

---

# COSE474-2024F: Final Project

## “Does a Good Beginning Really Guarantee a Good Ending?”

---

Yurim Lee / 2020390472 / Computer Science

### 1. Introduction

#### 1.1. Motivation & Problem Definitions

본 연구는 DAPT(Distribution-Aware Prompt Tuning) (Helber et al., 2018) 모델에서 발견된 비일반적인 학습 성능을 바탕으로 시작되었습니다. DAPT 모델은 EuroSAT(Helber et al., 2018)과 Food101(Kaur et al., 2017) 데이터셋에서 다른 데이터셋들에서와는 다른 학습 성능을 보였습니다. 특히, Food101에서는 1-shot에 비해 2-shot 성능이 감소하는 역설적인 결과가 나타났습니다. 이는 일반적으로 shot 수가 증가함에 따라 성능이 증가하는 다른 데이터셋의 실험 결과와 상반되며, 단순히 샘플 수를 늘린다고 해서 모델 성능이 선형적으로 개선되지 않음을 시사합니다. EuroSAT 데이터셋에서는 2-shot과 4-shot 간의 성능 차이가 거의 없었습니다. 이 또한 추가적인 샘플이 모델 성능 향상에 기여하지 못하고 있음을 나타냅니다. 이러한 현상을 바탕으로, 두 데이터셋에서 모델이 비정상적인 패턴을 보이는 이유를 탐구하고, Food101의 2-shot과 EuroSAT의 4-shot에서 성능을 높일 수 있는 방법을 모색하기로 하였습니다. 이를 위해 RPO(Read-only Prompt Optimization)(Lee et al., 2023)에서 제안된 Initialization 기법을 DAPT 프레임워크에 적용하고 프롬프트 토큰의 초기 설정이 모델 성능에 미치는 영향을 심층적으로 분석하고자 했습니다. 이러한 작업을 통하여 궁극적으로는 극소수 샘플 환경에서의 모델 성능 안정성을 높이고자 합니다.

구체적으로, Text Prompt에서 RPO의 ST-initialization을 활용하여 random token initialization, same ST-token initialization, different ST-token initialization라는 세 가지 방법론을 적용하였습니다. Same ST-token Initialization은 CLIP 모델의 EOS 토큰을 기반으로 하여 모든 프롬프트 토큰을 동일한 값으로 초기화합니다. Different ST-token Initialization은 CLIP 모델의 EOS 토큰을 기반으로 하며, 각 토큰에 서로 다른 노이즈 패턴을 적용하여 토큰 다양성을 제공합니다. 두 방법론 간의 비교를 통해 토큰 간 차별성이 없을 때 모델 성능에 미치는 영향을 평가할 수 있습니다.

이 연구의 Contribution은 다음과 같습니다:

- 극소량 데이터 환경에서 텍스트 프롬프트 토큰 초기화 전략의 성능을 체계적으로 분석하여,

DAPT(Domain Adaptive Pre-training) 모델의 불안정성 문제에 대한 새로운 해결 접근법 제시

- RPO 초기화 접근법을 활용하여, Random Initialization, Same ST-token Initialization, Different ST-token Initialization 등 세 가지 초기화 방법론의 상세한 비교 및 성능 평가
- 프롬프트 토큰의 초기화 방식이 모델 성능에 미치는 영향에 대한 심층적 연구를 통해, 토큰 간 차별성과 다양성이 극소수 샘플 학습 환경에서 미치는 영향에 대한 실증적 분석

### 2. Methods

#### 2.1. Figure

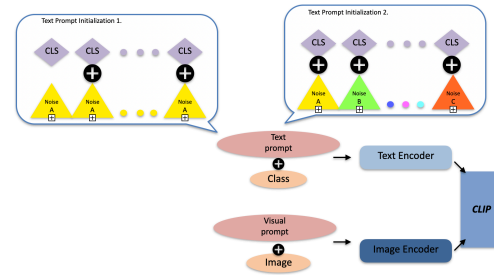


Figure 1. Initialization Method

#### 2.2. Novelty

연구에서는 프롬프트 토큰 Initialization이 모델 성능에 미치는 영향을 명시적이고 체계적으로 실험하였습니다. 구체적으로, random token initialization, same ST-token initialization, different ST-token initialization이라는 세 가지 방법을 통해 시작 토큰에 약간씩 변화를 주었습니다. 세 가지 초기화 전략을 심층적으로 분석하여 few-shot 학습에서 프롬프트 토큰 설정의 미세한 변화가 모델 성능에 미치는 영향을 탐구하였습니다. 이를 통해 프롬프트 학습의 메커니즘을 더욱 깊이 이해하고, DAPT 모델의 성능을 개선할 수 있는 resource efficient한 새로운 학습법을 제시하는 데 의의를 두고 있습니다

