: PPR between DVI, UMAP, and PCA

DVI (0.01s) vs UMAP (58.31s)

Results: Better Temporal Continuity

Table: Temporal Results, i.e., $temporal_{pv}$ value (k=15)

Solution	CIFAR-10		MNIST		FMNIST	
	train	**test**	**train**	**test**	**train**	**test**
UMAP-T	-0.453	-0.448	-0.581	-0.578	-0.622	-0.613
DVI-T	-0.442	-0.460	-0.463	-0.466	-0.291	-0.286
DVI	**-0.463**	**-0.498**	**-0.609**	**-0.611**	**-0.626**	**-0.632**

*Pearson correlation between moving distance and #preserved neighbours

Results: Runtime Efficiency

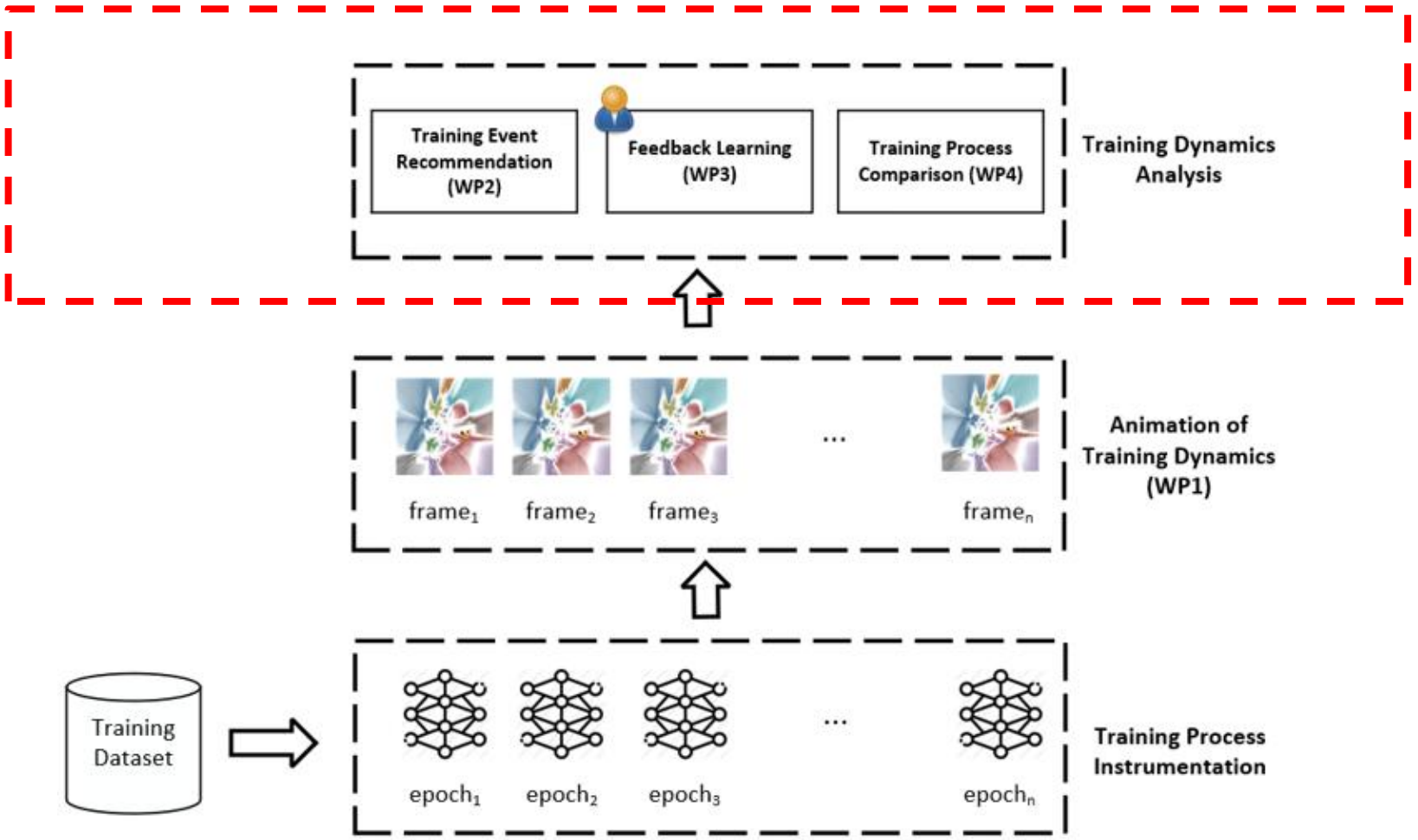
Table: Visualization Overhead (in seconds)

Solution	Overhead Type	CIFAR-10	MNIST	FMNIST
DVI	Offline	792.784	914.921	896.296
	Online	0.016	0.010	0.010
UMAP	Offline	50.170	58.311	58.748
	Online	1819.598	2187.888	2150.703
tSNE	Offline	207.757	286.068	282.725
	Online	/	/	/
PCA	Offline	0.803	0.958	0.951
	Online	0.035	0.036	0.035

