

경력기술서: 전제영

MLOps/DevOps and AI Engineer (Updated at 2024-12-18)

NOTE

저는 서울, 경기에서 활동하는 MLOps 엔지니어 전제영[Jeayoung Jeon]입니다. 저는 다음의 전문성을 가지고 있습니다:

- 클라우드 네이티브 환경의 **MLOps (API, Pipeline)** 워크로드와 **AI/LLM 플랫폼**을 구현/운영할 수 있습니다.
- 고가용성**과 **GPU 비용 절감**을 위한 **하이브리드 쿠버네티스 클러스터**를 개발/운영할 수 있습니다.
- 머신러닝, 컴퓨터비전, 차량기술** 연구원 경험을 활용하여 **MLOps/DevOps** 관련 의사 결정에 기여할 수 있습니다.

저는 **ICT, AI** 및 **차량기술**을 포함한 다양한 분야의 도전을 꿈꾸고 있습니다. 제가 기여할 수 있는 팀이 있다면 연락 부탁드립니다.

제 경험에 대해 더 궁금하신 점이 있다면 **포트폴리오(영어)** (<https://jyje.live/works>), **포트폴리오(한국어)** (<https://blog.jyje.live>) 를 참고해 주세요.

✉: jyjeon@outlook.com

🏠: <https://jyje.live>

in: [LinkedIn: jyje](https://www.linkedin.com/in/jyje) (<https://www.linkedin.com/in/jyje>)

🐙: [Github](http://github.com/jyje) (<http://github.com/jyje>)

🔗: [StackShare](https://stackshare.io/jyje/jyje-pro-stack) ([https://stackshare.io/jyje/jyje-pro-](https://stackshare.io/jyje/jyje-pro-stack)

stack)

Work

Mar 2024 – Nov 2024 (9개월)

👤 **Intermediate Software Engineer [책임연구원] at MAXST** (<https://maxst.com>)

Roles: 리드 MLOps/DevOps 엔지니어 at 맥스트 기술사업부

- MLOps** 사내 연구소와 협업하기 위해 오픈 소스를 활용하여 ML API, 데이터 파이프라인, AI 플랫폼을 개발
- LLMOps** 사내에 자체 구축한 RAG+LLM 시스템을 활용하여 사내 문서를 학습한 챗봇 개발
- SRE** 웹 서비스 안정성 엔지니어링과 머신러닝 워크로드 안정성 엔지니어링 담당

Jan 2021 – Feb 2024 (3년 2개월)

👤 **Software Engineer [선임연구원] at MAXST** (<https://maxst.com>)

Roles: 연구원 & DevOps 엔지니어 at 맥스트 기술사업부

- Algorithm Research** 컴퓨터 비전 알고리즘 연구, 최신 논문 리뷰 및 프로토타입 구현
- Hybrid Clusters** 고가용/고효율의 디지털 트윈 플랫폼을 위한 하이브리드 클러스터(AWS EKS + 온프레미스) 구현
- DevOps** 사내외 프로젝트를 위한 온프레미스 쿠버네티스 클러스터와 데이터 파이프라인 개발

Mar 2012 – Aug 2020 (8년 6개월)

👤 **Graduate Student Researcher [대학원 연구원, 컴퓨터 비전 전공] at POSTECH** (<https://eee.postech.ac.kr/>)

Roles: 통합박사과정 at 포항공과대학교 전기전자공학과

- Computer Vision** 고성능/고효율의 컴퓨터 비전 알고리즘을 위한 하이퍼파라미터 추정 연구
- Automotives** 컴퓨터 비전 기술을 이용한 차량자율주행 기술 (ADAS, SLAM) 연구; 자율주행기술을 지원하는 국가과제 다수 참여
- FPGA** 실시간 동시 행렬계산을 이용한 컴퓨터 비전 및 머신러닝 알고리즘 효율적 구현; GPU/NPU 가속을 위한 단일칩 시스템 (SoC) 개발



Jan 2024 – Oct 2024 (10 개월)

프로젝트 와이드스: 디지털 트윈 & AR 콘텐츠 플랫폼 at 와이드스, 맥스트 (<https://widearth.world>)**Roles:** 리드 ML/Infra 담당 ~ MLOps/DevOps + ML Backend + SRE [기여도 75%]

- DevOps & SRE** IaC, GitOps, CI/CD 파이프라인, 모니터링, 로깅, 알림, 다중 배포 시나리오, 긴급 대응
- Hybrid Clusters** AWS EKS + 베어메탈 쿠버네티스, API 게이트웨이 패턴, 동적 VM 관리, GPU 비용 최적화
- ML Workloads** ML APIs, ML 파이프라인, 데이터레이크, 도커라이징, 모델 CI/CD

