

07

| 발표기술 |

초연결네트워크

대규모 딥러닝 고속처리를 위한 분산 딥러닝 플랫폼 기술

- 특허명 : 파라미터 공유 장치 및 방법 (10-2017-0060400)
- 보유기관 : 한국전자통신연구원
- 상태정보 : 출원 '17.05.16 → 공개 '18.11.26
- 기타정보 : 관련특허 포트폴리오 구축(총 6건)



기술개요

- 대규모 데이터와 big size 딥러닝 모델의 학습시간을 효율적인 분산 처리를 통해 향상시키는 기술로, 인피니밴드 기반 중저가/이종 규격의 GPU 클러스터 환경에서도 운용 가능함
- enabling 기술로서 특히 고해상도 이미지/영상의 객체 분류, 인식, 세그멘테이션과 같이 파라미터 크기가 매우 크고, 대규모 데이터를 학습해야 하는 딥러닝 플랫폼에 적용 가능함

기존 문제점

- 최소 파라미터의 통신량을 줄이는 분산 딥러닝 플랫폼, 빅데이터 처리 플랫폼과 연계하는 방식으로 분산 딥러닝 플랫폼을 개발하고 있으나, 범용성/성능 면에서 한계를 보이고 있음

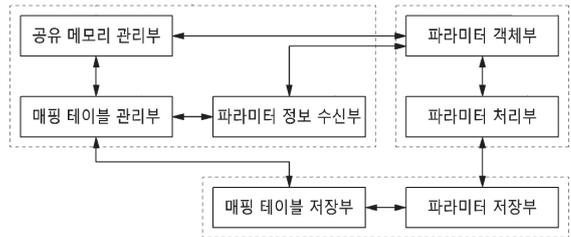


기술 차별점

- 분산 컴퓨터의 원격 메모리를 모아서 공유 메모리 형태로 딥러닝 분산 프로세스에 제공하여 이 공유 메모리를 딥러닝 파라미터 통신을 위해 사용함으로써 분산 학습 시간을 획기적으로 개선함

세부내용

- 메모리 박스 접근부와 메모리 박스의 일 예를 세부적으로 나타낸 블록도
- 분산 딥러닝 학습에서 파라미터 서버를 메모리 박스로 대체 지원함에 있어서 딥러닝 프레임워크가 가진 원래의 기능과 사용자가 사용하는 딥러닝 모델 개발 및 학습 인터페이스에 수정을 가하지 않음



- 한국전자통신연구원 김호민(042-860-1804, hominkim@etri.re.kr)
- 공동마케팅사무국 서원우(042-862-6018, sww93@wips.co.kr)