

학부생 연구기회 프로그램 (UROP) 공고

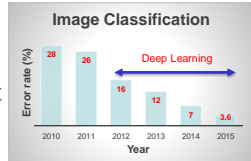
◆ 담당교수 : 이영기	◆ 연구실명 : 모바일 및 유비쿼터스 컴퓨팅 연구실
◆ UROP 연구 과제명 : 다수의 DNN 기반 애플리케이션의 동시 실행을 위한 모바일 딥러닝 시스템	
◆ 모집대상 : 학부 3학년 혹은 4학년	
◆ 모집기간 : ~ 2018년 10월 중순	

모바일 AI / 딥러닝 시스템을 함께 연구할 학생을 찾습니다! 근 미래의 모바일 디바이스는, 다수의 DNN 기반의 추론 연산을 수행하면서, 사용자의 행동, 감정, 활동, 위치 등 다양한 정보를 실시간으로 파악하고, 이를 바탕으로 능동적 애플리케이션을 제공하게 될 것입니다. 하지만 현재의 모바일 디바이스는 자원 제약으로 인해 하나의 DNN 기반 추론을 실행시키는 데에도 오랜 시간이 걸리는 문제가 있습니다. 본 연구에서는, 연산 공유, 캐싱, 스케줄링 등 다양한 시스템 최적화 기술을 고안, 구현하고자 합니다. 이를 바탕으로 DNN 기반 추론 연산을 가속화하고 다수의 DNN 기반 추론 연산의 동시 실행을 가능하게 하는 것을 목표로 하고 있습니다. 관심 있는 학생은 주저하지 말고 아래의 제 이메일이나 전화로 연락 주시기 바랍니다.

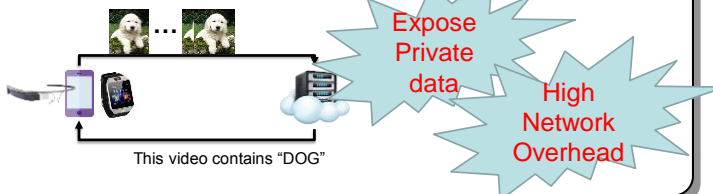
Motivation

• Use of Deep Learning

- High accuracy
- High computational requirement
 - AlexNet (2012): 725M FLOPS



• Current Practice: Processing on Cloud

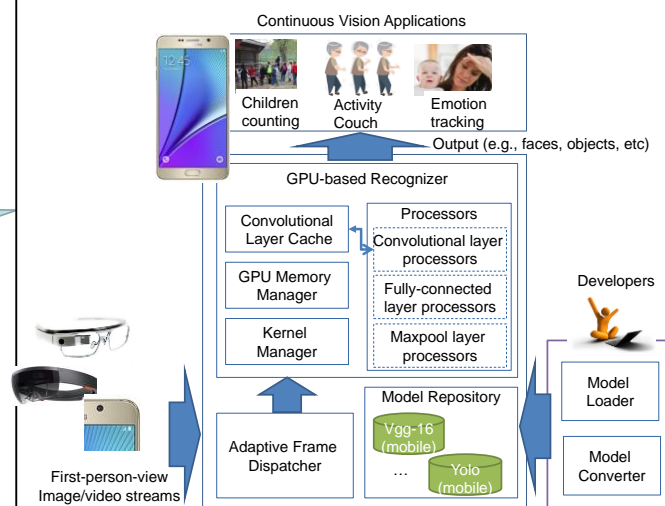


Optimizations

• In this project, we plan to design and develop various optimization techniques tailored to mobile devices.

- Intermediate result caching for mobile GPUs
- Layer decomposition
- Shared processing across multiple DNNs

Proposed Architecture



서울대학교 컴퓨터공학부
Seoul National University
Dept. of Computer Science and Engineering