

# Xingyu Li

Email: [yu1925280097@gmail.com](mailto:yu1925280097@gmail.com) Mob: 15541911139  
GitHub: [StarUniversus \(Xingyu Li\) \(github.com\)](https://github.com/StarUniversus) [HomePage](#)

## EDUCATION

**Northeastern University**

**Shenyang, China**

*Bachelor of Intelligent Medical Engineering*

*2019.09-2023.06*

*College of Medicine and Biological Information Engineering*

## PUBLICATIONS

- Global contrast-masked autoencoders are powerful pathological representation learners* Cited : 21  
Hao Quan\*, **Xingyu Li\*** (co-first author), Weixing Chen, Qun Bai, Mingchen Zou, Ruijie Yang, Tingting Zheng, Ruiqun Qi, Xinghua Gao, Xiaoyu Cui  
[Pattern Recognition \(PR\)\(2024\)](#) [\[PDF\]](#) [\[BibTeX\]](#) [\[Source Code\]](#)  
**Highlights:**
  - We have designed two self-supervised pretext tasks: masking image reconstruction and contrastive learning, which can train the encoder to have the ability to represent local-global features.
  - We discuss the mask ratio, which is suitable for pathology-specific training methodologies based on the masked image modeling paradigm.
  - We selected three pathological image datasets and proved the effectiveness of GCMAE algorithm through extensive experiments.
  - An automatic pathology image diagnosis process was designed based on the GCMAE to improve the credibility of the model in clinical applications.
- CorrectFlow: On-the-Spot Correction for Multimodal Reasoning with Multi-Agent Collaboration*  
Xiao Dong, Pan Zhou, **Xingyu Li**, Zheng Chong, Yuhao Cheng, Jianxing Yu, Jian Yin, Xiaodan Liang  
*CVPR2025 (submission)* [\[PDF\]](#)  
**Highlights:**
  - We present CorrectFlow, a novel multi-agent collaboration framework designed to overcome the limitations of a single MLLM in mitigating intrinsic errors and verification biases. By separating knowledge extraction from reasoning validation, CorrectFlow enhances robustness and minimizes internal reasoning errors.
  - Three core strategies are proposed: (1) Confidence Check 106 which stimulates the MLLM's ability to self-evaluate; (2) Path Pruning to facilitate the expansion of thought paths; and Expert Intervention for providing real-time correction for reasoning paths. These strategies work collectively to ensure reliable and accurate reasoning.
  - CorrectFlow pioneers a dynamic escalation mechanism that enables lower-level agents to transfer control to higher-level agents upon identifying potential reasoning errors, thereby surpassing traditional passive validation methods. This active intervention leads to more refined and robust reasoning outcomes.
  - We extensively evaluate CorrectFlow on public benchmark datasets, demonstrating its superior performance compared to existing baseline methods, effectively addressing both intrinsic biases and verification limitations in multimodal reasoning tasks.
- ACTIONFILLER: FILL-IN-THE-BLANK PROMPTING FOR OS AGENTS*  
Xiao Dong, Zijun Zhang, **Xingyu Li**, Yuhao Cheng, Jianxing Yu, Jian Yin, Pan Zhou, Xiaodan Liang  
*ICLR 2025 (rebuttal)* [\[PDF\]](#)  
**Highlights:**
  - We focus on the often-overlooked issue of decision efficiency and propose a novel framework termed ActionFiller to streamline the generation of action sequences. This framework reduces reliance on cognitive decision-making processes, improves the utilization of memory packages, and enhances execution efficiency for operating system agents.

- To optimize action templates, we introduce the CohesiveFlow method, which optimizes unexecutable action sequences by dynamically updating parameters and leveraging environmental contexts, thereby facilitating more effective decision-making.
- We also present the EnduroSeq benchmark, specifically designed to evaluate long-horizon instruction execution, providing comprehensive validation of our approach.
- Our experimental findings demonstrate that ActionFiller not only increases task completion rates but also improves the adaptability of agents in diverse and complex scenarios, paving the way for more responsive AI-driven solutions.

4. *Decoding Expertise from Pathologists' Diagnostic Processes on Whole Slide Images*  
*Nature Communications (revision submission)* [\[PDF\]](#) [\[Source Code\]](#)

**Highlights:**

- Novel Human Expertise Decoding: The study developed the first DL model capable of decoding pathologists' expertise through their visual behavior and applying it to aid in WSI diagnosis.
- Enhanced Classification Performance and Data Efficiency: By integrating pathologists' diagnostic processes, the model significantly improved classification performance and data collection efficiency, outperforming existing fully supervised and weakly supervised learning methods.
- Interpretability and Effectiveness in Human-like Diagnosis: The PEAN model can mimic the diagnostic process of pathologists, conducting step-by-step visual searches. The results validated the effectiveness and interpretability of its "human-like" diagnosis, enhancing the accuracy of the classification model.

