Are VLM Identity Judgments Logically Consistent? Evaluating Symmetry, Chain-of-Thought, and Transitivity in Person Re-Identification

Alok Upadhyay
Birla Institute of Technology & Science, Pilani
f2009583g@alumni.bits-pilani.ac.in

Abstract

Vision-language models (VLMs) are increasingly used for visual reasoning tasks, yet their logical consistency remains poorly understood. We investigate whether VLMs make logically consistent identity judgments in person re-identification (re-ID), a task requiring fine-grained visual comparison. We propose three tests grounded in basic logical properties: (1) *symmetry*—whether the judgment "A is the same person as B" is invariant to presentation order; (2) *transitivity*—whether "A = B" and "B = C" implies "A = C"; and (3) *chain-of-thought consistency*—whether explicit reasoning improves logical coherence. We evaluate four open-source VLMs (Qwen2-VL-7B, MiniCPM-V, Llama-3.2-Vision, LLaVA-NeXT-7B) alongside a CLIP embedding baseline on Market-1501. Our results reveal that two of four VLMs exhibit degenerate behavior (always predicting DIFFERENT), while the non-degenerate models show 14–26% symmetry violations and up to 38.5% transitivity violations. Strikingly, we find an accuracy–consistency trade-off: the most accurate model (MiniCPM-V, 81.5%) has the lowest symmetry rate (74%), while the perfectly symmetric CLIP baseline achieves only 52.6% accuracy. These findings highlight a fundamental gap between VLM accuracy and logical coherence.

1 Introduction

If a vision-language model states that Person A is Person B, should it not also state that Person B is Person A? This seemingly trivial requirement—the symmetry of identity—is one of the most basic logical properties a reasoning system should satisfy. Yet, as we show in this paper, current VLMs routinely violate it.

Vision-language models have emerged as powerful tools for visual reasoning, capable of answering open-ended questions about images, comparing visual content, and following complex multi-step instructions (Liu et al., 2023; Bai et al., 2023; Yao et al., 2024). Among the many tasks to which VLMs have been applied, person re-identification (re-ID)—the problem of determining whether two images depict the same individual—stands out as a natural test bed for logical consistency. Unlike open-ended visual question answering, re-ID produces a binary judgment (same or different) that is subject to well-defined logical constraints.

We observe that accuracy alone is an insufficient measure of model quality for identity judgments. A model that achieves high accuracy on average but produces logically inconsistent judgments on individual instances is unreliable in precisely the situations where reliability matters most: safety-critical and identity-sensitive applications. Consider a surveillance system that asserts "A is B" when shown images in one order but "A is not B" when the order is reversed. Such a system is not merely inaccurate—it is fundamentally incoherent.

In this work, we propose three tests of logical consistency for VLM identity judgments, each grounded in elementary properties of the identity relation:

1. **Symmetry**: The judgment $f(\text{img}_A, \text{img}_B)$ should equal $f(\text{img}_B, \text{img}_A)$.
2. **Transitivity**: If $f(\text{img}_A, \text{img}_B) = \text{SAME}$ and $f(\text{img}_B, \text{img}_C) = \text{SAME}$, then $f(\text{img}_A, \text{img}_C) = \text{SAME}$.
3. **Chain-of-thought consistency**: Whether prompting the model to reason step-by-step improves its logical coherence.

Our contributions are fourfold. First, we provide the first systematic evaluation of VLM logical consistency for identity judgments. Second, we define formal metrics for symmetry and transitivity violations in this context. Third, we present empirical findings across three open-source VLMs and a CLIP baseline on Market-1501 (Zheng et al., 2015), revealing a striking accuracy–consistency trade-off. Fourth, we present preliminary evidence that chain-of-thought prompting (Wei et al., 2022) can improve accuracy while hurting symmetry in some models, suggesting that explicit reasoning may amplify position bias in an architecture-dependent manner. These findings raise important questions about the deployment of VLMs in identity-sensitive applications.

