Resource-Efficient Time-Series Forecasting of Displacement Imagery using Koopman Autoencoders

Takayuki Shinohara and Hidetaka Saomoto
National Institute of Advanced Industrial Science and Technology
Tsukuba, Japan
shinohara.takayuki@aist.go.jp, h-saomoto@aist.go.jp

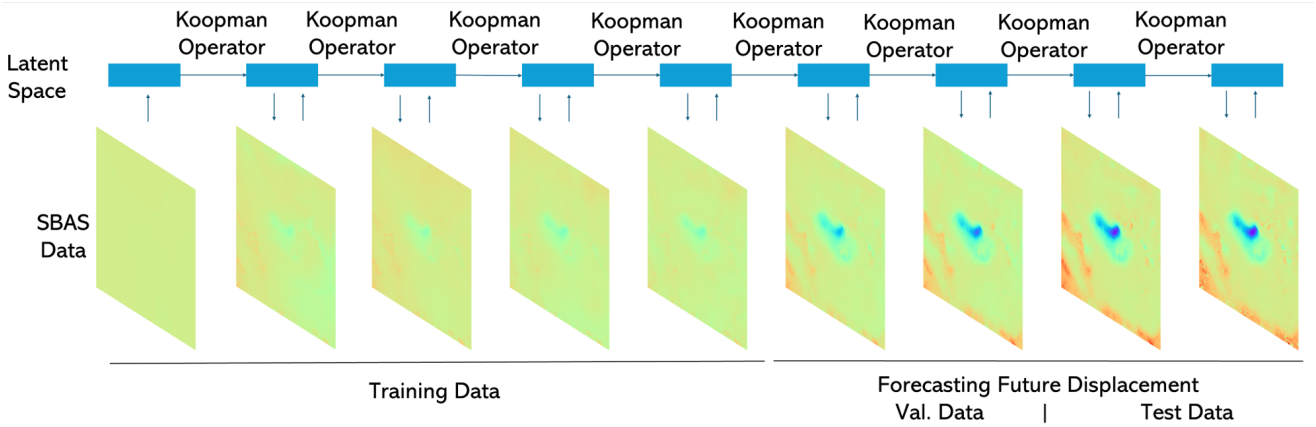

Figure 1. Overview of our method and visualization of one of the time series displacement satellite image samples in our dataset.

Abstract

Accurate yet lightweight forecasting of ground displacement is vital for real-time hazard response. We propose the Koopman Operator-based Autoencoder (KOA), a deep model that embeds a linear, physics-inspired Koopman operator into its latent space. A compact CNN encoder compresses each SBAS-InSAR frame; temporal evolution is then propagated by a single linear map, slashing parameter count and FLOPs relative to Transformer-style networks. Trained on nationwide Japanese SBAS archives and evaluated on unseen regions (Turkey, Italy, Hawaii), KOA matches state-of-the-art accuracy while cutting computational cost by orders of magnitude. This efficiency makes KOA practical for deployment on modest hardware in operational monitoring systems.

1. Introduction

Ground–surface deformation endangers infrastructure and populations, yet time series displacement images calculated by Small Baseline Subset (SBAS) algorithm [2] remain purely observational. Forecasting future displacement requires modelling highly nonlinear spatio-temporal dynamics across decades of imagery while staying computationally tractable.

Large sequence models such as Transformers and diffusion networks can predict complex dynamics, but their quadratic attention cost, heavy sampling, and weak physical guarantees make them ill-suited to SBAS archives. A lightweight, physics-aware alternative is therefore essential.

We introduce the *Koopman Operator-based Autoencoder* (**KOA**), which embeds a finite-dimensional Koopman operator [7, 8, 13] directly into the latent space of a compact convolutional autoencoder. A single linear map propagates latent states, while spectral regularisation enforces stability. KPA thus marries deep spatial encoding with physics-consistent temporal evolution, slashing parameters and FLOPs relative to Transformer baselines. Experiments on nationwide Japanese SBAS data—and transfer tests in Turkey, Italy, and Hawaii—show that KPA delivers state-of-the-art accuracy at a fraction of the computational cost, offering a practical tool for real-time hazard mitigation and sustainable urban planning.

