IEEE VR 2024
ORLANDO, FL USA

Generating Virtual Reality Stroke Gesture Data from Out-of-Distribution Desktop Stroke Gesture Data

Lin-Ping Yuan, Boyu Li, Jindong Wang, Huamin Qu, Wei Zeng

香港科技大学
THE HONG KONG UNIVERSITY OF SCIENCE AND TECHNOLOGY

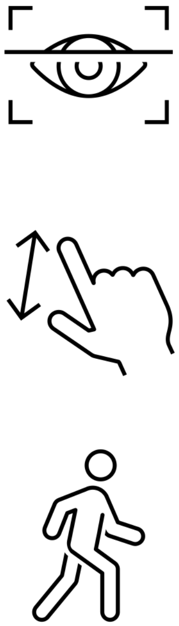
香港科技大学(广州)
THE HONG KONG UNIVERSITY OF SCIENCE AND TECHNOLOGY (GUANGZHOU)

Microsoft Research
微软亚洲研究院

THE 31st IEEE CONFERENCE ON VIRTUAL REALITY AND 3D USER INTERFACES

IEEE | IEEE COMPUTER SOCIETY | IEEE vgtc

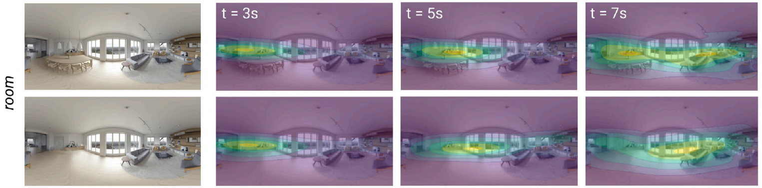
VR interaction data



VR interaction data has many usages.

UX designers can provide users with stroke gestures as intuitive user input for triggering commands [1].

Researchers can visualize users’ movement to get insights on space usage patterns [2].

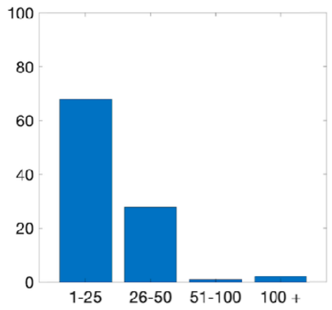


Storytellers can utilize the gaze data to refine their design of virtual scenes [3].

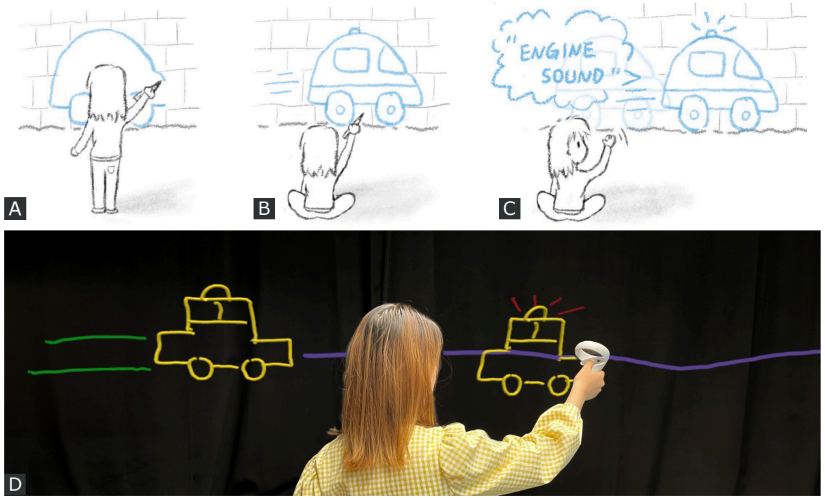
[1] Ousmer, Mehdi, et al. "Recognizing 3D trajectories as 2D multi-stroke gestures." ACM ISS 2020.
[2] Hubenschmid, Sebastian, et al. "Relive: Bridging in-situ and ex-situ visual analytics for analyzing mixed reality user studies." ACM CHI 2022.
[3] Martin, Daniel, et al. "Scangan360: A generative model of realistic scanpaths for 360 images." IEEE TVCG 2022.

Collecting VR interaction data is hard.

- Reasons
  - A small user base
  - Frequent deployment failures
  - Inconvenient collection setups
  - …



Number of participants [1]

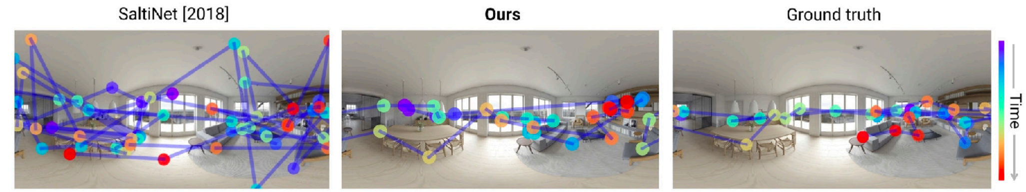


The current VR datasets often have limited quantity and diversity. → Fail to support downstream applications.

