

# Project Brainstorming Result

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

8조

202001156 김수영

202002510 송재현

지도교수: 이종률 교수님 (서명)

# Document Revision History

---

Rev#	DATE	AFFECTED SECTION	AUTHOR
1	2025/03/26	2, 3, 5번 문단 작성; Flow Chart 그림 추가	송재현
2	2025/03/28	1번 문단 작성, 2, 5번 문단 내용 추가; Mind Map 그림 추가	김수영
3	2025/03/29	최종 검토	김수영, 송재현

# Contents

---

1. 문제 정의.....	4
2. (문제 해결을 위한) 아이디어 발산.....	7
3. 아이디어 수렴.....	8
4. 시각화하기.....	9
5. AI 도구 활용 정보.....	10

# List of Figure

그림 1 아이디어 발산 회의 장면 캡처 .....	8
그림 2 연구 일정에 대한 FLOW CHART .....	9
그림 3 문제 정의 MIND MAP .....	9
표 1 아이디어 수렴 .....	8
표 2 AI 도구 활용 정보.....	10

## 1. 문제 정의

지난 발표 이후 우리 조의 기존 문제로 정의하였던 ‘Multi-Task Learning의 최적 공유 전략 연구 확장 및 실험적 유효성 분석’이라는 프로젝트 주제를 좀 더 구체화할 필요가 있다고 생각하였다. 우리 조의 연구 방향성은 Multi-Task Learning의 유효성 검증이 목표이므로 이러한 방향성을 바탕으로 구체적인 문제 정의를 다시 하기로 결정하였다. 문제 후보 선정 및 해당 문제에 대한 아이디어 수렴, 발산 관련 브레인스토밍을 하였다. 브레인스토밍을 통한 후보 문제는 총 4개가 나왔다

- 1) MTL을 활용하여 성능을 올린 논문 재현
  - [YOLOP: You Only Look Once for Panoptic Driving Perception](#) 논문에서 구체적으로 언급되지 않은 기존 STL모형을 직접 구현해보고 실제로 MTL이 STL에 비하여 성능 향상이 이루어지는지 논문 자체를 재현해본다.
- 2) 성능 향상이 입증된 모형을 바탕으로 MTL을 적용시켜 유효성을 검증
  - [PVT v2: improved baselines with pyramid vision transformer](#) 논문에서 제시한 프레임워크에 MTL을 적용시켜 성능 향상이 이루어지는지 확인해보고 MTL의 유효성을 검증해본다.
- 3) 기존에 알려진 모형을 MTL로 학습하여 MTL의 일반적인 성능이 높음을 검증
  - [Multi-task CNN Model for Attribute Prediction](#) 논문처럼 기존에 잘 알려진 모형(ResNet, MobileNet etc.)을 백본 네트워크로 활용하여 MTL이 어느 특정 모형에만 국한해서 성능 향상이 이루어지는 것이 아닌 일반적으로 성능 향상이 이루어짐을 확인해본다.

4) MTL에서 Task Sampling 방법 연구를 통한 성능 개선 방안 탐색

- [Efficiently Identifying Task Groupings for Multi-Task Learning](#) 논문에서 영감을 받아 MTL의 Task Sampling 방안을 연구하여 전체적인 성능을 높이는 방법을 탐색해본다.

지도 교수님과의 미팅 및 팀 내 회의를 통해 최종적으로 2번 문제를 우리 조의 연구 목표로 정하였다.

우선, 1번 문제는 논문 재구현이 최종 목표이기에 실험 결과를 해석하는데 있어 제한이 가해진다는 단점이 있다. 해당 논문에서는 MTL과 STL 간 성능 차이에 대한 단락이 굉장히 짧게 서술되어있다. 그렇기에 STL을 직접 구현해보고 정말 논문대로 결과가 나오는지 확인해보려 한 의도였다. 하지만 여기서 문제가 발생하는데 의도대로 논문 결과와 일치하는 - *MTL이 STL보다 성능이 좋다* 라는 결과가 나오면 그저 논문을 따라 읽고 구현한 것 외 연구라고 할만한 활동을 하지 않은 것이 되고, 논문 결과와 불일치하는 - *MTL이 STL보다 성능이 안좋다* 라는 결과가 나오면 이에 대한 추가적인 연구를 진행해야하기에 1년짜리 졸업프로젝트로는 적절하지만 종합설계1의 목적에는 맞지 않는다는 결론을 냈다. 물론 MTL이 더 나쁘더라는 결론이 나오더라도 문제 정의부터 실험까지 진행하는 프로세스를 밟아보는 종합설계1의 목적에는 충분히 부합하다고 생각한다. 그러나 ‘성능이 좋을 경우’에 대한 경우는 결국 기존 논문의 주장과 일치하는 결과를 단순히 확인하는 수준에 그치게 된다. 따라서 MTL과 STL의 성능 차이를 단순히 재현하는 것에 머무르기보다는, 성능이 좋은 나쁜 그 원인을 분석하고 확장 가능한 방향성을 탐색하는 것이 더 의미 있는 연구가 될 수 있고 이러한 방식의 연구는 10주 내 해결할 수 있기에는 힘들다 판단하였기에 제외하기로 하였다.

