Dahun Kim
Research Scientist, Google DeepMind

Google Gradient Canopy, Mountain View, CA 94043
mcahny@google.com | https://mcahny.github.io

Summary

- **Expertise:**   Vision-and-Language, Computer Vision, Machine Learning.

- **Research experience:**   Prior to Google DeepMind, I was a Research Scientist at Google Brain. I also worked at Adobe Research, Google Research and KAIST.

- **Research highlights:**   My recent research focuses on object-aware video understanding capability of the Large Multimodal Models (e.g., Gemini). The work on "Foundational vision-language model" is launched at Google Vertex AI model garden. The projects "Open-vocabulary detection", "Unified multimodal multitask model" have been featured by Google AI BlogPosts.

- **Oral/Spotlight publications:**   I have published 7 Oral/Spotlight papers in top-tier venues: CVPR, NeurIPS, ECCV, AAAI. They are highly prestigious with a selection ratio of top 2.5 - 6.5%.

- **Awards:**   Microsoft Research Fellowship, Global Ph.D. Fellowship (by the Ministry of Science and ICT), KAIST EE Best Thesis Award, 1st-place Award at ECCV 2018 Challenge Track, Two Bronze prizes from Samsung HumanTech Paper Award, Qualcomm Innovation Award, Outstanding Reviewer Awards.

Research Interest

My research interests lie in Vision-and-Language, Computer Vision, Machine Learning and applications.

- Understanding the interaction of vision and language, and enabling new capabilities with multimodal models.

- Learning algorithms for efficient and effective video understanding.

- Scene understanding, Video Editing, 3D avatar rendering.

Research Experience

- **Google DeepMind**, MTV, CA   Apr.2023 - Present
  Research Scientist, Vision-and-Language team

- **Google Brain**, MTV, CA   Jul.2022 - Apr.2023
  Research Scientist

- **Google Research**, MTV, CA   Jul.2022 - Apr.2023
  Research Intern, worked on "video mask transformer"

- **Google Brain**, MTV, CA   Jun.2020 - Nov.2020
  Research Intern, worked on "open-world detection - detect everything"

- **Adobe Research**, San Jose, CA   Jun.2019 - Sep.2019
  Research Intern, worked on "video panoptic segmentation"

Education

- **Ph.D.** in Dept. of **Electrical Eng., KAIST**   Mar.2018 - Feb.2022
Advised by Prof. In So Kweon
Thesis: "Learning Dense Pixel Features for Video Processing and Understanding"
Best Thesis Award from KAIST EE.

- **M.S.** in Dept. of **Electrical Eng., KAIST**   Mar.2016 - Feb.2018
Advised by Prof. In So Kweon
Thesis: "Reducing Human Supervision in Supervised Learning"

- **B.S.** in Dept. of **Electrical Eng., KAIST**   Mar.2012 - Feb.2016

- **High school: Korea Science Academy of KAIST**   Mar.2009 - Feb.2012

Academic Service

- **Area Chair** in CVPR 2024

- **Area Chair** in NeurIPS 2023

- **Area Chair** in CVPR 2023

- Journal Reviewer in TPAMI, TNNLS, TIP, EuroGraphics

- Conference Reviewer in CVPR [20, 21, 22], NeurIPS [20, 21], ECCV [20], ICCV [19, 21, 23], ICLR [21, 24], AAAI [20, 21, 22, 24],

Publications

- **Preprints - under reivew:**

033. **Dahun Kim**, A. Angelova, W. Kuo
"Detection-oriented Image-Text Pretraining for Open-Vocabulary Detection"
Launched at Google Vertex AI model garden

032. A. Piergiovanni, I. Nobel, **Dahun Kim**, M. Ryoo, V. Gomes, A. Angelova
"Mirasol3B: A Multimodal Autoregressive model for time-aligned and contextual modalities"
Featured in Google AI BlogPost

031. M. Kim, J. Choi, **Dahun Kim**, Y. M. Ro   "Many-to-Many Spoken Language Translation via Unified Speech and Text Representation Learning with Unit-to-Unit Translation"

- **Peer-Reviewed Conferences and Journals - Selected:**

030. **Dahun Kim**, A. Angelova, W. Kuo
"Contrastive Feature Masking Open-vocabulary Vision Transformer",
in **ICCV 2023**, Paris, France

029. **Dahun Kim**, A. Angelova, W. Kuo
"Region-Aware Pretraining for Open-Vocabulary Object Detection with Vision Transformers",
in **CVPR 2023 (Highlight - accept rate: 2.5%)**, Vancouver, Canada
Featured in Google AI BlogPost

028. W. Kuo†, A. Piergiovanni†, **Dahun Kim***, X. Luo*, B. Caine, W. Li, A. Ogale, L. Zhou, A. Dai, Z. Chen, C. Cui, A. Angelova
"MaMMUT: A Simple Vision-Encoder Text-Decoder Architecture for Multimodal Tasks",
in **TMLR 2023**: Transactions on Machine Learning Research
Featured in Google AI BlogPost

