

# CHENG, WENZHENG

+1 4435702555 | MD, USA | Email: [wcheng31@jh.edu](mailto:wcheng31@jh.edu) | Github: <https://wilsoncheng561.github.io/>

Research Interests: Surgical Computer Vision | Multimodal Learning | LLMS | Digital Twins | Robot learning

## Education

### **Johns Hopkins University (JHU)**

M.S. in Robotics | GPA: 3.7/4.0

MD, USA

Sep 2024-Present

### **Tianjin University (TJU)**

B.S. in Software Engineering | GPA: 84.8/100

Tianjin, China

Sep 2020-Jun 2024

## Publications

- [H Ding, Y Zhang, W Cheng. Towards Robust Algorithms for Surgical Phase Recognition via Digital Twin Representation.](#) arXiv preprint arXiv:2410.20026, 2024.
- [Zhang, H. & W Cheng. \(2023\). DeepHEN: quantitative prediction essential lncRNA genes and rethinking essentialities of lncRNA genes.](#) arXiv preprint 2309.10008.
- [W Cheng. \(2023\). Application and analysis of residual blocks in galaxy classification.](#) ACE Vol. 21: 143-152. DOI: 10.54254/2755-2721/21/20231135.

## Research Experience

### **Digital Twin Representation-based Surgical Phase Recognition (DT-SPR)**

MD, USA

*Graduate Researcher, Johns Hopkins University (LCSR), advised by Mathias Unberath*

12/2024-03/2025

- Co-author on DT-SPR, accepted oral at **DT4H@MICCAI 2025**
- Released a corrupted-data benchmark & Enhanced data augmentation and training strategies
- Improved top-1 accuracy by +13 pp and maintained OOD robustness via Multimodal fusion redesign : incorporate context + mask/depth feature to form two token streams; cross-modal fusion between streams

### **Multimodal and Foundation Model-based Surgical Video Understanding(extensions of DT-SPR)**

#### **MD, USA**

*Graduate Researcher, Johns Hopkins University (LCSR) , advised by Mathias Unberath*

03/2025-Present

- Curated an OOD benchmark for SPR by collecting and annotating 80 YouTube surgical videos, enabling rigorous robustness validation
- LLM-based surgical triplets extraction (GPT-4 prompting with CoT/memory/RAG + BERT embedding)
- Designed a zero-shot surgical tool segmentation pipeline: DINOv2 attention × Depth-Anything v2 → region proposals→ multi-strategy key-point picking. Using SAM2 generates high-quality masks proposals without labels, outperforming few-shot NO-TIME-TO-TRAIN on IoU@0.5:0.95 recall by + 9.2 pp
- Use SAM2 as a frozen feature backbone + ViT head and then a distillation-stabilized partially-unfrozen model, achieving strong multi-task instrument recognition (0.5679 mAP) while preserving segmentation quality
- Build a surgical action world model. Adapted the V-JEPA2 ViT-g video foundation model with a 7D tool-tissue state/action encoding to perform video-level surgical action prediction, reaching over 93% validation accuracy and demonstrating strong data-efficient transfer.

### **SurgiPose: Monocular Pose Estimation for Imitation Learning**

MD, USA

*Graduate Researcher, Johns Hopkins University (LCSR), advised by Axel Krieger*

05/2025-Present

- Computed the camera intrinsic matrix from consecutive, differently-angled surgical tool frames
- Improved the initial pose-estimation loss function, achieving ~60% reduction in optimization loss
- Curated additional OOD datasets (YouTube, SurgVu 2024) and generated masks with SAM2 for training/evaluation

### **Fast Diffusion-GAN based model for Medical Image Augmentation**

Tianjin, China

*Cooperated with Tianchi Lu, Collaboration between TJU and HKU*

11/2023-8/2024

- Developed a novel image generation model by combining diffusion-based generator with a GAN discriminator,

- improving image quality and stability compared to state-of-the-art models such as IDDPM, DDGAN, and Score-SDE
  - Integrated an autoencoder for dimensionality reduction of input images, accelerated the diffusion process by replacing the 1000-step Markov chain with 4-step fast inference method, significantly improving training efficiency and generation speed
  - Curated two datasets(CT and MRI). Evaluating the augmented data by downstream task classification using ResNet
- Cross-modal MRI Translation via Adversarial Diffusion (Undergraduate Thesis)** Tianjin, China  
*Undergraduate Researcher, advised by Xuebin Sun, TJU* 06/2023-04/2024
- Implemented Medical Diffusion GAN (MDG) for unpaired, bidirectional MRI synthesis ( $T2 \leftrightarrow PD$ ;  $FLAIR \leftrightarrow T2$ ) using fast conditional diffusion with an adversarial projector and cycle consistency between diffusion and non-diffusion branches
  - Introduced an adversarial projector and cross-branch cycle-consistent coupling to stabilize unpaired training across diffusion and non-diffusion modules
  - Benchmarked on IXI and BRATS: MDG outperformed DDPM & Cycle-GAN (evaluated with PSNR/SSIM/FID)

- DeepHEN: Predicting Essential lncRNA Genes and Rethinking Their Essentialities** Tianjin, China  
*2-person Project, advised by Pufeng Du, TJU* 06/2023-11/2023
- Quantitatively estimated long non-coding RNA (lncRNA) gene essentialities
  - Constructed a lncRNA-protein-protein reaction network and utilized variational graph auto-encoders (VGAE) and graph information passing to obtain feature vectors for each lncRNA gene
  - Employed the dna2vec model to generate sequence feature vectors
  - Proposed a method based on semi-supervised clustering to select negative samples, addressing the scarcity of known essential human lncRNAs and the absence of defined non-essential lncRNAs
  - Validated the effectiveness of negative sample selection strategy through genome-level enrichment analysis

## Selected Course Projects

- Development of a Java-web Project “Ele.me” Food Delivery App** Tianjin, China  
*Leader of 4 -person Project, Practice for Programming Courses* 08/2022-04/2023
- Implemented CRUD operations and managed primary and foreign key relationships using JDBC encapsulation
  - Developed enterprise-level front-end static pages with HTML5, CSS3, and JavaScript
  - Utilized front-end and back-end separation architecture with Vue and Servlet, incorporating AJAX, Session management, and MVC encapsulation
  - Implemented separation architecture application based on SpringBoot, integrating SpringMVC, Spring, and MyBatis
  - Developed Internet application using Vue and SpringCloud based on microservices architecture, involving service registration and discovery, load balancing, and service degradation, etc

## Professional Experience

- AutoBrain** Tianjin, China  
*Algorithm Research Intern* 04/2023-06/2023
- Conducted an extensive literature review to research point cloud segmentation and semantic occupancy perception models and their applications
  - Implemented point cloud processing methods (filtering, clustering, merging) using PCL and Open3D
  - Utilized LabelCloud to annotate point cloud segmentation dataset and generated the annotation results in KITTI format
  - Proposed a method to calibrate the extrinsic parameters between a camera and a LiDAR, and used Autoware tools to implement this calibration

## Skills

- Programming: •Python •Java •C •Matlab •SQL •HTML/CSS •JavaScript •HDL •Linux •LaTeX
- Core Courses: Computer Integrated Surgery, Computer Vision, Algorithms for Sensor-Based Robotics, Machine Learning: Deep Learning, Advanced Practice for Programming, Data warehouse and data mining, Artificial Intelligence