Results: 서비스 출시 성공 ~ 소수 정예, 모든 기능 구현, 가용성 향상, 비용 절약

- 서비스 출시** 15인 총원, 8인 개발, 1인 인프라 담당으로 기획과 개발에 참여하여 플랫폼 출시와 운영에 기여
- 고효율 ML** 하이브리드 인프라로 300+개 공간맵을 생성하였고, 기존 서비스 대비 클라우드 비용을 1.5천만원(70%) 절약
- 고가용 인프라** 하이브리드 클러스터 및 장애 대응으로 서비스 연간 가용성 96%, 다운타임 14일의 원만한 서비스를 구현

Skills: 프로젝트 와이드스를 위한 기술 스택AWS EKS Kubespray Python/FastAPI Argo Workflows Argo CD Bitbucket Pipelines Karpenter

Jan 2024 – Jun 2024 (6 개월)

MLOps: 오픈소스 기반 온프레미스 MLOps 도입 at 맥스트 (<https://maxst.com>)**Roles:** 리드 MLOps 담당 ~ 기획 + VoC + ML Workloads/Infra + PoC + 운영 [기여도 90%]

- Kubeflow** Argo Workflows와 통합; AutoML, 분산학습, 모델 레지스트리 지원, 최대 16 GPU 가속
- JupyterHub** IDE와 통합된 원격 개발 환경 구현; 원격 GPU 노트북 지원, 최대 4 GPU 가속
- VectorDB** Milvus, ChromaDB 등 고가용 벡터 데이터베이스 구현, RAG+LLM 챗봇 개발
- ML Infra** CI/CD, NAS, 데이터 레이크, 이미지 레지스트리, 모니터링 등 ML 워크로드를 지원할 인프라 구현

Results: 연구환경 개선, 리소스 통합 관리 ~ 리소스를 병합하여 가용성과 용량을 늘리고, 관리 자동화로 효율성은 유지

- 연구환경 개선** 연구원들이 개별 관리하던 분산 서버들을 쿠버네티스로 통합하여 인프라 용량 및 안정성 확보. PoC를 통해 구성원의 MLOps 도입 의사결정 성공
- AI 플랫폼** Kubeflow, JupyterHub 등 연구원 2명의 VoC에서 10명으로 점차 사용자 및 규모 확대. 지속적 MLOps 고도화로 연구환경 개선 및 기술부채 해소
- GPU 활용성 개선** GPU 사용률 3배 증가 및 800건 이상의 AutoML 실험 수행 결과로 상품화 성공

Skills: 온프레미스 MLOps 도입을 위한 기술 스택Kubeflow Katib Training Operator Model Registry JupyterHub Argo Workflows Milvus ChromaDB Ollama Open WebUI Grafana Stack
TensorBoard

Dec 2022 – Dec 2023 (13 개월)

DevOps: 사내외 프로젝트를 위한 하이브리드 클러스터 개발 at 맥스트 (<https://maxst.com>)**Roles:** DevOps 실무 담당 ~ 하이브리드 클러스터 개발 + CI/CD, 챗봇, 데이터 파이프라인 개발/운영 [기여도 50%]


- Hybrid Clusters** 퍼블릭 클러스터, 온프레미스 쿠버네티스, 멀티 클러스터, API 게이트웨이, IaC, GPU 오퍼레이터
- CI/CD** 퍼블릭 플랫폼 CI, 온프레미스 커스텀 CI, GitOps를 이용한 CD, 챗봇으로 결과/이슈 알림
- Pipelines** ML 연구를 위한 데이터 파이프라인, ML 추론을 위한 프로덕션 파이프라인

Results: 하이브리드 클러스터 도입 ~ 온프레미스 리소스 활용성 증가 + 퍼블릭 리소스 비용 감소 + DevOps 문화 전파

- 비용 절약** 퍼블릭의 가용성은 유지하면서, 온프레미스의 경제성을 이용해 순수 클라우드 대비 50% 이상의 비용 절약
- 자원 활용** 유휴 온프레미스 리소스의 90% 활용, 타부서 프로토타이핑을 위한 멀티클러스터 공급
- DevOps 문화** 클라우드 네이티브 개발환경 도입. 앱 현대화 및 CI/CD를 포함한 DevOps 문화 전파. 모니터링을 통한 의사결정 지원