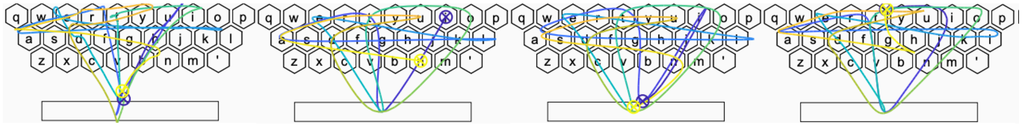
- E.g., the accuracy rate is only about 68% in [2].

[1] Lehman, Sarah M., et al. "ARCHIE++: A cloud-enabled framework for conducting AR system testing in the wild." IEEE TVCG 2022.
[2] Li, Eve Mingxiao, et al. "EnchantedBrush: Animating in Mixed Reality for Storytelling and Communication." Graphics Interface 2023.

Sythesizing VR interaction data



Scan path in 360 videos [1]



(a) Real (b) Jerk-Minimization (c) RNN (d) GAN-Transfer

Mid-air gesture typing [2]

Crowd motions [3]

Quantity can be increased efficiently.

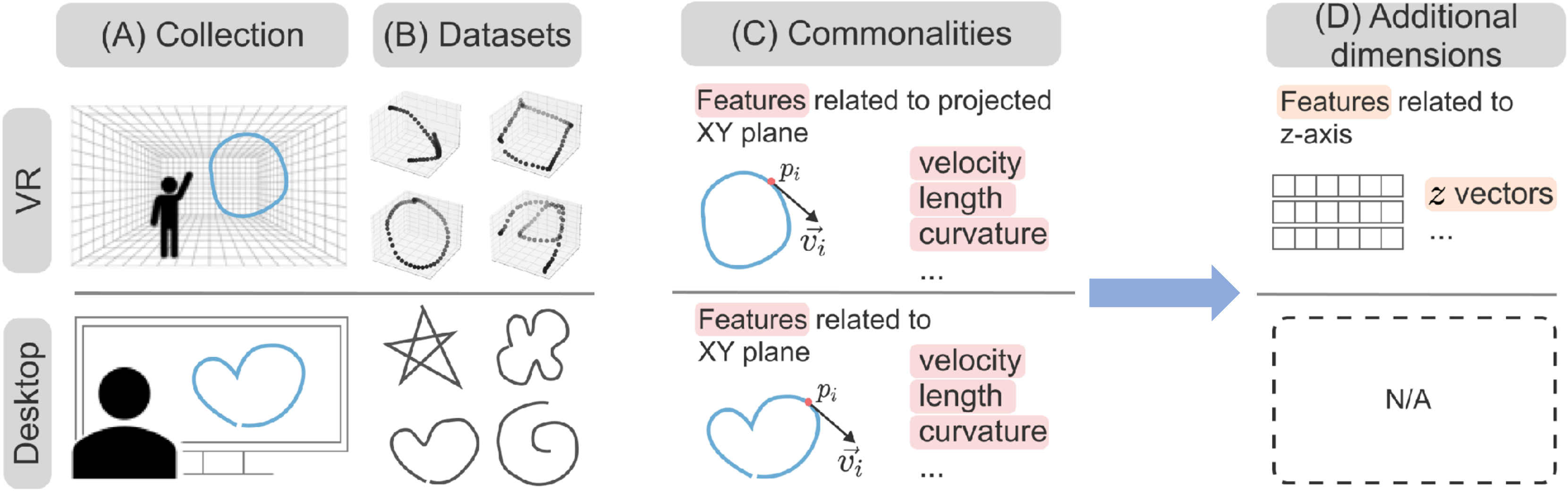
Diversity is still restricted because they only rely on existing VR interaction data as model **input**.

[1] Martin, Daniel, et al. "Scangan360: A generative model of realistic scanpaths for 360 images." IEEE TVCG 2022.
[2] Shen, Junxiao, John Dudley, and Per Ola Kristensson. "Simulating realistic human motion trajectories of mid-air gesture typing." IEEE ISMAR 2021.
[3] Yin, Tairan, et al. "The One-Man-Crowd: Single user generation of crowd motions using virtual reality." IEEE TVCG 2022.

Can desktop interaction data be an alternative?

Selected focus: planar stroke gestures

- Why desktop stroke gestures?
- Why is it possible to use desktop strokes to generate VR strokes?



Easy-to-collect desktop strokes.

Desktop strokes add diversity.

VR and desktop strokes share **commonalities**.

VR strokes present **additional dimensions**.

Research Question

How can desktop strokes enrich VR stroke datasets while preserving the original characteristics?