### 2.3. Algorithm

---

#### Algorithm 1 DAPT with 3 Different Initialization

---

Step 1: Initialization prompts

Method 1: CLIP’s EOS Token with Uniform Noise

for  $i = 1$  to  $K$  do

$p_t[i] \leftarrow CLIP\_EOS\_token + \mathcal{N}(0, \sigma^2 I)$

Method 2: CLIP’s EOS Token with Distinct Noise

for  $i = 1$  to  $K$  do

$p_t[i] \sim \mathcal{N}(CLIP\_EOS\_token, \sigma^2 I)$

Method 3: Random Initialization

for  $i = 1$  to  $K$  do

$p_v[i] \sim \mathcal{N}(0, \sigma^2 I)$   
     $p_t[i] \sim \mathcal{N}(0, \sigma^2 I)$

Step 2: Feature extraction

$z_i \leftarrow f(x_i)$

$s_c \leftarrow \frac{1}{N} \sum_{(x_i, y_i) \in D_c} z_i, \forall c$

$\tilde{z}_i \leftarrow f(q_i)$

$\tilde{w}_j \leftarrow g(p_j)$

Step 3: Loss calculation for each data point

for  $(x_i, y_i) \in D$  do

$L_{CLIP} \leftarrow -\frac{1}{B} \sum_{j=1}^B \log \frac{\exp(\tilde{z}_i^\top \tilde{w}_{y_i} / \tau)}{\sum_{j=1}^C \exp(\tilde{z}_i^\top \tilde{w}_j / \tau)}$

$L_{inter} \leftarrow -\frac{1}{B} \sum_{m \neq n} \exp(-t \|\tilde{w}_m - \tilde{w}_n\|_2^2)$

$L_{intra} \leftarrow c_i[y_i = c] \|\tilde{z}_i - s_c\|_2^2$

$L \leftarrow L_{CLIP} + \beta_t L_{inter} + \beta_v L_{intra}$

$L.backward()$

Step 4: Update prompts

$\tilde{z}.update()$

$\tilde{w}.update()$

---

Algorithm Based on DAPT (Distribution-Aware Prompt Tuning)(Cho et al., 2023)

### 3. Experiments

모든 실험은 Initialization 부분을 제외하고는 DAPT 논문과 동일한 설정을 따르며, 1, 2, 4, 8, 16 shot 실험을 진행했습니다. 각 실험은 각각 50, 100, 100, 200, 200 epoch로 학습되었습니다. Text Prompt 초기화 방법으로는 RPO에서 ST-Initialization을 기반으로 변형하여 실험하였습니다.

### 3.1. Data & Resource

연구의 실험 설계는 Initialization 차이가 모델 성능에 미치는 직간접적인 영향을 탐구하고, 다양한 few-shot 샘플 시나리오에서의 성능 개선을 검증하는 데 중점을 두었습니다. 특히 EuroSAT과 Oxford Pets 두 데이터셋을 통해 기존 DAPT 모델과의 결과를 비교하여 새롭게 제안된 방법론의 robustness를 평가하고자 했습니다. 이러한 접근은 자원이 부족한 환경에서 학습이 잘 이루어지지 않을 때, 초기 설정 변화를 통해 모델 성능을 개선하려는 고민에서 비롯되었습니다. EuroSAT 데이터셋은 위성 이미지 분류를 위한 대표적인 벤치마크로, 다양한 토지와 지표 유형을 포함하고 있습니다. DAPT 모델이 2-shot과 4-shot 간 성능 차이를 보이지 않아 실험 대상으로 선정하였으며, extremely few shot 학습 환경에서의 모델 성능 평가에 적합한 데이터셋입니다. 연구에서는 특히 4-shot에서의 성능 향상을 목표로 하였습니다. Oxford Pets(Parkhi et al., 2012) 데이터셋은 반려동물 이미지 분류를 위한 데이터셋으로, 다양한 품종의 고양이와 개 이미지를 포함하고 있습니다. 이 데이터셋은 기존 DAPT 모델이 비교적 안정적인 성능을 보이기에, 새롭게 제안된 방법론의 기존 성능 유지를 검증하는 데 중요한 역할을 합니다. 두 데이터셋을 선택함으로써 제안된 프롬프트 학습 방법의 generalizability와 robustness를 종합적으로 평가하고자 하였습니다. 실험에 사용된 컴퓨팅 자원은 Google Colab Pro에서 T4 GPU와 PyTorch 프레임워크를 활용하였습니다.