2. Method

2.1. Problem Statement

We aim to predict future frames from ground surface displacement imagery. Each displacement image $\mathbf{x}_i \in \mathbb{R}^{H \times W}$

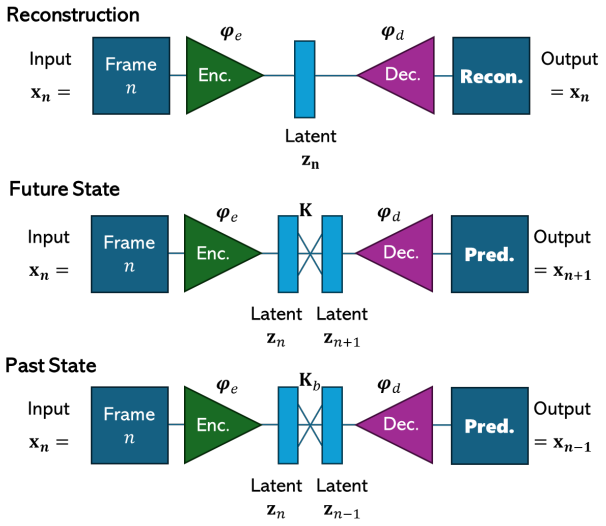

Figure 2. Our network architecture comprises three key components: Reconstruction, Future State, and Past State. All encoders and decoders are weight-shared.

is treated as a discrete, time-invariant system:

$$\mathbf{x}_{i+1} = \mathbf{f}(\mathbf{x}_i), \tag{1}$$

where $\mathbf{f}$ is unknown and possibly nonlinear. Instead of learning $\mathbf{f}$ directly, we use Koopman theory, which finds a transformation φ_e such that the system becomes linear:

$$\mathbf{z}_{k+1} = \mathbf{K}, \mathbf{z}_k, \quad \mathbf{z}_k = \varphi_e(\mathbf{x}_k), \tag{2}$$

allowing us to predict l steps as $\mathbf{z}_{k+l} = \mathbf{K}^l \mathbf{z}_k$ and reconstruct via $\mathbf{x}_{k+l} = \varphi_d(\mathbf{z}_{k+l})$.

2.2. Koopman Operator-based Autoencoder(KOA)

KOA combines an autoencoder with linear latent dynamics:

- **Encoder:** CNN-based encoder φ_e maps each frame to a latent vector $\mathbf{z}_n$.
- **Koopman Layer:** Latent vectors evolve linearly, i.e., $\mathbf{z}_{n+k} = \mathbf{K}\mathbf{z}_n$.
- **Decoder:** CNN-based decoder φ_d maps $\mathbf{z}_{n+k}$ to predicted frames $\hat{\mathbf{x}}_{n+k}$.

Additionally, a backward operator $\mathbf{K}_b$ predicts past states and is trained to approximate $\mathbf{K}^{-1}$. Orthogonal $\mathbf{K}$ allows $\mathbf{K}_b = \mathbf{K}^\mathsf{T}$ for consistency.

The loss consists of reconstruction, forward and backward prediction, and latent consistency:

$$\mathcal{L} = \gamma_{\text{rec}}\mathcal{L}_{\text{rec}} + \gamma_{\text{fwd}}\mathcal{L}_{\text{fwd}} + \gamma_{\text{bwd}}\mathcal{L}_{\text{bwd}} + \gamma_{\text{lat}}\mathcal{L}_{\text{lat}}. \tag{3}$$

2.3. Network and Training

Our architecture uses a 3-stage FasterNet encoder and symmetric decoder with depthwise convolutions and LayerNorm. Latent codes $\mathbf{z}_n$ are projected from the final feature map and reconstructed via upsampling. The Koopman matrix $\mathbf{K}$ operates on these latent vectors.

We train on N_{in} input frames and predict N_{out} future frames. Initially, only reconstruction and prediction loss are active. After epoch e_s, latent consistency loss is added with weight γ_{lat}.

3. Experimental Results

3.1. Training Details

Key tunables are the learning rate l_r (with decay l_{rd} and schedule), maximum predict roll-out k_m, loss weights $\{\gamma_{\text{recon}}, \gamma_{\text{fwd}}, \gamma_{\text{past}}, \gamma_{\text{lat}}\}$, and N_l set as 16. Input- and output-frame counts ($N_{\text{in}}, N_{\text{out}}$) are fixed by the task. All models are trained for 600 epochs; we adopt Adam, batch size 32, cosine decay, and early stopping on validation MSE.

