Jianxiang Dong

Stony Brook, NY dongjianxiang1995@gmail.com (202)-280-0072 homepage

EDUCATION

Stony Brook University, Ph.D. in Computer Science — Aug 2020 – Present

George Washington University, M.S. in Computer Science — Aug 2018 – May 2020

Area of Focus

Multimodal deep learning for vision–language tasks, focusing on language-based video retrieval and temporal grounding; large-scale vision–language models (e.g., multimodal LLMs); Biomedical Image Analysis.

EXPERIENCE

Research Project Assistant, Stony Brook University — Jan 2022 – Current

- **Grounded Video Question Answering:** Construct a Grounded VQA dataset by developing a semi-automated labeling pipeline using Qwen2.5-VL. Fine-tuning the Multimodel LLM to localize the relevant video segments and produce answers to a given question. (Skills: Multimodal LLM such as Qwen2.5-VL, Prompt Engineering, Instruction Tuning)
- **Text Temporal Grounding in Videos:** Propose novel deep learning models to localize the start/end timestamps of temporal moments in untrimmed videos that are relevant to sentences. Achieve gains of more than $\sim 5\%$ recall rate. (Skills: Vision Transformer, CLIP, Graph Neural Networks, Contrastive Learning)
- **Annotation-efficient Training:** Develop an annotation-efficient training framework for the Text Temporal Localization in Videos, which uses only 30% full labels to achieve comparable performance with fully-supervised models using 100% full labels. (Skills: Weakly/Semi-Supervised Learning, Active Learning, Curriculum Learning)
- **Text-to-video Retrieval:** Propose novel deep learning models to perform video retrieval using queries that match only part of video segments. Results outperform state-of-the-art methods by $\sim 4\%$ recall rate. (Skills: Vision Language Model, Multiple Instance Learning, Representation Learning)

Collaborative Researcher, Brookhaven National Lab — Jan 2025 – Current || Jan 2023 – Apr 2024

- **Efficient Spot Detection:** Develop a fast, memory-efficient U-Net–based deep learning model to segment diffraction spots in X-ray images, achieving robust detection under severe noise and ice-ring artifacts. (Skills: U-Net, Biomedical Image Analysis, Bioinformatics)
- **Lossy Data Compression:** Designing an ML-based data compression and reconstruction system for macromolecular crystallography data, aiming to efficiently compress large-scale data with minimal information loss and enable accurate reconstruction when needed. (Skills: Autoencoder, Data Compression, Computational Biology)

Research Internship, Xeo Air — Jul 2022 – Dec 2022

- **Bridge Inspection AI:** Design deep learning pipelines for multi-class segmentation and detection of bridge defects. Collect the bridge inspection dataset to evaluate the performance of the models. (Skills: Object Detection, Semantic Segmentation)
- **Interactive AI Training Framework:** Develop an expert-interactive training framework for the AI, where the AI highlights uncertain detections for expert validation and incorporate feedback to iteratively improve model accuracy. (Skills: Human-in-the-loop, Interactive Training)

Research Project Assistant, George Washington University — Feb 2020 – Jun 2020

- **Blood Oxygen Monitoring using Hyperspectral Images:** Construct a hyperspectral image dataset for blood oxygen estimation. Propose novel deep learning models for blood oxygen monitoring using hyperspectral images. (Skills: Hyperspectral Imaging)

TECHNICAL SKILLS

Python, JAVA, C++, MATLAB, Pytorch, TensorFlow, Git, SQL, ROS, Linux

PUBLICATIONS

- **RoCon: Relieving Over-concentration Problems for Partially Relevant Video Retrieval**
  Jianxiang Dong, Zhaozheng Yin
  Submitted to AAAI 2026
- **Multi-scale Token-aware Contrastive Learning with Hard Negatives Mining for Long Video TSG**
  Jianxiang Dong, Zhaozheng Yin
  Under review, IEEE Transactions on Multimedia (TMM)
- **Annotation-efficient Hybrid Learning for Temporal Sentence Grounding**
  Jianxiang Dong, Zhaozheng Yin
  IEEE Transactions on Circuits and Systems for Video Technology (TCSVT), 2025
- **Weakly Semi-supervised Temporal Sentence Grounding in Videos with Point Annotations**
  Jianxiang Dong, Zhaozheng Yin
  IEEE Transactions on Multimedia (TMM), 2025
- **Bragg Spot Finder (BSF): Machine-learning-aided Diffraction Spot Filtering**
  Jianxiang Dong, Zhaozheng Yin, Dale Kreitler, Herbert J. Bernstein and Jean Jakoncic.
  Journal of Applied Crystallography (JAC), 2024
- **Graph-based Dense Event Grounding with Relative Positional Encoding**
  Jianxiang Dong, Zhaozheng Yin
  Computer Vision and Image Understanding (CVIU), 2024
- **Boundary-aware Temporal Sentence Grounding with Adaptive Proposal Refinement**
  Jianxiang Dong, Zhaozheng Yin
  Asian Conference on Computer Vision (ACCV), 2022

ACADEMIC SERVICE

- **Teaching Assistant**: Intro. to Java Programming, Intro. to Computer Network
- **Reviewer**: NeurIPS, AAAI
- **Mentor**: CSIRE program at Stony Brook University (deep learning project)