### 3.2. Analysis

#### 3.2.1. Analysis 1

Table 1을 보면 EuroSAT 데이터셋 실험에서 16-shot 및 8-shot과 같이 충분한 샘플이 제공되는 경우 EOS 토큰을 사용하는 Initialization 방식이 오히려 성능 저하를 초래하는 현상이 관찰되었습니다. 4-shot에서 EOS 토큰에 각각 다른 노이즈를 추가한 방식은 random token initialization에 비해 3.9의 성능 향상을 보였지만, 여전히 2-shot 성능을 능가하지는 못했습니다. 반면, 1-shot 설정에서는 EOS를 사용한 same ST-token initialization에서는 0.1의 성능 증가를 보였고, different ST-token initialization에서는 6.0의 큰 상승을 나타냈습니다. 이에 비해 Oxford Pets 데이터셋에서는 모든 설정에서 랜덤한 값을 주는 것이 더 나은 성능을 보였습니다. 훈련 구성은 1-shot은 50 epoch, 2&4-shot은 100 epoch, 8&16-shot은 200 epoch으로 설정되었습니다. 이는 훈련 샘플 수와 학습 횟수가 줄어들수록 좋은 Initialization이 모델 성능 향상에 더 중요한 역할을 한다는 것을 시사합니다. 그러나 epoch 수가 증가하거나 더 많은 few shot 샘플이 도입되면 성능이 오히려 감소하는 경향이 있으며, 이는 overfitting으로 인한 결과일 가능성이 있습니다. 이러한 결과는 다양한 맥락에서 ST-token initialization 기반 전략의 유효성에 대한 의문을 제기하며, 제한된 데이터로 학습할 때는 효과적인 초기화가 개선 효과를 가져올 수 있지만, 충분한 데이

	eurosat	oxford_pets
original-16	91.1	91.8
original-8	86.9	92.0
original-4	56.7	92.2
original-2	67.4	90.8
original-1	38.9	90.6
n_init_same16	91.1	88.2
n_init_same_8	83.2	86.6
n_init_same_4	56.2	89.1
n_init_same_2	54.2	80.1
n_init_same_1	39.0	88.0
n_init_diff_16	90.4	88.0
n_init_diff_8	83.9	87.0
n_init_diff_4	60.6	89.2
n_init_diff_2	55.5	79.8
n_init_diff_1	45.9	88.2

Table 1. Changes in Model Performance Based on Initialization - Seed 1

터가 주어지거나 모델이 이미 높은 성능을 보일 때는 그렇지 않을 수 있음을 보여줍니다.

### 3.2.2. Analysis 2

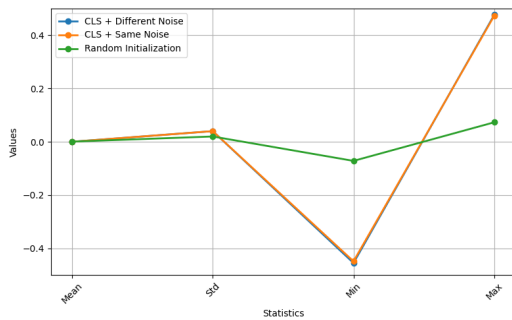


Figure 2. Statistics of the first layer of text prompts according to initialization

텍스트 프롬프트의 첫 번째 토큰에 대한 통계 분석을 진행하였습니다. 분석 결과, CLIP의 EOS 토큰에서 서로 다른 노이즈를 추가한 경우와 동일한 노이즈를 추가한 경우는 mean, std, min, max 측면에서 거의 유사한 패턴을 보였습니다. 그러나 이러한 통계적 유사성에도 불구하고, Table 1에서는 성능상 유의미한 차이가 확인되었습니다. 특히 random initialization의 경우, min과 max에서 큰 차이를 보였지만, mean과 std는 상대적으로

로 EOS 토큰을 활용한 Initialization과 유사하였습니다. 흥미로운 점은 개별 토큰 수준에서의 미세한 차이가 전체 모델 성능에 누적적으로 상당한 영향을 미칠 수 있다는 것입니다. 이는 초기화 방법의 선택이 모델 성능에 미묘하지만 동시에 중대한 영향을 끼칠 수 있음을 시사합니다.

