

Multi-UAV 의 협업을 통한 DNN 서비스 분산 처리 기법

진민규¹, 이찬민², 서민석², 박주성², 최시은², 조안나², 이수경²

¹연세대학교 전기전자공학과, ²연세대학교 컴퓨터과학과

wlsalsrb1236@yonsei.ac.kr, noel98@yonsei.ac.kr, tat1218@naver.com, pjs@yonsei.ac.kr,

sieunchoi@yonsei.ac.kr, annacho6@yonsei.ac.kr, sklee@yonsei.ac.kr

Distributed DNN Service through Multi-UAV Collaboration

Min-Gyu Jin¹, Chan-Min Lee², Min-Seok Seo², Ju-Seong Park²,

Si-Eun Choi², An-Na Cho², Su-Kyoung Lee²

¹Dept. of Electrical & Electronic Engineering, Yonsei University

²Dept. of Computer Science, Yonsei University

요 약

유연한 이동성, 쉬운 배치, 저렴한 비용 등의 장점을 가진 Unmanned Aerial Vehicle(UAV)를 이용해 Deep Neural Network(DNN) 서비스를 제공하는 기술이 연구되고 있다. 하지만 UAV는 메모리와 컴퓨팅 능력, 배터리가 제한되어 있어 DNN 서비스의 요구사항을 만족시키기 위해서는 다수의 UAV 간의 협업이 필요하다. 본 논문에서는 다수의 UAV 협업 환경에서 DNN 서비스의 처리 지연시간을 줄이기 위해 UAV 들의 작업량을 고려한 서비스 분산 처리 기법을 제안하고, 시뮬레이션을 통해 DNN 서비스 처리 지연 시간을 분석한다.

1. 서론

Unmanned Aerial Vehicle(UAV)는 유연하고 신속하게 이동할 수 있다는 특징이 있어 통신 분야를 비롯해 교통, 치안, 재난 등 다양한 상황과 분야에서 사용되고 있다. 최근 스마트 카메라를 탑재한 UAV가 촬영한 항공 영상을 분석하기 위한 Deep Neural Network(DNN) 모델 기반 서비스들(예, object detection, classification)을 UAV의 컴퓨팅 자원을 사용해 처리하는 연구가 진행되고 있다[1],[2]. 하지만 이러한 서비스들은 실시간 처리 및 고연산 컴퓨팅 자원을 요구하며, 제한된 배터리 및 연산 능력을 가진 UAV가 다수의 DNN 기반 서비스들을 처리하기에는 한계가 있다. 이를 해결하기 위해 [3],[4]는 다수의 UAV들과의 협업을 통해 서비스를 처리하는 방법을 제안하였다. 하지만 각 UAV에서 현재 수행되고 있는 작업량을 고려하지 않아 특정 UAV에 작업량이 집중될 경우, 협업 시스템 전체 처리량이 저하되고 병목현상이 발생해 서비스 처리 지연시간이 증가할 수 있다[5].

이에 본 논문에서는 다수의 UAV 협업 환경에서 DNN 서비스의 처리 지연시간을 최소화하기 위해 UAV들의 작업량을 고려한 서비스 분산 처리 기법을 제안하고, 시뮬레이션을 통해 제안 기법이 서비스 처리 지연시간을 감소시키는데 효과적임을 입증한다.

2. 시스템 모델

본 논문에서는 스마트 카메라가 장착된 I 개의 UAV들이 DNN 서비스를 수행하는 환경을 고려하였다. DNN 서비스는 여러 개의 작업량(workload)으로 구성되어 있으며, UAV는 무선 통신을 통해 이웃 UAV들의 현재 작업량 상태를 알고 있다고 가정한다.

i 번째 UAV가 전체 작업량 S 를 처리하는데 소요되는 처리 지연시간 (T_i^{pr})은 다음과 같다.

