

PetLens 2.0

프로젝트 결과 보고서

반려동물 이미지·영상 멀티모달 AI 분석 서비스

Pet Detection

Multi-pet

Dog-130

SAM2

CLIP Retrieval

Video Tracking

Animal Pose

Action Timeline

프로젝트 핵심

- 사진 분석

Grounding DINO 탐지 → SAM2 분할 → 개체별 품종 분석과 유사 이미지 검색

- 품종 분류

강아지는 직접 학습한 Dog-130 ViT, 고양이는 Pet-37 classifier 사용

- 영상 분석

SAM2 Video tracking을 기반으로 개체별 품종·이동·행동 타임라인 분석

- 고급 분석

SigLIP2, DINOv2, ViTPose++를 선택형 비교·포즈 분석에 활용

수도권ICT이노베이션스퀘어 디지털신기술 운영사무국 제출용

제출일 2026. 09. 08.

프로젝트 목적 및 개요

PetLens 2.0은 반려동물 사진과 영상을 입력하면 개체를 탐지하고, 품종·유사 이미지·포즈·움직임 및 행동 정보를 하나의 웹 흐름에서 분석하는 멀티모달 Pet Vision 서비스입니다.

주요 수행 내용

Grounding DINO와 SAM2를 이용해 여러 반려동물을 개체별로 탐지·분할하고, species에 따라 Dog-130 또는 Pet-37 classifier로 분기했습니다. CLIP 기반 의미 검색과 유사 이미지 검색, SigLIP2-DINOv2 비교, ViTPose++ 포즈 분석을 연결했으며 영상에서는 SAM2 tracking과 zero-shot 행동 타임라인까지 구현했습니다.