### 3.2.3. Analysis 3

Eurosat 1-shot 시나리오에서 서로 다르게 초기화된 토큰을 사용하여 1 epoch 학습 후, 텍스트 feature embedding 분포를 t-SNE 차원 축소 기법을 통하여 시각화 하였습니다. 이를 통해 각기 다른 initialization 방법이 초기 학습 단계에서 텍스트 feature embedding의 representation과 구조에 어떠한 영향을 미치는지 탐구하였습니다. (Figure 3) Random Initialization은 가장 덜 구조화된 embedding을 생성하여, data point가 밀집되어 있고 feature label 간 분명한 구분이 없습니다. 이는 Random Initialization이 텍스트 feature의 구조적 정보를 거의 포착하지 못함을 의미합니다. Same ST-Initialization은 중간 수준의 data point 분포를 보여줍니다. CLIP 기반 initialization이 균일 노이즈에도 불구하고 일부 semantic structure를 유지하며, Random Initialization 보다는 더 나은 feature label 분포를 보입니다. Different ST-Initialization은 가장 좋은 학습 결과를 보입니다. data point가 2D visualization space에 넓게 퍼져 있어 텍스트 feature embedding의 다양성이 높고, distinct한 그룹으로 클러스터링되어 텍스트 특징 간의 의미적, 구조적 관계를 효과적으로 포착합니다. 이러한 초기 text embedding의 구조적 차이는 epoch과 샘플 수가 제한된 1-shot 시나리오에서 명확한 결과 차이를 야기합니다. Different ST-초기화는 다른 초기화 방법들에 비해 현저히 우수한 성능을 보였는데, 이는 사전 학습된 CLIP 모델의 풍부한 의미 정보를 효과적으로 활용했기 때문입니다. 이러한 결과는 초기 embedding space의 구조화된 특성이 extremely few shot 환경에서 성능에 결정적인 영향을 미칠 수 있음을 시사합니다. Initialization 방법의 선택은 특정 작업, 데이터 특성, 그리고 원하는 텍스트 임베딩의 목표에 따라 달라지며, 연구 목적과 응용 분야의 고유한 요구사항에 따라 최적의 접근 방식이 달라질 수 있습니다.

### 3.2.4. Results & Discussion

연구에서는 Text Prompt에서 RPO의 초기화 접근법을 활용하여 DAPT의 성능을 개선하는 것을 목표로 했습니다. 그리고 이를 위해 두 가지 주요 초기화 방법을 설정했습니다. 첫 번째 방법은 완전한 random initialization입니다. 이 접근법은 zero 또는 빈 tensor로 시작하여 균일 분포나 정규 분포를 통해 컨텍스트 토큰을 초기화합니다. 오로지 수학적 계산을 통해 초기 값을 결정하며, 기존 모델의 학습된 지식에 의존하지 않고 완전히 새로운 표현을 생성합니다. 두 번째 방법은 CLIP 모델의 EOS 임베딩을 재활용하는 ST-token Initialization 전략입니다. 이 접근법은 사전 학습된 모델의 지식을