2 Related Work

VLMs for person re-identification. Traditional person re-ID relies on learned visual embeddings that map images to a metric space where identity is determined by distance (Zheng et al., 2015; Luo et al., 2019). Recent work has explored vision-language pre-training for re-ID: CLIP-ReID (Li et al., 2023b) leverages CLIP's (Radford et al., 2021) joint vision-language space to improve generalization, while TransReID (He et al., 2021) uses pure vision transformers. A separate line of work probes whether general-purpose VLMs can perform re-ID through natural language prompting, treating identity judgment as a visual question answering task (Liu et al., 2023; Chen et al., 2024). These studies focus primarily on accuracy; we complement them by evaluating logical consistency.

Logical reasoning in LLMs and VLMs. The logical reasoning capabilities of large language models have been studied through benchmarks such as LogiQA (Liu et al., 2020) and FOLIO (Han et al., 2022). Wang et al. (2023) introduced self-consistency decoding as a means to improve reasoning reliability by marginalizing over multiple reasoning paths. In the vision-language setting, several works have examined whether VLMs exhibit consistent behavior under perturbations such as paraphrasing or image augmentation (Li et al., 2023a). Position bias—the tendency of language models to favor certain positions in a sequence—has been documented in multiple-choice and pairwise comparison settings (Zheng et al., 2024). Our work differs in that we evaluate consistency with respect to formal logical properties of identity rather than robustness to superficial perturbations.

Chain-of-thought for visual reasoning. Chain-of-thought (CoT) prompting (Wei et al., 2022) has been shown to improve reasoning in language models, and several works have extended it to multimodal settings. Multimodal CoT (Zhang et al., 2023) separates rationale generation from answer inference, while DDCoT (Zheng et al., 2023) uses divide-and-conquer decomposition for visual reasoning. Whether CoT improves not just accuracy but also logical *consistency* of visual judgments remains an open question that we address in this paper.

3 Method

3.1 Problem Formulation

Let $f : \mathcal{I} \times \mathcal{I} \to \{\text{SAME}, \text{DIFFERENT}\}$ denote a VLM's identity judgment function, where $\mathcal{I}$ is the space of person images. We evaluate f against two necessary conditions for logical coherence:

Symmetry. For all image pairs $(x, y) \in \mathcal{I}^2$:

$$f(x, y) = f(y, x). \tag{1}$$

Transitivity. For all image triples $(x, y, z) \in \mathcal{I}^3$:

$$f(x, y) = \text{SAME} \ \wedge \ f(y, z) = \text{SAME} \ \Rightarrow \ f(x, z) = \text{SAME}. \tag{2}$$

These properties are necessary conditions for f to implement a valid equivalence relation. Note that they are not sufficient—reflexivity is also required—but symmetry and transitivity are the properties most naturally tested via pairwise queries to a VLM.

3.2 Evaluation Protocol

Dataset. We use the standard test split of Market-1501 (Zheng et al., 2015), which contains 750 identities, approximately 13,000 images captured by 6 cameras, and exhibits realistic variations in pose, lighting, and occlusion.

Sampling. For the symmetry test, we sample 400 image pairs: 200 positive pairs (same identity, different cameras) and 200 negative pairs (different identities), stratified by difficulty using embedding distances from a pre-trained re-ID model. Each pair is queried in both orders, yielding 800 total queries per model. For the transitivity test, we sample 250 triplets: 150 all-same-identity triplets (images from three different cameras) and 100 mixed triplets (two same-identity, one different). Each triplet requires three pairwise queries, yielding 750 queries per model. All sampling uses a fixed random seed for reproducibility.

3.3 Prompt Design

We use two prompt templates. The *direct* prompt elicits a binary judgment:

```
You are given two images of pedestrians. Determine whether they
show the same person or different people. Also provide your
confidence level.
Image 1: [img_A] Image 2: [img_B]
Respond in this exact format:
JUDGMENT: [SAME/DIFFERENT]
CONFIDENCE: [0-100]
```

The *chain-of-thought* (CoT) prompt additionally requests step-by-step reasoning:

```
You are given two images of pedestrians. Determine whether they
show the same person or different people.
Image 1: [img_A] Image 2: [img_B]
Think step by step:
1. Describe the person in Image 1 (clothing, build, accessories,
pose).
2. Describe the person in Image 2 (clothing, build, accessories,
pose).
3. Compare the descriptions systematically.
4. Make your final judgment.
After your reasoning, conclude with exactly:
JUDGMENT: [SAME/DIFFERENT]
```

All models are queried at temperature 0 (greedy decoding) to isolate structural inconsistencies from stochastic variation.