可视化模型训练调试框架示意图

意图检测

可观测性

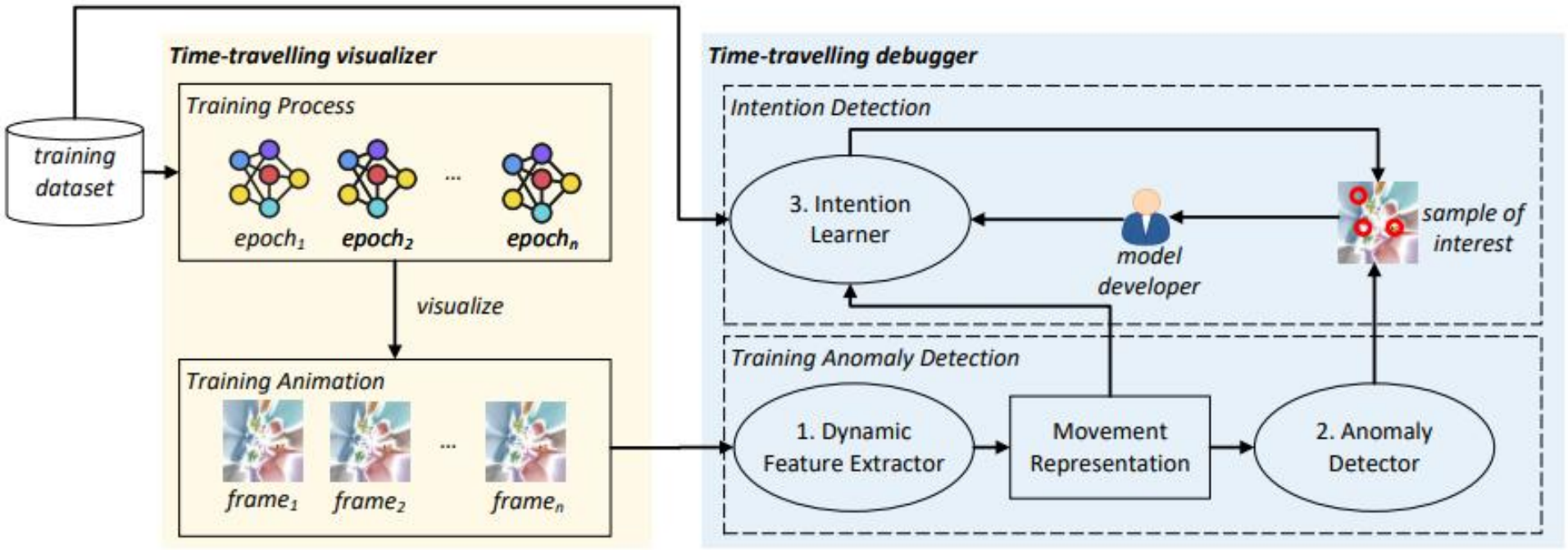


FSE’23
NeurIPS’22

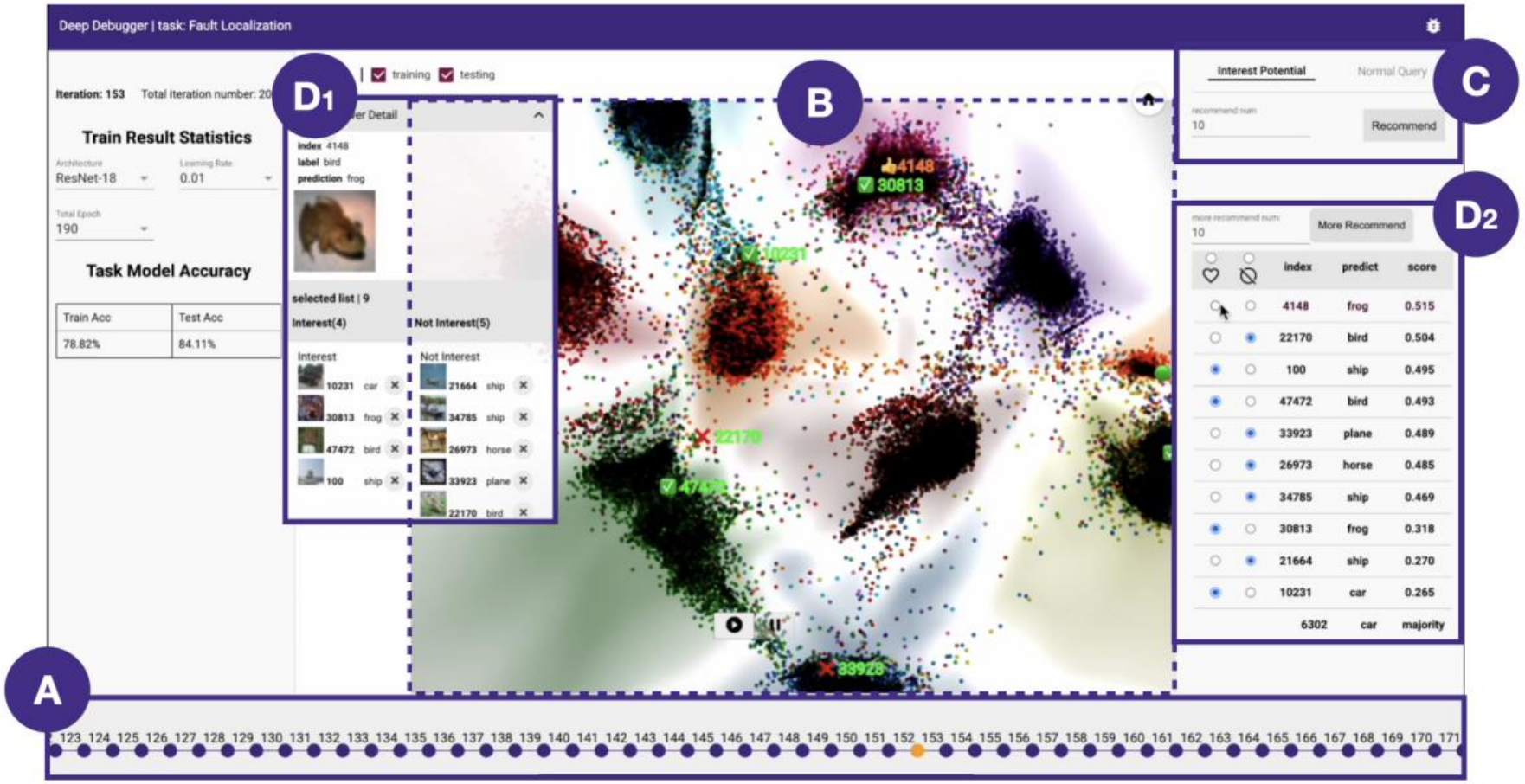
AAAI’22,
IJCAI’22
ChinaSoft’22
（原型竞赛一等奖）