Preliminary Studies

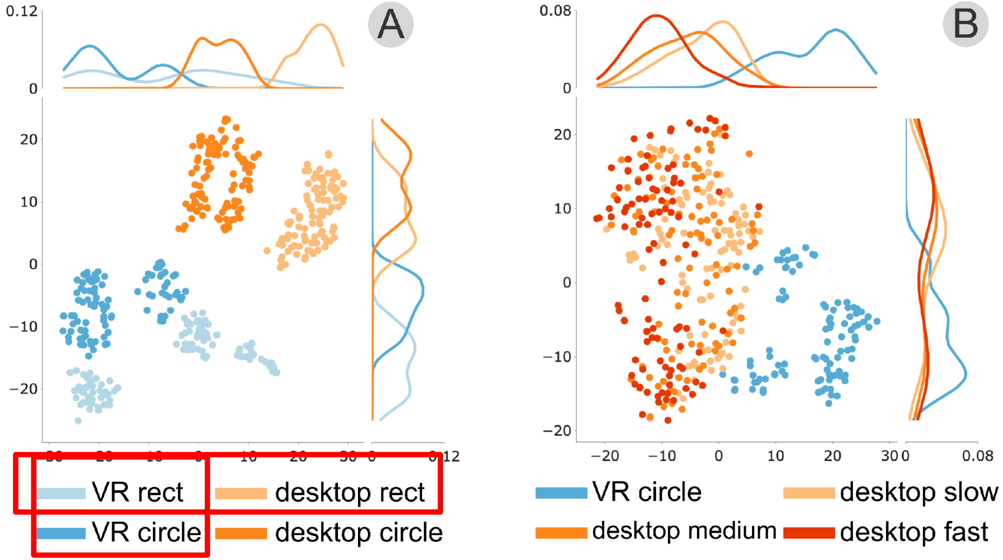
Datasets	VR dataset: 3DMadLabSD [1] Desktop dataset: $1 [2]
Pre-processing	Temporal resampling Normalization
Feature extraction [3]	Point-level features Segment-level features
Data analysis	Commonalities Additional dimensions

Point-level features	Length
	Turning angle
	Curvature
	Velocity
	Acceleration
	Jerk
Segment-level features	Path length
	Starting and ending point distance
	Line similarity
	Area of bounding box
	Length of bounding box diagonal
	Angle of bounding box's diagonal
	Total tuning angle
	Overall sharpness
	Overall curvature

[1] Huang, Jinmiao, et al. "Gesture-based system for next generation natural and intuitive interfaces." AI EDAM 2019.
[2] Wobbrock, Jacob O., et al. "Gestures without libraries, toolkits or training: a $1 recognizer for user interface prototypes." ACM UIST 2007.
[3] Tu, Huawei, et al. "A comparative evaluation of finger and pen stroke gestures." ACM CHI 2012.

Preliminary Studies – Findings

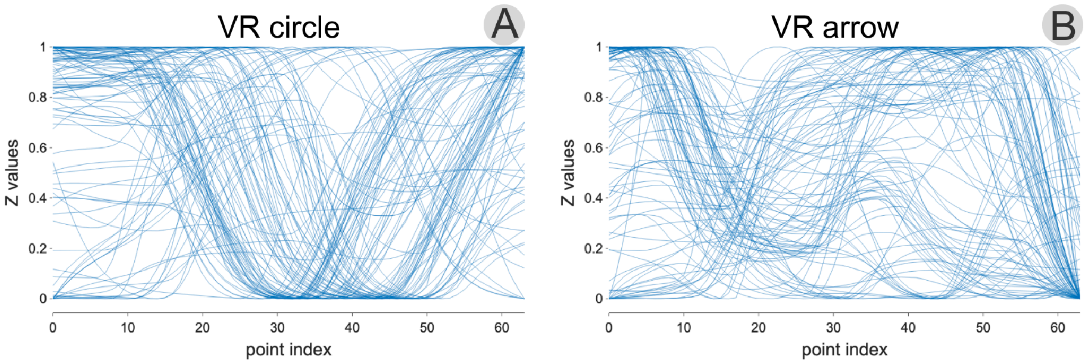
- Commonalities
  - Distribution shifts
    - **Between** VR and desktop datasets
    - **Within** VR or desktop datasets
  - Possible causes
    - input environments (i.e., VR or desktop)
    - stroke shapes
    - drawing speeds
    - other unknown factors



→ Challenge 1: It is hard to generalize the models trained on VR strokes to desktop strokes that comes from unseen distributions (i.e., out-of-distribution).

Preliminary Studies -- Findings

- Additional Dimensions
  - Z vectors spread out the entire output space and overlap between different stroke types.

