단계적 파인튜닝에 의한 비디오 프레임 순서 복원

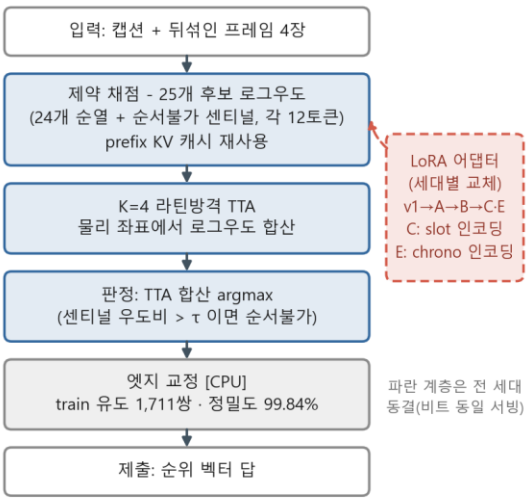
팀 코스닥삼천

1. 전체적인 방법론 및 모델 구조

캡션과 함께 뒤섞여 제시된 비디오 프레임 4장의 원본 순서, 즉 24가지 순열 중 하나를 예측하는 것이다(일부는 순서가 정의되지 않는 순서불가 클래스). 지표는 부분 점수가 없는 exact match이다.

방법론의 골격은 자유 생성 대신 후보 전수 채점을 쓰는 제약 추론이다. 출력을 24개 순열과 순서불가 센티널의 25개 후보로 제한하고, 각 후보 답 문자열의 조건부 로그우도를 채점한다. 숫자 0–4가 단일 토큰이라 모든 답(예: `[3, 1, 2, 4]`)이 정확히 12토큰이며 길이 편향이 없다. 답 규약은 대회 CSV와 같은 순위 벡터다(A[j]는 슬롯 j에 제시된 프레임의 원본 위치). 채점 위에 네 가지 제시 순서의 로그우도를 합산하는 TTA, 센티널과 최량 순열의 로그우도비를 임계값 τ와 비교하는 순서불가 판정, 프레임 쌍 제약으로 모순만 교정하는 엣지 후처리가 차례로 얹힌다(Figure 1).

Figure 1. 추론 파이프라인



모델은 Qwen3-VL Instruct(가중치 공개 2025/10)다. NF4 4-bit로 동결한 베이스에 LoRA 어댑터만 학습해 얹고 비전 계층에도 LoRA를 적용하며, 서빙은 어댑터를 병합 없이 얹어 학습·추론 경로를 비트 단위로 동일하게 유지한다. 8B로 검증한 방식을 32B로 확장했다.

2. 아이디어의 동기와 핵심 기여

핵심 선택은 세 가지다. 첫째, 출력 공간이 25개뿐인 점을 이용해 추론을 후보 전수 채점으로 바꿨다. 둘째, 개선을 세대당 가설 하나로 제한하고 채택 기준을 결과 관측 전에 정했다. 셋째, 표본 외 강건성을 우선해 리더보드를 참고하지 않았고, 학습과 평가 분포의 차이를 계량했다. 이 규율에는 팀의 배경인 퀀트 리서치 방식을 섞었다. 여러 후보에서 리더보드 최고치를 고르지 않는 것은 백테스트를 여러 번 돌려 최고치를 보고할 때 생기는 선택 편향(과적합)이고, 신호가 채점 시 노이즈를 넘길 때만 학습

기법을 채택한 것은 거래비용보다 못한 알파를 버리는 기준이며, 오류 상관이 낮은 두 인코딩을 나란히 실험하려던 상호보완적 구성은 상관 낮은 자산으로 분산을 줄이는 포트폴리오의 발상이다. 하단의 네가지 방식은 모두 이 규율 위에서, 문제의 성질 하나씩을 겨냥해 설계됐다.

