

문제정의서(연구개발계획서)

Project Name	Multi-Task Learning을 활용한 PVT v2 프레임워크 성능개선
-----------------	--

08 조

202001156 김수영

202002510 송재현

지도교수: 이종률 교수님 (서명)

Document Revision History

REV#	DATE	AFFECTED SECTION	AUTHOR
1	2025/04/01	1,2 문단 작성	김수영
2	2025/04/02	3,4 ,6문단 작성	김수영
3	2025/04/04	5문단 작성	송재현
4	2025/04/04	최종 검토 및 Reference 작성	김수영, 송재현

Table of Contents

1. 연구 개발의 필요성	5
2. 연구 개발의 목표 및 내용	6
3. 이해당사자 인터뷰/ 설문 인사이트	8
4. 기대 효과 및 향후 확장 가능성	12
5. 연구 개발의 추진전략 및 방법	13
6. AI 도구 활용 정보	14
7. 참고문헌(REFERENCE)	15

List of Figure

그림 1 브레인스토밍	8
그림 2 설문1.....	9
그림 3 설문2.....	10
그림 4 설문3.....	10
그림 5 설문4.....	11

1. 연구 개발의 필요성

Multi-Task Learning(다중 작업 학습, 이하 MTL)은 여러 작업을 동시에 학습시켜 작업 간의 공통점과 차이점을 활용해 최종적으로 각각의 작업들에 대한 성능을 높이려는 기계학습 분야의 패러다임이다. MTL의 주요 특징으로는 다음과 같다.

- MTL은 귀납적 전이(inductive transfer)이다.
- 병렬적으로 학습이 진행된다.¹

MTL은 관련된 작업 간에 표현 혹은 지식을 공유함으로써(shared representation) 원래 작업에서 모델의 일반화 능력을 향상시킬 수 있으며, 이러한 접근 방식이 MTL의 주요 개념이다.

MTL의 주요 장점으로는 “일반화”와 “경량화”이다. 앞서 말하였듯이 MTL은 관련된 작업 간에 공유된 표현을 학습하는 것이 주요 개념임을 말하였다. 또한, MTL은 일반화라는 특징 덕분에 단일 작업 학습 보다 모형 구축 시 필요한 파라미터의 수를 줄일 수 있다.²

AI서비스가 현대인들에게 필수적인 서비스로 자리매김한 결과 많은 연구자들은 모형의 성능 개선을 위해 더 크고 많은 파라미터를 지닌 모형을 개발하고 있는 중이다. 이미 상용화된 GTP나 GROK, DeepSeek와 같은 모형들은 수천에서 수만대의 GPU를 이용해 학습을 시키고 모형을 돌리는 실정이다. 이렇게 중량화된 모형은 많은 에너지를 소비하며 임베디드 시스템과 같은 분야에는 적용하기 어렵다는 한계가 있다. 더군다나 생성되는 데이터의 양도 기하급수적으로 증가함에 따라 광범위한 작업을 수행해야할 일반화된 모형의 필요성 역시 강조되고 있다. MTL은 이러한 AI의 주요 문제점인 일반화와 경량화를 동시에 이룰 수 있다는 장점이 있다.

이러한 장점이 있음에도 불구하고 아직까지 MTL의 개발 실적은 미비한 상황이다. 우선적으로 MTL을 적용하는 것이 무조건 단일 작업 학습보다 더 적은 파라미터로 더 높은 성능을 낼 수 있는 기적의 학습방법이 아니라는 것을 인지하여야한다. MTL을 적용시키기 위해서는 적절한 작업들을 선정해야하고(Grouping) MTL의 목표 함수를 최적화하기 위해 적절한 가중치 방식을 사용하여야 함을 지적한다.³ 이처럼 아직까지 최적화된 MTL을 위한 구체적이고 명시적인 방법론이 제시된 것이 아닌 “적절한” 작업선정과 가중치 함수를 찾기 위해 MTL 연구자들은 여러 방법론들을 개발하고 증명하고 있다. 그렇기에, MTL이 정말 유효한 학습방법인지