4번 문제는 MTL의 Sampling 기법을 탐색 및 연구해보자는 취지로 제시한 문제이긴 하나 우리 조가 앞서 제시했던 것과 마찬가지로 구체적이지 않고 문제 해결이라는 취지 보다는 순수한 연구 성격이 강한 문제이기에 역시 종합설계1의 목적에는 맞지 않는다는 결론을 냈다. 또한, MTL의 유효성 검증이라는 우리의 연구 방향성과도 다소 어긋나 있는 문제이기에 적절하지 않다 판단하였다.

3번 문제의 경우, 인용한 논문 자체가 2016년에 출간된 다소 오래된 논문이고 사실상 문제 자체가 2번 문제를 좀 더 포괄적으로 표현한 것이기에 배제하였다.

2번 문제에 인용된 논문에서는 PVT라는 모형을 개선한 PVT v2 모형에 대한 개선 방향점과 그에 대한 성능 향상을 실험적으로 확인한다. 개선된 모형에 이미지 분류, 객체 탐지, 그리고 의미론적 분할이라는 작업을 따로따로 수행하여 그 성능을 평가한다. 이 점에 착안하여

그렇다면 똑같은 모형에 다른 작업들을 따로 학습시키고 성능을 평가하는 것이 아닌 MTL을 통해 전부 학습하는 것이 성능 개선에 더 도움이 되지 않을까? 라는 아이디어를 내게 되었다. 대상 논문과 모형, 수행할 작업들, 작업에 필요한 데이터셋이 모두 마련되어 있기고, 실험 역시 기존에 알려져 있는 모형을 MTL로 학습시켜 성능 평가를 진행하면 되기에 종합설계1 수업에서 진행할 프로젝트로 활용하기에 무리가 없을 것이라 판단하였다. 또한, MTL에 대한 유효성 검증이라는 면에서도 적절하다 판단하였다. 또한, 비록 실험 결과가 원하는 대로 나오지 않았다 하더라도 이러한 결과는 1번 문제에서 언급하였듯이 이러한 절차를 밟는 것 자체를 의의로 둘 수 있다. 즉, 2번 문제는 1번 문제의 문제점을 개선한 주제이기에 지금까지 설정한 문제들 중 가장 적절한 문제라고 판단하여 최종 프로젝트 주제를 2번 문제 해결로 설정한 것이다.

## 2. (문제 해결을 위한) 아이디어 발산

본 연구에서는 Pyramid Vision Transformer v2(PVT v2)를 기반으로 Multi-Task Learning(MTL)의 유효성을 실험적으로 검증하고자 한다. PVT v2는 기존 Pyramid Vision Transformer v1(PVT v1) 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. PVT v2는 기존 모델 대비 연산 효율성이 높으며, 다양한 컴퓨터 비전 작업에서 우수한 성능을 보인다. 해당 모형을 활용하여 이미지 분류, 객체 탐지, 그리고 의미론적 분할(Semantic Segmentation)을 데이터셋으로 MTL을 수행하고, MTL을 사용하지 않는 기존 모형과의 성능 비교, MTL의 성능 향상을 검증한다.

