2022 Tencent AI Lab Rhino-Bird Focused Research Program

Research Topics

1. Machine Learning for Life Science & Chemistry

Machine Learning, especially, Deep Learning has been successfully applied to fields including computer vision, speech recognition and natural language processing. Meanwhile, recent research shows the potential of machine learning in solving scientific problems, such as protein folding, molecular dynamics simulation and protein / molecular docking. The goal of this project is to develop cutting-edge machine learning algorithms for solving hard problems in the life science and chemistry area. Some possible research topics include:

1.1. The Methology for Drug Discovery & Chemistry

1.1.1. Graph generation algorithms for molecular de novo design

Molecular design and molecular synthesis based on machine learning have achieved great success, improving the efficiency and accuracy of molecular design and synthesis. The research directions of this topic include:

- Molecule generation according to the density data of binding protein target and molecule or other related data.
- Molecule generation according to the chemical synthesis path.
- Novel molecule generation paradigm design.

1.1.2. Physics-driven multi-body interaction with deep graph learning

For rigid-body / flexible docking between proteins and small molecules, the performance of deep learning-based models may be limited due to insufficient training data obtained through experiments. The research directions of this topic include:

- Data-efficient deep learning method by integrating physics prior knowledge into the model, to further boost its performance in rigid-body / flexible docking problems.
- Large-scale unsupervised pretraining for the rigid-body / flexible docking between proteins and small molecules.
- Efficient simulation of systems containing various rigid-body objects, such as the motion simulation or force simulation of rigid bodies containing sticks, hinges, rotatable rods, and other mechanical structures under certain force field conditions.

1.1.3. Knowledge transfer and meta learning

In the structure-based molecular virtual screening problem, it often fails to learn a well-generalized model for the target assay given only a few labeled ligands. The research directions of this topic include:

- Design transfer learning, meta-learning, or multi-task learning algorithms to improve the performance of ligands activity prediction models for target assays by transferring knowledge learned from relevant assays that have large amounts of labeled ligands in public datasets.
- Design algorithms to use the context information of the target assay to reduce overfitting of training data that may be noisy or biased.

1.1.4. Out-of-distribution (OOD) learning

The problem of distribution shift is prevalent in various tasks of AI-aided drug discovery. For example, for the task of structure-based virtual screening, the models are often trained on data of known protein targets but have to be tested on unknown targets. Meanwhile, the current model backbone of Drug AI is the graph neural networks. The main research directions of this topic include, but not limited to the following:

- Design OOD learning algorithms and theory, such as algorithms for domain generalization and domain adaptation scenarios, so that Drug AI algorithms could work efficiently in the scenarios of distribution shift.
- The combination of OOD learning and deep graph learning. On one hand, one could improve the generalization ability of GNN models in the OOD scenarios, on the other hand, one could utilize the strong modeling capabilities to design OOD algorithms.

1.1.5. The applications in drug discovery

The central goal of drug discovery is to identify chemically synthesized molecules that can specifically bind to a target molecule – usually a protein or enzyme – involved in a disease. Machine learning approaches can help speed up the drug discovery process. Suggested topics include but are not limited to:

- Prediction algorithms to automate or assist in the retrosynthesis analysis.
- Computational techniques for virtual screening of molecules to bind to a drug target.
- Deep learning algorithms for scaffold hopping to discover structurally novel compounds starting from known active compounds by modifying the central core structure of the molecule.
- Machine learning methods for ADMET (absorption, distribution, metabolism, excretion, and toxicity) prediction of molecules.

1.2. The Applications in Life Science

1.2.1. Machine learning for bioinformatics

Bioinformatics involves the processing of biological data, modeling the biosystem, building algorithms and tools for biological and biomedical purposes. Suggested topics include but are not limited to:

- Next-generation deep-learning-based algorithms and tools for analysis and mining of state-of-the-art single-cell sequencing, single-cell multi-omics data and spatial transcriptomics data.
- Next-generation deep-learning-based algorithms and tools analysis and mining for immune repertoire data.