분석 파이프라인

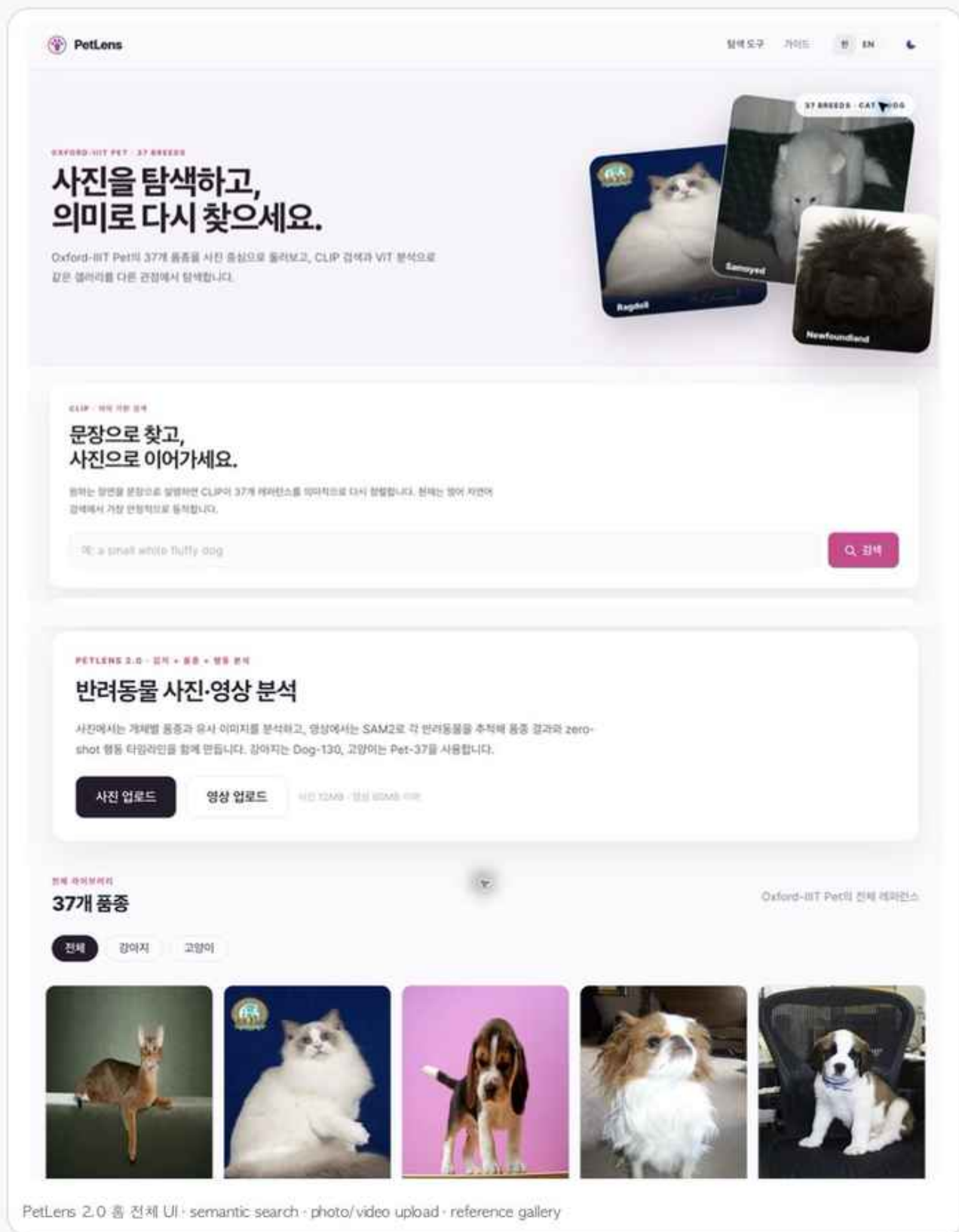


결과 및 기대효과

단순 품종 분류를 넘어 이미지와 영상에서 개체별 분석 결과를 제공하는 실제 웹 서비스로 배포했습니다. Dog-130 모델은 130개 강아지 품종에서 Validation Accuracy 86.79%, Macro F1 86.64%를 기록했습니다. 향후 행동 모니터링, 사진·영상 기반 검색, 반려동물별 상태 분석과 같은 서비스로 확장할 수 있습니다.

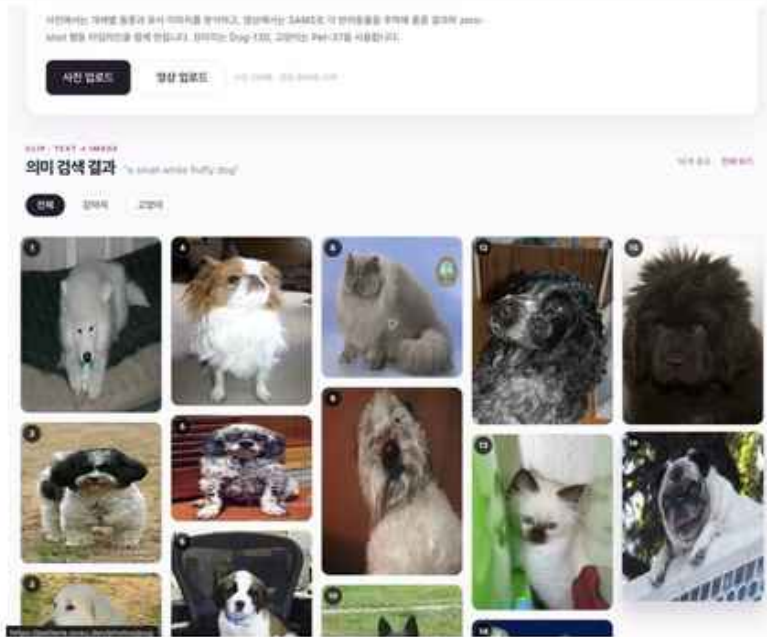
전체 서비스 화면

홈 화면에서 자연어 의미 검색, 사진 업로드, 영상 업로드, 37-reference gallery를 한 흐름으로 사용할 수 있습니다.