Skills: DevOps 및 하이브리드 클러스터를 위한 기술 스택Kubernetes AWS EKS IaC Ansible Terraform CI/CD Bitbucket Pipeline Runners Argo CD Argo Workflows Python/FastAPI Python/Bolt (Slack)

Jan 2021 – Dec 2022 (2 년)

 **컴퓨터 비전 엔지니어 at 맥스트** (<https://maxst.com>)

Roles: 연구원 ~ 디지털 트윈 시스템을 위한 알고리즘 개발 및 프로토타이핑 [기여도 50%]

- **Digital Twins** 360 영상을 3D 공간으로 변환하는 알고리즘 응용, 디지털 트윈 시스템 구현
- **AR/XR** 다양한 스마트글래스에 대한 카메라 정합 및 AR/XR 프로토타입 개발
- **Automation** 스크립트 기반, 영상 데이터 취득 및 분석을 위한 자동화 파이프라인 개발
- **Military Service** 대학원 전공과 관련된 컴퓨터 비전 직군에 종사하는 전문연구요원으로 군대체복무 수행

Results: 컴퓨터 비전 알고리즘 개발 및 디지털 트윈 시스템 구축

- **디지털 트윈** 디지털 트윈 시스템을 위한 Visual-SLAM 및 ICP 알고리즘 프로토타이핑
- **자동화** 데이터 취득 및 분석을 위한 자동화 파이프라인 공급 및 산출물 활용

Skills: 컴퓨터 비전 연구를 위한 기술 스택

Computer Vision Visual-SLAM SfM ICP Python OpenCV .NET/C# Unity

Jan 2012 – Aug 2020 (8 년)

 **컴퓨터 비전 및 ADAS 연구원 (통합박사과정) at POSTECH** (<https://eee.postech.ac.kr/>)

Roles: 대학원생 연구원 ~ 카메라 기반 컴퓨터 비전 및 자동차 기술 연구/개발 [full-time]

- **2018-2020** 제어 및 컴퓨팅 연구실 (Prof. 한수희)
디지털 트윈 및 동시적 위치 추정 및 지도 작성 방법(SLAM) 연구
 - 자율주행을 위한 다중 카메라 기반 SLAM 연구
 - ADAS과 SLAM을 위한 디지털 트윈 프로토타이핑 연구/개발
 - 실제 환경과 가상 환경을 위한 동시적 위치 추정 및 지도 작성 방법
- **2012-2018** 첨단 신호 처리 연구실 (Prod. 정홍)
운전자 보조 시스템(ADAS) 및 고성능-소형화(Edge) 컴퓨터 비전 연구
 - 고성능, 교호율의 FPGA를 이용한 실시간 ADAS 구현
 - FPGA를 이용한 실시간 교통 표지판 및 도로 탐지 시스템 개발
 - 영상깊이추정을 위한 스테레오 비전 알고리즘 연구
 - 스테레오 비전을 이용한 차량 카메라 온라인(실시간) 정합 개발
 - 비용누적테이블(Cost Aggregation Table)을 이용한 컴퓨터 비전 최적화 알고리즘 연구

Results: 프로젝트 및 연구논문 ~ 가상 환경에서의 자동차 시뮬레이션 및 ADAS On-Edge에 대한 연구를 수행

- **Digital Twins** Virtual Visual-SLAM: 실제 환경과 가상 환경을 위한 동시적 위치 추정 및 지도 작성 방법
- **Edge ADAS** FPGA를 이용한 실시간 운전자 보조 시스템 개발 및 교통표지판 검출 알고리즘 연구

Skills: 컴퓨터 비전 및 ADAS 연구를 위한 기술 스택

Computer Vision Digital Signal Processing Automotives Autonomous Driving Advanced Driver Assistance Systems (ADAS)
Finite Programmable Gate Array (FPGA) Traffic Sign Detection Lane Terrain Detection MATLAB/Simulink C/C++

Skills



NOTE

제 스킬 중 현업에 바로 쓸 수 있는 것은 강조하였습니다.