1.2.2. Machine learning for precision medicine

Precision medicine is an emerging approach to clinical research and patient care that focuses on understanding and treating disease by integrating multi-modal or multi-omics data from an individual to make patient-tailored decisions. Suggested topics include but are not limited to:

- Non-invasive early disease detection using electronic health record (EHR), medical images, and multi-omics data.
- Disease prognosis using EHR, medical images, and multi-omics data.
- Personalized treatment planning and disease management using EHR, medical test data, and wearable devices.

1.2.3. Machine learning for computer assisted intervention (CAI)

CAI is a field of research and practice, where medical interventions are supported by computer-based tools and methodologies. Suggested research topics include but not limited to:

- Development and optimization of CAI, including pre-surgical planning, image processing, intraoperative decision supports, medical robotics, and surgery navigation.
- Clinical Research for CAI, including system validation, clinical feasibility, performance evaluation, and workflow study.

1.2.4. Machine learning for neuroscience

Neuroscience studies the architecture and function of the brain and maps how each individual neuron operates. Efficiently processing large scale multi-modality neuro data needs advanced machine learning algorithms and big data platform. Suggested topics include but are not limited to:

- Development of computational models for analyzing and modeling neuroscience data.
- Building cloud-based data platform for management and sharing of neuroscience data at all levels of analysis.

1.3. Trustworthy AI

In order to apply machine learning especially deep learning to sensitive fields, such as medical care, finance, data market and bioinformatics, many researchers have begun to pay attention to the direction of trustworthy AI. The research of trustworthy AI focuses on how to develop corresponding interpretability analysis, fairness embedding and evaluation, privacy protection and robust optimization algorithms for AI models. This subject includes the following sub-directions:

- Transparency and Interpretability for Machine Learning including Interpretable Methods for Deep Neural Models, Interpretable Deep Neural Models, Evaluation Metrics for Interpretability.
- Fairness in cooperative learning scenarios including Participant valuation algorithm in cooperative learning scenarios， Valuation that satisfies social equity in the cooperative learning scenario， Data hegemony in cooperative learning scenarios.
- Robustness and reliability of deep learning including Uncertainty estimation of deep model, Distributionally robust optimization and Adversarial robustness of deep learning.

2. Deep Reinforcement Learning for Robotics

2.1. Advanced theories, algorithms and applications in simulation to reality (sim2real) problems

We are interested in developing novel theories that can reveal the gap between simulation and reality from the very fundamental perspective of reinforcement learning, benefiting design of novel algorithms that can take advantage of highly efficient simulators while interacting with the reality as infrequent as possible. Data efficient reinforcement learning approaches, such as off-policy methods, efficient exploration methods, sampling algorithms, etc., are demanded as well. We are also interested in applications in real world problems, especially on robotics control, locomotion, manipulation, etc.

2.2. Model-based reinforcement learning and its theory, algorithms and generalization in real-world tasks

We are interested in investigating novel model-based reinforcement learning theories and algorithms that can efficiently learn a dynamics model and use it for interaction, reducing the frequency of accessing the real-world environment. The dynamics model is encouraged to be a combined system with the physics processes behind the dynamics, instead of a black-box neural network. The dynamics model is also required to be able to make corrections of the applied physics processes which are possibly misspecified. Efficient implementation of dynamics modeling, architecture, and model-based RL systems in real-world applications are demanded as well, especially on robotics control, locomotion, manipulation, etc.

2.3. Highly efficient physics engine for robotics

Modern deep learning based AI technologies are increasingly successfully being used to solve problems in many fields. We also notice that these approaches normally require a lot of data. For robotics, it is expensive to acquire such data from real world. Although a lot of existing physics simulation engines, including MuJoCo, Bullet, ODE and PhysX can be used to provide the demanding data, their efficiency is often limited for many applications. Therefore, the following fields are worth studying.

- Fully utilize the potential of the existing physics engines, such as how to efficiently vectorize and parallelize the numerical calculation of physics processes.
- Develop new methods for simulating physical system with fast speed, high accuracy and easy parallelization, especially for modeling the contact and friction. Moreover, the method should be able to corporate with AI algorithm to achieve the overall high efficiency. For example, many researchers try to use modern differential programming language to model the physical dynamics to obtain the automatic differentiation and CUDA acceleration.

