•VAEs are strong for image generation, dimensionality reduction, and latent space regularization.
•Video dynamic textures are continuous/moving; Euclidean methods cannot model their manifold structure.

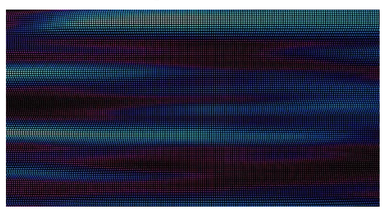
IOCMA 2026 Conference
The 2nd International Online Conference on Mathematics and Applications
10–12 June 2026 | Online
MDPI

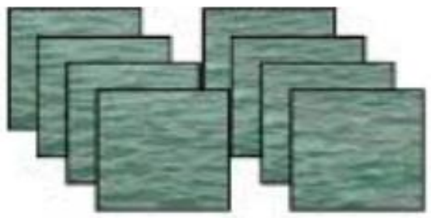
Variational AutoEncoder (VAE) and Lie-SVM Approaches in capturing Static and Dynamic Generative-Discriminative Features of Visual Datasets

Hugo Wai Leung MAK
1 Department of Mathematics, The Chinese University of Hong Kong, Hong Kong, China
2 Department of Mathematics, The Hong Kong University of Science and Technology, Hong Kong, China
Contact Email: hwlmak@ust.hk

INTRODUCTION & AIM

- Analyzing video and animation requires **quantification of static and dynamic textures**
- The integration of **geometry, probability, machine learning** and **artificial intelligence** becomes particularly crucial
- VAE is strong at **image generation, dimensionality reduction**, and **latent space regularization**; however Euclidean approaches fail to model the manifold structure of video dynamic textures



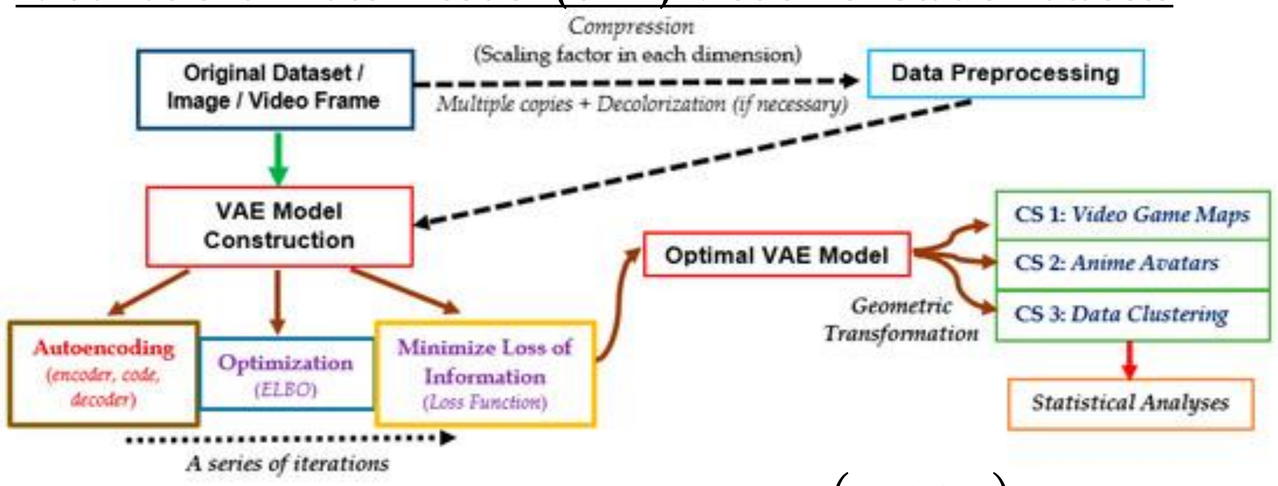


Key Aims

- Combine **VAE probabilistic modeling** (*generating video frames & latent inference*) with **Lie group geometry** (*dynamic texture classification*)
- Seek an approach to **formalize dynamic structure** and build **Lie-SVM classifier** with appropriate kernels
- Create a unified pipeline for **generative latent modeling** and **manifold-aware classification** for future interactive user interface (UI) applications

METHODOLOGIES

1. Variational AutoEncoder (VAE) Model for Static Datasets



Encoder **maps input to latent distribution** $\left(q_\phi(z|x)\right)$

Decoder **reconstructs** $\left(p_\theta(x|z)\right)$

Total Loss of Information (*to be **minimized***)

$$L = D_{KL}(q_\phi(z|x)\|p(z)) + E_{q_\phi(z|x)}[-\log p_\theta(x|z)]$$

KL Divergence: Regularize latent space
Reconstruction Loss: Preserve original features

METHODOLOGIES (Cont'd)

2. Dynamic Texture Modeling & Lie-SVM Classifier

- ARMA model describes **video dynamic textures** and extracts **matrix shape of Gaussians (SOG) descriptors**
- **Lie group manifold distance** (Frobenius norm):