从可视化技术走向调试技术

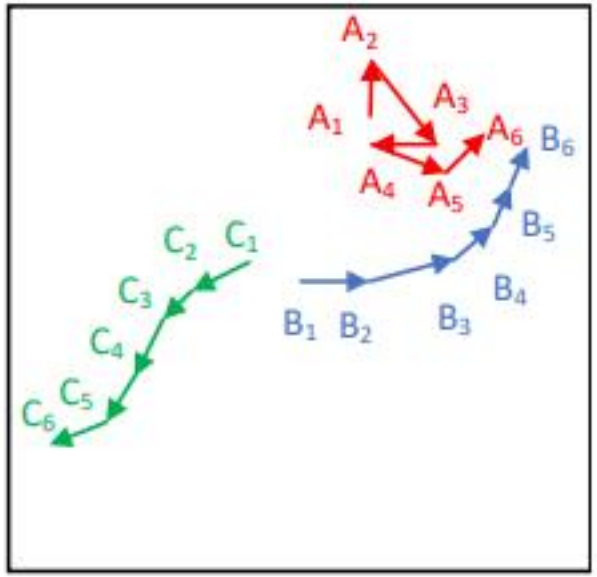
Yang, X., and Lin, Y., et al. (2023). *DeepDebugger: An Interactive Time-Travelling Debugging Approach for Deep Classifiers.* ***FSE'23***



调试技术工具化


Deep Debugger | task: Fault Localization
D1
B
C
D2
A
training
testing
Iteration: 153
Total iteration number: 20
Train Result Statistics
Architecture
ResNet-18
Learning Rate
0.01
Total Epoch
190
Task Model Accuracy
Train Acc
Test Acc
78.82%
84.11%
index 4148
label bird
prediction frog
selected list | 9
Interest(4)
Not Interest(5)
Interest
10231 car
30813 frog
47472 bird
100 ship
Not Interest
21664 ship
34785 ship
26973 horse
33923 plane
22170 bird
Interest Potential
Normal Query
recommend num
10
Recommend
more recommend num
10
More Recommend
index predict score
4148 frog 0.515
22170 bird 0.504
100 ship 0.495
47472 bird 0.493
33923 plane 0.489
26973 horse 0.485
34785 ship 0.469
30813 frog 0.318
21664 ship 0.270
10231 car 0.265
6302 car majority
4148
30813
33928