3.2. SBAS Dataset

We employ the nationwide Japanese SBAS archive of [9], comprising 191 deformation stacks generated from Sentinel-1 via LiCSBAS [10](Table 3 in the supplemental material 6.2). Interferograms are unwrapped, denoised, outliers removed, and frames down-sampled to 64×64 pixels. Each series is chronologically split: the first 30 frames form the training context (N_{in}), the next 30 frames the validation targets (N_{out}), and the most recent 20 frames the test horizon(Figure 5 in the supplemental material 6.2). To probe data-scarce regimes we additionally consider $(N_{\text{in}}, N_{\text{out}}) = (10, 50)$ and $(20, 40)$ while keeping the same test set. This strict forward-time split mimics operational forecasting. The details are described in the supplemental material 6.2.

3.3. Forecasting Performance

Table 1 summarises mean-squared displacement error (mm) over the 20 test frames. KOA consistently outperforms or matches deep baselines while using two orders of magnitude fewer parameters. **KOA** yields MSEs of 39.19, 35.93, 34.71 for $N_{\text{in}} = 10, 20, 30$, respectively, rivaling ConvLSTM and beating Transformers on short contexts. Adding an autoencoder reduces DMD and Mamba errors by $\sim 70\%$, underscoring the value of compact latent representations. KOA surpasses its non-AE Koopman variant by up to 65%, confirming that coupling Koopman dynamics with learned compression is crucial.

Transformers underperform on this regional-scale dataset, reflecting their data hunger [4]. By contrast, KOA's physics-based prior delivers strong accuracy from limited training sequences, making it well suited to geographically constrained or rapidly deployed monitoring systems.

3.4. Computational Efficiency Analysis

To evaluate the computational efficiency of our proposed KOA model, we conduct comprehensive comparisons with