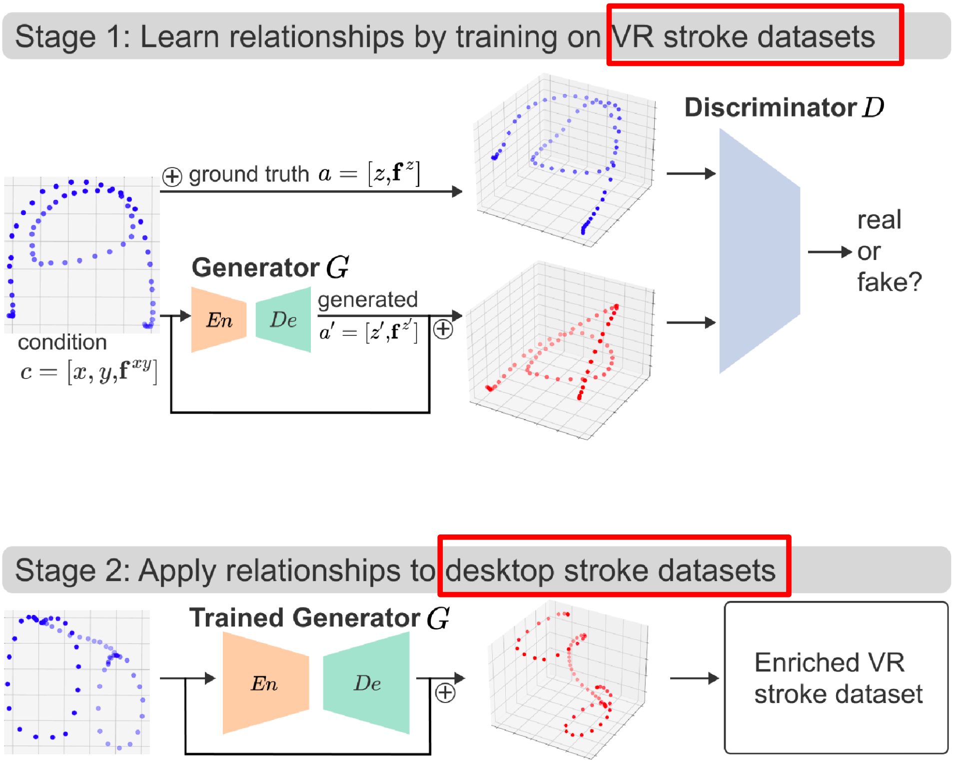


→ Challenge 2: It is hard to capture relationships between commonalities and additional dimensions from small original VR datasets.

Proposed Methods

We formulate the problem of generating VR strokes based on desktop strokes as a conditional time series generation problem.

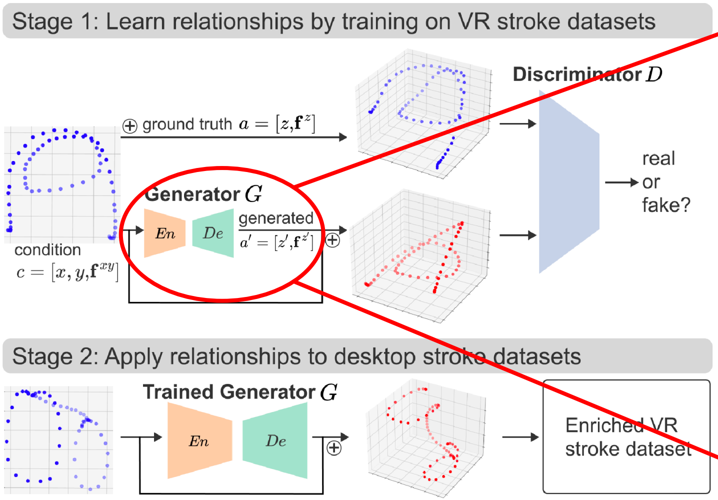
→ We generate additional dimensions conditioning on commonalities.

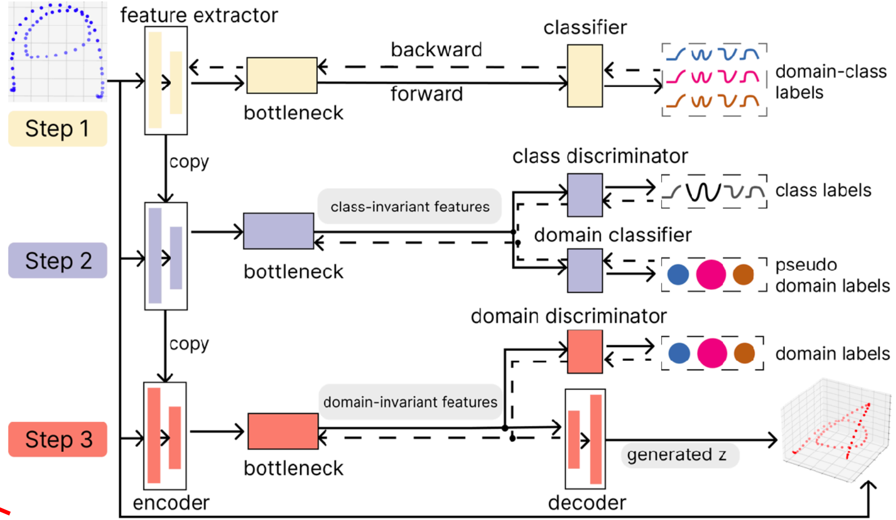


Proposed Methods

To address the first challenge, we further formulate the problem as a conditional time series generation problem under out-of-distribution circumstances.