$$T_i^{pr} = \frac{cS}{f_i} \quad (1)$$

c 는 1 bit의 작업량을 처리하는데 필요한 CPU 사이클 수를 의미하고, f_i 는 UAV의 CPU 클럭 주파수를 의미한다. UAV가 직접 처리할 작업량의 비율을 $\alpha_i \in [0,1]$ 라고 할 때, i 에서 처리되는 작업량의 처리 지연시간은 $\alpha_i \cdot T_i^{pr}$ 이다.

i 번째 UAV에서 α_i 비율의 작업량을 처리하면 잔여 작업량 비율 $(1 - \alpha_i)$ 은 이웃 UAV에서 처리되어야 한다. i 는 이웃 UAV들 중 현재 작업량이 가장 적은 $j \in I$ (단, $j \neq i$) 번째 UAV를 선택하고 j 에게 잔여 작업량을 전송한다. 이때 i 에서 j 로 작업량 전송 시 발생하는 전송 지연시간(T_{ij}^{tr})은 다음과 같다.

$$T_{ij}^{tr} = \frac{(1 - \alpha_i)S}{R} \quad (2)$$

R 은 데이터 전송률을 의미하고, Shannon-Hartley formula 에 의해 $R = B \log_2(1 + \frac{hp}{\sigma})$ 로 계산된다. B 는 채널 대역폭, h 는 채널 이득, P 는 UAV 의 송신 전력, σ 는 UAV 의 잡음 전력을 의미한다. j 번째 UAV 가 전송 받은 작업량을 처리할 때 소요되는 처리 지연시간은 $(1 - \alpha_i) \cdot T_j^{pr}$ 이다. 따라서 서비스를 처리하는데 소요 되는 총 처리시간(T^{compl})은 다음과 같다.

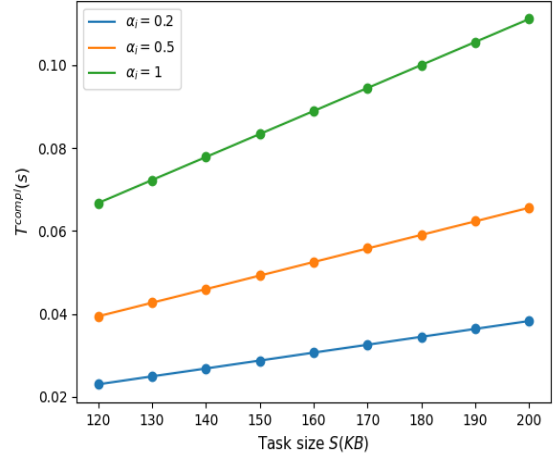
$$T^{compl} = \alpha_i \cdot T_i^{pr} + (1 - \alpha_i) \cdot T_j^{pr} + T_{ij}^{tr} \quad (3)$$

3. 실험 및 성능평가

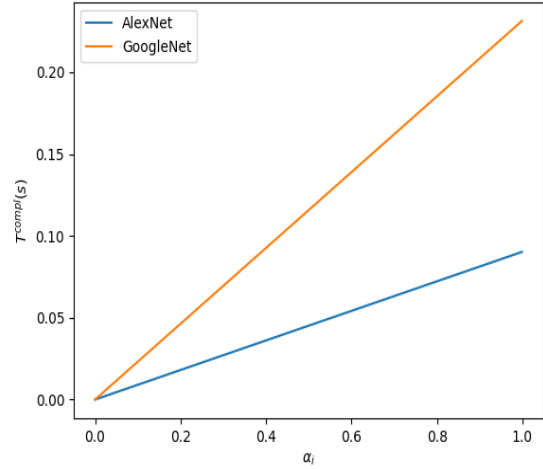
본 논문에서 제안하는 기법의 성능을 평가하기 위해 5 개의 UAV 가 배치된 환경의 시뮬레이터를 구성하여 실험을 진행하였다. 채널 대역폭은 5MHz, UAV 의 CPU 클럭 주파수는 2GHz 로 설정하였다. 먼저, UAV 가 직접 처리할 작업량 비율이 서비스의 총 처리시간에 영향을 미치는지 확인하기 위한 실험을 진행하였고, 실험 결과는 (그림 1)과 같다. α_i 를 0.2, 0.5, 1 의 값으로 설정한 경우, 작업량 S 에 따른 총 처리시간(T^{compl}) 변화를 나타낸다. 작업량이 증가함에 따라 총 처리 시간이 증가하고, UAV 가 직접 처리하는 작업량 비율이 클수록 총 처리시간이 가파르게 증가하는 것을 확인할 수 있고, 이를 통해 UAV 가 직접 처리할 작업량 비율이 DNN 서비스의 총 처리 시간에 영향을 미치는 것을 확인하였다.