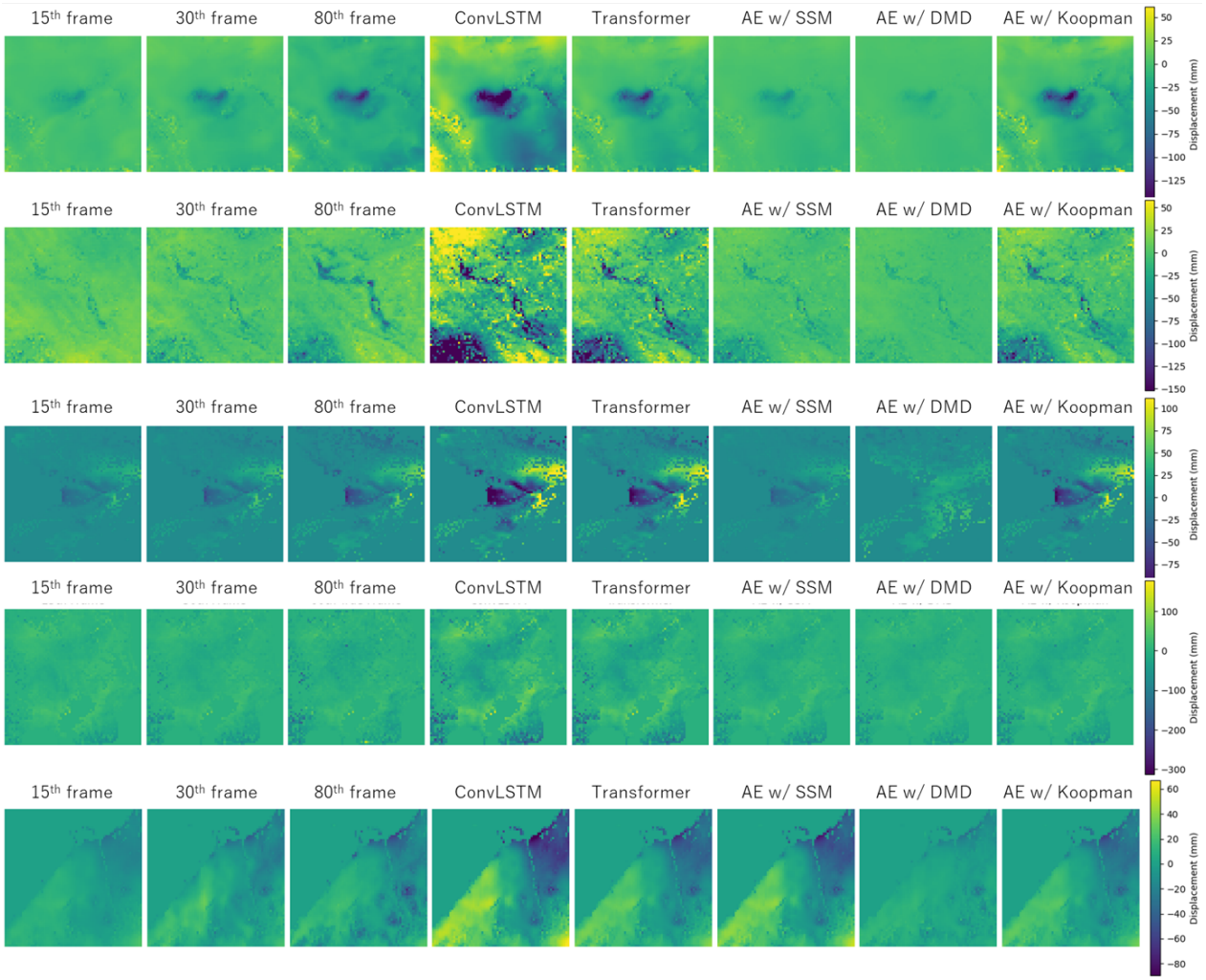

Figure 3. Starting the initial SBAS image(the first frame of training data), we compared our model, Dynamic Mode Decomposition-based AutoEncoder and State Space Model-based Auto Encoder to forecast the future displacement of SBAS data for the 80th frames(the last frames of test data). All models use the first 30 frames as context data for the training process.

Method	$N_{\text{in}} = 10$	$N_{\text{in}} = 20$	$N_{\text{in}} = 30$
SSM w/o AE	143.348	141.438	138.156
SSM w AE	41.782	39.238	36.593
DMD w/o AE	153.854	151.289	143.662
DMD w AE	45.721	42.299	38.327
Koopman w/o AE	129.101	102.716	100.243
Koopman w AE(Ours)	39.192	35.929	34.708
ConvLSTM	40.122	35.524	33.182
Transformer	83.316	45.625	40.537

Table 1. Prediction displacement of Mean Square Error (mm) comparison at test data: Our Koopman Operator-based Auto Encoder(KOA) and baseline algorithms on SBAS dataset. N_{in} is used as input frames of SBAS dataset.

state-of-the-art models, including Transformers and ConvLSTM networks. All experiments were performed on a single NVIDIA RTX 3080 GPU to ensure fair comparison. All models use the first 30 frames as context data for the training process. Specific experimental details are described in the supplemental material 7.1.

As shown in Table 2, our KOA achieves remarkable efficiency improvements across all computational metrics. The parameter count of our model is merely 0.2M, which represents a reduction of 99.3% compared to the Transformer baseline (28M parameters) and 98.0% compared to ConvLSTM (10M parameters). This dramatic reduction in model complexity is attributed to our efficient Koopman operator design and the utilization of spectral methods through FFT.

In terms of computational complexity, KOA demonstrates exceptional efficiency, requiring only 19M FLOPs compared to 31.3G for Transformers and 76.5M for ConvLSTM. This represents a substantial 99.9% reduction in computational requirements compared to the Transformer baseline. The

Model	Parameters	FLOPs	Inference Time
Transformer	28M	31.3G	15-30ms
ConvLSTM	10M	76.5M	5-10ms
KOA (Ours)	**0.2M**	**19M**	**1-3ms**

Table 2. Computational Resource Requirements.All models use the first 30 frames as context data for the training process.

inference speed of our approach demonstrates a particular advantage in real-time applications, processing each frame in 1- 3ms compared to 15- 30ms for Transformers and 5-10ms for ConvLSTM. This translates to a 10× speedup over Transformers and 3× over ConvLSTM, while maintaining competitive accuracy as shown in our previous experiments. This efficiency stems from the spectral representation and the linear nature of operations in the Koopman-embedded latent space, eliminating the need for complex attention mechanisms or recurrent computations.

3.5. Robustness Evaluation on Unseen Regions

To assess the robustness of our proposed Koopman Operator Autoencoder (KOA) in predicting ground deformation in unseen regions, we conducted evaluations on three independent volcanic areas: Mauna Loa in Hawaii (Sentinel-1 frame 087D_07004_060904), Mount Etna in Italy (Sentinel-1 frame 124D_05291_081406), and Mount Ararat in Turkey (Sentinel-1 frame 152D_04960_131313) (shown in Supplementary material 6.3). These locations were not included in the training dataset, providing a rigorous test of the model's generalization capability.