→ Conditional domain-invariant generator with out-of-distribution generalization techniques [1] to deal with the distribution shifts.

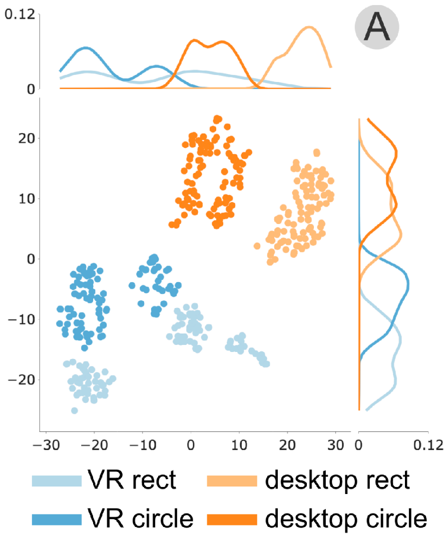




[1] Wang, Lu, et al. "DIVERSIFY: A General Framework for Time Series Out-of-distribution Detection and Generalization", IEEE TPAMI 2024

Proposed Method

- Conditional domain-invariant generator
  - Characterize latent distributions



[1] Wang, Lu, et al. "DIVERSIFY: A General Framework for Time Series Out-of-distribution Detection and Generalization", IEEE TPAMI 2024

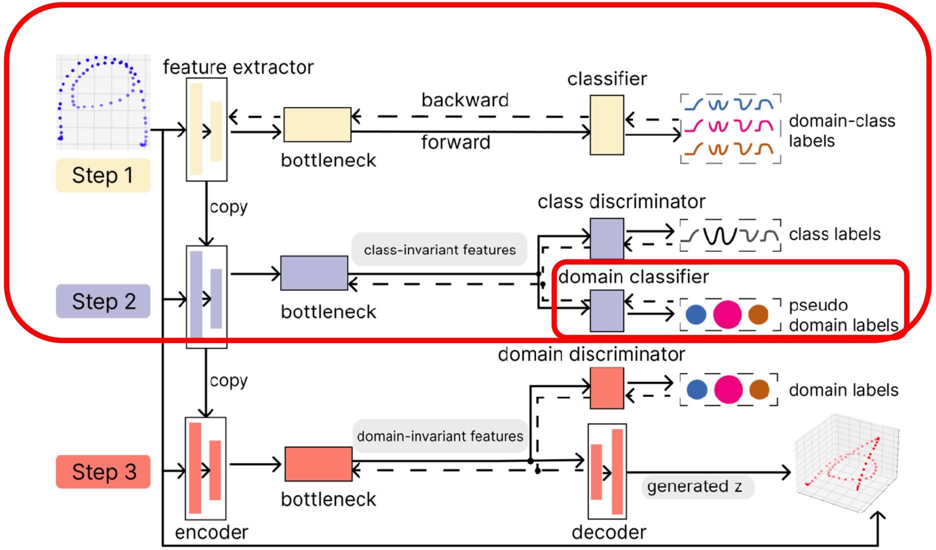
Proposed Method

- Conditional domain-invariant generator
  - Characterize latent distributions

Group all the VR strokes into several **latent** domains, whose distribution gaps are maximized [1].

✗ Individual factors (e.g., shapes, speeds)

✓ Domain-classifier



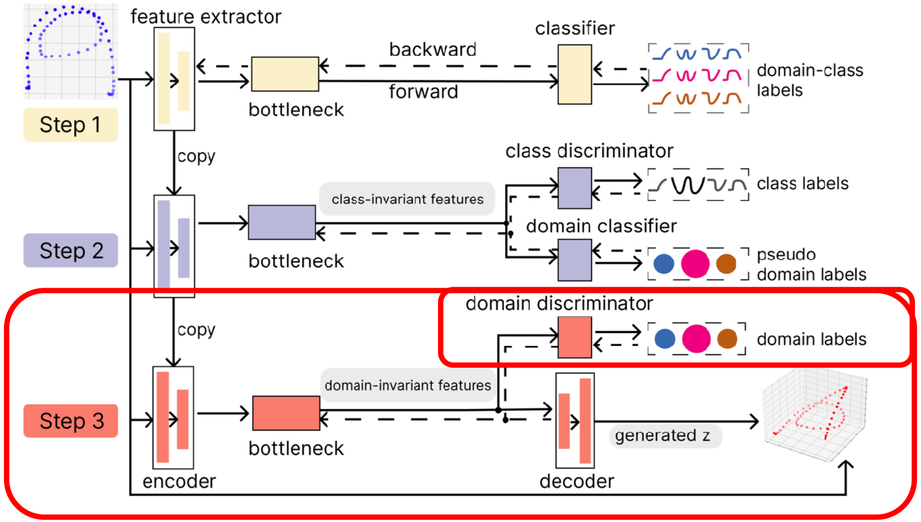
[1] Wang, Lu, et al. "DIVERSIFY: A General Framework for Time Series Out-of-distribution Detection and Generalization", IEEE TPAMI 2024

Proposed Method

- Conditional domain-invariant generator
  - Learn conditional domain-invariant representations

Utilize adversarial learning to fool a domain discriminator that classifies domains.

