

AI3CT.2026

DINOv3 임베딩 기반 이미지 매칭

2026.01.07.
강남규



CONTENTS



I 연구목표

II 환경 구축 및 데이터 수집

III 모델별 유사도 점수 분포

IV 모델별 유사도 검색 결과

V 모델 및 순위별 평균 점수

VI 모델별 참조영상 검색 시간 분포

VII 참조영상별 총 소요시간

VIII 결론 및 논의

“UAM 항공 영상에서 DINOv3 patch 임베딩은 기체 회전 상황에서 얼마나 안정적인 표현을 유지하는가?”

DINOv3 임베딩 FAISS 검색 결과 요약

DINOv3로 생성된 원본데이터–참조데이터의 임베딩에 대해 FAISS(IndexFlatIP, L2 정규화)로 Top10 검색을 수행하고, 각 참조데이터별 결과를 CoarseLocalizationInfo (JSON)로 저장.

사용한 모델/설정

매칭 파라미터: TopK 10

메트릭: 내적(IP) 기반 코사인 유사도

백엔드: FAISS GPU

백본 모델(weights):

모델	가중치 파라미터 수(M)	사전학습 데이터셋
ViT-B/16 distilled	86	LVD-1689M
ViT-H+/16 distilled	840	LVD-1689M
ViT-L/16 distilled	300	SAT-493M
ViT-L/16 distilled	300	LVD-1689M
ViT-S/16 distilled	21	LVD-1689M
ViT-S+/16 distilled	29	LVD-1689M

LVD-1689M: 약 16.89억 장 규모의 웹 이미지(Instagram 공개 포스트 기반 대규모 풀에서 선별)
SAT-493M: 약 4.93억 장의 위성 타일(Maxar RGB 정사영상 등)

TopK 검색은 DINOv3 기반 전역 임베딩(GlobalToken)만 사용

DINOv3 모델이 한 장의 이미지를 전체적으로 요약해 만든 단일 벡터

내용:

- 이미지 전체의 시맨틱/장면 정보를 압축한 고차원 벡터.
- 위치 좌표나 패치별 분포 같은 공간 정보는 없음.
- 오직 한 벡터로 전체 이미지를 대표.
- 이미지 전체를 대표하는 1개 토큰($1 \times D$ 벡터)이라 위치 정보가 없는 전역 표현

차원: GlobalToken 차원(= 모델 hidden size):

- vits16, vits16+: 384차원
- vitb16: 768차원
- vitl16, vitl16sat: 1024차원
- vith16+: 1280

차원생성 경로: 이미지 리사이즈/정규화 후 모델의 전역 토큰을 추출.

후처리: TopK 검색 시에는 이 벡터를 L2 정규화한 뒤 FAISS IndexFlatIP로 코사인 유사도 검색에 사용.

테스트 검색 방향:

([비행데이터] → [참조데이터] 검색)

비행데이터: 251124155712(3Dtilesx), 251124160703

참조데이터: 300고도 중첩도 50의 0, 45, 90, 135, 180, 225, 270, 315도 회전 변환 이미지

각 모델별 유사도 검색: vitb16, vith16+, vitl16, vitl16sat, vits16, vits16+

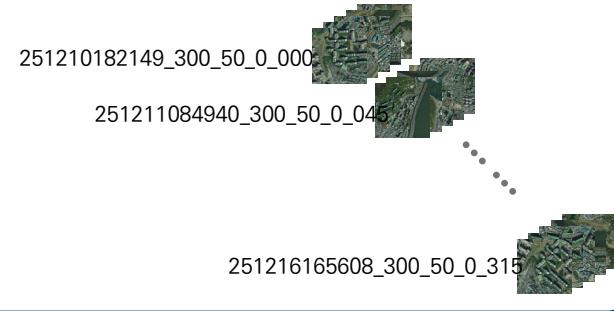
비행데이터 (쿼리)
(300고도)



참조 데이터
(300고도)

회전 변환 (0, 45, 90, 135, 180, 225, 270, 315)