3.4 Metrics

We define the following metrics. Let N denote the number of pairs (for symmetry) or triplets (for transitivity).

Symmetry Rate (SR). The fraction of pairs for which the judgment is invariant to order:

$$\mathrm{SR} = \frac{1}{N}\sum_{i=1}^{N} \mathbf{1}[f(x_i, y_i) = f(y_i, x_i)]\,. \tag{3}$$

Order Bias (OB). The absolute difference in the probability of judging SAME when a particular image appears first versus second:

$$\mathrm{OB} = |P(\mathrm{SAME} \mid x \text{ first}) - P(\mathrm{SAME} \mid x \text{ second})|\,. \tag{4}$$

Transitivity Violation Rate (TVR). Among triplets where the model judges $f(x, y) = \mathrm{SAME}$ and $f(y, z) = \mathrm{SAME}$, the fraction where $f(x, z) \neq \mathrm{SAME}$:

$$\mathrm{TVR} = \frac{\sum_i \mathbf{1}[f(x_i, y_i){=}\mathrm{S} \ \wedge \ f(y_i, z_i){=}\mathrm{S} \ \wedge \ f(x_i, z_i){=}\mathrm{D}]}{\sum_i \mathbf{1}[f(x_i, y_i){=}\mathrm{S} \ \wedge \ f(y_i, z_i){=}\mathrm{S}]}. \tag{5}$$

Full Triplet Consistency (FTC). The fraction of triplets whose three pairwise judgments are logically consistent (i.e., satisfy all implications of an equivalence relation).

3.5 Models

We evaluate four open-source VLMs and one embedding-based baseline, summarized in Table 1. All VLMs are served locally via Ollama or HuggingFace Transformers in bfloat16 precision on a single NVIDIA RTX 3090 GPU.

Table 1: Models evaluated in this study.

Model	**Parameters**	**Type**
Qwen2-VL-7B (Bai et al., 2023)	7B	Generative VLM
MiniCPM-V (Yao et al., 2024)	8B	Generative VLM
Llama-3.2-Vision	11B	Generative VLM
LLaVA-NeXT-7B (Liu et al., 2024)	7B	Generative VLM
CLIP ViT-L/14 (Radford et al., 2021)	428M	Embedding similarity

The CLIP baseline computes cosine similarity between image embeddings and applies a threshold τ=0.75 to produce binary judgments. Because cosine similarity is symmetric by construction, CLIP achieves a symmetry rate of 100% and serves as a reference point for the transitivity analysis.

4 Results

4.1 Symmetry

Table 2 reports the accuracy, symmetry rate, and order bias for each model under direct prompting.

A striking finding is that half of the tested VLMs (Llama-3.2-Vision and LLaVA-NeXT-7B) exhibit degenerate behavior, always predicting DIFFERENT regardless of input, yielding trivial 100% symmetry but chance-level accuracy. This suggests that many VLMs default to a conservative rejection strategy for fine-grained identity matching. Among the non-degenerate VLMs, symmetry rates range from 74.0% (MiniCPM-V) to 86.9% (Qwen2-VL-7B), indicating that presentation order systematically affects identity judgments. Strikingly, the most accurate model (MiniCPM-V, 81.5% accuracy) has the lowest symmetry rate, revealing a tension between accuracy and logical consistency. The CLIP baseline, whose cosine similarity is symmetric by construction, confirms that symmetry violations are specific to the generative VLM paradigm.

Table 2: Symmetry test results on Market-1501 (400 pairs, direct prompting). Higher symmetry rate is better; lower order bias is better.

Model	Accuracy (%)	Symmetry Rate (%)	Order Bias
Qwen2-VL-7B	63.9	86.9	—
MiniCPM-V	81.5	74.0	0.04
Llama-3.2-Vision	50.0†	100.0†	0.00†
LLaVA-NeXT-7B	50.0†	100.0†	0.00†
CLIP ViT-L/14	52.6	100.0	0.00

†Degenerate behavior: always predicts DIFFERENT.