3. Computer Vision & Graphics

3.1. Single/Multi-modality Learning with Advanced Deep Networks

Algorithms, models, network architectures for large-scale single / multi-modality recognition, large-scale video understanding, and spatio-termpoal reasoning. We mainly focus on designing novel network architectures (e.g., Transformers, MLP) and algorithms for higher classification accuracy, more general representations, and reasoning in the spatio-temporal domains for real-world applications.

Recommend topics:

- Self-supervised visual representation learning, including images, videos, and multi-modality data (text, audio, knowledge).
- Network architecture/backbone design for image and video understanding, including image classification, object detection, semantic segmentation, and action recognition.
- Spatio-temporal applications, including video captions, video-text retrieval, video grounding, visual tracking, video object segmentation, action detection, action localization.
- Reasoning, decision making, and planning related to image and video in real-world applications.

3.2. Image and Video Generation

Theories, models, and algorithms for visual content generation. We mainly focus on controllable cross-modality image and video generation, especially for human face and body. The modalities such as text, audio, and video are used as guidance for modeling the structure and motion of objects.

Recommended topics:

- Text-to-image synthesis: Text-guided portrait generation/editing. Text-guided scene image generation/editing. Text-guided motion generation.
- Text-to-video synthesis: Given a script, generate a video or a storyboard.

3.3. Neural Implicit Representations & Neural Avatar Modeling

Theory and applications of deep generative models, neural rendering and human animation. We primarily focus on modeling view-consistent and photorealistic generations of an animatable full-body (clothed) avatars with well-trained deep neural networks, where we encourage to combine traditional computer graphics techniques with the latest deep generative models and neural implicit representations e.g. NeRF or SDF.

Recommended topics:

- Photo-realisc avatar: Modeling photorealisti, view-consistent and animatable full-body avatars of certain person and cloth, by using panoptic multi-camera system.
- Realtime rendering for High-fidelity free-view video of human performances.
- Driving avatars with hand-object interactions.
- Control of the avatar with (realtime) video-based performance capture.
- Editable garment models based on generative implicit representations, such as controllable size, color, style and geometry.
- Compositional clothed human implicit representations enabling garment retargeting.

3.4. 3D Digital Human

Research problems related to 3D digital human applications, including 3D face reconstruction and animation, facial mesh tracking, 3D human body motion synthesis and retargeting, audio/text driven 3D facial avatars, etc. The main concerns of this topic are to efficiently produce high-quality 3D digital assets that are readily to be used inside game engines or other 3D applications.

Recommended topics:

- Personalized 3D human facial expression blendshape generation for different characters.
- Text/audio driven 3D human facial expression animation generation for digital humans.
- High fidelity 3D face reconstruction and mesh tracking with camera arrays.
- 3D human body motion synthesis for digital humans.
- 3D human body motion retargeting with considerations of self-contact and self-penetration.
- High-fidelity and real-time 3D human rendering preserving real garment textures.
- Single/sparse-view 3D clothed human reconstruction in the wild.

3.5. Embodied AI via 3D object and scene understanding

We are interested in any 3D vision technologies that are related to the creation of an intelligent embodied agent (Embodied AI). In general, it consists of two directions, scene-level understanding and object-level understanding, including scene reconstruction, camera localization, robot navigation, object grasping/pulling/pushing, etc. It is also highly related to the topic of active vision, where the agent is enabled with the capability of active movement to achieve the goal in a more efficient and effective manner.

Recommended topics:

- Robot navigation: In an unknown environment, provided the goal represented as point/image/language, how the robot navigates to the goal in minimum time steps with the obstacle avoidance ability.
- Cooperative robot navigation and manipulation: Given a movable robot mounted with a robotic arm, how to efficiently move the robot to achieve more accurate object manipulation.
- Articulated object manipulation: Given the articulated objects, such as laptops, glasses, microwaves, how to detect the part pose/affordance for more effective articulated object manipulation.
- Cloth unfolding by robotic manipulation: How to design the single/dual-arm platform to unfold a cloth with smart robotic manipulation.