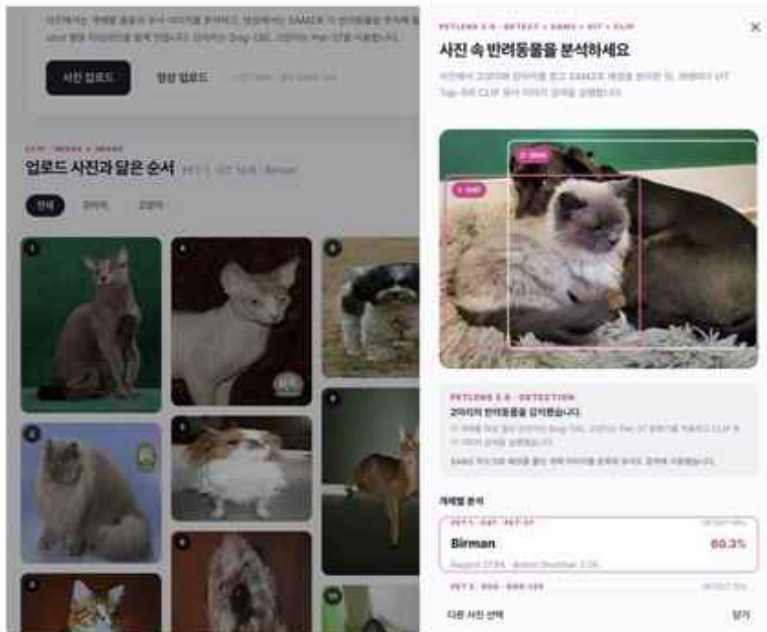


의미 검색과 멀티펫 분석

사용자가 자연어로 원하는 반려동물을 검색할 수 있고, 한 사진에 여러 마리가 있어도 개체별로 선택해 독립 분석합니다.



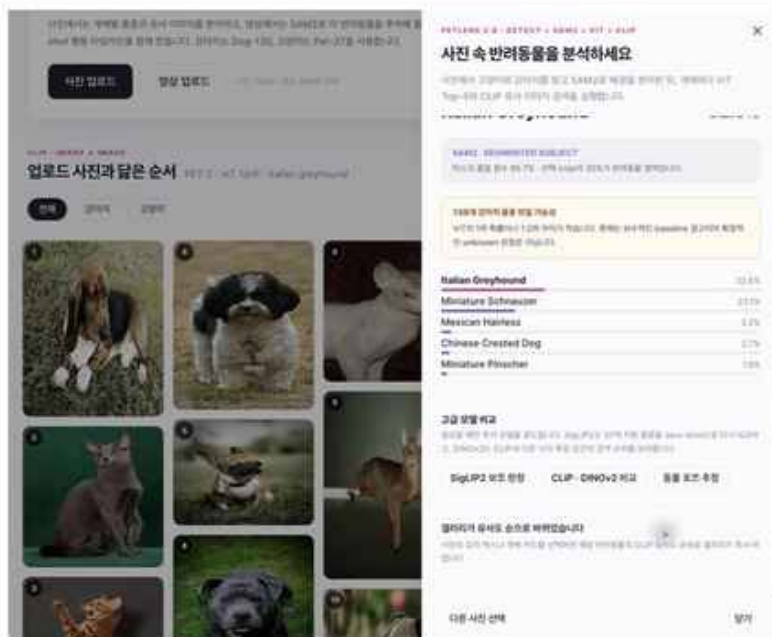
자연어 → CLIP embedding → reference gallery 의미 기반 재정렬