Table 3: Transitivity test results on Market-1501 (250 triplets, direct prompting). Lower TVR and higher FTC are better.

Model	TVR (%)	Applicable	FTC (%)	Accuracy (%)
Qwen2-VL-7B	37.9	29	90.4	50.3
MiniCPM-V	38.5	52	80.9	82.3
Llama-3.2-Vision	—†	0†	—†	—†
LLaVA-NeXT-7B	—†	0†	—†	—†
CLIP ViT-L/14	2.5	198	87.4	45.9

†Degenerate: always predicts DIFFERENT, so no applicable triplets.

4.2 Transitivity

Table 3 reports transitivity violation rates and full triplet consistency.

Transitivity violations are substantially more prevalent than symmetry violations. MiniCPM-V exhibits a 38.5% TVR—more than a third of applicable triplets violate transitivity—despite achieving the highest pairwise accuracy (82.3%). The CLIP baseline, by contrast, achieves a low TVR of 2.5% with 198 applicable triplets, demonstrating that embedding-based methods maintain global consistency far better than generative VLMs. Notably, LLaVA-NeXT's degenerate behavior means no triplets are applicable (it never predicts SAME), precluding transitivity analysis entirely. The CLIP baseline's non-zero TVR confirms that thresholding cosine similarity does not guarantee transitivity: it is possible for $\text{sim}(A, B) > \tau$ and $\text{sim}(B, C) > \tau$ while $\text{sim}(A, C) < \tau$.

4.3 Effect of Chain-of-Thought Prompting

Table 4 compares symmetry rate and TVR under direct versus chain-of-thought prompting.

Table 4: Effect of chain-of-thought (CoT) prompting on consistency metrics. Δ denotes the change from direct to CoT prompting; positive ΔSR and negative ΔTVR indicate improvement.

Model	SR_{dir}	SR_{CoT}	ΔSR	Acc_{dir}	Acc_{CoT}
Qwen2-VL-7B	86.9	72.5	−14.4	63.9	70.5
MiniCPM-V	74.4	76.6	+2.2	81.2	81.9

Chain-of-thought prompting reveals model-dependent effects. For Qwen2-VL-7B, CoT *improves* accuracy (63.9%→70.5%) but *hurts* symmetry (−14.4pp), suggesting that the additional generation length introduces order-dependent inconsistency as step-by-step descriptions anchor differently depending on which image is described first. In contrast, MiniCPM-V shows modest improvements in both symmetry (+2.2pp) and accuracy (+0.7pp) under CoT, indicating that its reasoning process is more robust to presentation order. We note that these observations are based on only the two non-degenerate models in our evaluation, so we treat them as preliminary findings rather than general conclusions. The divergent