MLOps & LLMOps :

Ollama OpenAI API RAG AutoRAG Kubeflow AutoML Katib Training Operator JupyterHub Data Pipelines

DevOps & SRE :

Kubernetes On-Premise AWS EKS GCP GKE Hybrid Clusters ARM64 IaC Kubespray Terraform Ansible Istio Grafana Stack Karpenter

CI/CD/CT/CT :

Argo Projects Bitbucket Pipelines GitHub Actions Self-Hosted Runner Kaniko Buildah Locust Litmus

ML Backend :

Python/FastAPI Ollama Milvus PostgreSQL Redis

Computer Vision :

Automotives SLAM PyTorch OpenCV FPGA

UI/UX :

Slackbot Python/FastUI Open WebUI Chainlit .NET/MAUI .NET/WPF Unity

FinOps & BizOps :

Kubecost Continuous BI

Programming languages :

Python .NET/C# C/C++ MATLAB

Education



Mar 2012 – Aug 2020

🎓 석사학위(통합과정수료) in 전기전자공학과, 컴퓨터비전 및 신호처리 from Pohang University of Science and Technology (POSTECH) with GPA of 3.2/4.3

- Thesis: 실제 환경과 가상 환경을 위한 동시적 위치 추정 및 지도 작성 방법

Mar 2008 – Feb 2012

🎓 학사학위 in School of Electronic Engineering, Electronic Communication from Kumoh National Institute of Technology (KIT) with GPA of 4.3/4.5

- Thesis: LED 조명을 이용한 수중 환경에서의 가시광 통신 연구, 2011

Awards



May 2014

🏆 알테라 디자인 컨테스트 2014, 우수상 from Intel-Altera Korea

[System] FPGA, Vision-Based Driver Support Navigation System

May 2014

🏆 워크샵 최우수 포스터 세션 from KYUTECH-POSTECH 연합 워크샵

[Poster] Iterative Polygon Detection using Harris Corner Space Method for Finding Traffic Signs

May 2013

🏆 알테라 디자인 컨테스트 2013, 2위 입상 from Intel-Altera Korea

[System] FPGA, Vision-Based Traffic Sign Recognition System

Feb 2012

🏆 학부 수석 졸업 from 금오공과대학교

[수석 졸업] 전자공학부 최우수 졸업

Jan 2012

🏆 NAVER 파워지식IN 2011 (<https://m.site.naver.com/1y6qP>) from NAVER

[Activity] 지식봉사활동 `전기전자공학, 수학, 프로그래밍 분야`. 2009-2011 활동, 2012 선정 / 답변수 723, 채택 98.1%

Publications



Jul 2020, POSTECH, 학위논문 (1저자)

실제 환경과 가상 환경을 위한 동시적 위치 추정 및 지도 작성 방법 (<http://postech.dcollection.net/common/orgView/200000341295>) by 전제영

Nov 2014, ISVC, Advances in Visual Computing, 10th International Symposium (2저자)

Cost Aggregation Table: Cost Aggregation Method Using Summed Area Table Scheme for Dense Stereo Correspondence (https://doi.org/10.1007/978-3-319-14249-4_78) by JeongMok Ha, Jeayoung Jeon, GiYeong Bae, SungYong Jo & Hong Jeong

Oct 2014, ICCAS, 14th International Conference on Control, Automation and Systems (1저자)

Polygonal symmetry transform for detecting rectangular traffic signs (<https://doi.org/10.1109/ICCAS.2014.6987934>) by Jea Young Jeon, JeongMok Ha, Sung Yong Jo, Gi Yeong Bae, Hong Jeong

Apr 2011, ICS-KIEE (공동 1저자)

LED 조명을 이용한 수중 환경에서의 가시광 통신 연구 (<https://www.dbpia.co.kr/Journal/articleDetail?nodeId=NODE01951197>) by 이대희, 박기성, 전제영, 양연모

Certifications



Nov 2024 (Expired in Nov 2027)

GitHub Foundations (<https://www.credly.com/badges/876fa6b3-0b27-4ddf-bbb3-a9d853918566>) from **GitHub**

Sep 2024 (Expired in Sep 2026)

CAPA: Certified Argo Project Associate (<https://www.credly.com/badges/ee42c2c7-2ac3-411f-8713-cc26cbec8022>) from **The Linux Foundation**

Jun 2024 (Expired in Jun 2026)

CKAD: Certified Kubernetes Application Developer (<https://www.credly.com/badges/9e072a3a-57d0-403e-8bef-5831d618675c>) from **The Linux Foundation**

Mar 2024 (Expired in Mar 2027)

CKA: Certified Kubernetes Administrator (<https://www.credly.com/badges/d944bde7-222a-4ce5-b4e6-4e6c84df0ef8>) from **The Linux Foundation**

Interests



Research/Dev :

Agentic RAG Digital Twins AMD-to-ARM Transition Hybrid Clusters

DevOps Culture :

Coop First, Tech Next Automate as Possible Internal Development Platform

Home Clusters :

Raspberry Pies Personal RAG Live Demo

Languages



Korean :

English :

Native

Working Proficiency