

■ 기술명 : [클라우드-엣지 연동 AI 서비스 지원 기술]

[영문 : Software module for AI service composition on edge device]

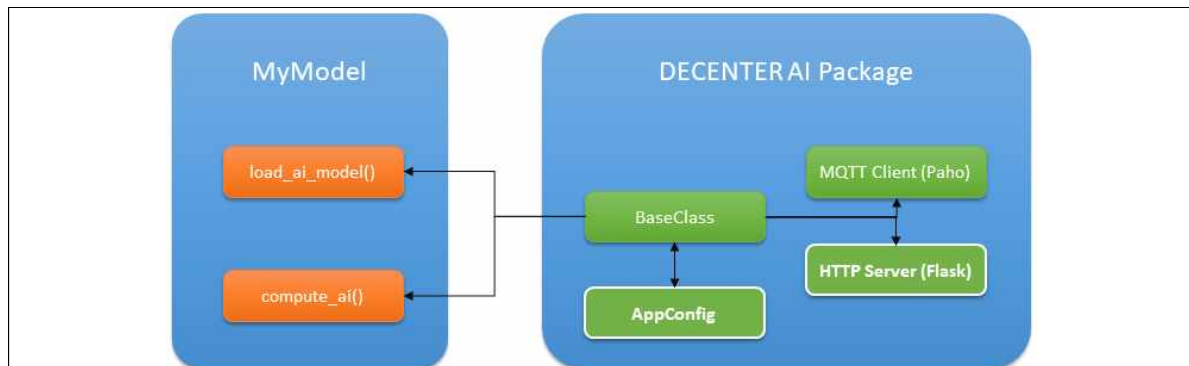
산업기술분류	정보통신-소프트웨어-소프트웨어 솔루션
Key-word(국문)	클라우드 컴퓨팅, 엣지 컴퓨팅, 협업 기술
Key-word(영문)	cloud computing, edge computing, edge collaboration

■ 기술의 개요

- (배경) 엣지 컴퓨팅은 다양한 엣지 하드웨어의 발전에 힘입어 클라우드 컴퓨팅과 함께 AI 및 분석 서비스를 실장할 수 있는 환경으로 발전하고 있으며, 다양한 경량화 엔진들이 개발되어 자체 분석 엔진을 실장하는 형태로 발전하고 있음
- AI 서비스를 구성하기 위한 Deep Learning 기술의 경우 대용량의 데이터를 활용한 학습을 통하여 그 성능의 비약적으로 발전하고 있으나, 이를 엣지상에서 서비스하기 위해서는 다양한 제약이 존재함
- AI 서비스의 구성은 Deep Learning 관련 기술과 함께 네트워크 프로토콜, 아키텍처 등 서비스 플랫폼 기술에 대한 지식이 요구되어 구현에 대한 다양한 선지식이 필요함
- (개요) 본 기술은 다양한 대용량 데이터로부터 학습된 Deep learning model을 기반으로, 클라우드 혹은 엣지에 서비스를 구성하기 용이한 소프트웨어 플랫폼을 제공.
 - 기존 학습된 모델 기반의 AI 서비스 구성 기술
 - Docker Container 기반의 AI 서비스 컨테이너 구성
 - Kubernetes 등 resource orchestration 도구를 통한 서비스 배포 지원
 - 클라우드 VM 혹은 edge device baremetal에 적용 가능



<기술 개요도 >

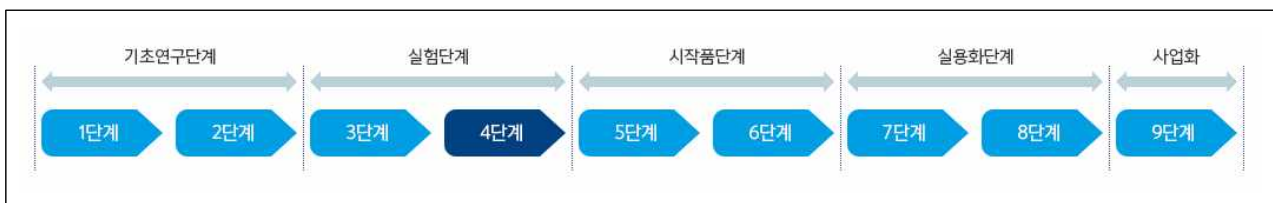


< AI package 구조도 >

- Docker, Kuberentes 호환
- Python 언어 기반 패키지 구성
- AI - 네트워크 관리 모듈 분리를 통한 서비스 구성
- 기존 Deep Learning application의 re-factoring을 통한 손쉬운 구성 지원

[명칭: 마이크로서비스 기반 인공지능 및 계학습 서비스 구성 시스템 및 방법]
(관련출원특허 : 10-2020-0039361)

■ 기술의 구현수준(TRL) : (4)단계



■ 구현기술 명세 및 장점(경쟁기술과의 차별성)

○ 당해단계 기술명세

- 학습된 딥 러닝 모델을 활용한 서비스 구성
- 서비스 구성 프로토콜 지원: HTTP, MQTT
- 서비스 구성 지원 모델: VGG16, Yolo v3
- 서비스 구성 지원 H/W: Nvidia 1080ti



○ 추후단계 확보대상 기술명세

- 서비스 구성 프로토콜 확장: gRPC
- 서비스 구성 지원 모델 확장: 영상 및 음성 모델 지원
- 서비스 하드웨어 지원 확장: Nvidia Jetson, AGX

○ 보유기술 장점 및 차별성(경쟁기술대비)

- 네트워크 아키텍처 및 딥 러닝의 분리를 통한 유연한 설계의 지원
- 마이크로 서비스 아키텍처 활용을 통한 유지 보수 용이
- 기존 모델의 활용을 통한 서비스 활용도 증대
- 다양한 네트워크 아키텍처 지원을 통한 서비스 구성 용이

■ 활용범위 및 응용분야

- 다양한 영상 분석 모델의 실장을 지원하는 엣지 카메라 모듈
 - 단일 엣지 카메라 하드웨어에 대하여 상황에 따른 분석 모델 실장
 - 열영상 감지용 모델, 침입자 감지 모델, 성별 구분 모델 등
- 영상 기반 서비스 구성을 지원하는 스마트 팩토리, 스마트 팜, 스마트 홈 용 엣지 서버 구성
 - 상황에 따른 모델 실장을 통한 스마트 서비스 구성 모듈로 활용

■ 지식재산권 현황

구분	발명의 명칭	출원번호 (출원일)	등록번호 (등록일)
특허	마이크로서비스 기반 인공지능 및 기계학습 서비스 구성 시스템 및 방법] (관련출원특허 : 10-2020-0039361)	10-2020-0039361 (2020.03.31.)	