탐색



유사도
결과



Rank: 1 2 3 ... 8 9 10

모델별 유사도 검색 결과



결과: 비행데이터: 251124155712_00057.jpg

vitb16



vith16+



vitl16



vitl16sat



vits16



vits16+





결과: 비행데이터: 251124155712_00091.jpg

vitb16



vith16+



vitl16



vitl16sat



vits16



vits16+





결과: 비행데이터: 251124155712_00100.jpg

vitb16



vith16+



vitl16



vitl16sat



vits16

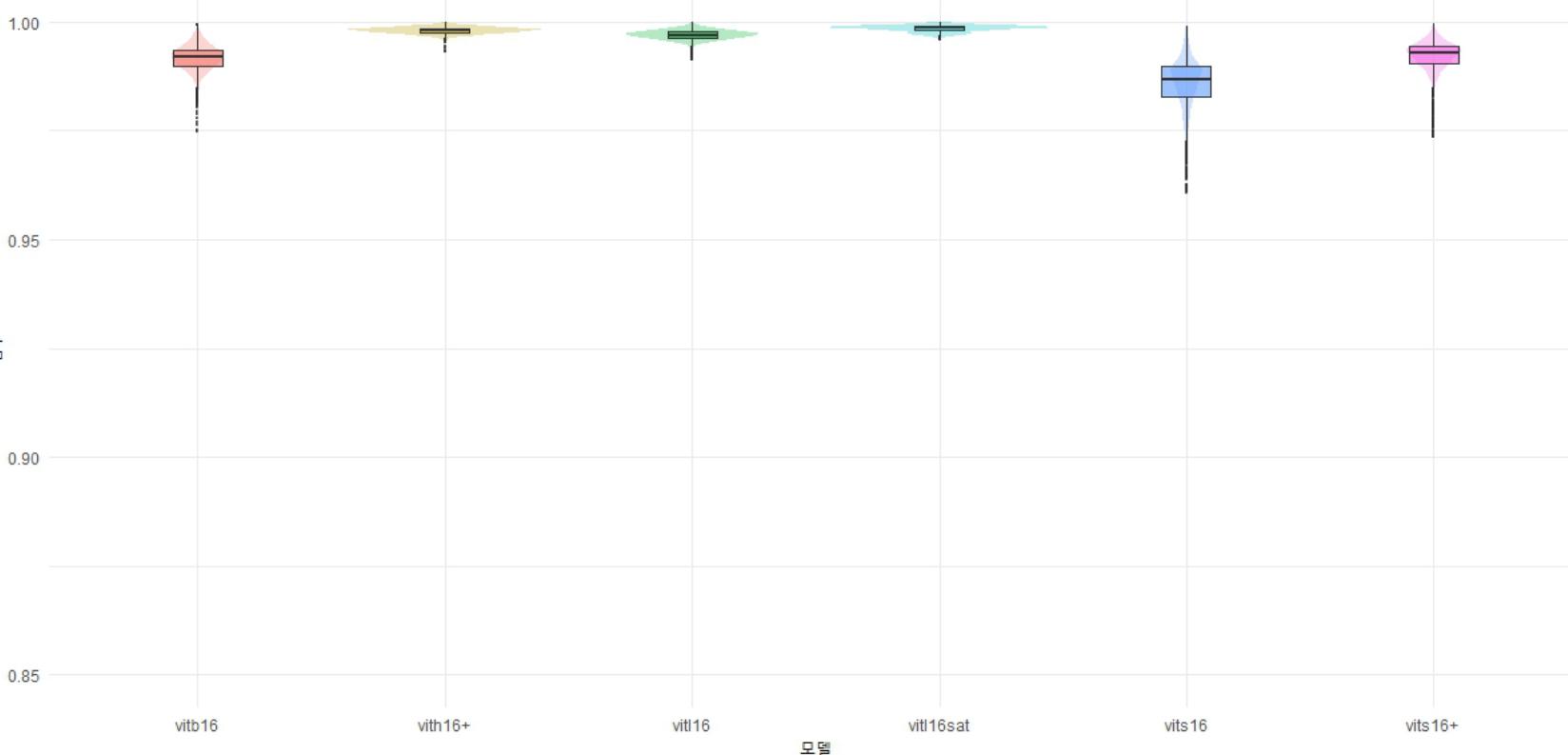


vits16+

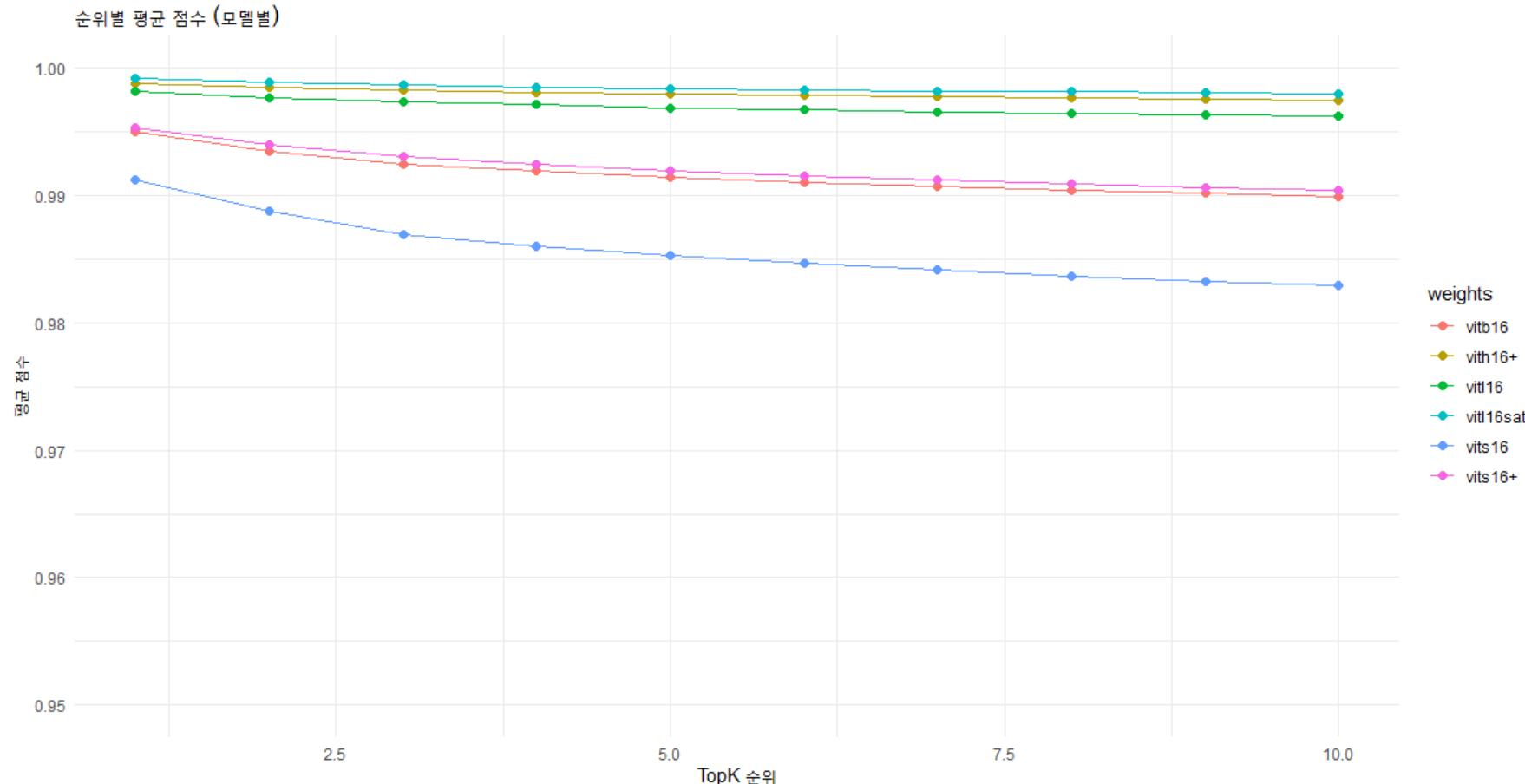


각 모델의 유사도 점수 전체 분포

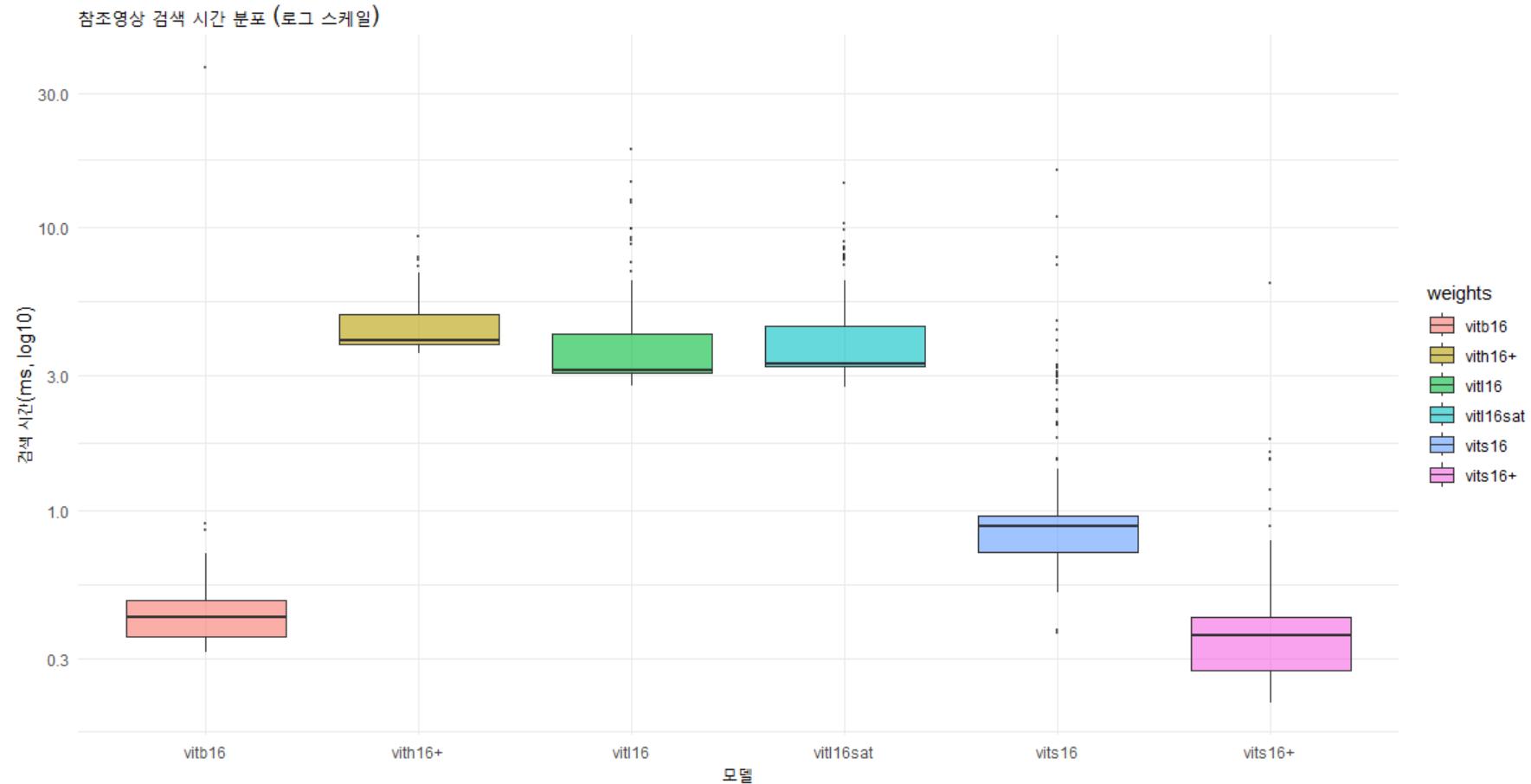
모델별 유사도 점수 분포



Top1 ~ Top10 순위에 따른 모델별 차이 파악
상위권 점수 급락 여부 확인

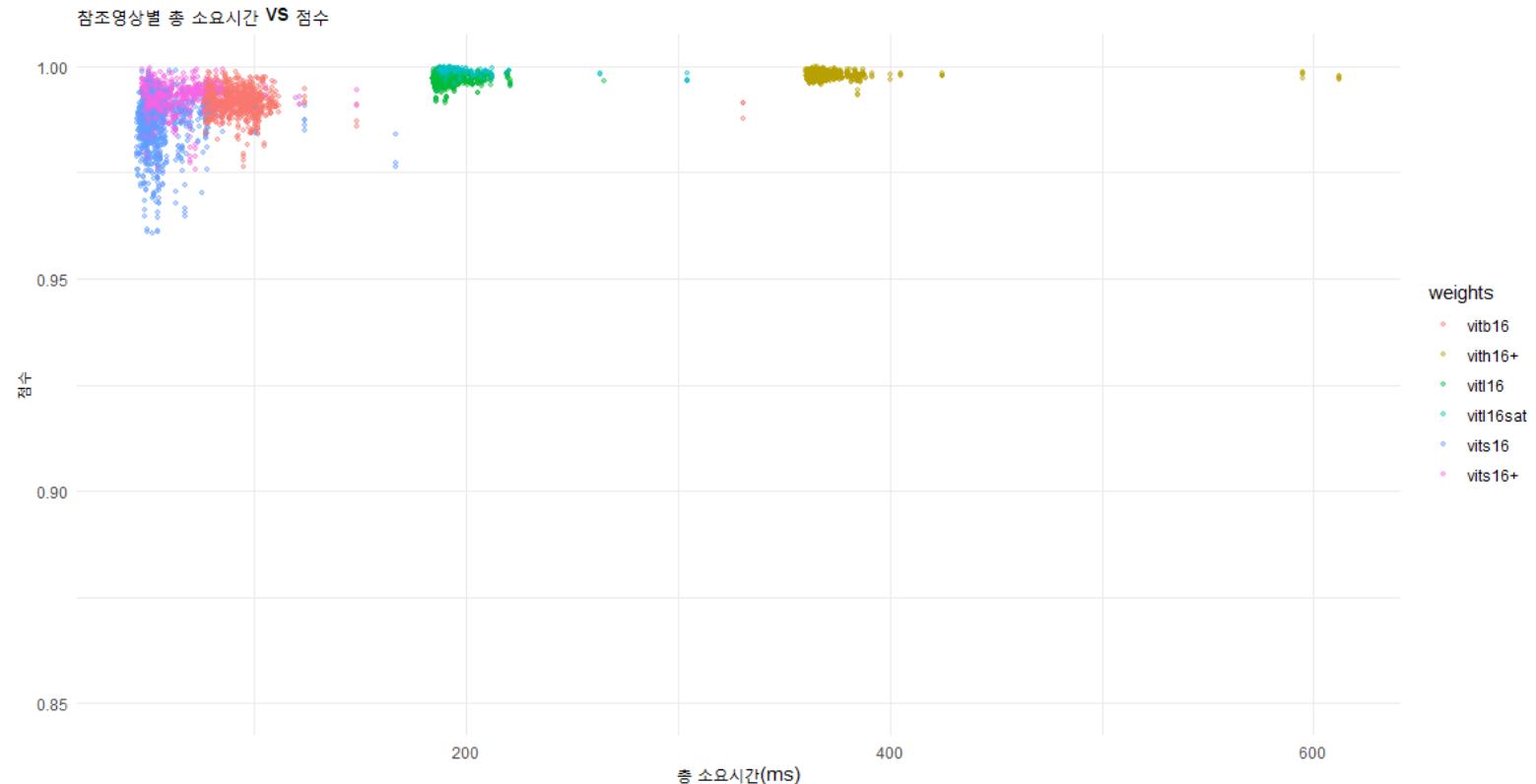


모델별 검색 지역 특성; 분포 및 이상치 파악
로그 스케일



(임베딩+검색) - 점수 산점도

응답속도와 점수 사이 관계 탐색
 유사도 검색 점수에 대한 병목현상의 영향 평가
 분포 패턴 확인



(임베딩+검색) - 점수 산점도