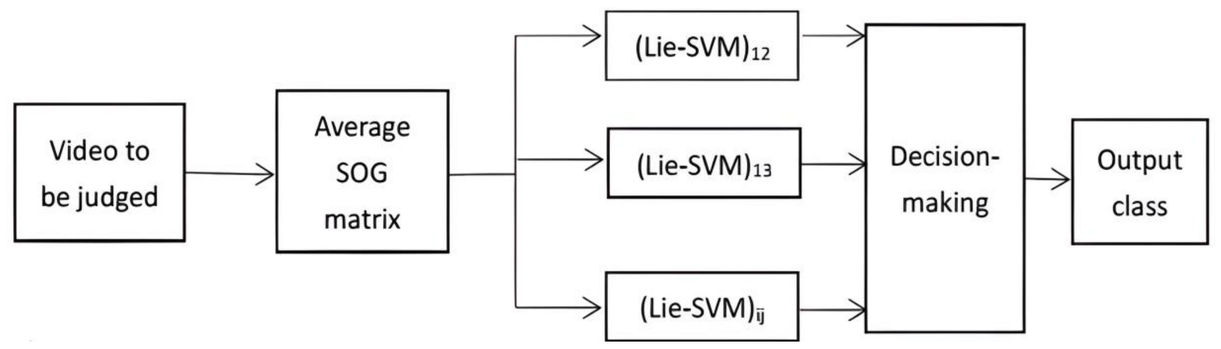
$$d(g_i, g_j) = \left\|\ln\left(g_i^{-1} g_j\right)\right\|_F$$

- Embed **manifold distance** into SVM as kernel function:

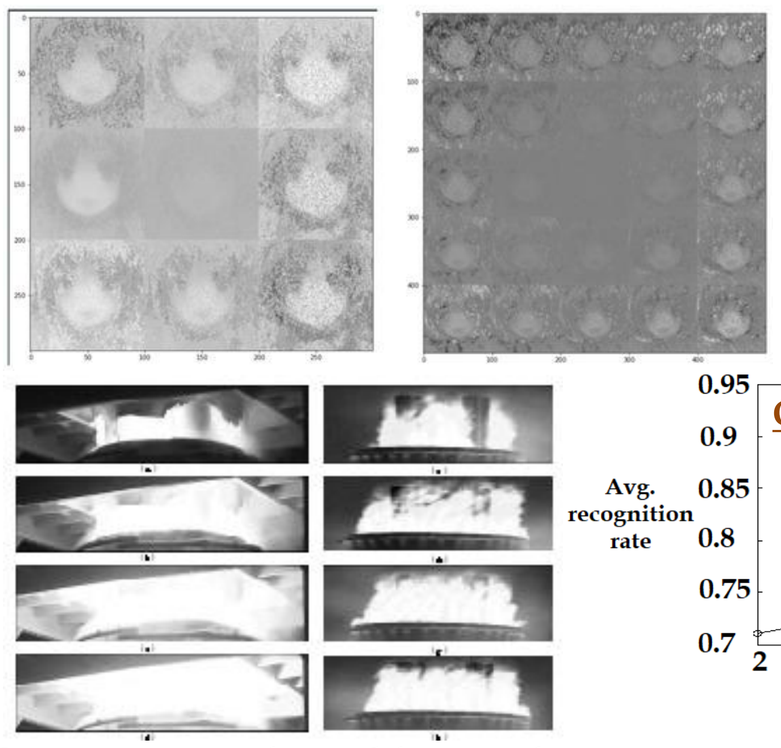
$$K(g_i, g_j) = \exp\left[\lambda\left(\left\|\ln\left(g_i^{-1} g_j\right)\right\|_F + \left\|\ln\left(g_j^{-1} g_i\right)\right\|_F\right) + h\right]$$

$0 < \lambda < 1$: *scaling factor;* ***h****: offset factor* → *by user or ML*

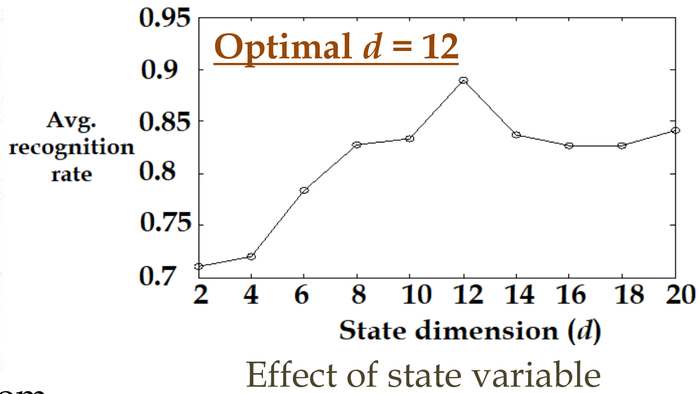
- Adopt **one-vs-one multi-class Lie-SVM strategy** + voting rule for **classification**
- Apply **geometric regularization** of scaling parameters for optimizing classification performance



RESULTS & DISCUSSION



Two sample output images obtained from the VAE model (*anime-girl dataset*)



Reconstruction of visual features from video – average SOG matrix is obtained

Effect of state variable dimension (d) on recognition rate

1) VAE clustering accuracy (MNIST): **85.4%** (50 training epochs)
2) Lie-SVM overall accuracy: **11% higher** than Martin-distance SVM
3) Lie-SVM runtime: **2.3% lower** than traditional SVM

CONCLUSION

- *The proposed Lie-SVM achieves **superior accuracy** and **efficiency** for video dynamic texture classification as compared to conventional algorithms*
- ***Combining VAE and Lie group manifold learning** provides a **robust solution** for static visual generation and dynamic video analytics*