1. 생성 문제를 채점 문제로 바꾼 제약 추론. 정답 후보 25개를 전부 채점하는 방식으로, 모델이 자유롭게 답을 생성하게 두면 형식이 어긋난 무효 출력이 나올 수 있고, 학습은 토큰 단위 교차엔트로피로 하면서 평가는 문자열 전체로 받는 불일치도 생긴다. 다행히 가능한 답이 25개뿐이라, 생성 대신 25개 후보의 로그우도를 전부 계산해 가장 높은 것을 고르는 방식으로 바꿨다. 무효 출력이 원천 차단되고, 학습이 낮추는 값과 추론이 고르는 값이 정확히 같아진다. 25번 채점하는 비용은 후보들이 프롬프트를 공유하고 마지막 12토큰만 다르다는 구조 덕분에(KV 캐시 재사용) 거의 들지 않는다.

2. 제시-불변 센티널. "순서 없음"을 위한 전용 클래스로 항등 순열 같은 특정 순열로 답하게 하는 게 자연스러워 보이지만, 프레임 제시 순서를 바꿔가며 여러 번 채점하는 TTA와 만나면 오류가 생긴다. 제시가 바뀌면 항등 순열이 가리키는 배열도 매번 달라져서, 같은 "순서 없음" 판단이 네 개의 다른 후보로 흩어지고 신호가 1/4로 희석된다. 그래서 제시 순서와 무관한 별도의 센티널 문자열을 뒀다. 어떻게 제시하든 증거가 한 후보에 쌓이며, TTA 제시 집합도 라틴방격으로 짜서 각 프레임이 각 슬롯을 정확히 한 번씩 지나가게 했다. 그 결과 슬롯 위치에 대한 편향이 합산 과정에서 정확히 상쇄된다.

3. 데이터 생성 구조를 역이용한 엣지 후처리. 학습 데이터는 프레임들이 한정된 클립 풀에서 재조합되고 있었다(교차 샘플 재사용 50.7%). 같은 4장 세트가 다시 나와도 정답 순열이 유지되는 경우는 4.3%뿐이라 세트를 통째로 외우는 건 소용없지만, 두 프레임 사이의 앞뒤 관계는 어디서 다시 나오든 변하지 않았다. 재조합이 클립의 시간 순서는 건드리지 않기 때문이다. 그래서 2회 이상 같이 등장하고 순서가 한 번도 어긋나지 않은 쌍만 제약으로 삼고, 모델 예측이 이 제약과 모순될 때만 가장 가까운(Kendall-τ 기준) 일관된 순열로 교체한다. 개입 조건이 좁고 제약의 정밀도가 측정돼 있어, 틀릴 위험이 건당 0.16%로 유계다. 이기는 수가 아니라 지지 않는 엣지를 추가했다.

4. 답 인코딩의 방향 전환(chrono). 답을 말하는 방향 바꾸는 것으로, 같은 정답도 두 방향으로 말할 수 있다. "1번 슬롯의 프레임이 원래 몇 번째였나"(slot)와 "첫 장면은 무엇이고 그다음은 무엇인가"(chrono)다. 둘은 서로 역순열이라 담긴 정보는 완전히 같다. 하지만 slot은 첫 토큰부터 역참조를 풀어야 하는 반면, chrono는 내용을 시간 순서대로 따라가는 질문의 연쇄가 되고, 문장 정렬 연구(Re-BART, NAON)의 방출 방향도 이쪽이다. 실제로 학습 전 캐시 재분석에서 chrono가 짧은 캡션의 앞자리 정확도에서 우위였고, 위치별 정확도 총합은 그대로였다. 정보가 늘어난 게 아니라, 배우기 쉬운 자리로 옮겨진 것이다.

요컨대 이 접근의 독창성은 새 모델이나 새 손실이 아니라 문제의 성질을 읽고 그 자리에 맞게 조립한 데 있다. 지표의 all or nothing 성질은 제약 채점으로, 출력 공간의 유한성은 KV 캐시 전수 채점으로, TTA와 클래스 표현의 충돌은 제시-불변 센티널로, 데이터의 재조합 구조는 하방 유계 후처리로, 자기회귀의 조건화 순서는 인코딩 방향으로 각각 번역했고, 채택 여부는 전부 표본 외 규율로만 판정했다.

