

System Model (Sequence Diagram)

Document

Project Name	Multi-Task Learning을 활용한 PVT v2 프레임워크 성능 개선
-----------------	---------------------------------------------

08 조

202001156 김수영

202002510 송재현

지도교수: 이종률 교수님 (서명)

Document Revision History

REV#	DATE	AFFECTED SECTION	AUTHOR
1	2025/04/29	1.1, 1.2, 1.3, 2	김수영
2	2025/05/01	1 문단 보충 및 3,4 작성	김수영
3	2025/05/01	보고서 검토	김수영, 송재현

Table of Contents

1.	INTRODUCTION	5
1.1.	연구 배경	5
1.2.	연구 목적	6
1.3.	연구 질문/가설.....	7
2.	USE CASE DIAGRAM.....	8
2.1.	소프트웨어 활용 사례.....	8
2.2.	문제 해결에 대한 사용 사례 DIAGRAM.....	9
3.	SEQUENCE DIAGRAM.....	10
3.1.	해결 방법에 대한 알고리즘 순서도	10
4.	AI 도구 활용 정보	11

List of Figure

그림 1 소프트웨어 활용 사례 DIAGRAM	8
그림 2 문제 해결에 대한 사용 사례 DIAGRAM	9

1. Introduction

1.1. 연구 배경

Multi-Task Learning(MTL)은 여러 작업을 동시에 학습하여 각 작업 간의 관련성을 활용, 기존의 Single Task Learning(단일 작업 학습, STL)에 비해서 모델의 일반화 성능을 향상, 또는 모델의 경량화를 달성할 수 있는 학습 전략이다. 실제로 테슬라에서는 자율주행기술 연구에 HydraNet이라는 MTL이 적용된 모형을 접목시켜 자율주행을 위한 주변 환경 감지에 활용하고 있다. 하지만 Multi-Task Learning이 단일 작업 학습보다 유리하다는 실증적인 근거가 아직까지는 부족한 실정이며, 실제 적용 가능성에 대한 평가도 미흡한 상황이다. 잘못된 작업들을 학습에 이용할 경우 부정적 전이(Negative Transfer)가 발생할 수도 있기에 적절한 작업 선정에 대한 명확한 기준도 제시되어야 한다. 또한, 여러 작업들을 동시에 처리하다보니 성능 평가의 기준에 대한 불확실성도 존재한다는 문제점이 존재한다.

MTL 모형을 개발 및 발전시키는 방법론은 크게 두 가지로 MTL에 최적화된 이미 연구된 모형들을 개선하는 방향과 기존에 STL에서 성능이 입증된 모형을 MTL로 확장시키는 방향이 있다. 실제로 MulT¹, M3ViT², IPT³와 같은 모형들이 존재하며 이들의 유효성은 이미 실험적으로 충분히 검증되었다. 하지만 기존에 Single Task Learning에 최적화된 모형을 Multi-Task Learning으로 확장하는 연구 역시 중요하다.

그 이유로는 첫 번째, STL을 MTL로 확장하는 행위의 당위성이다. MTL은 여러 Task를 Shared Knowledge로 학습시키는 방법론인 만큼 초창기에는 기존에 존재했던 STL 모형에 약간의 변형을 통해 MTL 방법론을 구현하였다. 실제로 초기 MTL 모형인 Multi-layer feedforward neural network는 기존에 있는 입력층, 은닉층, 출력층으로 구성된 학습 모형에 출력층의 유닛의 개수를 단일이 아닌 여러 개로 설정한 모형이다.⁴ 이는 STL을 MTL로 확장시키는 연구가 충분히 합리적이라는 당위성을 부여해준다.

두 번째로, STL 모형의 성능 향상이다. Swin MTL은 기존에 알려져 있는 Swin Transformer를 MTL으로 확장한 모형으로 추론 시 메모리 사용량이 획기적으로 감소하여

¹ MulT: An End-to-End Multitask Learning Transformer - Deblina Bhattacharjee, Tong Zhang, Sabine Susstrunk and Mathieu Salzmann. CVPR, 2022, pp. 12031-12041

² M3ViT: Mixture-of-Experts Vision Transformer for Efficient Multi-task Learning with Model-Accelerator Co-design - Hanxue Liang, Zhiwen Fan, Rishov Sarkar, Ziyu Jiang, Tianlong Chen, Kai Zou, Yu Cheng, Cong Hao, Zhangyang Wang. arXiv:2210.14793v1 [cs.CV] 26 Oct 2022

³ Pre-Trained Image Processing Transformer - Hanting Chen, Yunhe Wang, Tianyu Guo, Chang Xu, Yiping Deng, Zhenhua Liu, Siwei Ma, Chunjing Xu, Chao Xu, Wen Gao. CVPR, 2021, pp. 12299-12310

⁴ A Survey on Multi-Task Learning - Yu Zhang, Qiang Yang. arXiv:1707.08114v3 [cs.LG] 29 Mar 2021

임베디드 환경에서 대단히 효과적인 모형이라는 것이 입증되었다.⁵ 이는 기존 모형을 단순히 MTL로 재학습시키는 것이 기존 모형의 성능 자체를 향상시킬 수 있다는 것을 의미한다. 이는 기존에 알려진 모형에 간단한 구조적 변경을 적용하여 상호 보완적인 성능 향상을 얻을 수 있음을 의미한다.