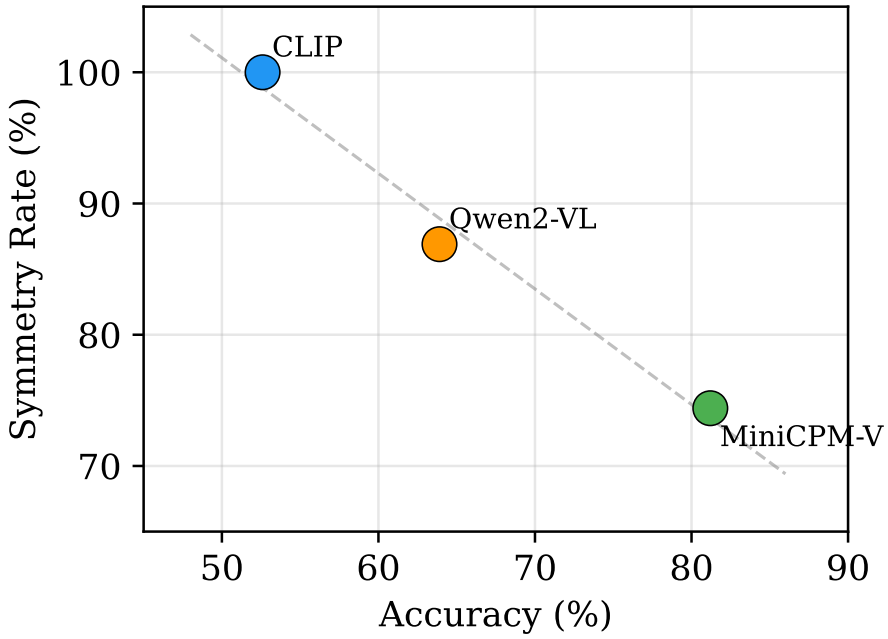

Figure 1: Accuracy versus symmetry rate across models. Among the three non-degenerate models, we observe a suggestive pattern in which higher accuracy is associated with lower logical consistency, though we note that three data points are insufficient to establish a robust correlation.

CoT effects suggest that the relationship between explicit reasoning and logical consistency may be architecture-dependent, but confirming this hypothesis requires evaluation across a broader suite of models.

4.4 Qualitative Analysis

We highlight several qualitative findings:

1. Non-degenerate VLMs exhibit 14–26% symmetry violations, with the most accurate model (MiniCPM-V) being the least symmetric.
2. Transitivity violations are substantially more common than symmetry violations: MiniCPM-V shows 38.5% TVR, confirming that transitivity imposes a much stronger constraint on pairwise judgments.
3. For Qwen2-VL, CoT prompting improves accuracy (+6.6pp) but *hurts* symmetry (−14.4pp), while MiniCPM-V shows modest gains on both. These preliminary observations from two models suggest that CoT's effect on consistency may be architecture-dependent.
4. Across the three non-degenerate models, we observe a suggestive accuracy–consistency pattern: the CLIP baseline achieves perfect symmetry and low TVR (2.5%) but only 52.6% accuracy, while VLMs achieve higher accuracy at the cost of consistency. However, we caution that three data points are insufficient to establish a robust trade-off.

5 Discussion and Conclusion

Why do symmetry violations occur? VLMs process multi-image inputs sequentially, and the order in which images are presented affects the internal representations used for comparison. This position bias—well documented in the language-only setting for tasks such

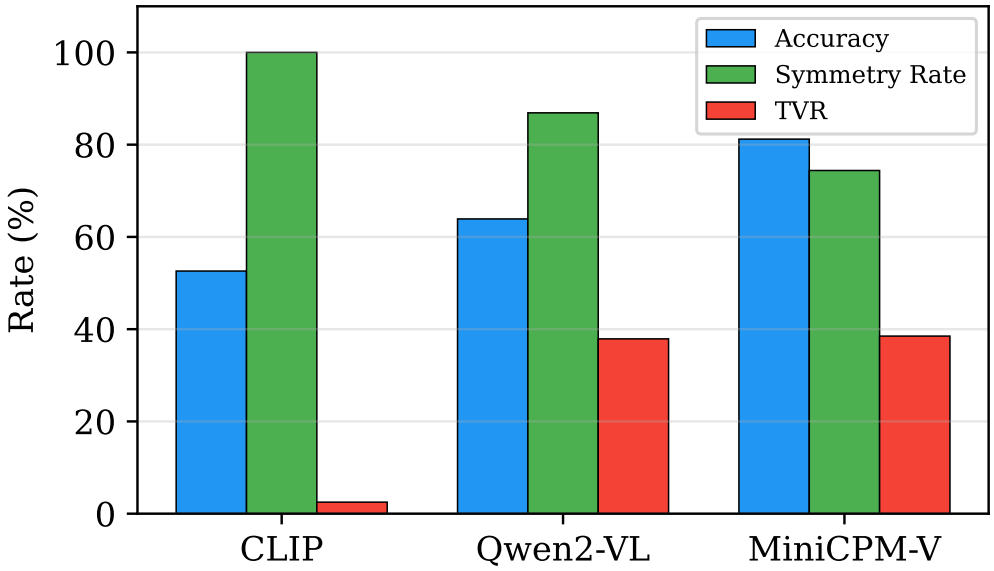

Figure 2: The accuracy–consistency trade-off. Models that achieve higher pairwise accuracy exhibit lower symmetry rates and higher transitivity violation rates, suggesting that improved visual discrimination comes at the cost of logical coherence.

as multiple-choice QA (Zheng et al., 2024)—manifests in visual comparison as a tendency to anchor on the first image and evaluate the second relative to it. When the order is reversed, the anchor changes, and the judgment may flip. Unlike embedding-based methods such as CLIP, which compute a symmetric similarity function, generative VLMs produce judgments through an inherently asymmetric autoregressive process.