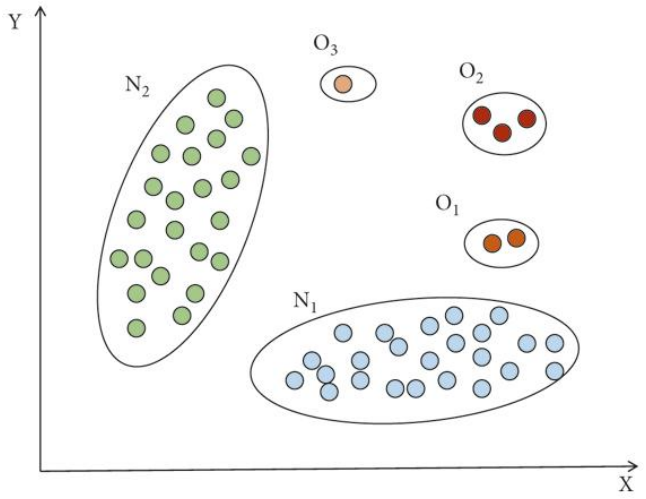
运动信息提取



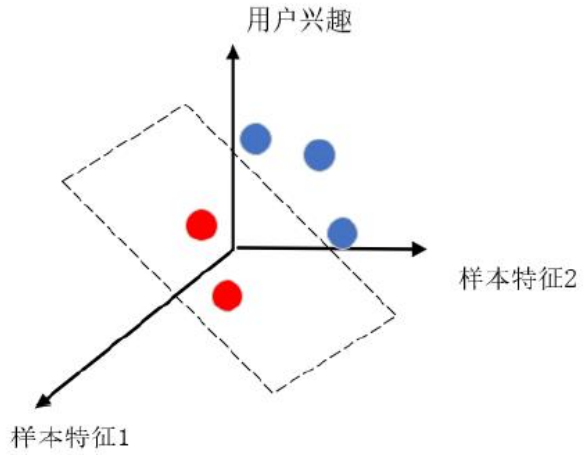
- 位置变化信息
- 速度变化信息
- 加速度变化信息
- 置信度变化信息

→ 样本运动信息特征

异常检测



交互式推荐



$$IE(s) = IE(a_1, a_2, ..., a_n) = \sum_{i=1}^{n} c_i \cdot a_i + c_0$$

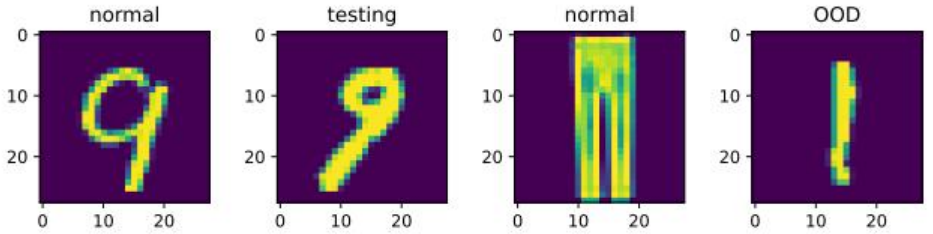
实验部分

- 异常检测实验
- 意图学习实验
- 错误反馈注入实验

异常检测实验

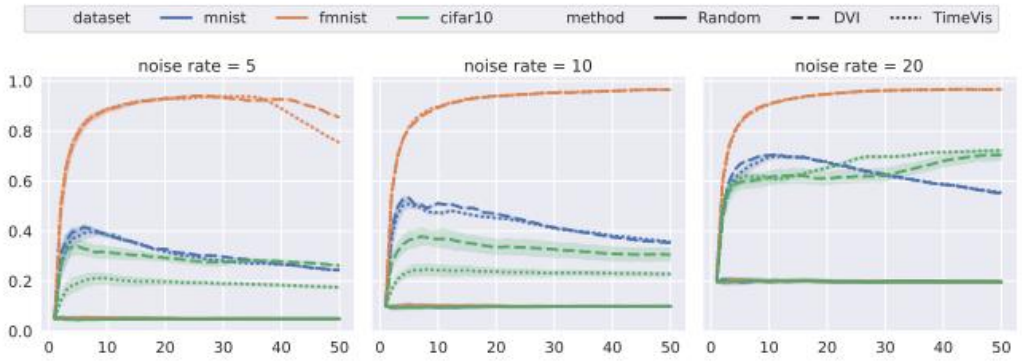
- 异常轨迹定义：
  - 随机轨迹生成
  - 其他数据集中样本的轨迹

Dataset	MNIST		FMNIST		CIFAR-10	
Visualizer	DVI	TimeVis	DVI	TimeVis	DVI	TimeVis
Precision	92.6%	94.7%	80.2%	85.5%	80.1%	72.4%
Recall	67.4%	78.4%	89.0%	83.8%	56.4%	56.6%
F1 score	78.0%	85.8%	84.4%	84.6%	66.2%	63.5%