¹ Multitask Learning. – Rich Caruana. Machine Learning 28: 41-75

² An Overview of Multi-Task Learning in Deep Neural Networks. – Sebastian Ruder. arXiv:1706.05098v1 [cs.LG] 15 Jun 2017

³ A Comparison of Loss Weighting Strategies for Multi task Learning in Deep Neural Networks. - Ting Gong; Tyler Lee; Cory Stephenson; Venkata Renduchintala; Suchismita Padhy; Anthony Ndirango. IEEE Access Volume 7, p.141627 - 141632

에 대한 검증 역시 주요 연구 대상이다. MTL에 대한 많은 연구들은 MTL에 대한 적절한 Grouping 전략, 학습 전략 개발을 위주로 진행하고 있으나 MTL이라는 학습 방법 그 자체의 유효성에 대한 검증은 부진한 상황이다. 즉, 기존에 잘 사용되는 학습 방법인 단일 작업 학습보다 나은지에 대한 실험적인 근거가 부족한 것이다.

현재 MTL 관련 기술은 PyTorch, TensorFlow 등 주요 프레임워크에서 지원되며, Google Research, Meta AI, Tesla 등 글로벌 빅테크 기업들도 연구를 진행 중이다. 하지만, MTL의 효용성, 적용 가능성, 성능 측정 기준 등에 대한 명확한 가이드라인 부재가 문제로 지적되고 있다. 이러한 현실을 고려할 때, MTL의 유효성을 실험적으로 검증하고, 최적화된 학습 전략을 도출하는 연구개발은 시급히 추진되어야 할 과제라 할 수 있다.

2. 연구 개발의 목표 및 내용

우리 조는 브레인스토밍을 통해 PVT v2라는 프레임워크에 MTL을 접목시켜 해당 프레임워크를 개선하고 최종적으로 MTL의 유효성을 검증할 것이다. 이후 Tesla의 Hydranet과 자율주행 시 필요한 사람인식, 차선 감지 등의 영역에서 비교를 해보며 자율주행 인식 성능을 개선할 것이다.

우리 조의 연구 목표는 크게 두가지이다. 첫번째로는 MTL의 유효성을 검증하는 것이고 두번째로는 MTL을 통해 검증된 모형을 실제 문제에 적용시켜 그 성능 향상을 입증하는 것이다.

첫번째 목표를 실현하기 위해 계획한 것이 바로 PVT v2 프레임워크에 MTL을 접목시키는 것이다. PVT(Pyramid Vision Transformer) v2는 기존 PVT 프레임워크를 개선하여 새로운 비전 Transformer 네트워크를 제시한 것으로, 주요 개선 사항으로는 선형 공간 축소 어텐션(Linear Spatial Reduction Attention, Linear SRA), 겹치는 패치 임베딩(Overlapping Patch Embedding), 합성곱 기반 FFN(Convolutional Feed-Forward Network, CFFN)이 있다. 해당 프레임워크는 이미지 분류(Image Classification), 객체 탐지(Object Detection), 의미론적 분할(Semantic Segmentation)에서 기존에 알려진 모형들보다 더 나은 성능을 보여줄 것을 실험적으로 증명하였다.⁴ 우리는 이 논문에서 제시한 프레임워크를 백본 네트워크로 활용하여 각 작업들을 MTL로 학습시켜 성능향상을 확인해볼 것이다. 논문에서 각각의 작업을 수행할 때 사용한 데이터셋으로는 ImageNet, COCO, ADE20K로 우리 역시 동일한 데이터셋을 활용해 그 성능을 평가할 것이다.

⁴ 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. Computational Visual Media, 16 March 2022, Volume 8, p 415-424

두 번째 목표로, MTL 기반 모델이 실제 자율주행 인식 문제에 적용 가능한지 확인하기 위해 Tesla의 HydraNet을 벤치마킹하고, 도로 인식, 차선 감지, 객체 탐지 등의 과제에서 비교 실험을 수행한다.

프로젝트를 통해 해결하고자 하는 핵심 문제는 다음과 같다.