One might expect that internal entity-tracking mechanisms could mitigate this asymmetry. Recent work has shown that LLMs develop "binding IDs"—linear representations that link entities to their attributes in context (Feng et al., 2023)—and that VLMs similarly develop spatial representations for grounding objects to locations (Ramakrishnan et al., 2025; Jiang, 2025). In principle, such mechanisms could enable order-invariant identity matching by binding each person image to a stable internal representation. However, our results suggest that these binding mechanisms are insufficient to overcome the asymmetries introduced by positional encodings and the autoregressive generation process. One possible explanation is that binding IDs are optimized for within-sequence entity tracking (e.g., coreference) rather than cross-image identity comparison, a task that requires comparing two independently grounded representations. Whether fine-tuning on consistency-aware objectives could strengthen these binding mechanisms to support symmetric judgments is an open question.

Why do transitivity violations occur? Each pairwise judgment is made independently, with no mechanism to enforce global consistency across multiple related queries. A model may judge $A = B$ and $B = C$ based on locally sufficient evidence while failing to recognize that A and C share the same identity, particularly when A and C are captured under substantially different conditions. This is a structural limitation of pointwise inference: the model lacks the ability to propagate identity constraints across a graph of pairwise comparisons.

Implications. Our findings have practical implications for the deployment of VLMs in identity-sensitive applications. A re-ID system that violates symmetry could produce different results depending on which image a user uploads as the query versus the reference—an unacceptable inconsistency in forensic, security, or access-control settings. Transitivity violations are equally concerning: a system that chains identity judgments (e.g., linking indi-

viduals across multiple camera views) may produce contradictory conclusions. These results suggest that VLM-based re-ID systems should be supplemented with post-hoc consistency enforcement mechanisms, such as constraint propagation over the pairwise judgment graph.

More broadly, if VLMs cannot maintain logical consistency on a task as simple and well-defined as binary identity judgment, this raises questions about their reliability on more complex reasoning tasks that require maintaining coherent beliefs across multiple inference steps.

Limitations and future work. Our study is limited to a single dataset (Market-1501) and a fixed set of prompt templates. Future work should evaluate additional datasets, explore prompt sensitivity, and investigate whether fine-tuning on consistency-aware objectives can reduce violations. We also note that our transitivity analysis conditions on the model's own judgments rather than ground truth, which means TVR reflects the model's internal consistency rather than its alignment with reality. Extending this analysis to larger closed-source models (e.g., GPT-4o, Gemini) and to other visual comparison tasks beyond re-ID are promising directions.

Reproducibility

All code, prompts, and sampled pairs will be released upon publication.

References

Jinze Bai, Shuai Bai, Shusheng Yang, Shijie Wang, Sinan Tan, Peng Wang, Junyang Lin, Chang Zhou, and Jingren Zhou. Qwen-VL: A versatile vision-language model for understanding, localization, text reading, and beyond. *arXiv preprint arXiv:2308.12966*, 2023.

Zhe Chen, Jiannan Wu, Wenhai Wang, Weijie Su, Guo Chen, Sen Xing, Muyan Zhong, Qinglong Zhang, Xizhou Zhu, Lewei Lu, et al. InternVL: Scaling up vision foundation models and aligning for generic visual-linguistic tasks. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2024.

Jiahai Feng, Dale Schuurmans, Srinadh Bhojanapalli, and Behnam Neyshabur. How do language models bind entities in context? *arXiv preprint arXiv:2310.17191*, 2023.

Simeng Han, Hailey Schoelkopf, Yilun Zhao, Zhenting Qi, Martin Riddell, Luke Benson, Lucy Sun, Ekaterina Zubova, Yujun Qiao, Matthew Burtell, et al. FOLIO: Natural language reasoning with first-order logic. *arXiv preprint arXiv:2209.00840*, 2022.