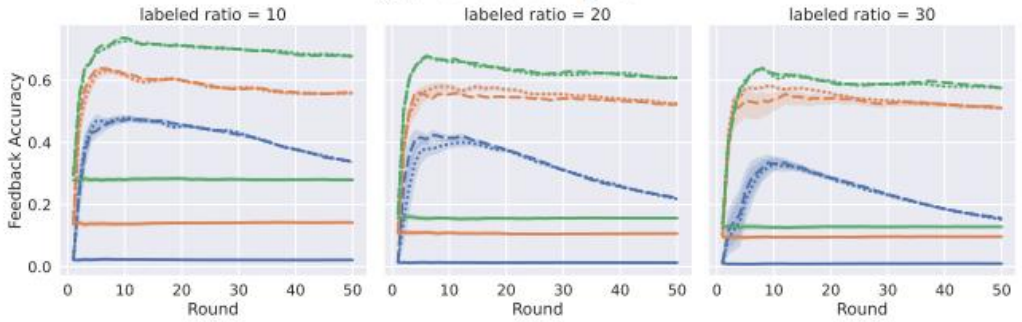


反馈学习实验

- 意图1：噪音样本
- 意图2：可能预测错误的未见样本

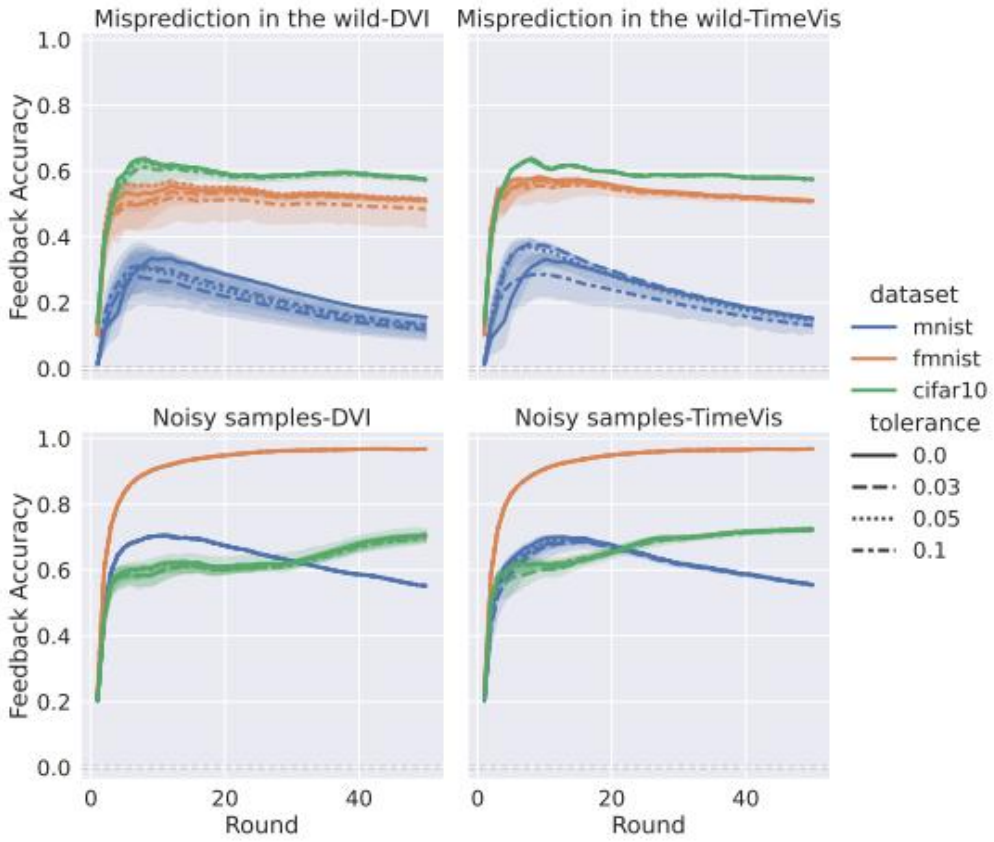


(a) Noise samples

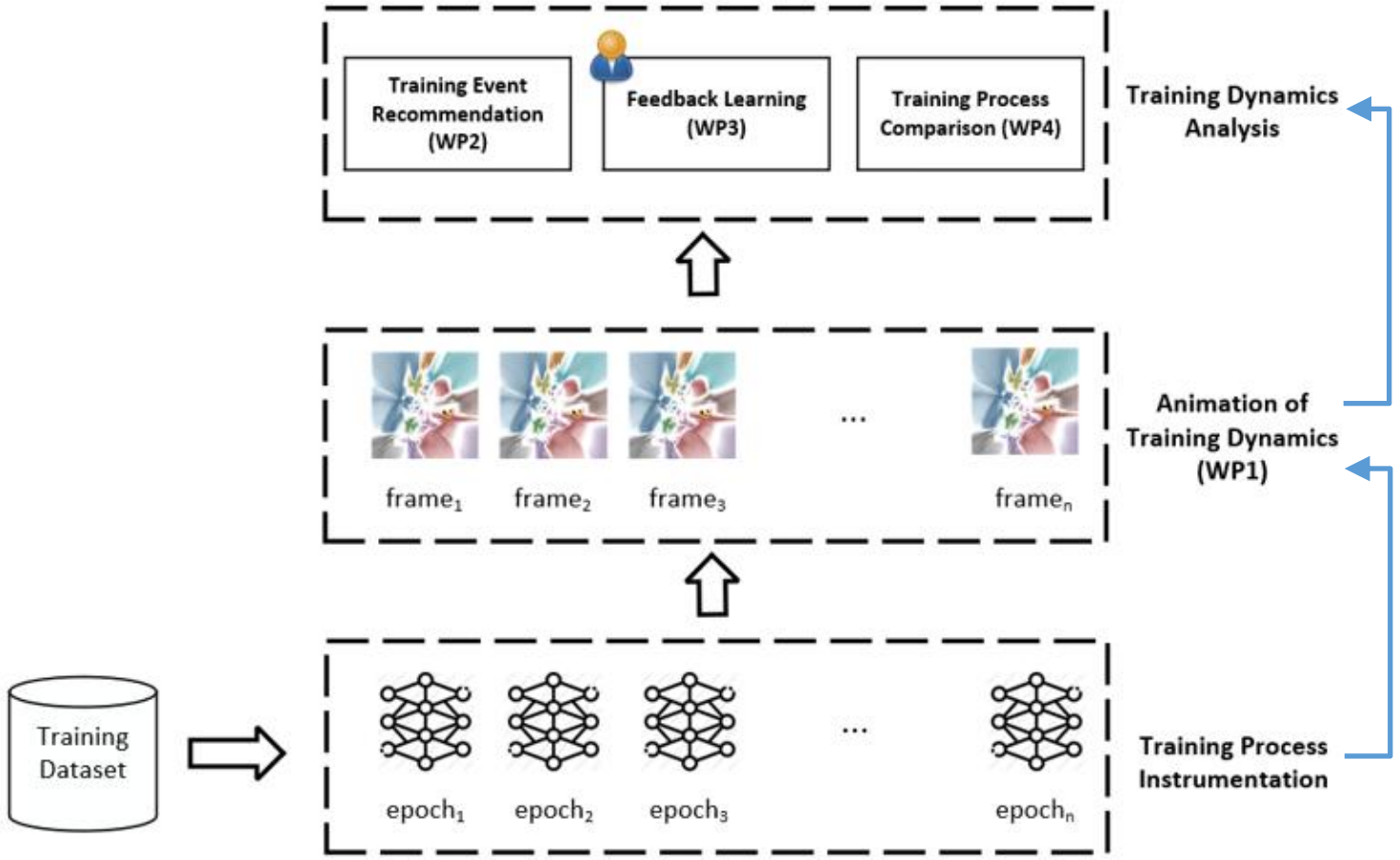


(b) Misprediction samples in the wild

错误反馈注入实验


Misprediction in the wild-DVI
Misprediction in the wild-TimeVis
Noisy samples-DVI
Noisy samples-TimeVis
Feedback Accuracy
Round
dataset
mnist
fmnist
cifar10
tolerance
0.0
0.03
0.05
0.1