3.6. Towards scene acquisition, understanding and synthesis

Our specific aim is to be able to capture scenes from the physical world, analyze the constituents that exist in the acquired scenes, and create not only duplications but at scale diverse and realistic variations in virtual worlds. The subjects of interest would span over small-scale indoor scenes to extensive terrains. Such ability is of great benefit to application scenarios that require strong connections between the reality and the virtual, e.g., virtual/augmented/mixed reality, robot-enviroment interaction, etc. This goal necessitates at present the effort and exploration in the following areas:

Recommended topics:

- Scene acquisition, which seeks to develop effective representations and efficient setups for acquiring the low-level features of the scene, including the geometry, appearance, dynamics, etc.
- Scene understanding, which aims to interpret the acquired scenes using machine learning setups, the effort would cover the distribution of the geometry, appearance, dynamics, semantic instances, functionality, interactivity, affordance, relationship across constituents, etc.
- Scene synthesis, which, given the interpretation obtained above, targets creating a mirror of the reality or variants with highly plausible geometry, appearance, functionality, interactibility, affordance, and high-level relationship to offer highly immersive experience in virtual worlds.

4. Natural Language Processing

4.1. Natural Language Understanding

NLU is to process, interpret and analyze both formal and social texts with necessary techniques that can help human or downstream systems understand them.

Suggested research topics:

- Novel model frameworks for morphology, syntax and semantics.
- Exploring new tasks related to text understanding.
- Improving ultra-fine grained entity typing with less human annotation.
- Representation, construction and reasoning of knowledge graphs.
- Novel pretraining methods and models for cross-modal language understanding.
- Novel model architectures for unsupervised pre-training.
- Novel model architectures for neuro-symbolic reasoning in NLU.
- Incorporating external background knowledge in language understanding.
- Leverage multiple sources of teaching or educational materials for improving subject-area/examination-style question answering.
- Improve the robustness of machine reading comprehension (especially extractive) models in real-world settings.
- Theoretical understanding of self-supervised learning / pretraining.
- Pathways-like sparse network with self-learning/self-improving to accomplish multiple tasks.

4.2. Natural Language Generation

NLG aims at transforming all kinds of data and information into natural language for various purposes.

Suggested research topics:

- Large language models and other generation models, including novel model architectures and efficient training.

- Long text generation, such as stories and news.
- Abstractive long text summarization, including long text modeling and multi-sentence summaries generation.
- Controllable text generation, including conditioning on attributes, prompts, tables, retrieved sentences, images, or videos.
- Exploiting knowledge, semantic, and other reasoning information in generation tasks.
- Model analysis for generation models, including interpretability analysis, robustness analysis, attack and defense analysis.
- Automatic evaluation methods for open-ended generation tasks.

4.3. Dialogs

Dialog systems, including open-domain chitchat and task-oriented settings, are the key ability for enabling backend artificial intelligence system to interact with people through language to assist, enable, or entertain.

Suggested research topics:

- Leveraging knowledge in dialogue system, including dialog related knowledge base/graph construction, novel methods to incorporate knowledge into dialog systems, knowledge-grounded dialog generation.
- Multi-turn dialog systems, including retrieval and/or generation based response prediction, topic sticking and recommendation, and end-to-end task-oriented dialog systems.
- Personalized dialog models, including personalized dialog corpus construction and novel personalized dialog models.
- Multi-modal dialog systems, including flexible use of multiple input and/or output modes.
- Semantic parsing in single and multi-turn dialog including the logical form generation and deeper reasoning over them.

4.4. Machine Translation

Machine translation research focuses on improving machine translation from amateur to professional, by conducting fundamental research on machine translation, as well as bridging the gap between machine translation systems and human translators.

Suggested research topics:

- Interactive translation that bridges the gap between machine translation systems and human translators, including designing new human-machine interactive actions and evaluation on efficiency for interactive machine translation.
- Pre-Training for NMT: exploit existing pre-trained model or design specific pre-training algorithm for NMT models to improve translation performance.
- Adequacy-oriented NMT models that include various techniques such as advanced architectures and learning strategies, to alleviate the key problem of NMT – inadequate translation.

- Semantically coherent translation, including novel model architectures and training mechanisms.

