

Improving Human Activity Recognition through Diffusion Models

Kosuke Ukita^{*1} Tsuyoshi Okita^{*2}

^{*}Department of Artificial Intelligence, Kyushu Institute of Technology, Iizuka

¹ukita.kosuke299@mail.kyutech.jp, ²tsuyoshi@ai.kyutech.ac.jp

Keywords: generative model, diffusion model, transformer, human activity recognition

Introduction: This paper proposes a classifier based on a diffusion model for human activity recognition. To this end, we introduce three architectures: (1) the representation-conditional diffusion transformer, (2) the first classifier, RepcondFormer, and (3) the second classifier, RepcondClassifier. Experimental results show that RepcondFormer outperforms in all datasets, while RepcondClassifier performs better in three datasets when fine-tuned.

Approach: Our proposal model is based on the class-conditional diffusion transformer provided by DiT[1], which uses a Transformer architecture instead of UNet. The representation-conditional diffusion transformer integrates a transformer encoder that processes sensor data input, encoding a sliding window into a representation. After training out representation-conditional diffusion transformer, we explore two approaches in the subsequent experiments. The first approach involves building a classifier, named RepcondFormer, by fine-tuning the internal Transformer encoder. The second approach involves constructing the classifier RepcondClassifier, which generates classification results from a diffusion transformer tuned with class-conditional labels.

Results and Discussion:

Table 1: Figure shows the results in downstream task of human activity recognition.

Method	ADL		Opportunity		PAMAP2		REALWORLD		WISDM	
	acc	f1	acc	f1	acc	f1	acc	f1	acc	f1
Random Forest ¹	.8346	.6830	.6814	.4260	.6602	.5467	.6618	.5712	.4630	.4224
Transformer scratch ²	.8503	.8230	.6276	.6209	.7682	.7594	.8554	.8663	.8219	.8194
SENVt-u4[2]	.9078	.8780	.7489	.7411	.8488	.8454	.9015	.9124	.8693	.8679
SENVt-u7[2]	.8992	.8684	.7429	.7448	.8477	.8445	.9058	.9163	.8875	.8867
RepcondFormer	.9055	.8826	.7770	.7567	.8554	.8505	.9089	.9188	.9016	.9008
RepcondClassifier	20k: .8110	.6561	60k: .7587	.7138	60k: .8170	.8226	110k: .8558	.8366	100k: .7839	.7810
	30k: .9370	.9159	70k: .7423	.7008	70k: .8675	.8726	120k: .8546	.8443	200k: .8044	.8044
	40k: .8582	.8244	80k: .7412	.7155	80k: .8832	.8843	130k: .8546	.8377	300k: .7991	.8013
	50k: .8897	.8543	90k: .7482	.7343	90k: .8571	.8578	140k: .8594	.8423	400k: .7989	.8005
	60k: .9055	.8599	100k: .7693	.7658	100k: .8414	.8380	150k: .8570	.8259	500k: .7823	.7854

RepcondFormer outperforms in all datasets. Specifically, the F1 score improved by 3.29% for the WISDM dataset. RepcondClassifier performs better in three datasets: ADL, Opportunity, and PAMAP2 (Table1).

Fig1 illustrates signals generated for specific class labels from the perspective of signal generation. When using RepcondClassifier, we fine-tune the diffusion transformer with class labels. Subsequently, signals are generated conditioned on these class labels. The figure presents results using the PAMAP2 dataset. The activity labels are plotted from top to bottom. The left column shows randomly extracted signals from the original data, while the right column displays signals generated conditioned on the class labels. It appears that relatively distinct signals are being generated for each activity. The differences between the ground truth and the generated signals may be attributed to the fact that only a small portion of the overall data could be plotted.

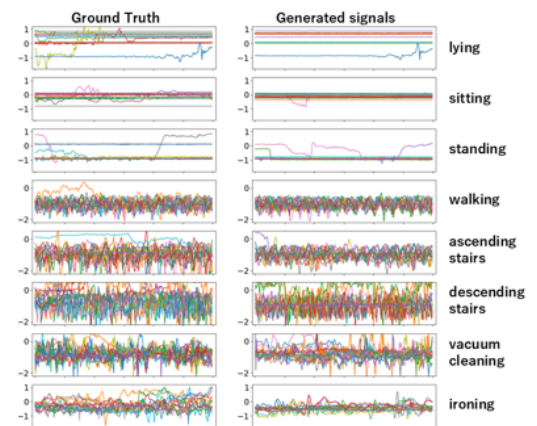


Fig 1: Figure shows signals generated for each class. 8 activity classes in the PAMAP2 dataset are plotted. The left figure shows the signals that are ground truth data. The right figure shows signals generated by conditioning on each class label.

- [1] Peebles & Xie, 2023, Scalable diffusion models with transformers, IEEE/CVF 4195-4205.
- [2] Okita and others, 2023, Towards LLMs for Sensor Data: Multi-Task Self-Supervised Learning, ACM 499-504.
- [3] Ukita & Okita, 2024, Analysis of Human Activity Recognition by Diffusion Models, ACM.