- MTL의 실질적인 성능 향상에 대한 실험적 근거 부족
- MTL 모델의 실제 응용 가능성에 대한 평가 미비

현재 MTL은 이론적으로 높은 가능성을 가진 학습 구조임에도 불구하고, 구체적인 성능 검증 및 산업 현장 적용 사례가 부족하다. 본 프로젝트는 이와 같은 공백을 채우고, MTL이 실제 문제 해결에 도움이 되는 기술임을 입증하고자 한다.

이를 해결하기 위한 해결 방법은 다음과 같다.

1. PVT v2를 기반으로 MTL 구조를 설계하고 구현
2. 세 가지 작업을 동시에 학습하여 일반화 성능과 효율성 비교
3. 자율주행 분야에서 사용되는 HydraNet과 성능 비교 분석

본 연구의 가치는, 보다 경량화된 고성능 모델을 통해 다양한 비전 작업을 동시에 수행할 수 있는 모델을 제시함으로써, 자율주행과 같은 환경에서의 안전성과 효율성을 높일 수 있다는 점이다. 따라서 MTL을 통해 성능향상이 입증된 기존의 모형을 자율주행분야에 활용하여 그 성능개선을 확인해보는 활동은 MTL의 유효성 입증이라는 학계에서의 측면과 실제 산업에서 요구하는 문제를 동시에 처리할 수 있는 양 측 모두에게 의미있는 결과를 가져다줄 수 있는 연구가 될 것이라 기대한다.



그림 1 브레인스토밍

3. 이해당사자 인터뷰/설문 인사이트

연구주제에 대한 설문조사는 크게 두 가지 관점에 대한 설문을 실시하였고 각 관점 당 두 가지 문항을 넣어서 총 네 문항에 대한 설문을 실시하였다.

첫 번째 관점은 Multi-Task Learning에 대한 인식을 물어보는 것으로 Multi-Task Learning의 상용화 가능성과 기존에 알려져 있는 학습방법인 Single Task Learning에 비해 더 유용한 학습방법일 것이라 생각하는지를 물어보았다.

두 번째 관점은 Multi-Task Learning의 연구 방향 및 필요성 평가에 대한 것이다. Multi-Task Learning의 유효성 검증에 관한 연구의 필요성과 해당 Multi-Task Learning이 향후 AI 성능 발전에 필요한 기술인지 물어보았다.

조사 방법은 Google Form을 이용하여 온라인으로 진행하였고 조원의 지인이나, 학교 커뮤니티에 홍보하는 등의 방식으로 설문을 진행하였으며 총 25명이 응답해주었다. 엄밀하게 표집을 진행한 것은 아니나 과목과 과제의 특성 그리고 시간의 한계와 같은 여러 요인을 고려하

면 이러한 방식이 그나마 단기간 내 많은 응답을 얻어낼 수 있는 방법이라 생각하였다. 25명이라는 응답 수치가 적다면 적은 수치라고 볼 수도 있으나 여건과 기간의 한계를 고려하면 이 정도의 수치만 되어도 설문을 진행한 가장 큰 목적인 대중들의 인식 파악은 달성한 수치라고 본다.

모든 문항들은 1점(매우 그렇지 않다)부터 5점(매우 그렇다)까지의 점수로 평가를 받도록 구성되어있으며 엄밀한 표본 조사 결과 보고서는 아니기에 Google Form에서 제공해주는 그래프를 간단하게 분석하는 방향으로 결과 분석을 수행하였다.