5. *Discriminating Chromophobe RCC from Oncocytoma: A Transformer-based Approach Leveraging the Subtleties of Nuclear Structures within Kidney Tumors*

Jing Yang, **Xingyu Li**, Hongjiu Ren, Yanmei Zhu, Qimin Wang, Ruiqun Qi, Xiaoyu Cui and Huamao Jiang

[\[PDF\]](#)

**Highlights:**

- We conducted experiments on pathological slides at different magnifications. These experiments deeply analyzed the impact of pathological image features at different magnifications on the model, revealing the importance of high-magnification image features for diagnostic classification.
- We conducted cross-center experiments on high-magnification datasets from multiple centers. This allowed us to validate the generalization ability and robustness of deep learning models on pathological images collected under different geographical locations and equipment conditions.
- In diagnosing kidney tumors, particularly renal cell carcinoma and oncocytoma, these two lesions are difficult to distinguish due to their histological similarity. Doctors rely on observing subtle differences in nuclear structures to make a diagnosis. By integrating these key nuclear features, we significantly improved the performance of our diagnostic model. Specifically, the model achieved an average ACC of 92.50%, an AUC of 96.42, a Precision of 90.32, and a Recall of 100%.

6. *Attention-based Feature Distillation Multi-Instance Learning for Pathological Image Classification*

Tianh Nan, **Xy Li**, H Quan, K Yang, Xy Cui

*IEEE Journal of Biomedical and Health Informatics (under review) (JBHI) (2024)* [\[PDF\]](#) [\[Source Code\]](#)

**Highlights:**

- The integration of attention-based feature distillation with MIL for WSI classification, for the first time, addresses the challenge of excessive redundant features and achieves optimal performance in cancer classification tasks (Camelyon16 and NSCLC).
- AFD-MIL achieves high interpretability as it automatically selects regions similar to ground truth. Additionally, its compatibility with various existing models enhances performance, demonstrating high scalability.

7. *MedConvMamba: Enhancing Medical Image Classification by Integrating Convolutional Neural Networks with Mamba for Local Feature Extraction and Global Context Awareness*

Hui Qu, Jiayi Liu, Mingchen Zou, **Xingyu Li**, Canwei Dong, Xiaoyu Cui

*IEEE Journal of Biomedical and Health Informatics (under review) (JBHI) (2024)* [\[PDF\]](#)

**Highlights:**

- We propose MedConvMamba and explore the potential applications of CNN combined with SSM models in different modalities of medical images.
- A comprehensive experiment was conducted on two different modalities of datasets, and the results showed that MedConvMamba has considerable competitiveness while ensuring lightweight.
- This study has established a new baseline for medical image classification tasks, providing valuable insights for the development of more efficient and effective artificial intelligence algorithms and medical application systems based on CNN combined with SSM in the future.

## Project Experience

---

### Pathology Slide Review Software Platform "EasyPathology" Based on Gaze Analysis

Advised by Dr. Xiaoyu Cui.

Sep 2019 – Sep 2024

- Gaze Information Analysis and Eye Movement Metric Calculation
- Deployment of the Segment Anything Model Integrated with Gaze Information
- In collaboration with 19 pathologists and 13 hospitals

### Large Agent Model

Advised by Dr. Xiaodang Liang.

March 2024 – Now

- CorrectFlow: On-the-Spot Correction for Multimodal Reasoning with Multi-Agent Collaboration
- ACTIONFILLER: FILL-IN-THE-BLANK PROMPTING FOR OS AGENTS

## WORK EXPERIENCES

---

### Microsoft Opensource Study Community (MOSC), China

Aug 2021 - Sep 2021

- Research Sharing: Attention
  - ✧ Attention is all you need
  - ✧ An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale

### Microport Ltd, Shanghai, China

June 2022 - Sep 2022

- Research and Development Intern
  - ✧ Employing YOLOv7 for Object Detection of Lung Nodules in Chest CT Scans

### Tanka Ltd, Shenyang, China

Sep 2023 - Now

- Deploying Auto Speech Recognition(ASR) on Software
  - ✧ Alidamo: FunASR
  - ✧ Python, C#, Onnx, Pytorch, Github
- Gaze Information Analysis and Eye Movement Metric Calculation
  - ✧ C#, Github
- Deploying SAM Annotation Algorithm Service with FastAPI
  - ✧ Python, Fastapi, Linux, Openapi, Pydantic, Docker, Cloud, React, JS, HTML
- Multi-Instance Learning Approach for ChRCC and Oncocytoma Differentiation
  - ✧ Python, Pytorch, Mlflow, DVC, Huawei NPU, Cloud, Docker, HuggingFace, Gradio
- Multimodal Large Language Model for Skin Pathology
  - ✧ Data Collection: MySQL, Data Cleaning: ASR, GPT-4.0
  - ✧ Model Pre-training and Fine-tuning: PyTorch, Huggingface
  - ✧ Model Deployment: FastAPI, vLLM

## Competitions/Awards

---

- National Student Biomedical Engineering Innovation Design Competition: National Second Prize
- ISICDM 2023 Chest X-ray Posture Assessment Competition: The Second Prize
- Software Copyright: EasyPathology — Pathology slide analysis software
- Northeastern University Technology Innovation Individual