최대한 활용하면서, 원본 클래스 및 토큰 임베딩에 작은 규모의 랜덤 노이즈를 추가합니다. 더하여 여기에서 접근법을 다시 두 가지로 나누었습니다. 1. 모든 토큰을 동일한 값으로 설정. 2. 각 토큰에 서로 다른 노이즈를 적용하여 다양한 초기값 부여. 두 방법은 초기 임베딩 활용에 있어 근본적인 차이를 보입니다. 첫 번째 방법은 new representation 학습에 중점을 두는 반면, 두 번째 방법은 사전 학습된 지식의 보존과 활용에 초점을 맞춥니다. 연구 결과, 데이터와 학습 리소스가 제한적인 상황에서는 EOS를 활용한 초기화 방법이 모델 성능 개선에 중요한 역할을 할 수 있음을 확인했습니다. 그러나 충분한 학습 데이터와 시간이 주어진 경우, EOS를 활용한 초기화는 오히려 overfitting을 야기할 수 있습니다. 주요 인사이트는 초기화 전략의 효과가 학습 환경에 따라 크게 달라진다는 점입니다. 제한된 리소스에서는 효과적인 초기화 방법이 성능 향상의 방법이 될 수 있으며, 이는 앞으로 더 많은 연구가 필요한 흥미로운 접근법임을 시사합니다. 또한, 각 seed마다 성능 값이 다소 변동하기 때문에, 보다 정확한 성능 측정을 위해 세 가지 seed를 모두 사용하여 실험하고 평가하는 것이 필요합니다.

#### 4. Future Works

첫째, FGVC Aircraft 데이터셋에 DAPT를 적용하여 EOS 기반 초기화 방법의 효과성을 검증하고자 합니다. 이를 통해 데이터셋의 특성에 따른 초기화 전략의 적합성을 보다 심층적으로 분석할 수 있을 것입니다. 둘째, 2, 4, 8, 16 shot 실험에서 epoch 수를 조정하고 random initialization과 비교함으로써, 성능 저하의 근본 원인이 overfitting으로 인한 것인지를 명확히 규명하고자 합니다. 이 과정에서 학습 샘플 수에 따른 모델의 성능 변화를 체계적으로 관찰할 계획입니다. 셋째, 현재 DAPT의 손실 함수를 심도 있게 분석하고 개선하고자 합니다. 보고서에는 기재하지 않았지만 DAPT에서 제안한 새로운 손실 함수를 분석하고, 소폭 변형해 실험을 진행해보았습니다. 분석과 실험 결과, 이미지 loss는 매우 낮고 나머지 original loss 및 text loss가 매우 높았습니다. 이미지 loss, original loss, text loss의 불균형을 해소하기 위해 각 컴포넌트의 가중치를 세밀하게 조정하여 모델의 전체적인 성능을 향상시키고자 합니다. 마지막으로, 기존의 텍스트 토큰 중심 접근에서 나아가 RPO와 유사한 방식으로 이미지 프롬프트에도 CLS 토큰을 활용하는 실험을 계획하고 있습니다. 이러한 다각도의 접근은 단순히 현재 연구의 한계를 극복하는 것을 넘어, Vision-Language Model의 학습 메커니즘에 대한 깊이 있는 이해를 제공할 것으로 기대됩니다.