다음으로 α_i 값을 0 부터 1 까지 증가시켰을 때 서로 다른 모델을 사용하는 DNN 서비스의 총 처리시간을 비교하였고, 실험 결과는 (그림 2)와 같다. 본 실험에서는 UAV 가 수행하는 DNN 모델로 AlexNet 과 GoogleNET 을 고려하였고, 1KB 당 처리 사이클을 나타내는 computation intensity 를 각각 301(cycles/KB), 772(cycles/KB)로 설정하였다[6].

AlexNet 과 GoogleNet 은 각각 $\alpha_i = 1$ 일 때, 실제 테스트베드에서 측정된 모델 추론 시간인 89ms, 227ms 와 유사한 값을 보였으며[6], α_i 이 작아질수록 서비스의 총 처리시간이 줄어드는 것을 볼 수 있다. 이를 통해 제안하는 작업량 분산 처리 기법이 서비스 처리 지연 시간을 감소시킬 수 있음을 확인하였다.



(그림 1) S 와 α_i 변화에 따른 서비스 처리 지연시간



(그림 2) α_i 변화에 따른 DNN 서비스 처리 지연시간

4. 결론

본 논문에서는 스마트 카메라를 탑재한 다수의 UAVs 환경에서 DNN 서비스의 총 처리 시간을 줄이기 위한 UAV 들 간 서비스 분산 처리 기법을 제안하였다. 실험을 통해 UAV 가 직접 처리하는 작업량의 비율이 줄어들수록 DNN 서비스의 총 처리시간이 감소함을 확인하였다. 향후 연구에서는 UAV 의 이동성을 고려한 연구를 진행하고자 한다.

사사문구

이 연구논문은 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구결과임(No. 2022R1A2B5B01001683).

참고문헌

- [1] Yun, WonJoon, et al. "Cooperative multiagent deep reinforcement learning for reliable surveillance via autonomous multi-UAV control." *IEEE Transactions on Industrial Informatics*, vol. 18, no. 10, 7086-7096, (2022).
- [2] Saetchnikov, Ivan V, et al. "Object detection for unmanned aerial vehicle camera via convolutional neural

- networks." *IEEE Journal on Miniaturization for Air and Space Systems*, vol. 2, no. 2, 98-103, (2020).
- [3] Hu, Long, et al. "Ready player one: UAV-clustering-based multi-task offloading for vehicular VR/AR gaming." *IEEE Network*, vol. 33, no. 3, 42-48, (2019).
- [4] Bhandari, Sabin, et al. "Mobility and location-aware stable clustering scheme for UAV networks." *IEEE Access*, vol. 8, 106364-106372 (2020).
- [5] Zeng, Xiao, et al. "Distream: scaling live video analytics with workload-adaptive distributed edge intelligence." *Proceedings of the 18th Conference on Embedded Networked Sensor Systems*. (2020), 409-421.
- [6] Lickang Zeng, et al. "CoEdge: Cooperative DNN Inference with Adaptive Workload Partitioning Over Heterogeneous Edge Devices." *IEEE/ACM Transactions on Networking*, vol. 29, no. 2, pp. 595-608, (2021)