마지막으로 성능 평가에 대한 명확한 기준점 설정이다. PVT v2는 이미 Swin Transformer보다 우수한 성능이 입증된 모델로, 이를 MTL로 확장했을 때의 변화를 측정함으로써 MTL의 유효성을 객관적으로 평가할 수 있다. 즉, 이미 검증된 강력한 STL을 기준점(baseline)으로 삼아 MTL을 적용한 후의 성능 변화를 정량적으로 측정할 수 있다는 이점이 있다.

1.2. 연구 목적

우리 조의 연구 목표는 MTL의 유효성을 검증하는 것이다. 유효성이라는 것은 그 자체로 해당 학습 방법론의 성능이 유효한 것인지에 대한 검증일 수도 있지만, MTL의 특성상 경량화와 확장성 역시 유효성 검증에 있어 대단히 중요한 부분이라 할 수 있다.

PVT v2는 비전 태스크에서 우수한 특징 추출 능력을 보여주고 있으며 연구 배경에서도 언급된 Swin Transformer보다 높은 성능을 보여주는 Transformer이다.⁶ 그렇기에 이를 MTL로 확장할 경우 Swin MTL과 기존 PVT v2 모형보다 높은 성능을 보여줄 것이라 예상되며 자율주행 환경에서 요구되는 객체 탐지, 의미론적 분할, 이미지 분류 등 다양한 작업을 동시에 효율적으로 처리할 수 있을 것으로 기대된다. 테슬라의 HydraNet 사례에서 보듯, 실제 자율주행 환경에서는 여러 비전 태스크의 통합적 처리가 필수적이며, 우리의 연구는 더 효율적인 모델 구조를 제안함으로써 실제 적용 가능성을 높이는 데 기여할 수 있다.

이에 본 프로젝트에서는 PVT v2 기반 프레임워크에 MTL을 접목하여 성능을 검증하고, 자율주행 분야에서의 적용 가능성을 구체적인 실험을 통해 확인하고자 한다.

⁵ SwinMTL: A Shared Architecture for Simultaneous Depth Estimation and Semantic Segmentation from Monocular Camera Images - Pardis Taghavi, Reza Langari, Gaurav Pandey. arXiv:2403.10662v1 [cs.CV] 15 Mar 2024

⁶ PVT v2: Improved Baselines with Pyramid Vision Transformer - Wenhai Wang, Enze Xie, Xiang Li, Deng-Ping Fan, Kaitao Song, Ding Liang, Tong Lu, Ping Luo, Ling Shao - arXiv:2106.13797v7 [cs.CV] 17 Apr 2023

1.3. 연구 질문/가설

본 연구는 다음과 같은 연구 질문에 답하고자 한다:

- **RQ1**
PVT v2 프레임워크에 MTL을 적용한 모델은 단일 작업 학습 모델과 기존에 알려져 있는 MTL 모형에 비해 학습한 작업(이미지 분류, 객체 탐지, 의미론적 분할)에 대한 유의미한 성능 향상이 이루어지는가?
- **RQ2**
Multi-Task Learning을 이용해 학습시킨 모델은 자율주행 분야에서 기존 프레임워크 대비 어떤 장단점을 가지는가?

본 연구는 다음과 같은 가설을 설정할 수 있다:

- **H1**
MTL을 이용해 학습시킨 PVT v2 모델은 단일 작업 모델과 기존 MTL 모형보다 정확도, AP, mIoU, #Param 등의 성능 지표에서 유의미한 향상을 보일 것이다.
- **H2**
MTL 기반 모델은 자율 주행 분야에서 연구되는 모형인 HydraNet 대비 정확도 측면에서 유의미한 성능 향상을 보여줄 것이다.