Multi-Task Learning을 활용한 AI 모델 개발이 상용화될 가능성이 있다고 생각하십니까?
응답 25개

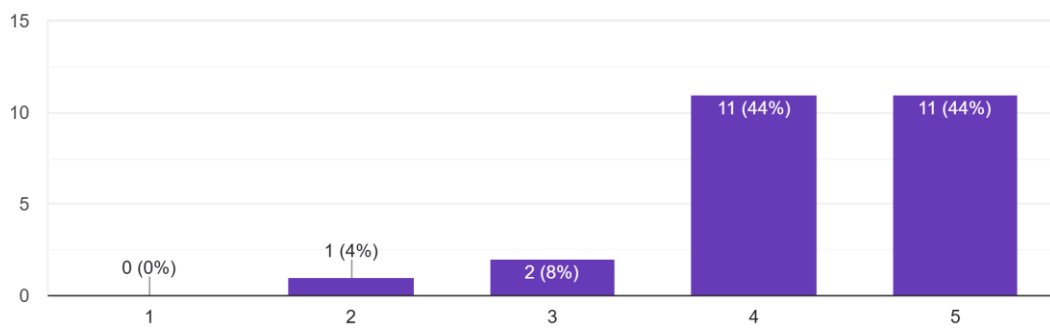


그림 2 설문1

첫 번째 문항으로 Multi-Task Learning의 상용화에 관한 문항이다. 대부분의 사람들이 Multi-Task Learning의 상용화에 긍정적인 답변을 해주었다. 그렇다(4점) 이상인 반응이 전체 응답 중 88%를 차지하고 있으며 부정적인 응답은 단 한 건만 존재하였다. 이는 Multi-Task Learning이라는 방법론을 활용한 모형의 상용화 가능성에 대한 밝은 전망을 나타낸다고 볼 수 있다. 또한, 대중들이 Multi-Task Learning이라는 학습 방법이 적용된 모형에 대해 대중들이 거부감 없이 수용할 준비가 되어 있으며, 실제 서비스로 접했을 때 활용 의사 또한 높을 수 있음을 시사한다.

Multi-Task Learning이 단일 작업 학습(Single Task Learning, 기존에 사용되는 학습 방법)보다 유용할 가능성이 있다고 생각하십니까?

응답 25개

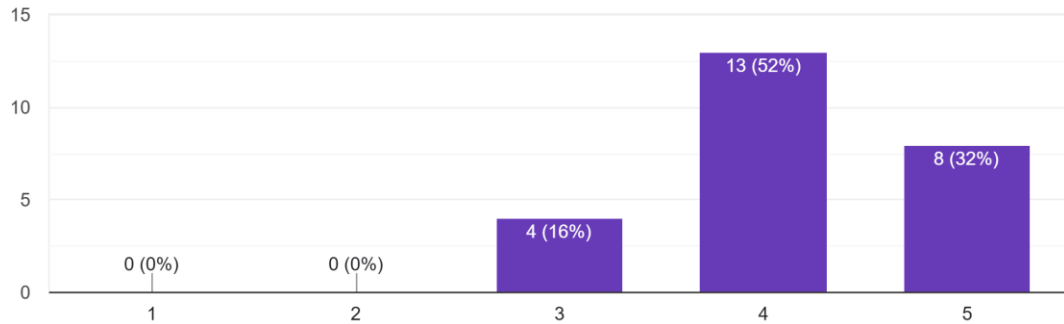


그림 3 설문2

두 번째 문항은 Multi-Task Learning이 단일 작업 학습(Single Task Learning)보다 유용하다고 생각하는지에 대한 문항이다. ‘그렇다(4점)’ 이상의 응답 비율이 84%, ‘보통이다(3점)’ 이상의 응답 비율이 100%로, 부정적인 응답은 전혀 없었다. 이는 Multi-Task Learning에 대한 대중의 인식이 전반적으로 긍정적임을 의미하며, 앞선 상용화 관련 응답과 연계해보았을 때, Multi-Task Learning이 Single Task Learning보다 더 높은 성능을 낼 것이라는 기대가 상용화에 대한 긍정적 평가로 이어졌음을 유추할 수 있다. 결과적으로, Multi-Task Learning은 대중에게 기대를 받고 있는 학습 방식이며, 현실적인 활용 가치가 높은 기술로 인식되고 있음을 알 수 있다.

Multi-Task Learning의 유효성 검증에 관한 연구가 필요하다고 생각하십니까?

