HCR-HP2: 순열 우도와 부모 분포 보존 학습을 이용한 네 프레임 시간 순서 복원

초코송이

초록

본 과제는 캡션과 순서가 섞인 네 장의 이미지로부터 전체 시간 순서를 복원하며, 네 위치가 모두 맞아야 정답으로 인정된다. 우리는 Qwen3-VL-8B-Instruct와 rank-32 LoRA 어댑터 하나를 사용하고, 자유 생성 대신 24개 합법 순열의 조건부 로그우도를 전수 비교하는 구조화 추론으로 문제를 정식화했다. 추가 학습은 시간 단서가 약한 표본을 다시 다루는 Hard-Caption Replay (HCR)와, 부모 분포를 보존하면서 정답이 상위 후보에 머문 사례만 국소적으로 보정하는 HP2로 구성했다. Public/Private Exact Match 정확도는 24순열 기준선 0.83595/0.82520, HCR 0.84816/0.83333, HP2 0.85340/0.83739였으며, 두 분할의 합산 정답 수는 682건에서 691건, 695건으로 증가했다. 동일한 모델과 어댑터의 공식 RTX 3090 실행은 819건을 2시간 43분 45초에 처리했고 최대 8,056 MiB를 사용했다.

1. 문제 정의와 핵심 기여

입력 x는 자연어 캡션과 무작위 순서의 이미지 네 장이고, 정답은 입력 위치 1-4의 순열 y이다. 가능한 답은 24개뿐이다. 평가 지표가 Exact Match이므로 일부 위치의 독립 분류보다 네 프레임의 공동 순서를 직접 비교하는 편이 문제 구조에 맞다.

자유 생성은 같은 순서를 여러 문자열로 표현하거나 숫자를 빠뜨리고 반복할 수 있다. 또한 캡션의 시간 단서도 균일하지 않다. then, before, after처럼 전개가 명시된 문장과 달리, 짧거나 미리 정한 시간 표현 어휘가 없는 문장은 이미지 사이의 상태 변화에 더 의존한다. 제공 학습 데이터 500건의 고정 점수 캐시에서는 정답 순위가 2-6위인 사례가 171건이었다. 이 세 문제를 각각 구조화 추론, HCR, HP2로 다루었다.

본 접근의 핵심 기여는 다음 세 가지다.

- 출력 공간을 24개 합법 순열로 고정해 형식 오류를 제거하고 Exact Match와 같은 단위로 후보를 비교했다.
- HCR은 제공 학습 데이터 안에서 시간 표현이 약한 표본을 다시 보여 주고 캡션 드롭아웃을 적용해 프레임 변화도 활용하도록 유도했다.
- HP2는 별도 교사 모델이나 후처리 모델 없이 HCR 체크포인트의 동결된 점수 분포를 보존 기준으로 삼아 상위 후보의 순위만 작게 조정했다.

2. 방법

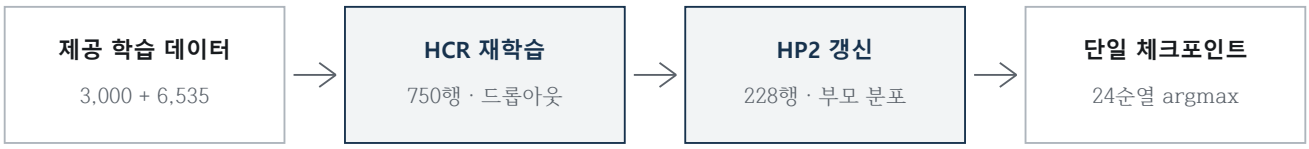

그림 1. HCR-HP2의 학습과 추론 경로. HCR 체크포인트에서 HP2를 연속 갱신하며, 추론에는 최종 LoRA 하나만 사용한다.

2.1 24순열 조건부 우도 추론

각 후보는 동일한 이미지-캡션 프롬프트 뒤에 답 문자열로 붙는다. 후보 순열 y의 점수는 답 토큰의 조건부 로그확률 합이며, 24개 중 최고점을 출력한다.

$$s_\theta(y|x) = \Sigma_t \log p_\theta(y_t \mid x, y_{<t}), \quad \hat{y} = \arg\max_{y \in S_4} s_\theta(y|x)$$

모든 후보가 S_4에서 나오므로 출력에는 항상 1, 2, 3, 4가 한 번씩 들어간다. 계산량은 후보 수에 비례하지만 네 프레임에서는 24개 후보 평가로 고정된다. 기반 모델은 Qwen/Qwen3-VL-8B-Instruct의 revision 0c351dd01ed87e9c1b53cbc748cba10e6187ff3b이다[1]. 기반 가중치는 NF4 4비트로 고정하고 rank=32, alpha=64인 LoRA[2]만 BF16으로 학습했다. 이는 QLoRA[3]의 메모리 절감 설정을 따른다.