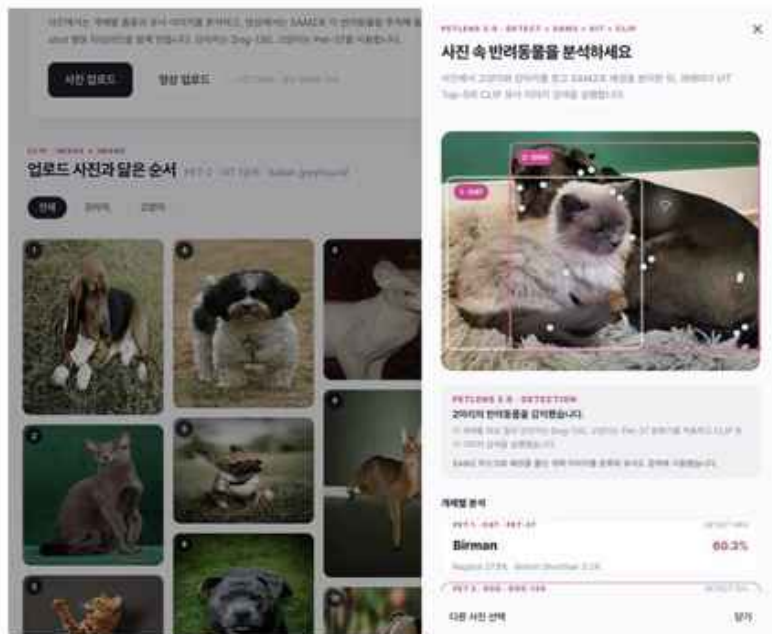
Grounding DINO 기반 dog / cat 개체 탐지와 멀티펫 분석

Dog-130 상세 분석과 Animal Pose

강아지는 130-class 전용 classifier로 분석하고, SAM2 foreground 분할 및 ViTPose++ keypoint 분석을 함께 제공합니다.



DOG-130 · 130 BREEDS / SAM2 segmentation / Top prediction



ViTPose++ AP-10K expert 기반 17-keypoint animal pose overlay

SigLIP2 · CLIP · DINOv2 비교

한 모델의 단일 출력만 제시하지 않고, 선택형 고급 분석에서 서로 다른 feature space의 결과를 비교할 수 있도록 구성했습니다.

사진에서는 개체별 품종과 유사 이미지를 분석하고, 영상에서는 SAM2로 각 반려동물을 추적해 zero-shot 행동 타임라인을 함께 제공합니다. 강아지는 Dog-130, 고양이 Pet-37을 사용합니다.

사진 업로드 **영상 업로드** 사진 1000장 | 영상 60MB 이하

CLIP · IMAGE + IMAGE **업로드 사진과 닮은 순서** PET-2 · VIT 1순위 · Italian greyhound

전체 강아지 고양이

PETLENS 2.0 · DETECT + SAM2 + VIT + CLIP

사진 속 반려동물을 분석하세요

사진에서 고양이와 강아지를 찾고 SAM2로 배경을 분리한 뒤, 개체마다 VIT Top-5와 CLIP 유사 이미지 검색을 실행합니다.

Top-5	CLIP 유사 이미지 검색
Chinese Crested Dog	2.7%
Miniature Pinscher	1.9%

고급 모델 비교

일회성 페인 추가 설정을 로드합니다. SigLIP2는 37개 품종을 zero-shot으로 다시 비교하고, DINOv2는 CLIP과 다른 시각 특징 공간의 검색 순위를 보여줍니다.

SigLIP2 **CLIP · DINOv2 비교** 동물 포즈 추정

VITPOSE++ · 4 · 49-100

16개 keypoint를 표시했습니다.

Nose 71% · L_Eye 78% · R_Eye 88% · L_Ear 45% · R_Ear 56% · L_Shoulder 54%

SIGLIP2 · ZERO-SHOT

VIT와 SigLIP2의 1순위가 다릅니다.

Rank	Species	Score
1.	Ragdoll	48.3%
2.	British Shorthair	20.6%
3.	Siamese	15.0%

CLIP · DINOv2 RETRIEVAL

Model	Rank	Species
CLIP	1.	Russian Blue
	2.	Persian
DINOv2	1.	Ragdoll
	2.	Persian

다른 사진 선택 **닫기**

SigLIP2 zero-shot 후보와 CLIP / DINOv2 retrieval 비교

CLIP similarity는 품종 확률이 아니라 embedding similarity이며, DINOv2는 시각적 feature retrieval 비교에 사용됩니다. SigLIP2 결과 역시 독립적인 zero-shot 후보로 해석하며 classifier의 절대 확률로 사용하지 않습니다.