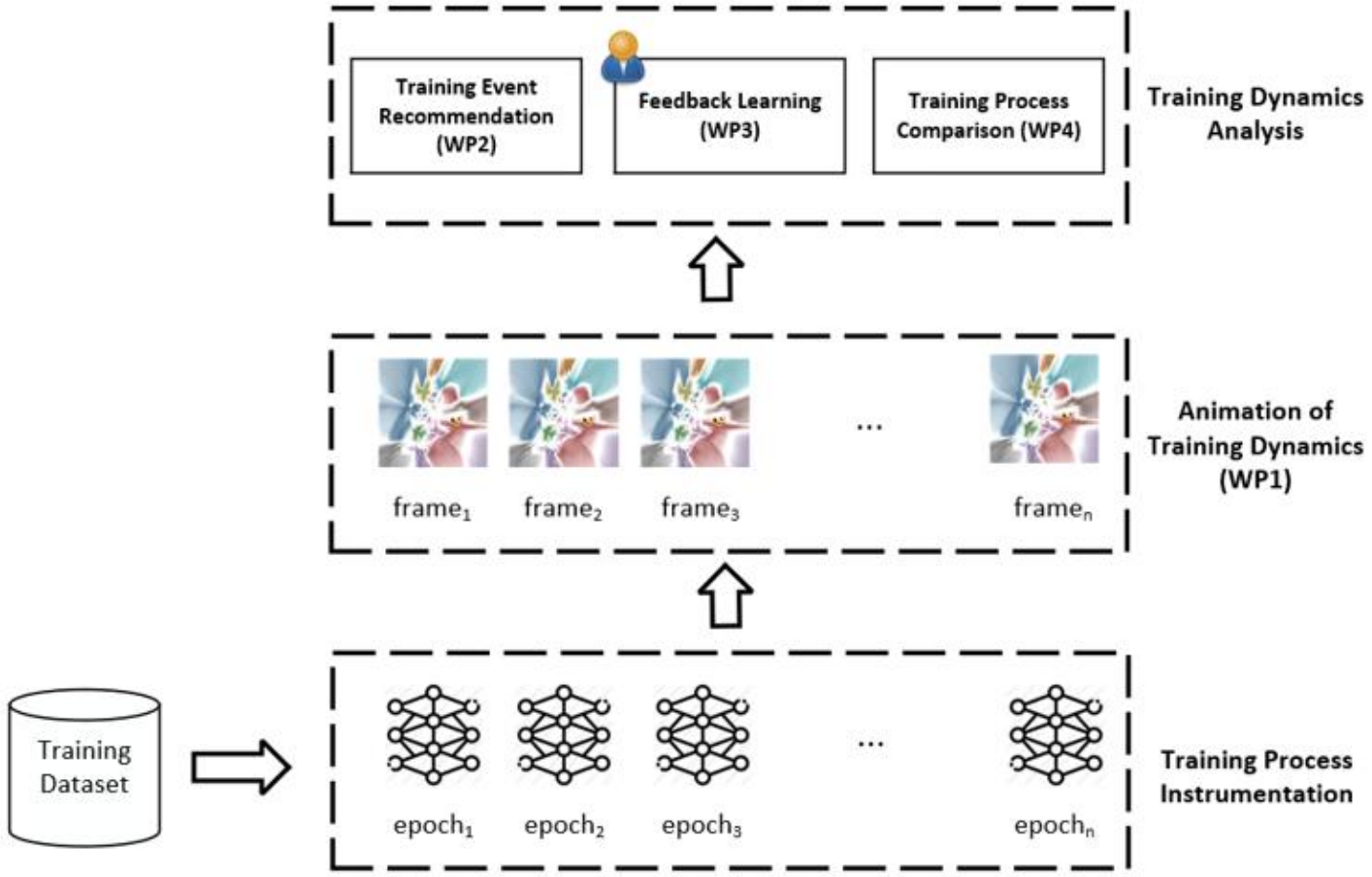
可视化调试框架示意图



可交互性：海量数据中的关键信息的提取和分析

可观测性：投影和展示高维空间上发生的现象

可视化调试框架示意图



1. 还有哪些其他的关键训练事件？

2. 如何将现象与训练代码关联？

3. 是否有更好的事件搜索技术？

。 。 。

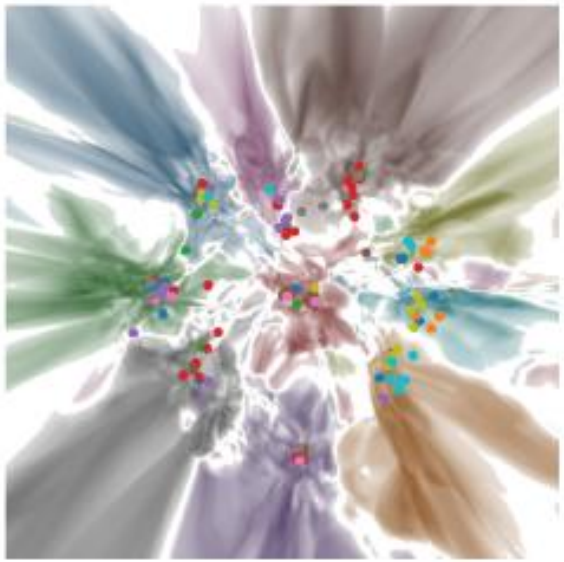
1. 是否有更加好的可视化策略来减少可视化误差？

2. 大量快照模型的管理和存储？

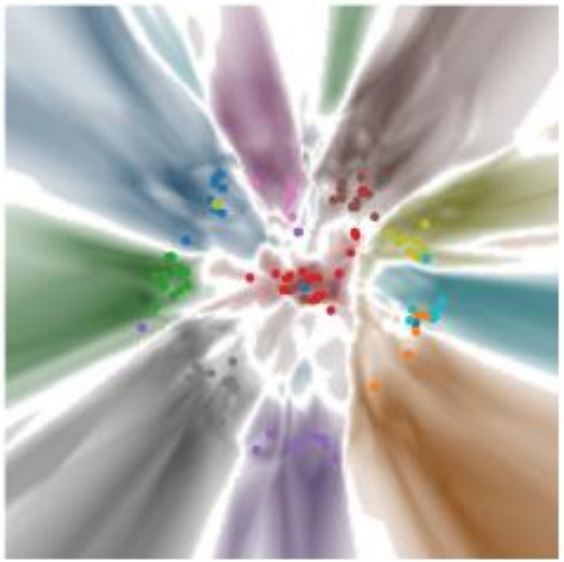
3. 如何对比两个训练过程？

。 。 。

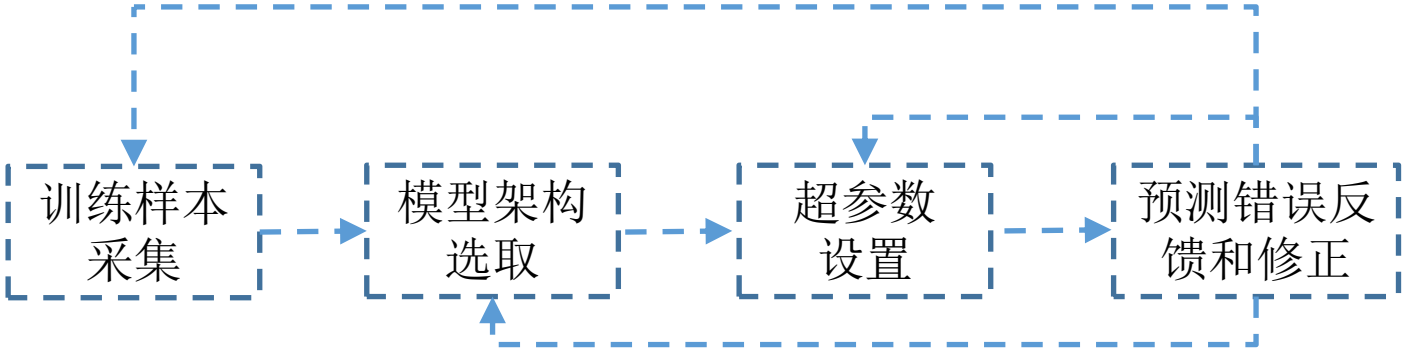
Ongoing Work: ContraVis

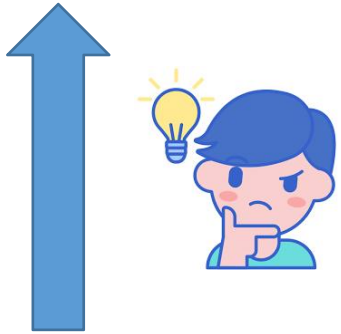


(a) A visualized classification landscape *without applying dropout*



(b) A visualized classification landscape *with dropout applied*



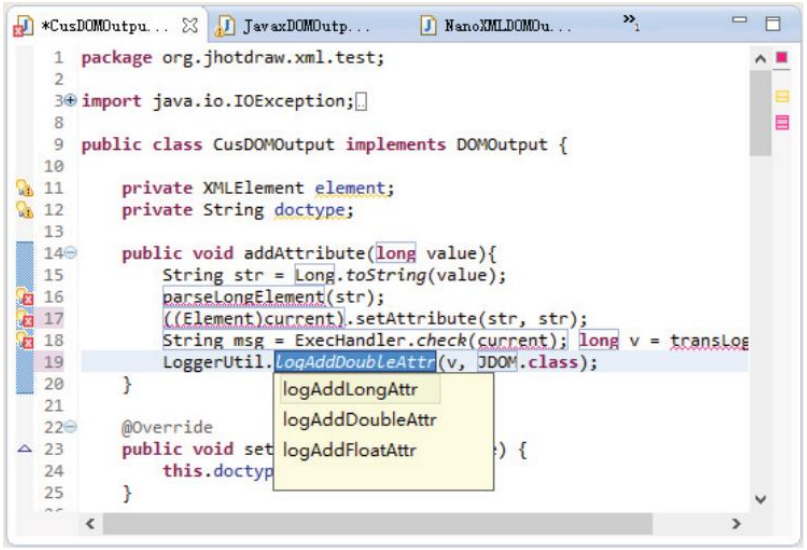


面向AI生态的软件工程技术

模型调试

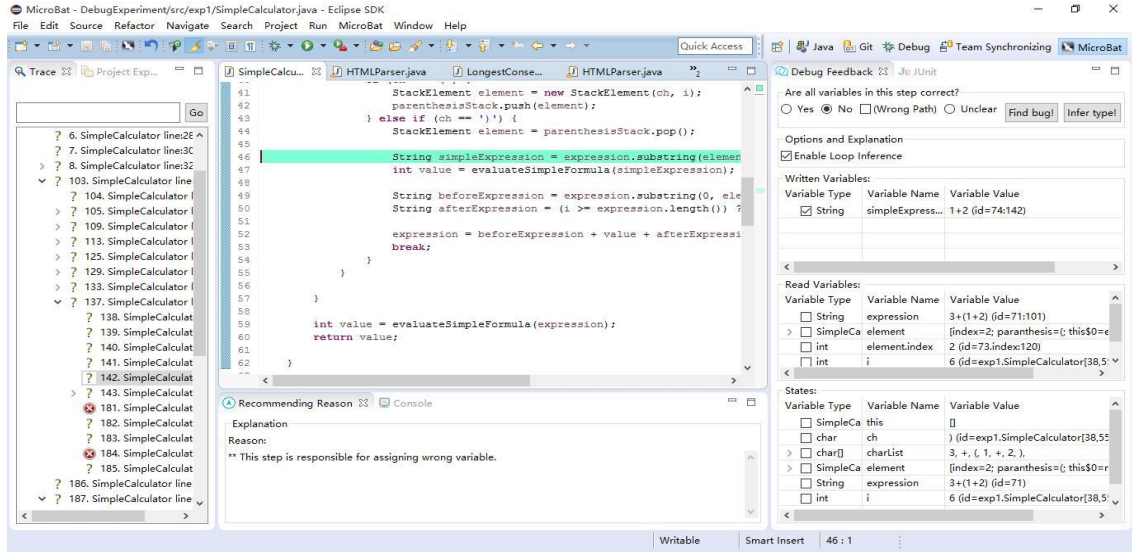
模型鲁棒性测试