응답 25개

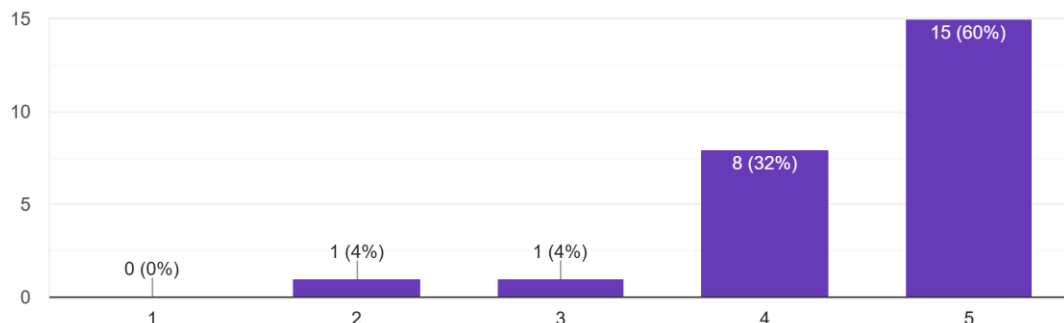


그림 4 설문3

세 번째 문항은 Multi-Task Learning의 유효성 검증 필요성을 묻는 항목이다. 응답자의 60%가 ‘매우 그렇다(5점)’ 라고 응답하였으며, ‘그렇다(4점)’ 이상 응답은 전체 비율 중 92%를 차지한다. 이는 Multi-Task Learning의 효과에 대한 검증이 필요하다는 데 대중들이 높은 공감대를 형성하고 있음을 보여준다. 특히, 이전 문항에서 파악한 Multi-Task Learning이 Single Task Learning보다 나은 성능을 낼 것이라는 기대와 맞물려, MTL의 유효성을 실증적으로 검증하는 과정이 매우 중요하다고 인식됨을 시사한다. 이러한 결과는 우리 연구의 핵심 주제인 “Multi-Task Learning의 유효성 검증”이 시의적절하며 학문적·산업적 가치를 동시에 가질 수 있는 주제임을 뒷받침한다.

Multi-Task Learning 연구가 향후 AI의 성능 발전에 필요한 기술이라고 생각하십니까?

응답 25개

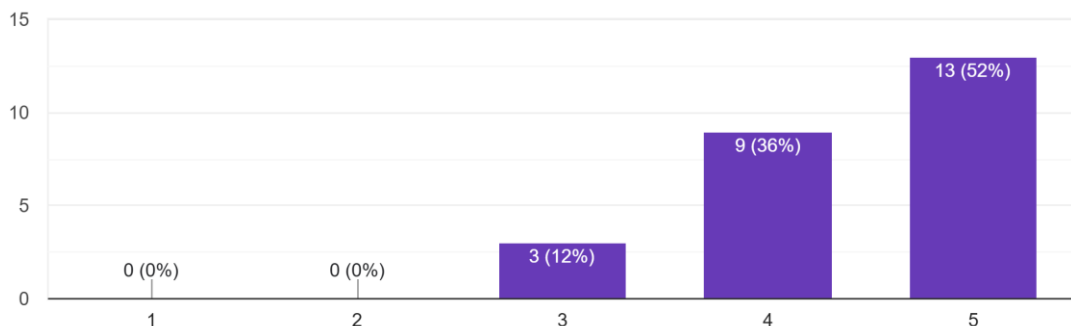


그림 5 설문4

마지막 문항은 Multi-Task Learning 연구가 AI 성능 발전에 필요한 기술인지를 묻는 항목이다. 역시나 ‘그렇다(4점)’ 이상인 비율이 88%를 차지하며 부정적인 응답은 단 한 건도 없음을 확인할 수 있다. 이는 Multi-Task Learning이 단순한 이론적 개념을 넘어, AI 기술의 실질적인 발전에 기여할 수 있는 핵심 기술로 인식되고 있음을 보여준다.

설문조사를 통해 얻은 인사이트로는 Multi-Task Learning의 효과를 입증할 수 있는 유효성 검증에 관한 연구의 필요성이 요구됨을 파악하였으며, 이는 우리 연구 주제의 방향성과 필요성에 필요한 당위성을 더욱 뒷받침해준다. 또한, 이러한 결과는 Multi-Task Learning을 주제로 한 본 연구가 기술적 유의미성과 사회적 수요를 모두 충족시키는 주제임을 다시금 확인시켜주는 계기가 되었다.

4. 기대 효과 및 향후 확장 가능성

이번 프로젝트를 통해 MTL의 유효성이 실험적으로 입증되고, PVT v2 기반 MTL 모델이 자율주행 인식 성능 개선에 기여할 수 있다는 결과가 도출될 경우, 다음과 같은 효과가 기대된다.