### 연구 키워드 정리

- Multi-Task Learning
- Pyramid Vision Transformer v2: 연구에서 사용할 백본 모형
- Image Classification: MTL 학습에 사용할 작업(Task) - 관련 데이터셋: ImageNet
- Object Detection: MTL 학습에 사용할 작업(Task) - 관련 데이터셋: COCO
- Semantic Segmentation: MTL 학습에 사용할 작업(Task) - 관련 데이터셋: ADE20K

### 주요 이론 및 연구 변수

- 독립변수(IV)
  - 1) Backbone의 공유 전략: 완전 공유, 부분 공유, 개별 Backbone 사용
  - 2) Decoder의 적용 방식: Task-Specific Decoder
- 종속변수(DV)
  - 1) 이미지 분류 성능
  - 2) 객체 탐지 성능
  - 3) 의미론적 분할 성능

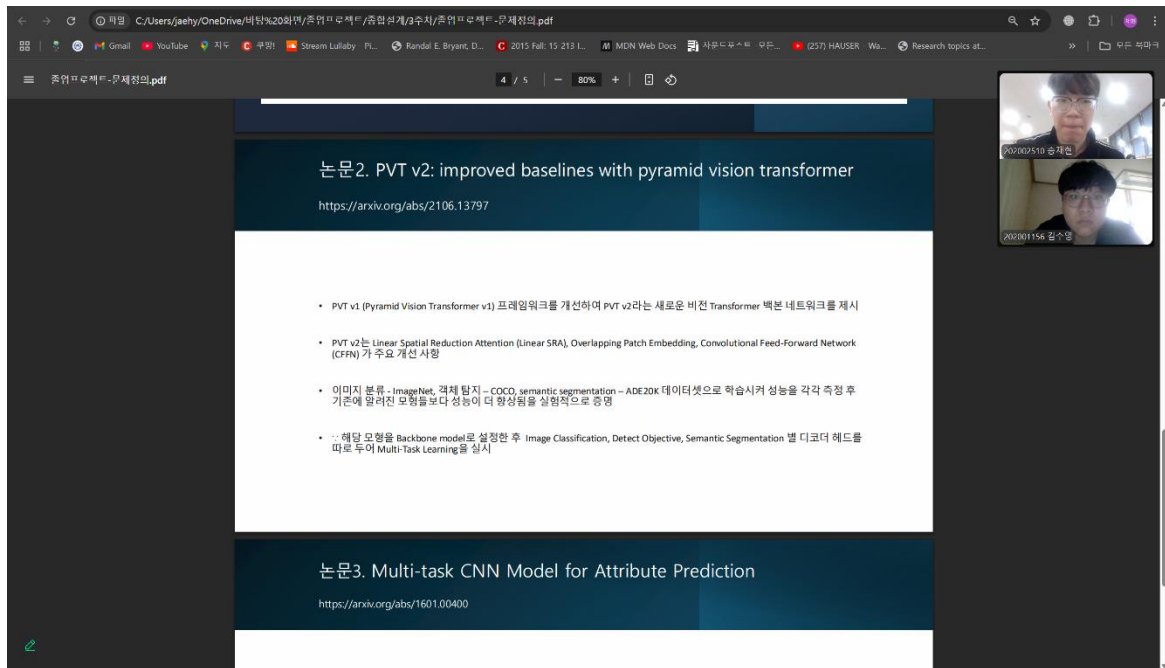


그림 1 아이디어 발산 회의 장면 캡처

### 3. 아이디어 수렴

핵심 개념	정의 및 하위 개념 목록
Multi-Task Learning	1. 하나의 모델이 여러 개의 작업을 동시에 처리할 수 있게 수행하는 학습 방법 2. 하나의 모델이 한 개의 작업을 수행하는 Single-Task Learning과 비교
PVT v2	본 연구에서 백본으로 설정하여 MTL 학습 작업을 수행할 모델 구조
데이터셋	본 연구에서 MTL 학습을 수행할 데이터로, ImageNet(이미지 분류), COCO(객체 탐지), ADE20K (의미론적 분할)이 있다
성능 평가 방법	모델의 정확도, 연산의 효율성, 연산량, 파라미터의 개수
Task-Specific Decoders	각 작업에 특화된 Decoder을 활용, 작업 별 최적의 성능 도출 목적