3. 학습 및 추론 과정

학습은 QLoRA 파인튜닝이다. 손실은 답 토큰 12개에만 부과해 추론이 최대화하는 양과 맞추고, 학습·추론 프롬프트가 토큰 단위로 같음을 실행 전 assert로 강제했다. 증강은 제공된 4장의 제시 순서 재배열뿐이다. 픽셀이 같아도 정답 문자열이 달라져 위치 지름길이 막히고 암기가 늦어지며, 샘플당 궤도 표본 4개로 34,140 예제를 구성했다. 학습률은 5e-5에서 증강 및 비전 학습 도입과 함께 1e-4로 올렸다. 체크포인트 선택 규칙은 항상 final이다. 무증강 학습에서는 곡선이 1.5 epoch에서 꺾여(암기 시작) 중간 선택의 여지가 있었지만, 증강을 켠 뒤로는 곡선이 종점까지 꺾이지 않아 선택 자체를 없앨 수 있었다. 32B는 전면 NF4 양자화로 24GB에 로드(18.2GB)했고, 36.5시간과 2,134스텝 동안 손실이 발산 없이 하강했다(0.515→약 0.10).

추론은 25개 후보가 약 1,000토큰의 프롬프트를 공유하고 12토큰 꼬리만 다르므로, prefix의 KV 캐시를 재사용해 1,000문항 채점을 92분에서 12.5분으로 줄였다(fp32 오차 5.9e-5, 순위 일치). TTA는 제시 순서를 {id, (2,3,4,1), (3,4,1,2), (4,1,2,3)}로 바꿔 채점해 물리 좌표에서 합산하고, τ는 홀드아웃에서 보정했다.

4. 실험 결과 및 분석

4.1 프로토콜

train 9,535건을 층화 분할해 홀드아웃 1,000건을 뗐다(순서불가율 15.50% 보존, 시드 고정). 모든 설계 및 임계값, 체크포인트 선택은 홀드아웃에서만 했고, 제출 전에 기대 밴드를 기록했으며, 리더보드 수치는 어떤 결정에도 쓰지 않았다. 근소한 차이는 McNemar 짝지은 검정으로 판정했고, 여러 시드나 체크포인트에서 최댓값을 고르는 것은 홀드아웃 1,000건에서 검증했다..

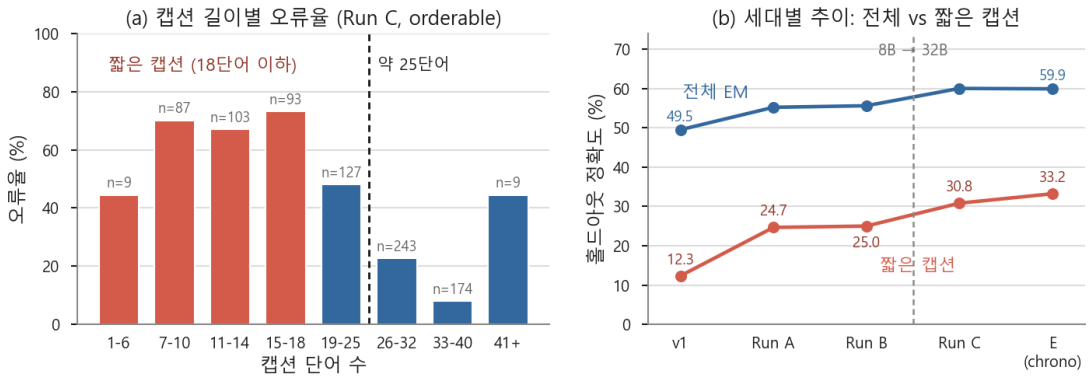
4.2 주 결과

Table 1. 세대별 성능

구성	홀드아웃 EM	짧은 캡션	public LB	private LB
제로샷 (non-tunning)	0.162	—	—	—
Claude Opus 4.8 제로샷 (참고용)	0.397	18.5%	—	—
v1 — 8B, 무증강·비전 동결	0.495	12.3%	0.80628	0.74796
v1 + 엣지	—	—	0.85165	0.81707
Run A — 8B, 증강+비전 학습	0.552	24.7%	0.91099	0.88211
Run B — 8B, +쌍비교 감독	0.556	25.0%	0.91273	0.89837
Run C — 32B, Run A 레시피	0.600	30.8%	0.93368	0.92276
Run D — 32B, 전 데이터 환입 리핏	—	—	0.92495	**0.93089**
Run E — 32B, chrono 인코딩†	0.599	33.2%	**0.93717**	0.91463