2.2 HCR: 시간 단서가 약한 표본의 재학습

제공된 학습 데이터 9,535건을 한 어댑터에 연속 학습했다. 처음 3,000건에서 과제 형식을 익히고 나머지 6,535건으로 범위를 넓힌 뒤, 마지막에 선택한 750건을 다시 학습했다. 세 단계는 각각 1 epoch이다.

표 1. HCR 단계별 학습 설정

단계	행 수	학습률	최대 이미지 픽셀	캡션 드롭아웃
초기 적응	3,000	5×10^{-5}	100,352	0
연속 학습	6,535	3×10^{-5}	100,352	0
HCR 재학습	750	5×10^{-6}	50,176	0.5

HCR의 난이도 조건은 영문 단어 수가 18개 미만이거나, {then, before, after, finally, first, next, while} 중 어느 어휘도 포함하지 않는 경우다. 이 조건을 만족한 3,692건과 나머지 행의 25%를 ID 해시로 추출해 재학습 후보 5,155건을 구성했다. 해시 추출에는 seed 20260711을, 데이터 분할·연속 학습과 최종 750건의 표본 추출에는 seed 42를 사용했다. 최종 재학습 집합은 난이도 표본 548건과 대조 표본 202건으로 구성해 특정 캡션 유형에만 치우치지 않도록 했다.

마지막 단계에서는 캡션을 확률 0.5로 비웠다. 텍스트가 약할 때 프레임 변화도 사용하도록 유도하려는 설정이다. 재학습 표본 선택과 캡션 드롭아웃을 분리한 통제 실험은 수행하지 않았으므로, HCR은 두 요소를 결합한 학습 단계로 해석한다.

2.3 HP2: 부모 분포를 보존하는 국소 보정

HP2는 HCR 뒤에 적용한 근접(proximal) 보정 단계의 내부 명칭이다. 행과 후보는 3,000행 초기 적응 체크포인트가 제공 학습 데이터 500건을 채점한 캐시에서 선정했다. 정답이 1위인 대조 표본 57건과 정답 순위가 2-6위인 난이도 표본 171건을 골랐고, 각 행에는 정답과 상위 오답 두 개만 남겼다. 부모 기준 점수는 첫 파라미터 갱신 전에 최종 HCR 체크포인트로 다시 계산했다. 따라서 초기 적응 체크포인트는 행과 후보 선택에만 쓰이고, 학습 목표의 부모 분포는 최종 HCR에서 나온다.

세 후보 집합 C(x)에 대한 동결 부모 분포와 현재 모델 분포를 다음과 같이 정의한다.

$$p_0(y|x) = \mathrm{softmax}_{y\in C(x)}(s_{\theta_0}/T), \quad p_\theta(y|x) = \mathrm{softmax}_{y\in C(x)}(s_\theta/T), \quad T = 1$$

대조 표본은 부모 분포를 그대로 목표로 삼고, 난이도 표본에서는 정답 원-핫 벡터를 0.15만 섞는다.

$$q_{\mathrm{control}} = p_0, \quad q_{\mathrm{hard}} = 0.85p_0 + 0.15\ \mathrm{onehot}(y^*)$$

주 손실은 q와 현재 세 후보 분포 사이의 교차엔트로피(cross-entropy)다. 정답 점수가 부모보다 낮아질 때만 제곱형 앵커 패널티를 더했다.

$$\mathscr{L} = -\Sigma_{y\in C(x)}\, q(y)\log p_\theta(y) + 0.1\,[\max(0,\, s_{\theta_0}(y^*) - s_\theta(y^*))]^2$$

표 2. HP2 학습 설정

설정	값	설정	값
선택 캐시	제공 학습 데이터 500건	최종 선택	228건
대조 / 난이도 표본	57 / 171	행별 후보 수	3
학습 반복 / 최적화 단계	1 / 29	학습률	10^{-6}
기울기 누적	8	후보 마이크로배치	2
T / 정답 이동량 / 앵커	1 / 0.15 / 0.1	최대 이미지 픽셀	50,176

이 목표는 별도 모델의 지식을 옮기는 방식이 아니라, HCR 체크포인트의 동결된 분포에서 멀어지지 않도록 제한하는 국소 갱신이다. 최종 추론에는 기반 모델과 누적 갱신이 끝난 HP2 LoRA 하나만 사용한다.