영상 Tracking과 Semantic Action Timeline

짧은 영상을 batch 방식으로 분석해 동일 개체를 프레임 간 추적하고, 품종·이동량·시간대별 행동 후보를 track 단위로 반환합니다.

PetLens 2.0 - VIDEO + SEMANTIC TRACKING

영상 속 반려동물들을 추적하세요

본 페이지에서 반려동물용 영상 SAM2의 성능을 다양한 개체들 추적하는 데
용 추적하는 동물 개체로 CLIP 용어 ID를 통해 검색합니다.

VIDEO SUMMARY:
20여 - 12개 프레임 분석
Clip: 1000 - 1000

TRACK 1 - 품종: West Highland White Terrier
종류: 품종 불
12.4%

행동	종류	빈도
앉아	품종 불	100
뛰어	품종 불	100
뛰어	품종 불	100

TRACK 2 - 품종: Toy Terrier
종류: 품종 불
20.4%

행동	종류	빈도
뛰어	품종 불	100
뛰어	품종 불	100
뛰어	품종 불	100

SAM2 Video tracking · track별 Dog-130 / Pet-37 분석

PetLens 2.0 - VIDEO + SEMANTIC TRACKING

영상 속 반려동물들을 추적하세요

본 페이지에서 반려동물용 영상 SAM2의 성능을 다양한 개체들 추적하는 데
용 추적하는 동물 개체로 CLIP 용어 ID를 통해 검색합니다.

TRACK 1 - 품종: West Highland White Terrier
종류: 품종 불
12.4%

행동	종류	빈도
앉아	품종 불	100
뛰어	품종 불	100
뛰어	품종 불	100

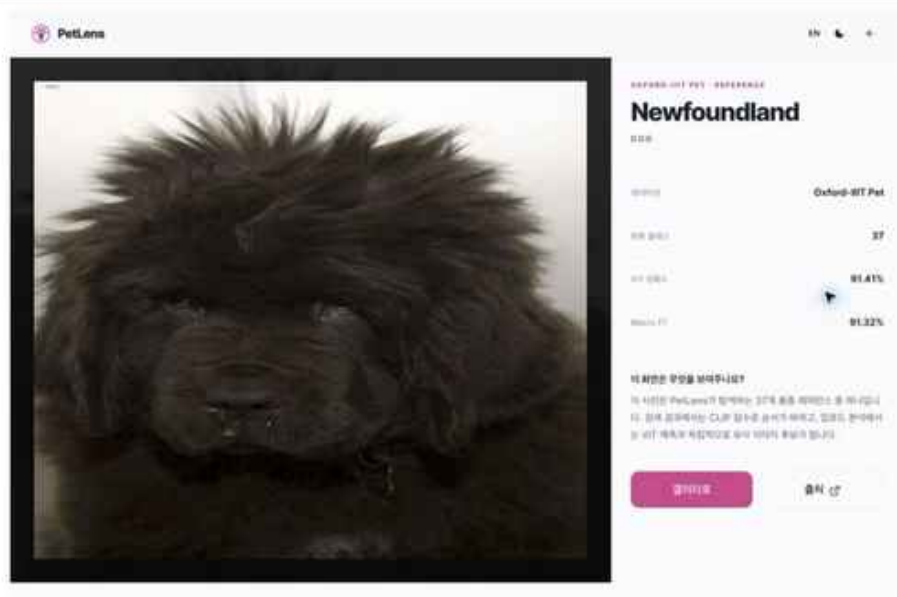
TRACK 2 - 품종: Toy Terrier
종류: 품종 불
20.4%

행동	종류	빈도
뛰어	품종 불	100
뛰어	품종 불	100
뛰어	품종 불	100

CLIP scene semantics + tracking motion 기반 zero-shot action timeline

품종 상세 화면과 Guide

분석 결과 이후 품종 상세 정보로 이어지고, Guide에서는 현재 PetLens 2.0의 이미지·영상 기능과 해석 방법을 안내합니다.