027. R. Li, **Dahun Kim**, W. Kuo
"RECLIP: Resource-efficient CLIP by Training with Small Images",
in **TMLR 2023**: Transactions on Machine Learning Research

026. Shin, **Dahun Kim**, Q. Yu, J. Xie, H.S. Kim, B. Green, I.S. Kweon, K.J. Yoon, L.C. Chen
"Video-kMaX: A Simple Unified Approach for Online and Near-Online Video Panoptic Segmentation",
in **WACV 2024 (Oral)** and **CVPRW 2023**: 'Transformers for Vision' Workshop

025. Y. Kwon, **Dahun Kim**, D. Ceylan, H. Fuchs
"Neural Image-based Avatars: Generalizable Radiance Fields for Human Avatar Modeling",
in **ICLR 2023**, Kigali, Rwanda

024. **Dahun Kim**, S. Woo, J.Y. Lee, I.S. Kweon
"Dense Pixel-level Interpretation of Dynamic Scenes with Video Panoptic Segmentation",
in **TIP 2022**: *IEEE Trans. on Image Processing*, IF=10.6

023. **Dahun Kim**, J. Xie, H. Wang, S. Qiao, H.S. Kim, H. Adam, I.S. Kweon, L.C. Chen
"TubeFormer-DeepLab: video mask transformer",
in **CVPR 2022**, New Orleans, USA
Ranked #1 on SemKITTI-DVPS, #3 on KITTI-STEP benchmark

022. Q. Yu, H. Wang, **Dahun Kim**, S. Qiao, M. Collins, Y. Zhu, H. Adam, A. Yuille, L.C. Chen
"CMT-DeepLab: dynamic clustering mask transformers for panoptic segmentation",
in **CVPR 2022 (Oral - accept rate: 4%)**, New Orleans, USA

021. **Dahun Kim**, T.Y. Lin, A. Angelova, I. S. Kweon, W. Kuo
"Learning open-world object proposals without learning to classify",
in **RA-L** and **ICRA 2022 (Oral)**; *IEEE Robotics and Automation Letters* , Philadelphia, USA
Invited paper talk at Open-World Segmentation (UVO) Workshop @ ICCV 2021
Received Qualcomm Innovation Award 2021

020. Y. Kwon, S. Petrangeli, **Dahun Kim**, H. Wang, V. Swaminathan, H. Fuchs
"Tailor Me: An Editing Network for Fashion Attribute Shape Manipulation".
in **WACV 2022 (Oral)**

019. Y. Kwon, **Dahun Kim**, D. Ceylan, H. Fuchs
"Neural Human Performer: learning generalizable radiance fields for human performance rendering",
in **NeurIPS 2021 (Spotlight - accept rate < 3.0%)**, Virtual
Received Bronze Prize, 28th Samsung HumanTech Paper Award
018. S. Woo, **Dahun Kim**, J.Y. Lee, I. S. Kweon,
"Learning to associate every segment for video panoptic segmentation".
in **CVPR 2021**, Virtual

017. S. Woo, **Dahun Kim**, J.Y. Lee, I.S. Kweon
"Global Context and Geometric Priors for Effective Non-Local Self-Attention".
in **BMVC 2021**
Received Bronze Prize, 27th Samsung HumanTech Paper Award

016. M. Kim, S. Woo, **Dahun Kim**, I. S. Kweon,
"The Devil is in the Boundary: Exploiting Boundary Representation for Basis-based Instance Segmentation".
in **WACV 2021 (Oral)**

015. Y. Kwon, S. Petrangeli, **Dahun Kim**, H. Wang, V. Swaminathan, H. Fuchs,
"Rotationally-Temporally Consistent Novel View Synthesis for Human Performance Video",
in **ECCV 2020 (Spotlight - accept rate: 5.3%)**, Virtual

014. **Dahun Kim**, S. Woo, J.Y. Lee, I.S. Kweon,
"Video panoptic segmentation",
in **CVPR 2020 (Oral - accept rate: 5.0%)**, Virtual

013. **Dahun Kim***, S. Woo*, J.Y. Lee, I.S. Kweon,
"Recurrent temporal aggregation framework for deep video inpainting",
in **TPAMI 2020**: *IEEE Trans. on Pattern Analysis and Machine Intelligence*, IF=24.314
Received KAIST-Samsung Industry-University Cooperation Best Paper Award

012. Y. Jung, **Dahun Kim**, S. Woo, K. Kim, S. Kim, I.S. Kweon,
"Hide-and-Tell: Learning to bridge photo streams for visual storytelling",
in **AAAI 2020**, New York, USA (Acceptance: $1591/7737 \approx 20.6\%$)

011. Y. Kwon, S. Petrangeli, **Dahun Kim**, H. Wang, H. Fuchs, V. Swaminathan,
"Rotationally-Consistent Novel View Synthesis for Humans",
in ACM **MM 2020**, Virtual (Acceptance: $472/1698 \approx 27.8\%$)

010. S. Woo, **Dahun Kim**, K. Park, J.Y. Lee, I.S. Kweon,
"Align-and-Attend Network for Globally and Locally Coherent Video Inpainting",
in **BMVC 2020** (Acceptance: $195/670 \approx 29.1\%$)