3. 실험 및 분석

3.1 단계별 Public/Private 결과

Private 점수는 제출 마감 뒤 공개되었으므로 학습과 모델 선택에는 사용하지 않았다. 공개 점수의 반올림 간격은 Public 573건과 Private 246건에 정확히 부합한다. 표의 괄호 안 정답 수와 합계는 이 간격에서 환산했다. 기준선은 3,000건 초기 적응과 6,535건 연속 학습을 마친 뒤, HCR 재학습을 적용하기 전 체크포인트다.

표 3. 단계별 Public/Private Exact Match 정확도

구성	Public (573건)	Private (246건)	합계 (819건)
기준선 (초기 적응 + 연속 학습)	0.83595 (479건)	0.82520 (203건)	682 / 819
기준선 + HCR	0.84816 (486건)	0.83333 (205건)	691 / 819
HCR + HP2	0.85340 (489건)	0.83739 (206건)	695 / 819

기준선에서 HCR로 이동할 때 Public과 Private의 정답 수는 각각 7건과 2건 늘었고, HCR에서 HP2로 이동할 때는 각각 3건과 1건 늘었다. 두 분할에서 개선 방향은 같았지만 HP2의 Private 추가 정답은 1건이다. 따라서 HP2는 큰 일반화 향상의 증거라기보다, 부모의 상위 후보 구조를 유지하면서 경계 사례를 제한적으로 바꾼 결과로 해석한다.

HP2는 HCR과 비교해 평가 데이터 819건 중 11건의 예측을 바꾸었고, 두 분할의 합산 정답 수는 4건 증가했다. 선정된 228건 중 정답 순위가 2-3위인 행은 108건(47.4%)이었다. 정답이 이미 상위권에 있는 오류의 비율이 세 후보만 사용한 설계의 근거다. 다만 228건은 의도적으로 고른 집합이어서 전체 학습 데이터의 오류 분포를 대표하지 않는다.

3.2 탐색 분기 비교

아래 결과는 부모 체크포인트와 목적 함수가 서로 달라 통제된 절제 실험(ablation)은 아니다. 최종 경로 밖에서 어떤 방향이 유지되지 않았는지 보여 주는 탐색 기록으로만 사용했다.

표 4. 탐색 분기 비교

탐색 분기	Public	Private	관찰
TATI-HCR (시간 방향 잔차 주입)	0.85165	0.83333	HCR 대비 Public +2건, Private 변화 없음
T3 (캡션 절 순서 복원 사전학습)	0.80279	0.78048	두 분할에서 기준선보다 낮음
R-Drop-S4 (대칭 KL 일관성 규제)	0.78010	0.76829	두 분할에서 기준선보다 낮음

TATI-HCR는 HCR보다 Public 정답이 2건 늘었지만 Private에서는 차이가 없었고, HP2의 성능에는 미치지 못했다. 캡션 절을 섞은 뒤 역순열을 예측하는 T3 사전학습과 R-Drop-S4의 대칭 KL 규제는 Public과 Private 모두 기준선보다 낮았다. 이에 따라 최종 시스템은 HCR-HP2의 단일 체크포인트 경로를 유지했다.

3.3 학습 노출 표본에서의 내부 진단

고정한 100건에서 HCR의 Exact Match/top-6 포함률은 0.33/0.78, HP2는 0.34/0.79였고 행별 변화는 개선 1건, 악화 0건이었다. 사전에 정한 통과 기준인 순개선 3건에는 미달했다. 두 모델이 이미 본 학습 표본이므로 독립 검증값은 아니며, 대규모 성능 붕괴가 없는지를 확인하는 내부 진단으로만 사용했다.

4. 효율성과 재현성

운영진의 RTX 3090 서버에서 최종 실행 경로를 819건 전체에 적용했다. 모델 revision, 어댑터 해시, ID 순서와 순열 형식을 실행 전후에 검사했다.