样本影响函数



AI辅助交互式代码编辑

FSE'15, EMNLP'22, NeurIPS'23

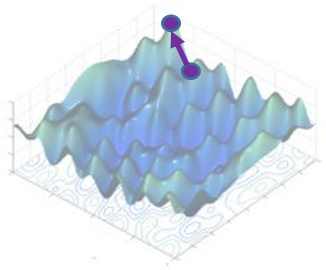


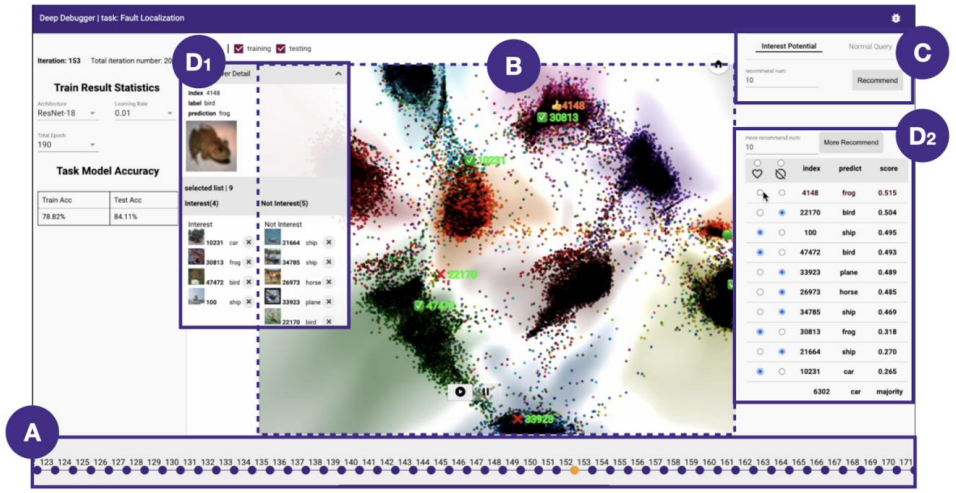
AI辅助自动调试 ICSE'17, ASE'18, TSE'19

EvoSuite

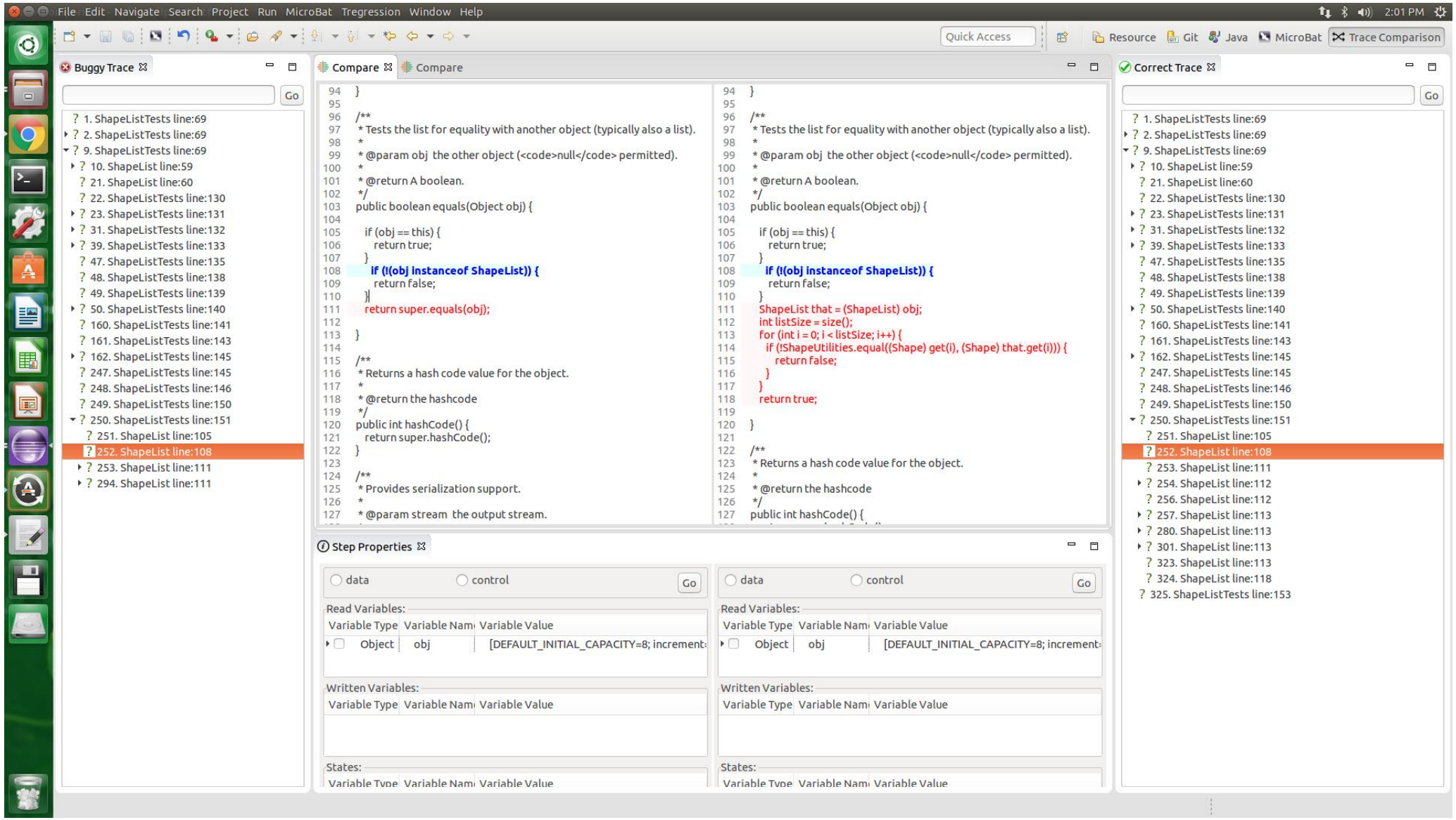
Evosuite++

5-7% coverage improvement





模型训练可视化解释 AAAI'22, IJCAI'22, FSE'22,


File Edit Navigate Search Project Run MicroBat Tregression Window Help
2:01 PM
Quick Access
Resource Git Java MicroBat Trace Comparison
Buggy Trace
Compare
Compare
Correct Trace
Go
1. ShapeListTests line:69
2. ShapeListTests line:69
9. ShapeListTests line:69
10. ShapeList line:59
21. ShapeList line:60
22. ShapeListTests line:130
23. ShapeListTests line:131
31. ShapeListTests line:132
39. ShapeListTests line:133
47. ShapeListTests line:135
48. ShapeListTests line:138
49. ShapeListTests line:139
50. ShapeListTests line:140
160. ShapeListTests line:141
161. ShapeListTests line:143
162. ShapeListTests line:145
247. ShapeListTests line:145
248. ShapeListTests line:146
249. ShapeListTests line:150
250. ShapeListTests line:151
251. ShapeList line:105
252. ShapeList line:108
253. ShapeList line:111
294. ShapeList line:111
325. ShapeListTests line:153
Step Properties
data
control
Read Variables:
Variable Type
Variable Nam
Variable Value
Object
obj
[DEFAULT_INITIAL_CAPACITY=8; increment=
Written Variables:
States:

AiDD AI+ 软件研发数字峰会
AI+ software Development Digital summit

THANKS