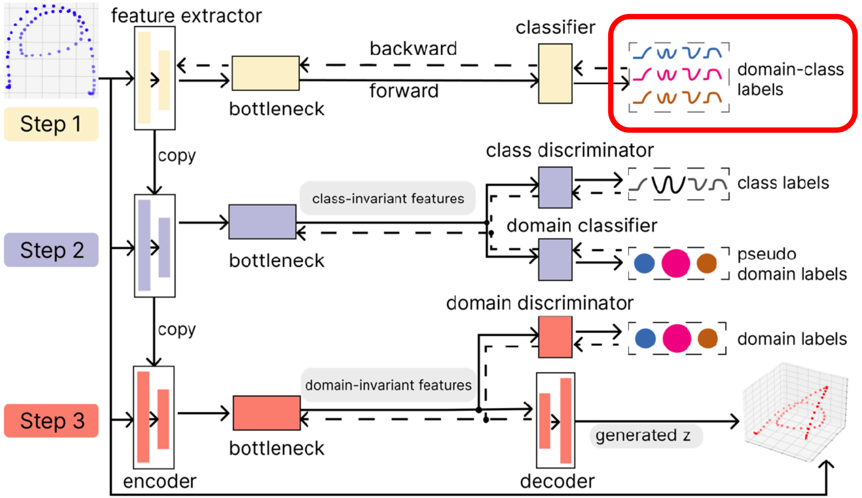
Make the discriminator unable to differentiate strokes from different latent domains [1].

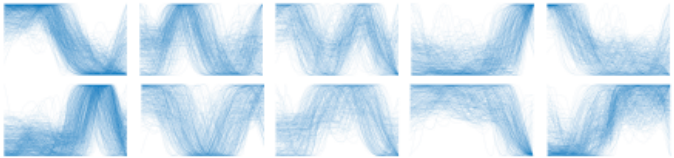


[1] Wang, Lu, et al. "DIVERSIFY: A General Framework for Time Series Out-of-distribution Detection and Generalization", IEEE TPAMI 2024

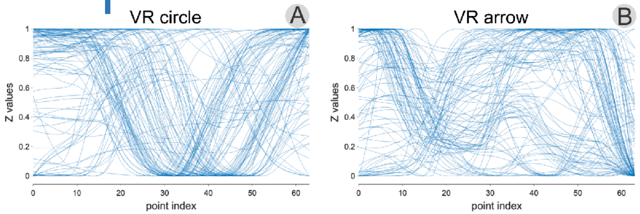
Proposed Method

- Conditional domain-invariant generator
  - Discretize output space to address the second challenge





Clustering Z vectors to reduce the burden of learning complex relationships

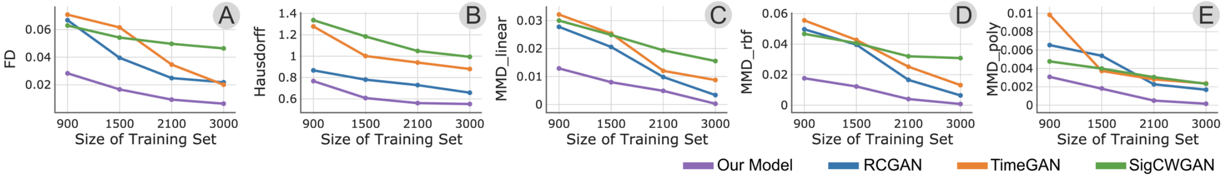


[1] Wang, Lu, et al. "DIVERSIFY: A General Framework for Time Series Out-of-distribution Detection and Generalization", IEEE TPAMI 2024

Evaluation

- Comparison with baselines
  - Purposes
    - assess the generalizability
    - examine the influence of training data size on model performance
  - Baselines
    - conditional time series generative models without integrating out-of-distribution generalization techniques
  - Training and testing sets
    - drawn from different distributions

	FD	Hausdorff	MMD_linear	MMD_rbf	MMD_poly
RCGAN	0.021814	0.657808	0.003368	0.006400	0.001685
TimeGAN	0.020063	0.879646	0.008755	0.013145	0.002367
SigCWGAN	0.046372	0.993806	0.015546	0.030862	0.002329
Our Model	**0.006323**	**0.551768**	**0.000272**	**0.000799**	**0.000160**