첫째, 사용자 관점에서 자율주행 시스템의 인식 정확도 향상과 오류 감소를 통해 안전성이 증대될 수 있다. 예를 들어, 차선 인식 정확도 및 객체 탐지 정밀도가 각각 5% 이상 향상될 경우, 실제 도로 주행 중 사고 위험을 실질적으로 줄일 수 있는 기반이 마련된다. 또한, 동일한 모델로 복수의 인식 작업을 동시에 수행할 수 있어 처리 시간 단축과 시스템 자원 절약 측면에서도 이점이 생길것이라 기대된다.

둘째, 사회적 관점에서는 자율주행차의 상용화 및 보급 확대를 앞당길 수 있는 기술적 신뢰 기반이 구축된다. 이는 교통사고 감소, 고령자·장애인 이동 편의성 향상, 물류·운송 산업의 자동화 등 다양한 사회적 이익으로 환원될 수 있다.

셋째, 산업계에서는 경량화된 고성능 AI 모델을 통해 자율주행뿐만 아니라 CCTV, 드론, 스마트 팩토리 등 다양한 영상 인식 기반 산업 분야로의 기술 확장이 가능하다. 특히 PVT v2와 같은 범용 백본 모델 위에 MTL 구조를 적용하는 방식은 다양한 산업용 AI 모델 설계 시 재사용성과 확장성을 높이는 데 기여할 수 있다.

추후에는 본 프로젝트에서 사용한 프레임워크를 기반으로, 의료 영상 진단, 재난 감지 시스템, 로봇 비전 등 고신뢰 인식이 필요한 다른 분야로도 응용 범위를 넓혀갈 수 있다. 또한, 작업 간 상호작용 분석을 기반으로 한 동적 가중치 조절 방법, 효율적 학습 전략 설계 등 후속 연구 주제도 다양하게 제시될 수 있다.

그러나 본 프로젝트의 1차 목표는 MTL의 유효성 자체를 검증하는 데에 있으며, 만약 기대한 수준의 성능 향상이 관찰되지 않는다면, 이는 오히려 MTL 구조 내에서의 그룹핑 전략 또는 학습 전략 설계가 어떻게 성능에 영향을 미치는지를 탐색할 수 있는 기회가 될 수 있으며 이러한 방향으로 후속연구를 진행할 수도 있다. 또한 해당 결과를 후속 연구에서 보다 효과적인 MTL 설계 방안을 모색하는 데 기초 자료로 활용할 수도 있다.

또한, 자율주행 분야에서 원하는 수준의 성능 개선이 나타나지 않을 경우, PVT v2 프레임워크가 자율주행 환경에 적합하지 않은 이유를 분석하는 방향으로 연구를 확장할 수 있다. 이는 모델 구조와 실제 응용 환경 간의 적합성 문제를 규명하는 의미 있는 접근이 될 수 있으며, 궁극적으로 보다 적합한 백본 구조나 학습 방식에 대한 제안으로 이어질 수 있다.

결론적으로, 본 연구는 MTL의 실용 가능성과 실제 적용 효과를 동시에 검증함으로써, 학술적 기여는 물론 산업적·사회적으로 의미 있는 성과를 도출할 수 있을 것이다.

5. 연구 개발의 추진전략 및 방법

본 프로젝트의 최종 목표는 다양한 시각 인식 작업(이미지 분류, 객체 검출, 장면 분할)을 하나의 통합 모델로 학습하는 멀티태스크 학습(MTL) 프레임워크를 개발하고, PVT v2(Pyramid Vision Transformer v2)를 백본으로 사용하여 기존 PVT v2 모델보다 효율적이거나, 높은 정확도를 달성하는 것이다.