응답속도와 점수 사이 관계 탐색
 유사도 검색 점수에 대한 병목현상의 영향 평가
 분포 패턴 확인

모델별 시간 요약

모델 이름	임베딩 생성 시간 평균값(ms)	검색 시간 평균값(ms)	총 시간 평균값(ms)
vitb16	91.39418821	0.568178767	91.96236698
vith16+	366.4572729	4.449860141	370.907133
vitl16	188.7161677	3.921877325	192.6380451
vitl16sat	189.1718083	4.125872577	193.2976809
vits16	53.90365644	1.187553334	55.09120977
vits16+	59.24270456	0.414413971	59.65711853

회전된 참조셋에서도 전 모델 TopK 유사도가 0.98~0.99+로 높음
Top1~Top10 점수 낙폭과 사례 분석으로 모델별 강건성 차이를 정밀 검토

정확도: vitl16sat/vith16+가 최상(0.9984/0.9979)이나, 추론시간이 길어 실시간성엔 불리.

vits16(+)=가장 빠르나 약간 낮은 유사도(0.9857~0.9921).

지연 병목: 임베딩 생성 시간이 대부분을 차지, 검색 시간은 ms 이하 수준
→ 최적화는 모델 경량화 등 전처리 측면에 집중 필요.

전역 토큰만 사용해 위치 정보가 없는 한계가 존재
→ 향후 패치 임베딩/로컬라이제이션 정보 결합, 회전/고도 변화 추가 실험 필요.

감사합니다.