Evaluation

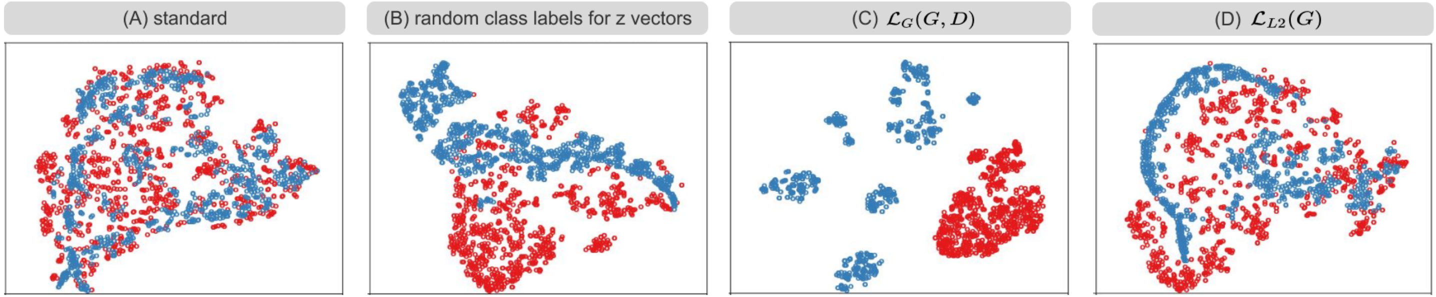
- Ablation Studies
  - Output space discretization
  - Loss functions of the generator

Table 2: Abalation studies on output space discretization.

	random class labels	z vector cluster labels
FD	0.030495	**0.006323**
hausdorff	0.667595	**0.551768**
mmd_linear	0.010434	**0.000272**
mmd_rbf	0.020506	**0.000799**
mmd_poly	0.001086	**0.000160**

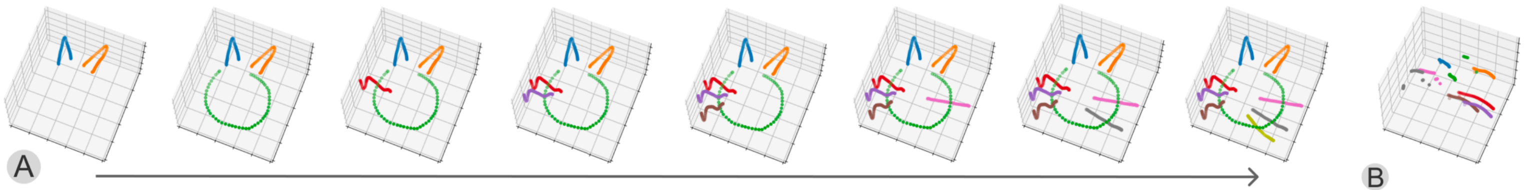
Table 3: Abalation studies on the loss functions of the generator G.

	$\mathcal{L}_G(G,D)$	$\mathcal{L}_{L2}(G)$	$\mathcal{L}_G+\mathcal{L}_{L2}$
FD	0.134043	0.013179	**0.006323**
hausdorff	1.121143	0.559428	**0.551768**
mmd_linear	0.071677	**0.000008**	0.000272
mmd_rbf	0.129512	0.001704	**0.000799**
mmd_poly	0.006074	0.000285	**0.000160**



Applications

- VR stroke prediction with CoSE [1] models



A: Trained on 5000 instances of synthesized VR cat sketches

B: Trained on 100 instances of real VR cat sketches

→ Our approach can make the prediction task possible with synthesized VR strokes, although the task is impossible with limited real VR strokes.

→ It reduces the burden to collect real VR strokes.

[1] Aksan, Emre, et al. "Cose: Compositional stroke embeddings." NeurIPS 2020.

Applications

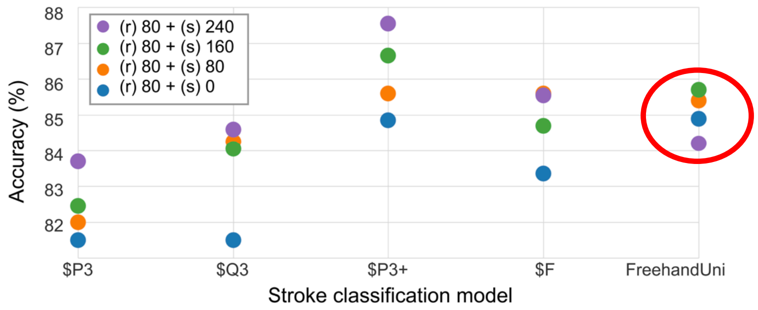
- VR stroke classification with different classifiers [1, 2]

Results with deep learning classifiers [1]

	Accuracy
800 real VR digits	96.67%
800 real VR digits + 1600 synthesized digits	99.63%
1600 synthesized digits	84.07%

The DL model can achieve satisfactory accuracy with synthesized datasets alone.