추진 일정

기간	주요 내용	세부 내용
5~6주차	데이터셋 조사 및 로드	ImageNet, COCO, ADE20K의 특성 파악 및 다운로드, 라이선스 확인, 분류/검출/세분화 용도 파악
7~8주차	데이터 전처리(증강 및 정규화)	데이터 증강(Augmentation), 정규화(Normalization) 전략 수립 및 파이프라인 구성
9~10주차	PVT v2 백본 설계, 구현	PVT v2 구조 분석 및 구현
11주차	Task-specific Decoder 설계	분류, 객체 검출, 세분화를 위한 Decoder 설계 및 멀티태스크 학습 구조 설계
12~13주차	MTL 학습 수행	Hard sharing을 이용한 MTL 학습 수행
14주차	성능 평가 및 테스트	각 Task 별 성능 평가(metric: Accuracy, mAP, mIoU 등), 기존 방식과 비교
15주차	결과 분석, 결론 도출	성능 개선 요인 분석, 한계점 파악 및 후속 연구 제안

팀 구성

이름	담당 역할
김수영	프로젝트 전체 기획, PVT v2 백본 구현, 실험 설계, 성능 분석, 결론 도출
송재현	데이터셋 수집, 데이터 증강 전략 설계 및 구현, Task-specific Decoder 구현, 성능 테스트

정량적, 정성적 목표

항목	목표
학회 발표	KCC, KAAI, KDM등의 학회 발표
산출물	MTL 모델 프로토타입 1개, 실험 보고서 1부
피드백 횟수	외부 피드백 최소 2회
서비스화 가능성	자율주행 차선 인식 시나리오 실험 구현

6.AI 도구 활용 정보

사용 도구 GPT-4o	
사용 목적	연구 계획의 방향성에 대한 질문
프롬프트	<ul style="list-style-type: none"> ● 내가 지금 Multi-Task Learning에 대한 주제를 바탕으로 졸업프로젝트를 진행중이야. 학부생 수준에서 10주 안에 수행해야하기에 굳이 어렵게 하지 않으려 해. 현재 구체적인 연구 계획은 정해지지 않았고 연구 방향성은 Multi-Task Learning의 유효성 검증에 관한 것을 바탕으로 할 예정이야. 그래서 연구를 어떻게 진행해야할지 구체적인 문제 정의를 하고 싶어. 일단 내가 원하는 방향성은 기존에 알려져 있는 모형들을 Multi-Task Learning으로 학습시켜 본 후 성능향상이 실제로 이루어지는지, Multi-Task Learning의 실질적인 유효성을 검증을 보여주는 것이야. 이러한 방향성을 바탕으로 연구 목표를 더 구체화해줘. 예를 들어 최신 모형을 바탕으로 Multi-Task Learning을 수행한다 할 시 써먹을 논문을 구체적으로 정한다거나, 실제 이용되고 있는 AI서비스의 모형들을 Multi-Task Learning으로 수행해보는 등 말이야. ● 1. 컴퓨터비전과 관련된 태스크를 염두에 두고 있어 2. 기본적인 성능 기준은 정확도(Accuracy)야 하지만, 만약 성능이 비슷하거나 약간 अच्छ게 나온다고 하여도 파라미터의 수가 유의미하게 줄었으면 그것 또한 경량화가 잘 됐다는 증거로써 성능에 포함시킬 예정이야. 3. 컴퓨팅자원과 데이터셋에 대한 제약은 없어 4. 잘 알려진 모델을 사용하는 것이 좋겠지? 실제 상용화된 모형이 Multi-Task Learning으로 학습될 경우 그에 대한 성능이 실제로 향상되는지 확인하는 것이 목적이니까
반영 위치	1. 연구 개발의 필요성 (p.5) 2. 연구 개발의 목표 및 내용 (p.6)
수작업 수정	있음 (사례 교체, 논리 보강 및 수정, 내용 및 자료 추가)

7. 참고문헌(Reference)

Multitask Learning. – Rich Caruana. Machine Learning 28: 41-75

An Overview of Multi-Task Learning in Deep Neural Networks. – Sebastian Ruder. arXiv:1706.05098v1 [cs.LG] 15 Jun 2017

A Comparison of Loss Weighting Strategies for Multi task Learning in Deep Neural Networks. - Ting Gong; Tyler Lee; Cory Stephenson; Venkata Renduchintala; Suchismita Padhy; Anthony Ndirango. IEEE Access Volume 7, p.141627 – 141632

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. Computational Visual Media, 16 March 2022, Volume 8, p 415-424