

# Yu-Cheng Chou

+1 (667) 406 6717 | johnson111788@gmail.com | [Google Scholar](#) | [Website](#) | [LinkedIn](#) | No Sponsorship Required

## SUMMARY

- CS Ph.D. at JHU, advised by Alan Yuille; Research Intern at NVIDIA; expected graduation May 2027.
- Research spans visual planning for robot action, 3D reasoning in VLMs, and memory-conditioned world modeling, with experience in VLM/VLA/WAM architecture design, SFT/RL post-training, and distributed training infrastructure.
- Seeking Research Scientist and Machine Learning Engineer roles in embodied AI and robotics, focused on translating vision-language reasoning into effective and efficient robot control.

## EDUCATION

|                                                                             |                     |
|-----------------------------------------------------------------------------|---------------------|
| <b>Johns Hopkins University, School of Engineering</b>                      | Baltimore, MD       |
| Ph.D. in Computer Science, Advisor: Prof. Alan Yuille                       | Jan 2023 – May 2027 |
| M.S.E. in Computer Science, Advisor: Prof. Alan Yuille & Prof. Zongwei Zhou | Jan 2023 – May 2024 |
| <b>Wuhan University, School of Computer Science</b>                         | Wuhan, China        |
| B.S. in Software Engineering, Advisor: Prof. Deng-Ping Fan                  | Aug 2018 – May 2022 |

## EXPERIENCE

|                                                                               |                                     |
|-------------------------------------------------------------------------------|-------------------------------------|
| <b>NVIDIA, Learning and Perception Research Group.</b> <i>Research Intern</i> | Jan 2026 – Present. Santa Clara, CA |
|-------------------------------------------------------------------------------|-------------------------------------|

### VFlash: Fast, Accurate System-1 Visual Planning for Physical AI

- Led the design and training of Flash Primitive Decoding, exploiting the known structure of visual primitives to predict them in parallel without reproducing the autoregressive target's exact tokens. Built a generalist drafter supporting bounding boxes, variable-length trajectories, and dense 3D motion while retaining verified speculative decoding for text.
- Designed and trained VLA/WAM architectures that condition action experts on visual plans, improving manipulation success from 62.6% to 70.2% on MolmoSpaces and from 40.2% to 47.0% on LIBERO-PRO.
- Enabled a lightweight drafter to outperform its autoregressive target VLM on 2D/3D robot and driving trajectory tasks; accelerated per-step robot inference by 5.2× and driving inference by 3.5× on Jetson Thor and DGX Spark.

### Eagle-Embodied: Seeing in 2D, Acting in 3D

- Led Sight-to-Action (S2A) development and post-training, establishing decoded 2D visual plans as a shared intermediate representation before lifting them into embodiment-specific 3D trajectories for manipulation and driving.
- Integrated decoded plan with manipulation and driving heads; compared with prefill-only conditioning using the same frozen VLM, explicit planning improved LIBERO+ success from 68.4% to 70.7% and SimplerEnv from 52.6% to 56.3%.

### NemoVLA: Scalable VLA Training Infrastructure for Robot Policies

- Independently integrated NeMotron training infrastructure with starVLA for 30B-class MoE VLA training using FSDP2, expert and context parallelism. Decoupled VLM backbones from flow-matching action heads through an autograd-aware bridge, preserving full-sequence conditioning and gradient flow under context parallelism.
- Designed an extensible system isolating VLA component in an overlay, simplifying maintenance and upstream upgrades.

|                                                                          |                                    |
|--------------------------------------------------------------------------|------------------------------------|
| <b>Johns Hopkins University, CCVL Research Group.</b> <i>PhD Student</i> | Jan. 2023 – Present. Baltimore, MD |
|--------------------------------------------------------------------------|------------------------------------|

### Captain Safari: 3D-Consistent World Generation with Pose-Guided Memory

**CVPR 2026**

- Developed a memory-guided video world model that conditions on previously observed scene geometry, preserving 3D consistency while following complex camera trajectories under strong parallax and aggressive 6-DoF motion.
- Designed a pose-aware retriever that compresses multi-view geometric memory into fixed-size world tokens, decoupling retrieval from iterative diffusion so denoising conditioning cost remains fixed as memory grows.
- Created OpenSafari to benchmark generation under strong parallax and rapid camera motion with verified trajectories; achieved leading 3D consistency and trajectory following, with 67.3% of 1,500 votes in a five-way human study.