009. **Dahun Kim***, S. Woo*, J.Y. Lee, I.S. Kweon,
"Deep video inpainting",
in **CVPR 2019**, Long Beach, USA (Acceptance: $1294/5160 \approx 25.2\%$)

008. **Dahun Kim***, S. Woo*, J.Y. Lee, I.S. Kweon,
"Deep blind video decaptioning by temporal aggregation and recurrence",
in **CVPR 2019**, Long Beach, USA (Acceptance: $1294/5160 \approx 25.2\%$)
1st place winner of ECCV 2018 Chalearn LAP Video De-Captioning Challenge

007. **Dahun Kim**, D. Cho, I.S. Kweon,
"Self-supervised video representation learning with space-time cubic puzzles",
in **AAAI 2019 (Oral - accept rate: 6.5%)**, Honolulu, USA

006. Y. Jung, D. Cho, **Dahun Kim**, S. Woo, I.S. Kweon,
"Discriminative feature learning for unsupervised video summarization",
in **AAAI 2019 (Oral - accept rate: 6.5%)**, Honolulu, USA
Received Honorable Mention, 25th Samsung HumanTech Paper Award

005. K. Park, S. Woo, **Dahun Kim**, D. Cho, I.S. Kweon,
"Preserving Semantic and Temporal Consistency for Unpaired Video-to-Video Translation",
in ACM **MM 2019**, Nice, France (Acceptance: $252/936 \approx 26.9\%$)

004. Cho, Y. Jung, F. Rameau, **Dahun Kim**, S. Woo, I.S. Kweon,
"Video Retargeting: Trade-off between Content Preservation and Spatio-temporal Consistency",
in ACM **MM 2019**, Nice, France (Acceptance: $252/936 \approx 26.9\%$)

003. S. Woo*, **Dahun Kim***, D. Cho, I.S. Kweon,
"LinkNet: relational embedding for scene graph",
in **NeurIPS 2018**, Montreal, Canada (Acceptance: $1011/4856 \approx 20.8\%$)

002. **Dahun Kim**, D. Cho, D. Yoo, I.S. Kweon,
"Learning image representations by completing damaged jigsaw puzzles",
in **WACV 2018 (Oral)**, Lake Tahoe, USA

001. **Dahun Kim**, D. Cho, D. Yoo, I.S. Kweon,
"Two-phase learning for weakly supervised object localization",
in **ICCV 2017**, Venice, Italy (Acceptance: $621/2143 \approx 28.9\%$)

Patents

P4. Electronic Device and Control Method of Same (US Patent App. 17/554,142)
P3. Video Panoptic Segmentation (US Patent App. 16/852,647)
P2. Panoptic Segmentation (US Patent 11,256,960)
P1. Method and Device for Hierarchical Learning of Neural Network Based on Weakly Supervised Learning (US Patent App. 16/758,089)

Awards and Honors

• Best Ph.D. Thesis Award, EE, KAIST	Apr.2022
• Bronze Award, 28th HumanTech Paper Award,	Feb.2022
Samsung Electronics Co., Ltd. ($5,000)	
• Qualcomm Innovation Award (Korea) 2021	Nov.2021
• Outstanding Reviewers Award, CVPR 2021	Aug.2021
• Outstanding Reviewers Award, ECCV 2020	Aug.2020
• Microsoft Research Asia (MSRA) Ph.D Fellowship 2019 Winner ($10,000)	Oct.2019
• 1-st Place Award in ChaLearnLAP 2018 Inpainting Challenge	Sep.2018
Track 2: video decaptioning (ECCV2018 Challenge)	
• Global Ph.D Fellowship, National Research Foundation of Korea	Mar.2018 - Feb.2021
(National Minister fellowship – $\approx$ $60,000 + 3-year full scholarship)	
• KAIST-Samsung Industry-University Cooperation, Best Paper Award ($3,000)	Jul.2020
• Bronze Award, 27th HumanTech Paper Award,	Feb.2021
Samsung Electronics Co., Ltd. ($5,000)	
• Honorable Mention, 25th HumanTech Paper Award,	Feb.2019
Samsung Electronics Co., Ltd. ($2,000)	
• Lab Student Representative (over 30 members),	Sep.2019 - Jul.2020
• Bronze Prize, Best Paper Award, 31th IPIU	Feb.2019
• International Computer Vision Summer School (ICVSS), Sicily, Italy	Jul.2018

References

Prof. In So Kweon: M.S. - Ph.D. advisor at KAIST
KEPCO Chair Professor, Dept. of EE, KAIST; Eemail: iskweon77@kaist.ac.kr
Dr. Joon-Young Lee: Internship mentor
Senior Research Scientist, Adobe Research; Email: jolee@adobe.com
Dr. Liang-Chieh Chen: Internship mentor
Senior Research Scientist, ByteDance Research (previously at Google); Email: lcchen@cs.ucla.edu
Dr. Tsung-Yi Lin: Internship mentor
Senior Research Scientist, Nvidia Research (previously at Google); Email: tsungyilin87@gmail.com