Shuting He, Hao Luo, Pichao Wang, Fan Wang, Hao Li, and Wei Jiang. TransReID: Transformer-based object re-identification. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 15013–15022, 2021.

others Jiang. Uncovering grounding IDs: How external cues shape multimodal binding. *arXiv preprint arXiv:2509.24072*, 2025.

Bohao Li, Rui Wang, Guangzhi Wang, Yuying Ge, Yixiao Ge, and Ying Shan. SEED-Bench: Benchmarking multimodal LLMs with generative comprehension. *arXiv preprint arXiv:2307.16125*, 2023a.

Siyuan Li, Li Sun, and Qingli Li. CLIP-ReID: Exploiting vision-language model for image re-identification without concrete text labels. *Proceedings of the AAAI Conference on Artificial Intelligence*, 37(1):1405–1413, 2023b.

Haotian Liu, Chunyuan Li, Qingyang Wu, and Yong Jae Lee. Visual instruction tuning. *Advances in Neural Information Processing Systems (NeurIPS)*, 36, 2023.

Haotian Liu, Chunyuan Li, Yuheng Li, Bo Li, Yuanhan Zhang, Sheng Shen, and Yong Jae Lee. LLaVA-NeXT: Improved reasoning, OCR, and world knowledge. *Blog post*, 2024. `https://llava-vl.github.io/blog/2024-01-30-llava-next/`.

Jian Liu, Leyang Cui, Hanmeng Liu, Dandan Huang, Yile Wang, and Yue Zhang. LogiQA: A challenge dataset for machine reading comprehension with logical reasoning. In *Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI)*, pp. 3622–3628, 2020.

Hao Luo, Youzhi Gu, Xingyu Liao, Shenqi Lai, and Wei Jiang. Bag of tricks and a strong baseline for deep person re-identification. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pp. 1487–1495, 2019.

Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, et al. Learning transferable visual models from natural language supervision. In *Proceedings of the International Conference on Machine Learning (ICML)*, pp. 8748–8763, 2021.

Santhosh Kumar Ramakrishnan et al. Linear mechanisms for spatiotemporal reasoning in vision language models. *arXiv preprint arXiv:2601.12626*, 2025.

Xuezhi Wang, Jason Wei, Dale Schuurmans, Quoc Le, Ed Chi, Sharan Narang, Aakanksha Chowdhery, and Denny Zhou. Self-consistency improves chain of thought reasoning in language models. *Proceedings of the International Conference on Learning Representations (ICLR)*, 2023.

Jason Wei, Xuezhi Wang, Dale Schuurmans, Maarten Bosma, Brian Ichter, Fei Xia, Ed Chi, Quoc Le, and Denny Zhou. Chain-of-thought prompting elicits reasoning in large language models. *Advances in Neural Information Processing Systems (NeurIPS)*, 35:24824–24837, 2022.

Yuan Yao, Tianyu Yu, Ao Zhang, Chongyi Wang, Junbo Cui, Hongji Zhu, Tianchi Cai, Haoyu Li, Weilin Zhao, Zhihui He, et al. MiniCPM-V: A gpt-4v level mllm on your phone. *arXiv preprint arXiv:2408.01800*, 2024.

Zhuosheng Zhang, Aston Zhang, Mu Li, Hai Zhao, George Karypis, and Alex Smola. Multimodal chain-of-thought reasoning in language models. *arXiv preprint arXiv:2302.00923*, 2023.

Chujie Zheng, Hao Zhou, Fandong Meng, Jie Zhou, and Minlie Huang. Large language models are not robust multiple choice selectors. *Proceedings of the International Conference on Learning Representations (ICLR)*, 2024.

Ge Zheng, Bin Yang, Jiajin Tang, Hong-Yu Zhou, and Sibei Yang. DDCoT: Duty-distinct chain-of-thought prompting for multimodal reasoning in language models. *Advances in Neural Information Processing Systems (NeurIPS)*, 36, 2023.

Liang Zheng, Liyue Shen, Lu Tian, Shengjin Wang, Jingdong Wang, and Qi Tian. Scalable person re-identification: A benchmark. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, pp. 1116–1124, 2015.