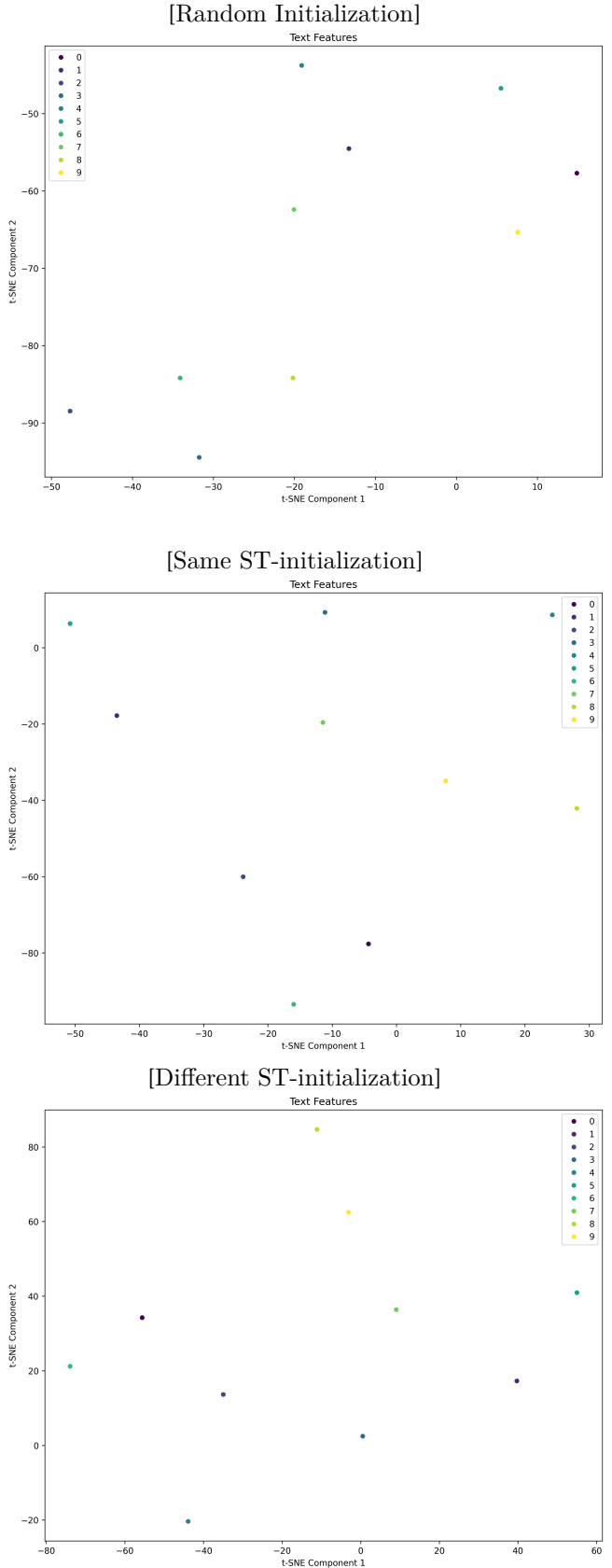


Figure 3. Comparison of Different Initialization Methods

## References

- Cho, E., Kim, J., and Kim, H. J. Distribution-aware prompt tuning for vision-language models. In Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 22004–22013, 2023.
- Helber, P., Bischke, B., Dengel, A., and Borth, D. Introducing eurosat: A novel dataset and deep learning benchmark for land use and land cover classification. In IGARSS 2018-2018 IEEE International Geoscience and Remote Sensing Symposium, pp. 204–207. IEEE, 2018.
- Kaur, P., Sikka, K., and Divakaran, A. Combining weakly and weakly supervised learning for classifying food images, 2017. URL <https://arxiv.org/abs/1712.08730>.
- Lee, D., Song, S., Suh, J., Choi, J., Lee, S., and Kim, H. J. Read-only prompt optimization for vision-language few-shot learning. In Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 1401–1411, 2023.
- Parkhi, O. M., Vedaldi, A., Zisserman, A., and Jawahar, C. V. Cats and dogs. In IEEE Conference on Computer Vision and Pattern Recognition, 2012.

Yesterday

**8th December, 9:41 pm**



Edited

example\_paper.bib

 You

**8th December, 12:14 am**

Edited

example\_paper.tex



Today

**9th December, 2:53 am**



Edited

example\_paper.tex

You

**9th December, 2:48 am**



Edited

example\_paper.bib

Edited

example\_paper.tex

You

**9th December, 2:42 am**



Edited

example\_paper.bib

Edited

example\_paper.tex

You

**9th December, 2:37 am**



Edited

example\_paper.tex

You

**9th December, 2:31 am**



Edited

example\_paper.tex

You

**9th December, 2:25 am**



Edited

Today

**9th December, 12:55 pm**



Edited

example\_paper.tex

 You

**9th December, 12:50 pm**



Edited

example\_paper.tex

 You

**9th December, 12:44 pm**



Edited

example\_paper.tex

 You

**9th December, 12:39 pm**



Edited

example\_paper.tex

 You

**9th December, 12:33 pm**



Edited

example\_paper.tex

 You

**9th December, 12:28 pm**



Edited

example\_paper.tex



Today

**10th December, 10:33 pm**



Edited

example\_paper.tex

 You

**10th December, 10:27 pm**



Edited

example\_paper.tex

 You

**10th December, 10:22 pm**



Edited

example\_paper.tex

 You

**10th December, 10:17 pm**



Edited

example\_paper.tex

 You

**10th December, 10:11 pm**



Edited

example\_paper.tex

 You






**10th December, 10:06 pm**



Edited

example\_paper.tex

 You

 딥러닝_final_1205.ipynb	Add files via upload	4 days ago
 딥러닝_final_1206.ipynb	Add files via upload	2 days ago
 딥러닝_최종시험.ipynb	Add files via upload	15 hours ago
 딥러닝_final.ipynb	Colab을 통해 생성됨	last week
 딥러닝_final_1204.ipynb	Colab을 통해 생성됨	last week