Results with template-based classifiers [2]



Accuracy may decrease when the amount of synthesized data exceeds a threshold.

→ The use of generated VR strokes in downstream applications needs to consider the characteristics of specific algorithms.

[1] Mohammadi, Seyed Saber, et al. "Pointview-gcn: 3d shape classification with multi-view point clouds." IEEE ICIP 2021.
[2] Ousmer, Mehdi, et al. "Recognizing 3D trajectories as 2D multi-stroke gestures." ACM ISS 2020.

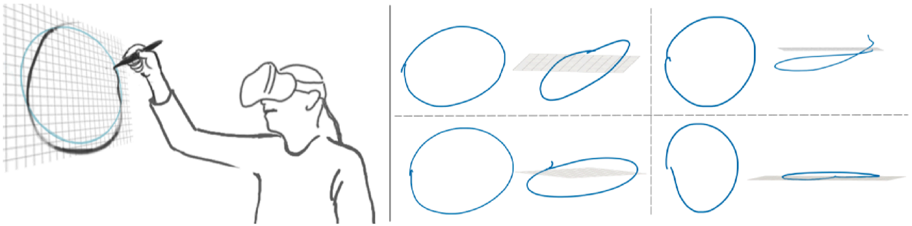
Take-home messages

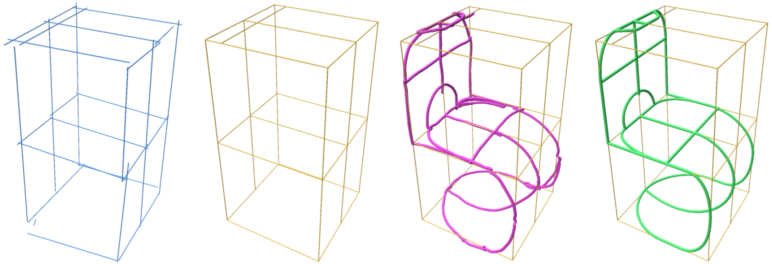
- Lessons learned for generating other types of VR interaction data?
  - Determining commonalities and additional dimensions.
  - Paying attention to distribution shifts.

- Reflections on the use of VR and desktop datasets
  - Our method does not require collecting real VR and desktop datasets under identical conditions thanks to its generalizability.
  - A limited real VR dataset that is insufficient for concrete applications might be adequate for training our generative model.
  - The use of generated VR strokes in downstream applications needs to consider the characteristics of specific algorithms.

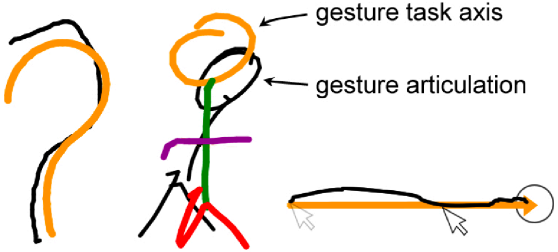
Future work

- From planar VR strokes [1] to non-planar VR strokes [2]





- Propose novel reference frames by adopting concepts such as gesture task axes [3] or scaffolds, rather than using the traditional Cartesian coordinate system.
- Reconsider the selection of commonalities and additional dimensions.

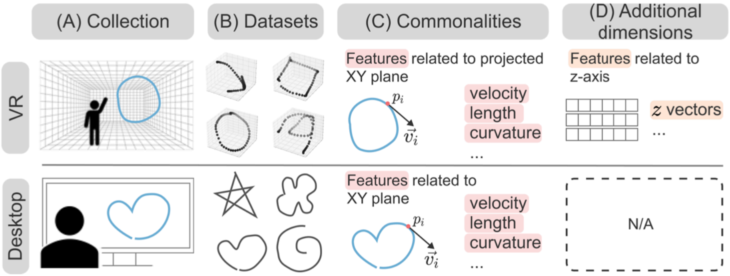


[1] Arora, Rahul, et al. "Experimental Evaluation of Sketching on Surfaces in VR." ACM CHI 2017
[2] Yu, Xue, et al. "Scaffoldsketch: Accurate industrial design drawing in VR." ACM UIST 2021.
[3] Vatavu, Radu-Daniel, et al. "Relative accuracy measures for stroke gestures." ACM International Conference on Multimodal Interaction 2013.

Generating Virtual Reality Stroke Gesture Data from Out-of-Distribution Desktop Stroke Gesture Data

Contributions:

- We explore generating VR stroke gesture data from desktop stroke gesture data as an alternative input source that is out-of-distribution.
- We propose a time series generative network with novel designs of output space discretization and conditional domain-invariant representation learning.
- We develop two applications that show the effectiveness and usefulness of the datasets enriched by our methods and demonstrate the potential opportunities opened by our methods.



- Code and datasets: https://github.com/yuanlinping/VRStrokeOOD