Call for Papers
Themed Series of APSIPA Trans. on Signal and Information Processing
"Recent Developments in Diffusion Models and Vision Transformers"

Introduction

The rapid advancement of generative artificial intelligence and deep learning has positioned diffusion models and vision transformers (ViTs) at the forefront of contemporary computer vision research. Diffusion models, renowned for their ability to generate high-fidelity data through iterative denoising, have significantly advanced generative modeling. Concurrently, vision transformers—leveraging self-attention mechanisms—have redefined visual representation learning, outperforming traditional convolutional neural networks across tasks such as image classification, segmentation, and multimodal understanding.

Recent breakthroughs underscore the suitability of transformer-based architectures as backbones for diffusion models, offering a scalable alternative to conventional U-Net designs and enabling unified frameworks for image, video, and multimodal synthesis. Despite these advancements, several key challenges remain, including efficient scaling to high-resolution data, enhancing controllability in conditional generation (e.g., text-guided editing), and adapting these models to real-world domains such as medical imaging and autonomous systems.

This themed series aims to showcase state-of-the-art research, innovative methodologies, and practical applications at the intersection of diffusion models and vision transformers. We welcome contributions on novel architectural designs, computational efficiency, and ethical considerations to advance scalable, robust, and deployable AI systems for both generative and discriminative tasks

Topics of Interest

We invite researchers to submit original research articles and reviews that address the following topics (but are not limited to):

- Theoretical advancements in diffusion models, e.g., stochastic differential equations, sampling algorithms
- Diffusion Models for controllable generation
- Diffusion models for image editing, enhancement, and restoration
- Efficiency and scalability design for diffusion models.
- Innovative Attention mechanism for Vision transformer
- Memory-efficient architectures for vision transformer
- Unified diffusion-transformer for scalable image/video generation
- Diffusion Transformers for joint visual and language generation
- Compression technique for vision transformers and diffusion models
- Other related topics

Submission Guidelines

Each paper submitted to this series will be reviewed using the first-come-first-serve principle. The target of the first round of decision-making is 5 weeks, and the period of the first round of revision is 2 weeks. The paper will be accepted between 8-12 weeks (depending on 1 or 2 revisions). Once the submission window has closed, accepted papers ready for publication will be published online. The series will be accompanied by an editorial written by the guest editorial team. If a paper cannot be accepted within the publication window, it will be considered as a regular paper.

For submission details, please refer to:

https://nowpublishers.com/Journal/AuthorInstructions/SIP

Important Dates

- **Submission Deadline: 1 October, 2025**

Guest Editorial Team

- Xinchao Wang (Lead), National University of Singapore, Singapore
  Email: xinchao.w@gmail.com
- Zunlei Feng, Zhejiang University, China
  Email: zunleifeng@zju.edu.cn
- Jiayan Qiu, Leicester University, UK
  Email: jq46@leicester.ac.uk
- Xingyi Yang, National University of Singapore, Singapore
  Email: xyang@u.nus.edu