v1은 홀드아웃 49.5%로 제로샷 16.2%를 크게 넘겼고, 캡션이 짧아 텍스트에 순서 단서가 없는 표본(18단어 이하, orderable의 약 35%)의 정확도가 12.3%로 지배적 병목이라는 진단을 남겼다(Figure 2). 이후 개선의 인과는 짧은 캡션 버킷(Figure 2b)이 설명한다. Run B는 순서 판단의 원자 단위인 "두 프레임 중 무엇이 먼저인가"를 직접 감독했지만 Run A와 동률(McNemar z=-0.11)에 그쳐, 병목을 훈련 방법이 아니라 표현력으로 귀속시키고 32B 확장을 정당화했다. 성분별 기여는 엣지를 동일 모델 A/B로, TTA를 K-곡선(56.9→60.0% 포화)으로, 인코딩을 동일 홀드아웃 페어드 비교로 분리했다.

Figure 2. 짧은 캡션 병목: 진단과 세대별 반응



(a) 오류율을 결정하는 것은 프레임의 시각 유사도가 아니라 캡션의 순서 정보량이며, 절벽 경계는 약 25단어다. (b) 전체 EM(청색)과 짧은 캡션 버킷(적색)의 세대별 정확도: 증강 및 비전 학습이 버킷을 증가시키고, 32B가 벽을 넘고, chrono 인코딩은 전체 동률인 채 이 버킷만 +2.4%p 올린다.

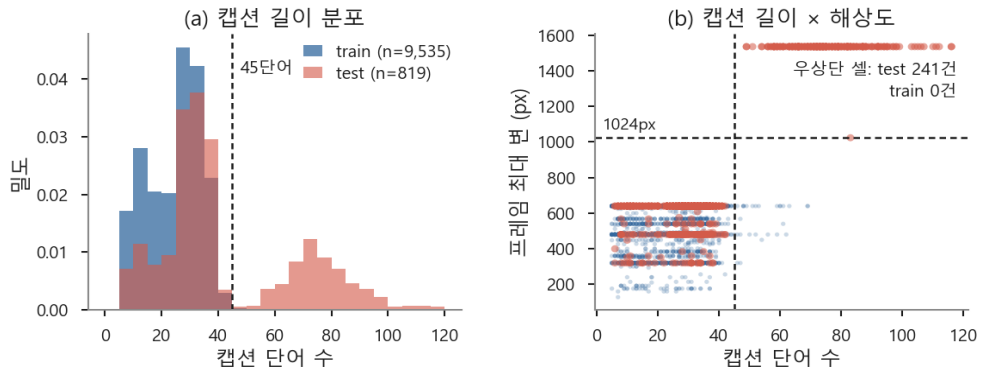
4.3 오류 분석: 짧은 캡션 병목의 해부

짧은 캡션 오답을 채점 때 저장해 둔 25후보 점수(Run C 캐시)로 분해하면 세 부류로 나뉜다. 첫째는 정답이 아깝게 2위에 그친 경우(경계 풀, 4.6%p)로, 1위와의 점수 차 중앙값이 0.24 nat에 불과해 채점 커널의 bf16 잡음과 같은 수준이다. 둘째는 네 번의 TTA 제시 어디에서도 정답이 1위에 오르지 못한 경우(무신호 풀, 15.3%p)로, 모델 안에 정답을 뒷받침할 신호 자체가 없다. 셋째는 순서불가 문제를 순열로 답한 경우(미탐지, 3.4%p)다. 순서불가 재현율은 긴 캡션 77.4%, 짧은 캡션 33.3%로 캡션 길이에 크게 의존했다. 인과를 확인하기 위해 모델이 맞히던 긴 캡션 표본에서 프레임과 정답은 그대로 두고 캡션만 연결어 경계에서 잘라 재채점했더니, 탐지율이 100%에서 25.4%로 떨어졌고 자르지 않은 대조군은 100%를 유지했다. 이 모델의 순서불가 판단은 프레임이 아니라 캡션과 프레임의 대조에서 나오며, 캡션이 짧으면 그 대조가 성립하지 않는 것이다. 같은 실험에서 순서가 있는 표본은 캡션을 잘라도 정답이 상위 2위 안에 67% 남아, 시각 채널에 아직 실현되지 않은 신호가 있음도 확인됐다.