5. Speech Technology

5.1. Far-field Signal Processing

In far-field microphone situations, the speech signal energy attenuation, the stationary and non-stationary noise, the reverberation, and the echo of the loudspeaker during the target sound propagation increase the difficulties of speech recognition and voice wake-up. Through the microphone array signal processing, noise reduction and speech/noise separation technologies, we could improve both speech quality and speech recognition performance.

Suggested research topics:

- Microphone array algorithm design to improve speech recognition with multiple speakers and interference sources.
- Ad-hoc, distributed microphones and array independent multi-channel speech enhancement and separation in the meeting and related scenarios.
- Joint training and optimization of front-end speech processing and back-end speech recognition acoustic models to upgrade both systems.
- Self-supervise/weakly-supervise and other learning methods for better utilizing large amount of unlabeled/weakly-labeled data and reducing domain/environment mismatch from training to testing.

5.2. Speech Recognition

Speech recognition, as one of the most natural way of human-computer interaction, plays a vital role in the AI era. Although human parity results have been reported on a clean benchmark dataset by models trained on affluent in-domain corpus, general domain ASR models still face challenges such as noise robustness, domain adaptation and long tail problems. We are interested in developing novel algorithms and models to address the above issues.

Suggested research topics:

- Novel neural network structures for low-latency streamingend-to-end ASR.
- Multi-channel multi-speaker ASR.
- Robust speech recognition against differrent accents, speaking styles, and environments.
- Multilingual speech recognition focusing on Mandarin-English code-switching.
- Speech recognition with weak-supervision and/or self-supervision training/pre-training on large amount of data.

5.3. Speech Generation

Speech generation technology, including both speech synthesis and voice conversion, is a key part of human-computer speech interaction. The user experience improves when generated voice is subjectively attractive to them. Personalized expressive speech generation technology aims to build generated voices that sounds familiar to the listeners, such as public figures, famous stars, friends and family members. However, the labeled data of the desired voices recorded in a clean environment is usually difficult to collect. Building high quality models with limited data remains a challenging task.

Suggested research topics:

- Multi-speaker multi-style, highly controllable and expressive speech synthesis.
- Multi-lingual and cross-lingual speech synthesis.
- Speech synthesis that exploits low-quality/ASR data.
- Singing voice synthesis/conversion.
- Many to one, any to any voice conversion in low-resource settings.
- Personalized voice clone with limited data.
- Multi-modal (speech/face/gesture) synthesis.
- Discourse-level TTS frontend for highly expressive speech synthesis (e.g., generation of speaker and speaking style information).
- Universal Vocoder.

5.4. Speaker Recognition and Diarization

Identifying a person by his or her voice is an important human trait usually taken for granted in natural human-to-human interaction/communication, yet it remains as a challenging task for computers. Recently, automatic speaker-recognition systems have emerged as an important means of verifying identity in many e-commerce applications as well as in general business interactions, intelligent housing system, forensics, and law enforcement. We are interested in building high-quality systems through advanced paradigms with both labeled and unlabeled data.

Speaker diarization is the process of partitioning an input audio stream into homogeneous segments according to the speaker identity which is now becoming important for many applications such as human-computer/android interaction, meeting transcription and surveillance systems, etc. We attempt to address the challenging task of tracking multiple moving speakers and identify who is speaking in both auditory-only and multi-modal settings. In some challenging scenarios when visual modality is available, proper fusion of multi-modal information is favoured in order to deal with corruption from audio data or visual data or both. While clustering of speaker embeddings from speech segments have been commenly used techniques for speaker diarization, an end-to-end neural diarization system has been drawing more and more attention recently where the whole system is jointly optimized with respect to the diarization error in an end-to-end fasion.

Suggested Research topics:

- Domain and environment robust speaker recognition.
- Speaker recognition with short utterances.
- post-processing.
- Self-supervised/weakly supervised learning for robust speaker representation.
- Joint training and optimization of single-channel/multi-channel front-end speech. processing and back-end speaker recognition models.

- Investigating spectrum, spatial, voiceprint and visual modalities fusion techniques and new paradigm for robust speaker tracking, diarization, separation and recognition.
- Research new neural network structures for End-to-end neural diarization.