### Spatial Reasoner: Explicit and Generalizable 3D Reasoning for VLMs

**NeurIPS 2025**

- Introduced explicit 3D locations and orientations as intermediate reasoning states, enabling VLMs to decompose spatial questions into perception, geometric computation, and reasoning within a shared calibrated coordinate system.
- Developed an SFT-to-GRPO pipeline that establishes explicit 3D reasoning through SFT and improves generalization through RL. Achieved 60.3% on 3DSRBench with a 7B model, outperforming Gemini 2.0 by 9.2%.
- Used intermediate 3D predictions to disentangle perception errors from reasoning failures, identifying 3D perception as the primary bottleneck and showing that final-answer rewards can improve QA accuracy but reduce 3D accuracy.

## Johns Hopkins University - Continued

### Selective Memory: Learning What to Retain from Streaming Data

MICCAI 2024

- Developed a streaming continual-learning method that combines feature-based deduplication with uncertainty-aware retention, preserving diverse and challenging examples within a bounded replay buffer to mitigate catastrophic forgetting.
- Achieved 47.2% average Dice across 16 sequential, partially labeled segmentation datasets within a 128-sample buffer, outperforming FIFO replay (32.7%) and a 100-epoch offline baseline (45.1%).

### Philips Research. *Research Intern*

Jun. 2024 - Aug. 2024. Boston, MA

- Developed diffusion-based data augmentation with spatial priors and structural conditioning to address class imbalance, improving recall by 22% on a minority subgroup (1.5% of data) at a matched false-positive rate.

### Microsoft Research Asia. *Research Intern*

Jan. 2022 - Jun. 2022. Beijing, China

- Developed a causal representation-learning framework for segmentation under domain shift, using backdoor adjustment to mitigate hidden confounding and reliance on spurious correlations; yielding +10% Dice over SOTA baselines.

### Tencent Wesure. *Intern of Computer Vision*

Oct. 2020 - Aug. 2021. Shenzhen, China

- Built a production OCR pipeline with geometric correction, layout clustering, and dual-model pseudo-labeling, improving tabular extraction accuracy by 23% and reducing labeling turnaround from three months to five days.

### KuaiShou. *Intern of Video Transmission Algorithm*

Aug. 2020 - Jan. 2021. Remote

- Developed multi-objective RL for application-aware congestion control, balancing throughput, latency, and packet loss. Increased throughput by 14% on stable Wi-Fi and reduced latency by 44% on cellular networks.

## SELECTED PUBLICATIONS

Asterisks (\*) denotes co-first authorship

1. **Captain Safari: A World Engine with Pose-Aligned 3D Memory** CVPR 2026  
Yu-Cheng Chou, Xingrui Wang, Yitong Li, Jiahao Wang, Cihang Xie, Alan L Yuille, Junfei Xiao
2. **SpatialReasoner: Towards Explicit and Generalizable 3D Spatial Reasoning** NeurIPS 2025  
Wufei Ma\*, Yu-Cheng Chou\*, Qihao Liu\*, Xingrui Wang, Celso de Melo, Jieneng Chen, Jianwen Xie, Alan Yuille
3. **Ultrasound Image Synthesis Using Generative AI for Lung Consolidation Detection** ISBI 2025  
Yu-Cheng Chou, Gary Y Li, Li Chen, Mohsen Zahir, Naveen Balaraju, <12 more authors>, Alvin Chen
4. **Embracing Massive Medical Data** MICCAI 2024  
Yu-Cheng Chou, Zongwei Zhou, Alan Yuille
5. **Video Polyp Segmentation: A Deep Learning Perspective** Machine Intelligence Research 2022  
Ge-Peng Ji\*, Guobao Xiao\*, Yu-Cheng Chou\*, Deng-Ping Fan, Kai Zhao, Geng Chen, Luc Van Gool
6. **Progressively Normalized Self-Attention Network for Video Polyp Segmentation** MICCAI 2021  
Ge-Peng Ji\*, Yu-Cheng Chou\*, Deng-Ping Fan, Geng Chen, Huazhu Fu, Debesh Jha, Ling Shao

## ACADEMIC SERVICE

- **Program Chair:** Embodied Spatial Reasoning (NeurIPS 2026, ICCV 2025), Body Maps (MICCAI 2024, ISBI 2024).
- **Peer Review:** CVPR, ECCV, ICLR, NeurIPS, ICML, MICCAI.

## TECHNICAL SKILLS

- **Multimodal learning:** VLM SFT, GRPO, RLVR, LoRA, 3D reasoning, video generation, world modeling.
- **Robot learning:** Visual planning, VLA/WAM architectures, flow-matching policies, plan-conditioned control.
- **Training systems:** PyTorch, Transformers, NeMo-Automodel, Megatron-LM, FSDP/FSDP2, expert and context parallelism, data packing, FlashAttention, verl, TRL.