2. Use Case Diagram

2.1. 소프트웨어 활용 사례

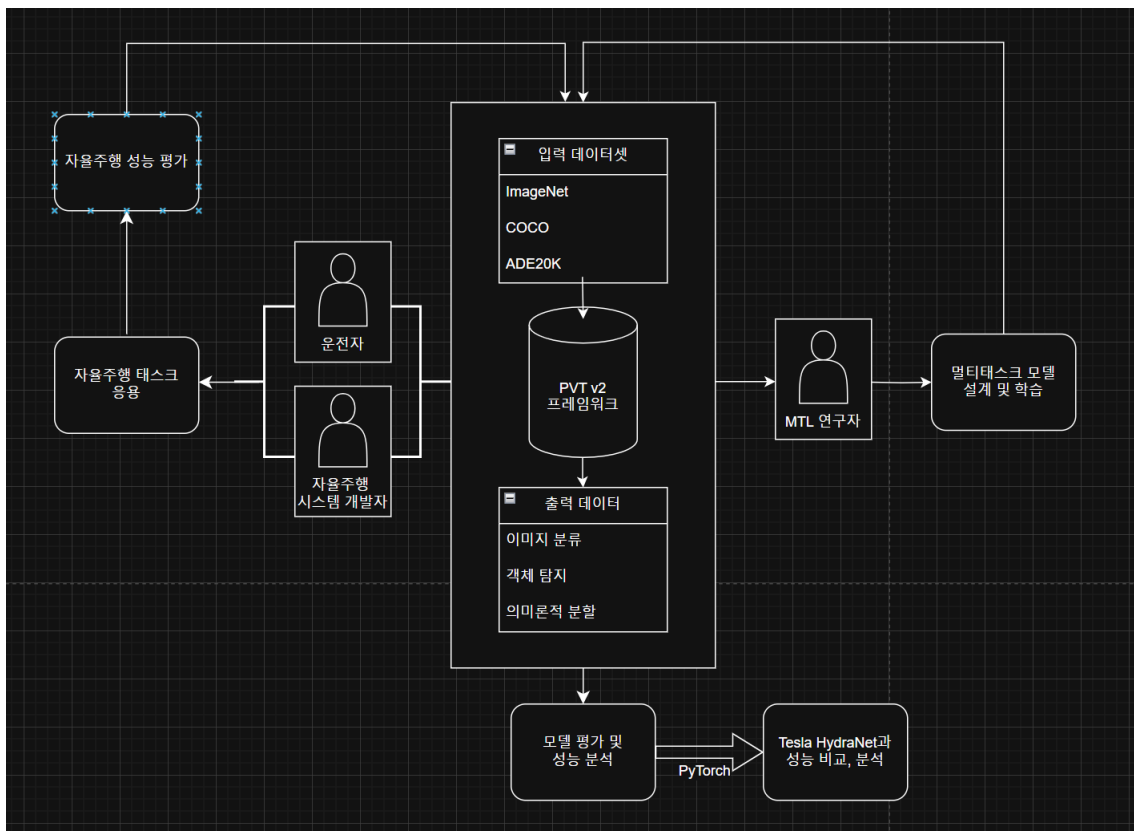


그림 1 소프트웨어 활용 사례 Diagram1

2.2. 문제 해결에 대한 사용 사례 Diagram

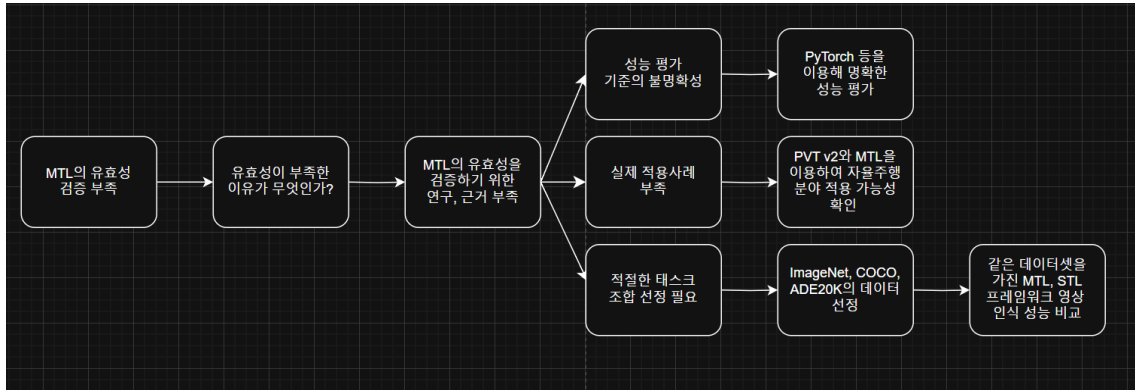
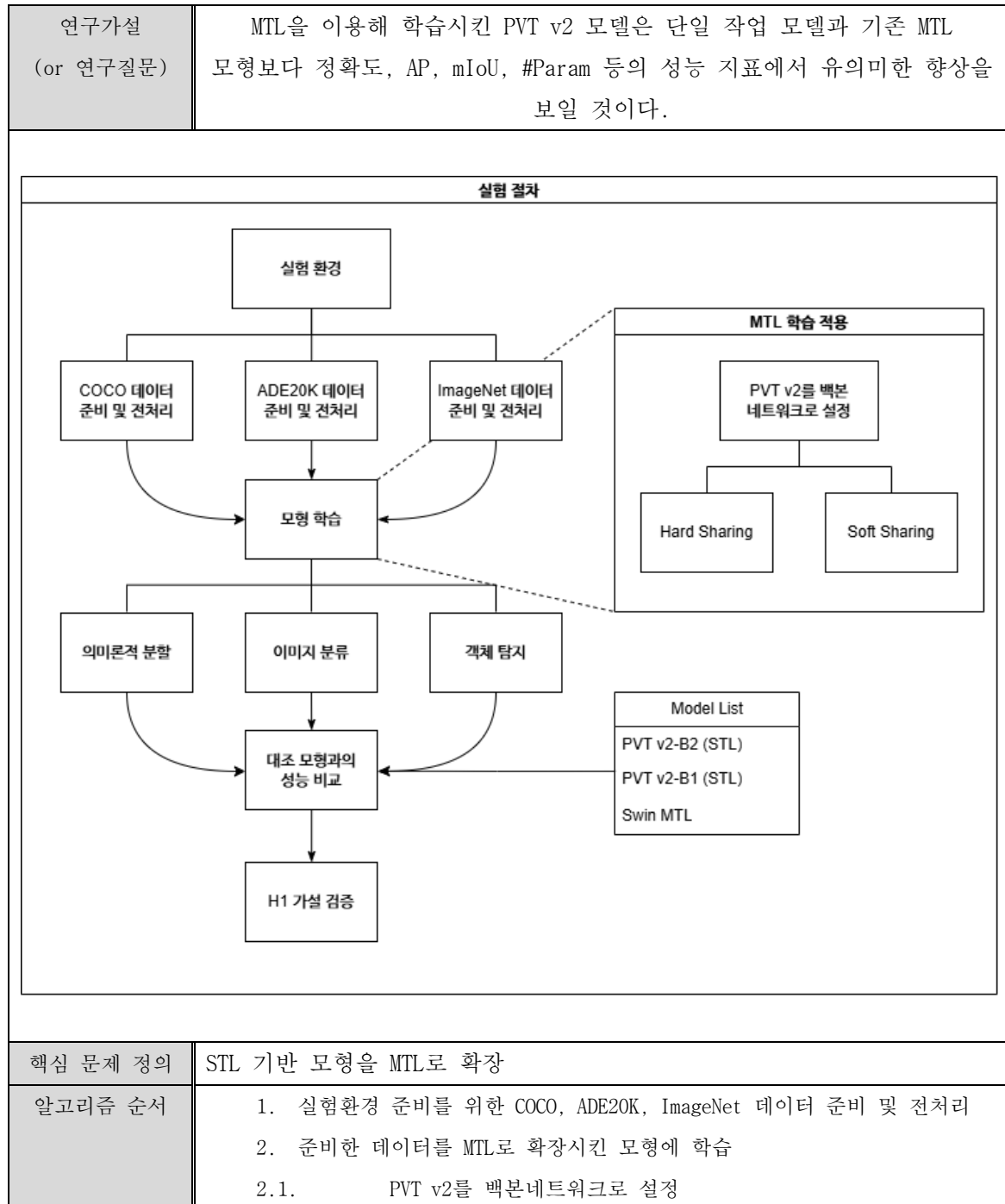


그림 2 문제 해결에 대한 사용 사례 Diagram

3. Sequence Diagram

3.1. 해결 방법에 대한 알고리즘 순서도



	2.2. Hard Sharing 구조로 모델 확장 2.3. Soft Sharing 구조로 모델 확장 3. 의미론적 분할, 객체 탐지, 이미지 분류 세 작업에 대한 성능 확인 4. 대조 모델과의 성능 비교 (PVT v2-B2, PVT v2-B1, Swin MTL) 5. 가설에 대한 검증 진행
--	--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

4. AI 도구 활용 정보

사용 도구	GPT-4.5
사용 목적	논문 탐색
프롬프트	<ul style="list-style-type: none"> ● MTL 기반 Transformer를 연구한 논문들을 찾아줘 ● 기존 STL 기반 모델을 MTL로 확장시킨 연구를 진행한 논문을 찾아줘
반영 위치	1. 연구배경 (p.5) 2. 연구목적 (p.6)
수작업	없음
수정	