표 1 아이디어 수렴



## 4. 시각화하기

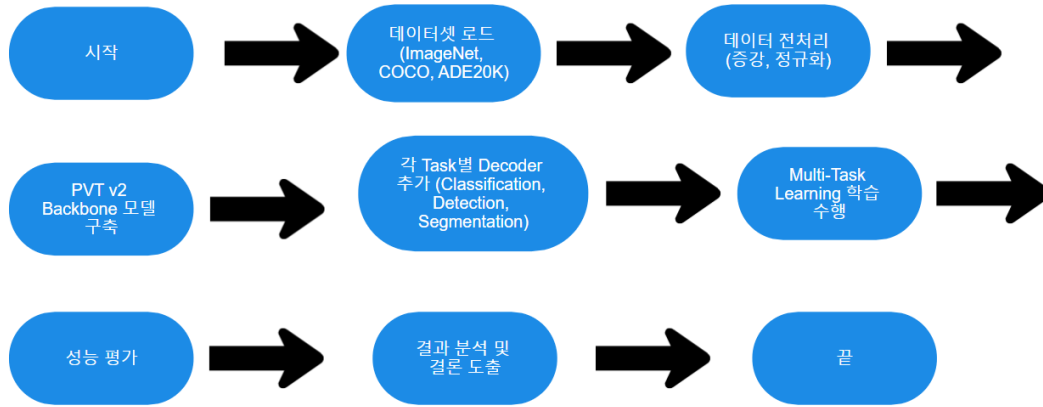


그림 2 연구 일정에 대한 Flow chart

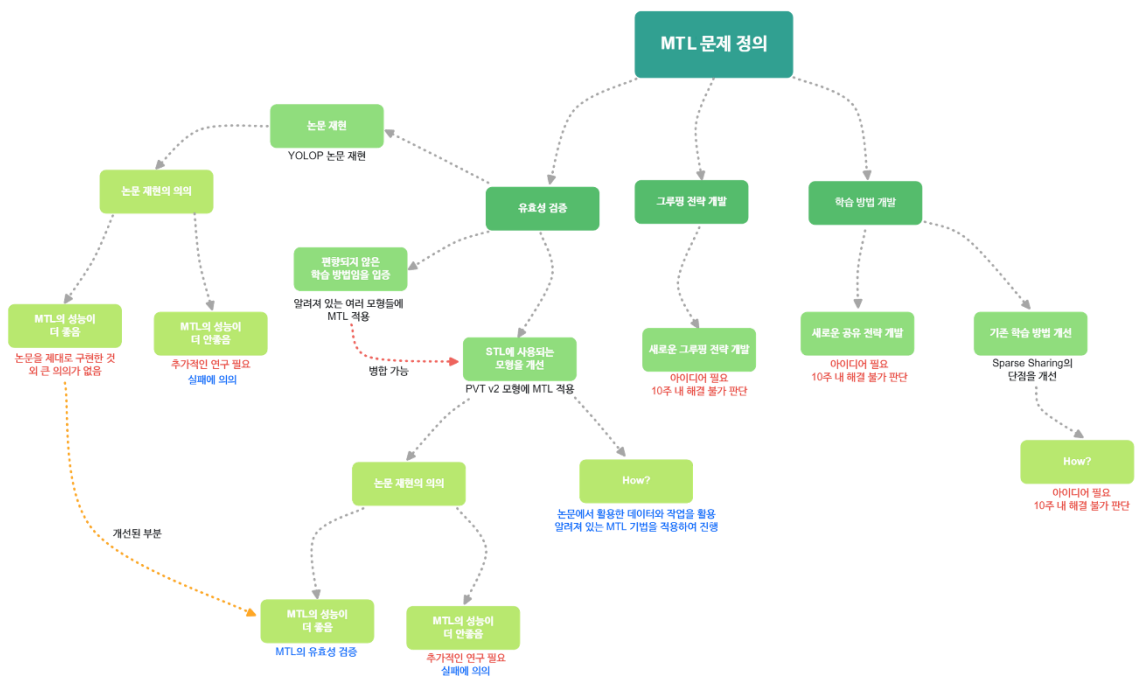


그림 3 문제 정의 Mind Map

## 5.AI 도구 활용 정보

사용 도구	GPT-4o
사용 목적	연구 계획의 방향성에 대한 질문
프롬프트	<ul style="list-style-type: none"> <li>● 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 씨먹을 논문을 구체적으로 정한다거나, 실제 이용되고 있는 AI서비스의 모형들을 Multi-Task Learning으로 수행해보는 등 말이야.</li> <li>● 1. 컴퓨터비전과 관련된 태스크를 염두에 두고 있어 2. 기본적인 성능 기준은 정확도(Accuracy)야 하지만, 만약 성능이 비슷하거나 약간 안좋게 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</li> </ul>
반영 위치	1. 문제 정의 (p.4) 2. 아이디어 발산 (p.6) 3. 시각화하기 (p.9)
수작업 수정	있음 (사례 교체, 논리 보강 및 수정, 내용 및 자료 추가)

표 2 AI 도구 활용 정보