4.4 평가 분포의 이질성

라벨 없이 입력 통계만 점검한 결과, test 819건 중 241건(29.4%)은 고해상도(1024×1536)이면서 45단어 이상의 장문 캡션인데, train에는 이런 표본이 하나도 없다(Figure 3). 즉 test의 3할은 train과 다른 공정에서 왔고, 홀드아웃은 이 부분집단을 원리상 담을 수 없으므로 절대 성능의 예측기가 아니라 상대 비교 도구다. 다만 이 부분집단은 만장일치율과 마진 지표상 오히려 쉬운 쪽이고, 위험 항목의 93%는 train과 동종인 나머지 578건에 있다.

Figure 3. 학습·평가 입력 분포의 이질성



4.5 기각된 시도

Run C 이후에도 성능을 더 끌어올릴 만한 방법들을 같은 기준으로 하나씩 시험했다. 오답의 상당수는 정답(1위)과 근소한 차이였기 때문에, 상위 2개 중에서만 제대로 골라도 이론상 +13%의 가능성이 있었다. 그러나 어떤 방법도 채택기준을 넘지 못했다. 저장된 점수를 다르게 해석하는 CPU 후처리 12종은 전부 효과가 없었고, 프레임을 배열해 놓고 캡션의 우도를 매기는 역방향 재순위도 신호가 없었다. 두 후보 배열을 나란히 보여주고 고르게 한 A/B test는 후보가 갈린 문제의 77.5%를 맞혀 신호는 분명했지만 전체 정확도는 +0.5%밖에 오르지 않았다. 판정 규칙의 남은 조합(+0.2%p)과 TTA 횟수 확장도 마찬가지였다. 훈련 쪽에서는 틀리는 유형을 더 자주 보여주는 오버샘플링이 효과가 없었고, 후보끼리 경쟁시키는 listwise 손실은 8B에서는 통했지만 32B에서는 구현의 수치 오차(0.7 nat)가 신호(0.33 nat)보다 컸으며, 캡션을 잘라 학습시키는 절단 증강은 프로브에서 신호를 확인하고도 실제 학습에서는 효과가 나지 않았다. 결국 남은 격차는 추론 방법의 문제가 아니라 캡션 정보량과 모델 크기의 문제였다.

5. 방법론의 장점과 한계

장점은 세 가지이다. 첫째, 후보 전수 채점이라 무효 출력이 구조적으로 없고, 파인튜닝이 최적화하는 양이 추론에서 비교하는 양과 같아 학습이 지표에 그대로 정렬된다. 둘째, 엣지 후처리는 예측이 제약과 모순일 때만 개입하고 오교정 확률이 정밀도(99.84%)로 묶여 있어 하방이 유계하다. 실제로 종료 후 공개된 private에서도 짝지어진 모든 비교에서 개선이 유지됐다. 셋째, 외부 데이터, API, 앙상블 없이 RTX 3090 한 대 예산 안에서 홀드아웃 exact match를 49.5%에서 61.6%로 올렸다.

한계로는 우선 홀드아웃은 train의 부분집합이라 다른 분포를 담지 못하므로, 상대 비교에는 유효하지만 절대 성능 예측에는 쓸 수 없다. 엣지는 train의 재편집 구조에 의존해 생성 공정이 다른 데이터에서는 커버리지가 줄어든다. 또한, 인코딩 지렛대는 학습 난이도를 재배치할 뿐 캡션의 정보량 자체를 늘리지 못해 집계 정확도는 동률에 그친다.

마지막으로 방법이 아니라 교훈 하나를 기록한다. 개발 내내 리더보드를 의사결정에서 배제했지만 최종 제출 하나만은 public 최고치로 골랐고, 종료 후 공개된 private 점수는 사전 규칙(final 체크포인트 채택)을 따랐을 때보다 낮았다. 반대로 규칙대로 지킨 선택들은 모두 표본 외에서 유지됐다. 후보가 몇 개 되지 않아도 선택 편향은 작동한다는 것을 깨달은 값진 경험이었다.