

MinJu Jeon

Hanyang University · Multimodal AI Lab · AI Researcher

Hi, I'm MinJu Jeon — a Master's student in Data Science at Hanyang University, currently a Research Intern at Naver Cloud(Voice Tech). I work on problems at the intersection of language and perception: multilingual speech (G2P/TTS), video-language understanding, and the messy data work that makes models actually usable in production.

- Multimodal Learning
- Video-Text Retrieval
- Speech & Language
- Data-Centric AI

NEWS

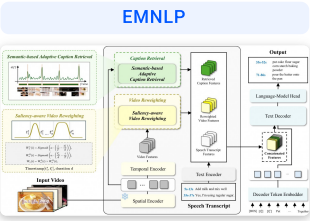
- Mar 2026Cap4Bridge accepted at IEEE Access
- Feb 2026Two papers accepted at CVPR 2026
- Dec 2025Started research internship at Naver Cloud, Voice Tech Team
- Aug 2025Sali4Vid accepted at EMNLP 2025 (Long, Main)



BACKGROUND

- Dec. 2025 – nowResearch Intern, Naver Cloud · Voice Tech Team
Multilingual G2P & robust TTS for non-canonical text
- Sep. 2024 – nowM.S. Data Science, Hanyang University
- March 2020 - Aug. 2024B.S. Industrial Engineering, Hanyang University

selected publications



Sali4Vid: Saliency-Aware Video Reweighting and Adaptive Caption Retrieval for Dense Video Captioning
MinJu Jeon, Si-Woo Kim, Ye-Chan Kim, HyunGee Kim, and Dong-Jin Kim
In Proceedings of the 2025 Conference on Empirical Methods in Natural Language Processing, 2025

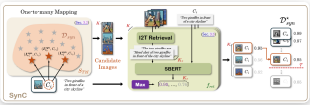
BIB

IEEE Access

Cap4Bridge: Caption-Guided Cross-Modal Contextualization with Stochastic Augmentation for Text-Video Retrieval
MinJu Jeon, Hyungee Kim, Si-Woo Kim, Youngtaek Oh, Soeun Lee, and Dong-Jin Kim
IEEE Access, 2026

BIB

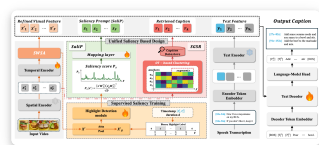
ACM MM



SynC: Synthetic Image Caption Dataset Refinement with One-to-many Mapping for Zero-shot Image Captioning
Si-Woo Kim, MinJu Jeon, Ye-Chan Kim, Soeun Lee, Taewhan Kim, and Dong-Jin Kim
In Proceedings of the 33rd ACM International Conference on Multimedia, 2025

BIB

CVPR



Follow the Saliency: Supervised Saliency for Retrieval-augmented Dense Video Captioning

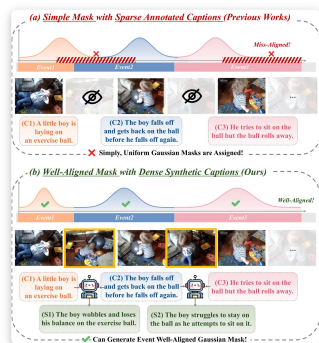
Seunghee Choi, MinJu Jeon, Hyunwoo Oh, Jihwan Lee, and Dong-Jin Kim

arXiv preprint arXiv:2603.11460, 2026

Accepted at CVPR 2026

BIB

CVPR



SAIL: Similarity-Aware Guidance and Inter-Caption Augmentation-based Learning for Weakly-Supervised Dense Video Captioning

Ye-Chan Kim, SeungJu Cha, Si-Woo Kim, MinJu Jeon, Hyungee Kim, and Dong-Jin Kim

arXiv preprint arXiv:2603.05437, 2026

Accepted at CVPR 2026

BIB



Interested in G2P, TTS, and Data Science. Feel free to reach out via email or LinkedIn!

publications

publications by categories in reversed chronological order. generated by jekyll-scholar.

2026

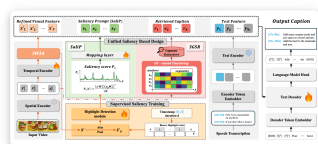
IEEE Access

Cap4Bridge: Caption-Guided Cross-Modal Contextualization with Stochastic Augmentation for Text-Video Retrieval

MinJu Jeon, Hyungee Kim, Si-Woo Kim, Youngtaek Oh, Soeun Lee, and Dong-Jin Kim

IEEE Access, 2026

CVPR



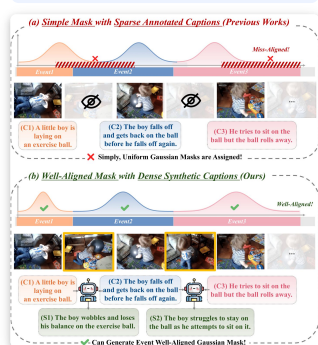
Follow the Saliency: Supervised Saliency for Retrieval-augmented Dense Video Captioning

Seunghee Choi, MinJu Jeon, Hyunwoo Oh, Jihwan Lee, and Dong-Jin Kim

arXiv preprint arXiv:2603.11460, 2026

Accepted at CVPR 2026

CVPR



SAIL: Similarity-Aware Guidance and Inter-Caption Augmentation-based Learning for Weakly-Supervised Dense Video Captioning

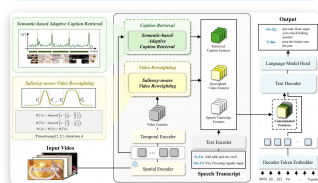
Ye-Chan Kim, SeungJu Cha, Si-Woo Kim, MinJu Jeon, Hyungee Kim, and Dong-Jin Kim

arXiv preprint arXiv:2603.05437, 2026

Accepted at CVPR 2026

2025

EMNLP

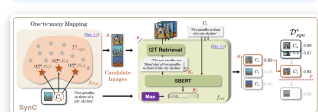


Sali4Vid: Saliency-Aware Video Reweighting and Adaptive Caption Retrieval for Dense Video Captioning

MinJu Jeon, Si-Woo Kim, Ye-Chan Kim, HyunGee Kim, and Dong-Jin Kim

In Proceedings of the 2025 Conference on Empirical Methods in Natural Language Processing, 2025

ACM MM



SynC: Synthetic Image Caption Dataset Refinement with One-to-many Mapping for Zero-shot Image Captioning

Si-Woo Kim, MinJu Jeon, Ye-Chan Kim, Soeun Lee, Taewhan Kim, and Dong-Jin Kim

In Proceedings of the 33rd ACM International Conference on Multimedia, 2025

BIB

projects

Industry-collaborated research projects.

industry



Retrieval-Augmented Dense Video Captioning & QA

End-to-end pipeline for automated highlight extraction and natural language QA over long-form videos — funded by Piaspace

DOOSAN

Risk State Prediction at Construction Sites

VLM-based risk prediction system for detecting unsafe worker behavior and equipment hazards — funded by Doosan Enerbility

HYUNDAI



Zero-Shot Captioning for Driver Status Reporting

Zero-shot image captioning for in-vehicle driver monitoring without task-specific labeled data — funded by Hyundai NGV



Your connection was interrupted

A network change was detected.

ERR_NETWORK_CHANGED

Reload