Our experimental results demonstrate that KOA trained on the first 30 frames as context data effectively generalizes to these unseen locations, accurately predicting ground deformation patterns derived from SBAS data(Figure 4). The model successfully reconstructs spatiotemporal deformation dynamics despite differences in geological characteristics, indicating its adaptability to diverse terrains. We evaluated on Mauna Loa the mean absolute error (MAE) was 525.2 mm, on Mount Etna it was 36.5 mm, and on Mount Ararat it reached 39.1 mm—each exceeding the Japanese SBAS data.

In the case of Mauna Loa, we observed instances where the predicted deformation deviated significantly from historical trends. This suggests that our approach could potentially be utilized for anomaly detection, as such deviations may indicate abnormal volcanic activity. This capability highlights the potential of KOA not only for standard forecasting tasks but also for early warning systems in volcanic monitoring applications. Overall, these findings confirm that our model maintains high robustness and applicability even in regions with no prior training data, supporting its effectiveness in real-world geophysical and remote sensing scenarios.

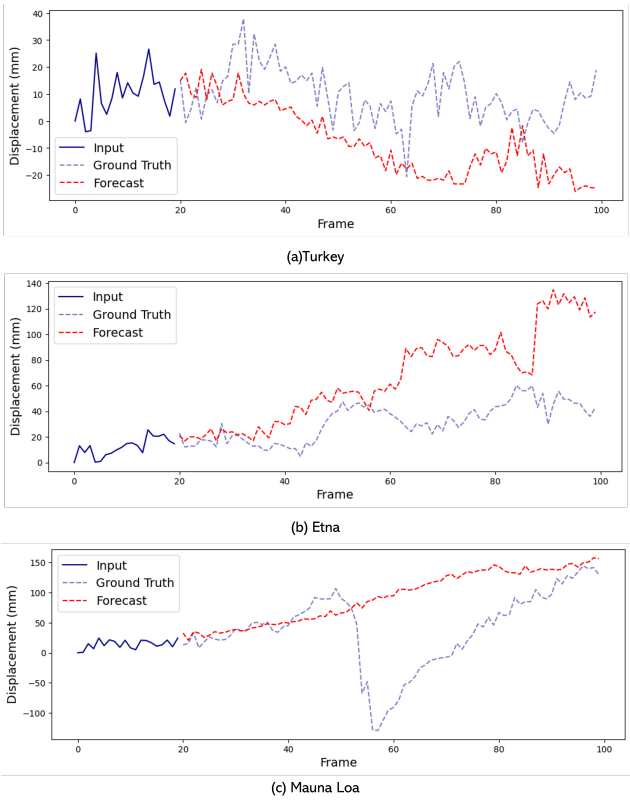

Figure 4. Predicted ground deformation on unseen regions using KOA. The model using the first 30 frames as context data for the training process successfully captures deformation trends across different geological environments.

4. Conclusion

In this paper, we introduced Koopman Operator-based Autoencoder (KOA), an efficient and lightweight model for forecasting time-series displacement of SBAS data. KOA leverages the Koopman operator to map the encoder-extracted nonlinear latent space of SBAS data into a linear latent space. A key aspect is enforcing temporal consistency in the latent variables by exploiting the time-invariance property of the Koopman operator for autonomous dynamical systems. We evaluated KOA on Japanese SBAS datasets, comparing it against physics-based techniques and recent time series prediction methods, including Diffusion models [6, 17]. Our method demonstrated superior performance with shorter training times and faster inference compared to state-of-the-art approaches. KOA's ability to handle limited data is significant for computationally demanding simulations of high-dimensional physical systems. Furthermore, this method can be extended to non-autonomous control systems using a bilinearly recurrent physics-based architecture based on [5].

Acknowledgements

This study makes use of displacement time–series data calculated by Sentinel-1 using LicSBAS. We used ABCI 3.0 provided by AIST and AIST Solutions[18].