Breed detail · 갤러리에서 품종별 상세 정보 확인



Guide · Dog-130 및 Video Action Timeline 사용 안내

Dog-130 모델 학습 과정

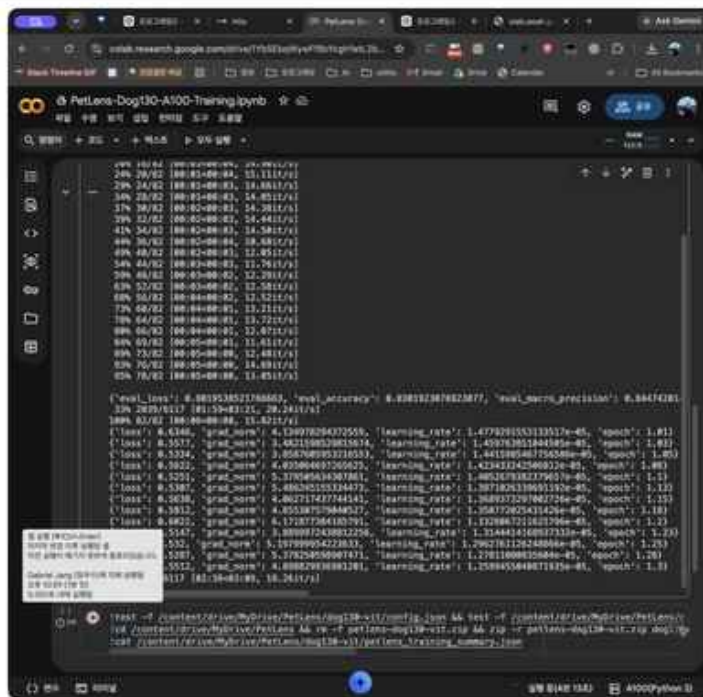
Tsinghua Dogs 130개 품종을 대상으로 google/vit-base-patch16-224 backbone을 fine-tuning했으며, NVIDIA A100-SXM4-40GB 환경에서 3 epochs 학습했습니다.



NVIDIA A100-SXM4-40GB 런타임 확인



Dog-130 학습 시작



학습 진행 상태: epoch / loss / validation 추적

Dog-130 학습 완료 및 평가 결과

최종 production artifact를 저장하고 AutoImageProcessor / AutoModel 로딩 검증까지 완료한 뒤 실제 API serving에 연결했습니다.



3 epochs 학습 완료



최종 metrics 및 artifact 확인

평가 지표

Classes

130

Train examples

65,228

Validation

5,200

Accuracy

86.79%

Macro Precision

88.13%

Macro Recall

86.79%

Macro F1

86.64%

Epochs

3

GPU · NVIDIA A100-SXM4-40GB | Backbone · google/vit-base-patch16-224

프로젝트 결과물

PetLens 2.0은 연구용 notebook에 머무르지 않고 웹 프론트엔드와 FastAPI 기반 ML service를 연결해 실제 공개 서비스로 배포했습니다.

최종 산출물

- 반려동물 사진 업로드 및 Grounding DINO 멀티펫 탐지
- SAM2 segmentation과 Dog-130 / Pet-37 개체별 품종 분석
- CLIP 자연어 검색·유사 이미지 retrieval 및 DINOv2 비교
- SigLIP2 zero-shot 보조 분석과 ViTPose++ Animal Pose
- SAM2 Video tracking 및 시간대별 zero-shot Action Timeline
- Dog-130 custom ViT 학습·평가·production serving

결과 확인

Live Demo	https://petlens.oosu.dev/
GitHub	https://github.com/oosuhada/petlens-ai
서비스 형태	Web + FastAPI ML Service

기대효과

반려동물 사진·영상의 탐지, 분류, 검색, 포즈, 행동 정보를 하나의 서비스에서 다루는 기반을 확보했습니다. 향후 장시간 행동 모니터링, 개인 반려동물별 기록, 검색·추천, 이상 행동 탐색 등의 기능으로 고도화할 수 있습니다.