표 5. 공식 RTX 3090 추론 자원 측정

항목	RTX 3090 측정값
처리 행 / 고유 ID / 잘못된 순열	819 / 819 / 0
경과 시간	9,824.59초 (2시간 43분 45초)
최대 VRAM	8,056 MiB
제출 실행 패키지	27.98 GB (26.06 GiB)

환경은 Python 3.10.20, PyTorch 2.5.1+cu121, Transformers 5.13.0, PEFT 0.19.1, bitsandbytes 0.49.2였다. 24 GB VRAM, 24시간, 80 GB 제한 안에서 완주했다. 외부 검증 데이터 3,884건에도 추가 학습 없이 같은 HP2 체크포인트와 추론 코드를 적용했고, RTX 3080에서 26,737.13초(7시간 25분 37초)에 완료했다. 출력은 행 수, ID 순서와 순열 형식 검사를 통과했다. 외부 검증의 전체 점수는 공개되지 않았고 리더보드도 운영진 설명상 전체의 극히 일부만 사용하므로 정량 비교에는 포함하지 않았다.

공식 RTX 3090 출력은 최종 선택한 RTX 4090 출력과 819건 중 12건이 달랐다. 공식 실행의 Public/Private Exact Match는 0.84816/0.84146, 선택 제출은 0.85340/0.83739였다. 즉 RTX 3090 출력은 Public에서 0.00524 낮고 Private에서 0.00407 높았다. 두 출력 모두 구조 검사를 통과했지만 비트 단위로 일치하지는 않았다. 재현 스크립트는 모델 revision, 어댑터 해시, 입력 순서와 출력 형식을 엄격히 검사하되, 서로 다른 지원 GPU 사이의 CSV 해시 일치는 요구하지 않는다.

본 대회를 위한 미세조정(fine-tuning)에는 운영진이 제공한 train만 사용했다. 최종 추론은 기반 모델과 누적 갱신이 끝난 HP2 LoRA 하나로 구성되며, 외부 API, 모델 앙상블, 생성형 모델로 만든 학습 표본이나 교사 모델의 추론 기록(trace), 수작업 테스트 라벨, 테스트 기반 적합 또는 조회를 사용하지 않았다.

5. 한계와 결론

Private에서도 단계별 개선 방향은 유지되었지만 개선 폭은 작아, 이를 폭넓은 일반화 성능으로 확대해석하기 어렵다. HCR의 짧은 캡션·시간 표현 어휘 규칙은 재현하기 쉽지만 암시적 시간 관계를 직접 판별하지 못하며, 재학습 표본 선택과 캡션 드롭아웃의 효과도 분리하지 못했다. HP2 학습은 세 후보의 국소 분포만 다루므로 나머지 21개 후보의 확률 보정(calibration)을 직접 개선하지 않는다. 또한 24개 후보 전수 평가는 네 프레임에서는 고정 비용이지만 프레임 수가 늘면 후보 수가 n!로 증가한다. 12건의 출력 차이는 점수 차가 작은 후보의 순위가 장치별 수치 경로에 민감할 수 있음을 보여 주지만, 원인을 별도로 분리해 확인하지는 못했다.

본 시스템은 네 프레임 순서 복원을 24개 합법 순열의 우도 비교로 정식화하고, 한 LoRA 어댑터를 HCR과 HP2로 연속 갱신했다. 모든 출력이 합법 순열이고 추론 경로가 단일 체크포인트로 끝난다는 점이 구조적 장점이다. Public/Private에서 기준선→HCR→HP2의 방향이 일치했고, 두 분할의 합산 정답 수는 682→691→695건으로 증가했다. HCR은 시간 단서가 약한 표본을, HP2는 상위 후보의 근접 오류를 겨냥한다. 이 두 학습 단계를 Exact Match에 맞춘 24순열 추론과 하나의 체크포인트 안에서 연결했고, 공식 평가 장비의 시간·메모리 제한 안에서 실행됨을 확인했다.

코드와 실행 문서는 GitHub 저장소에, HCR 및 HP2 가중치는 `hp2-final` release에 고정했다. HCR 아카이브는 HP2 학습 재현용이며, 최종 추론에는 HP2 LoRA만 사용한다. 선택 CSV의 SHA-256은 c3574bbb…f7c76f이며 전체 값은 저장소의 configs/artifacts.json에 기록했다.

참고문헌

[1] Qwen Team. Qwen3-VL-8B-Instruct, model card, revision 0c351dd01ed87e9c1b53cbc748cba10e6187ff3b, accessed 2026-07-27.

[2] E. J. Hu et al. “LoRA: Low-Rank Adaptation of Large Language Models.” ICLR, 2022. arXiv:2106.09685.

[3] T. Dettmers, A. Pagnoni, A. Holtzman, and L. Zettlemoyer. “QLoRA: Efficient Finetuning of Quantized LLMs.” NeurIPS, 2023. arXiv:2305.14314.