References

[1] A. M. Avila and I. Mezić. Data-driven analysis and forecasting of highway traffic dynamics. *Nature Communications*, 11 (1):2090, 2020. 1

[2] Paolo Berardino, Gianfranco Fornaro, Riccardo Lanari, and Eugenio Sansosti. A new algorithm for surface deformation monitoring based on small baseline differential sar interferograms. *IEEE Transactions on geoscience and remote sensing*, 40(11):2375–2383, 2002. 1

[3] Akshunna S. Dogra and William Redman. Optimizing neural networks via koopman operator theory. In *Advances in Neural Information Processing Systems*, pages 2087–2097. Curran Associates, Inc., 2020. 1

[4] Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, Jakob Uszkoreit, and Neil Houlsby. An image is worth 16x16 words: Transformers for image recognition at scale. In *International Conference on Learning Representations (ICLR)*, 2021. 2

[5] Debdipta Goswami and Derek A. Paley. Bilinearization, reachability, and optimal control of control-affine nonlinear systems: A koopman spectral approach. *IEEE Transactions on Automatic Control*, 67(6):2715–2728, 2022. 4

[6] Jonathan Ho, Tim Salimans, Alexey Gritsenko, William Chan, Mohammad Norouzi, and David J Fleet. Video diffusion models. *arXiv:2204.03458*, 2022. 4

[7] B. O. Koopman. Hamiltonian systems and transformation in Hilbert space. *Proceedings of National Academy of Sciences*, 17:315–318, 1931. 1

[8] Bethany Lusch, J Nathan Kutz, and Steven L Brunton. Deep learning for universal linear embeddings of nonlinear dynamics. *Nature communications*, 9(1):4950, 2018. 1

[9] Yu Morishita. Nationwide urban ground deformation monitoring in japan using sentinel-1 licsar products and licsbas. *Progress in Earth and Planetary Science*, 8(1):1–23, 2021. 2, 3

[10] Yu Morishita, Milan Lazecky, Tim J Wright, Jonathan R Weiss, John R Elliott, and Andy Hooper. Licsbas: An open-source insar time series analysis package integrated with the licsar automated sentinel-1 insar processor. *Remote Sensing*, 12(3):424, 2020. 2, 3

[11] Sriram S. K. S. Narayanan, Duvan Tellez-Castro, Sarang Sutavani, and Umesh Vaidya. Se(3) koopman-mpc: Data-driven learning and control of quadrotor uavs, 2023. 1

[12] Indranil Nayak, Fernando L Teixeira, and Mrinal Kumar. Koopman autoencoder architecture for current density modeling in kinetic plasma simulations. In *2021 International Applied Computational Electromagnetics Society Symposium (ACES)*, pages 1–3. IEEE, 2021. 1

[13] Samuel E Otto and Clarence W Rowley. Linearly recurrent autoencoder networks for learning dynamics. *SIAM Journal on Applied Dynamical Systems*, 18(1):558–593, 2019. 1

[14] Sebastian Peitz, Samuel E. Otto, and Clarence W. Rowley. Data-driven model predictive control using interpolated Koopman generators. *SIAM Journal on Applied Dynamical Systems*, 19(3):2162–2193, 2020. 1

[15] Clarence W. Rowley, Igor Mezić, Shervin Bagheri, Philipp Schlatter, and Dan S. Henningson. Spectral analysis of nonlinear flows. *Journal of Fluid Mechanics*, 641:115–127, 2009. 1

[16] Peter J. Schmid. Dynamic mode decomposition of numerical and experimental data. *Journal of Fluid Mechanics*, 656:5–28, 2010. 1

[17] Xiaoqian Shen, Xiang Li, and Mohamed Elhoseiny. Mostgan-v: Video generation with temporal motion styles. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 5652–5661, 2023. 4

[18] Ryousei Takano, Shinichiro Takizawa, Yusuke Tanimura, Hidemoto Nakada, and Hirotaka Ogawa. Abci 3.0: Evolution of the leading ai infrastructure in japan, 2024. 5

[19] Matthew O. Williams, Ioannis G. Kevrekidis, and Clarence W. Rowley. A data–driven approximation of the Koopman operator: extending Dynamic Mode Decomposition. *Journal of Nonlinear Science*, 25(6):1307–1346, 2015. 1

[20] Enoch Yeung, Soumya Kundu, and Nathan Hodas. Learning deep neural network representations for koopman operators of nonlinear dynamical systems. In *2019 American Control Conference (ACC)*, pages 4832–4839, 2019. 1