



Small Drafts, Big Verdict: Information-Intensive Visual Reasoning via Speculation

Yuhan Liu¹, Lianhui Qin^{2,*}, Shengjie Wang^{1,*}

¹New York University, ²University of California, San Diego

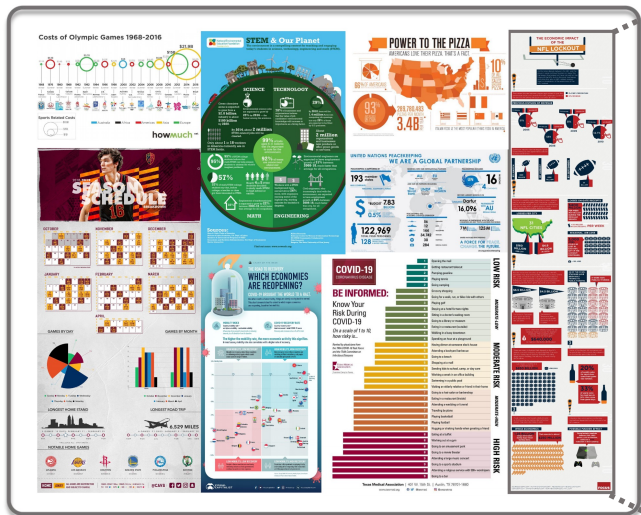


NYU

UC San Diego

Information-intensive visual reasoning is challenging!

Information-Intensive Images



Captions

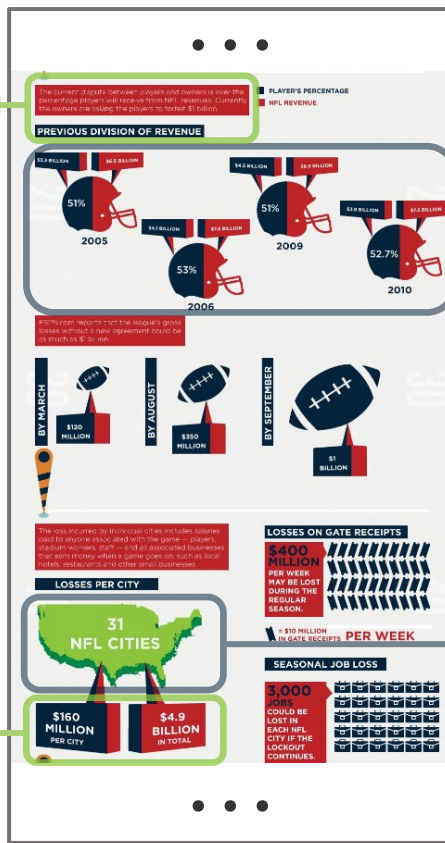


Chart panels

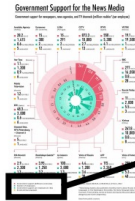
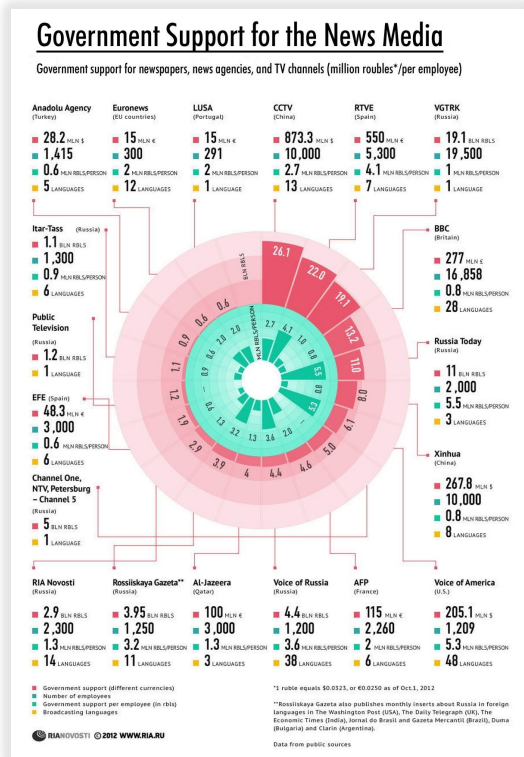
Map

Labels

Information-intensive visual reasoning is challenging!

Question: Which news media has the fewest employees?

Correct reasoning path:



Step 1: Keyword localization

- Government support (different currencies)
- Number of employees**
- Government support per employee (in rbls)
- Broadcasting languages

Step 2: Pattern alignment

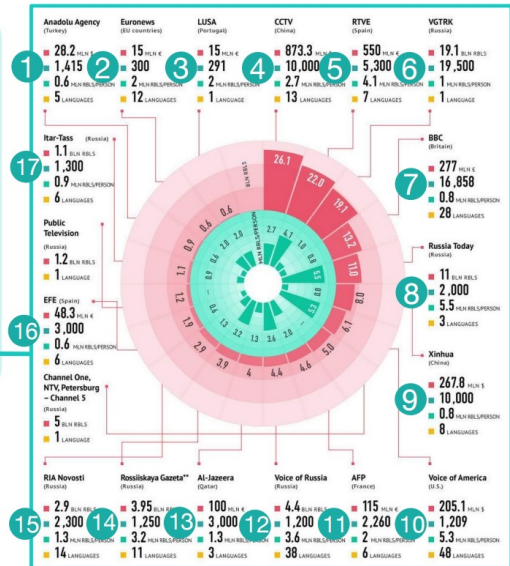


Step 3: Evidence aggregation

1 - 17

Step 4: Sorting for minimum

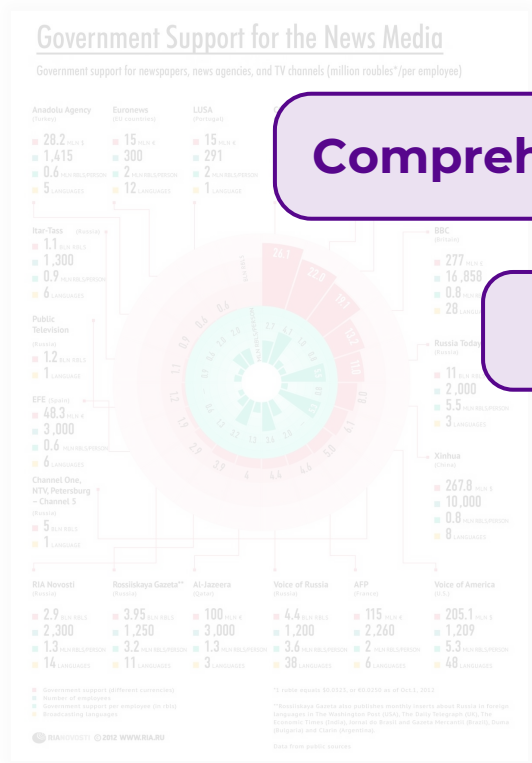
3 LUSA (291)



Information-intensive visual reasoning is challenging!

Question: Which news media has the fewest employees?

Correct reasoning path:

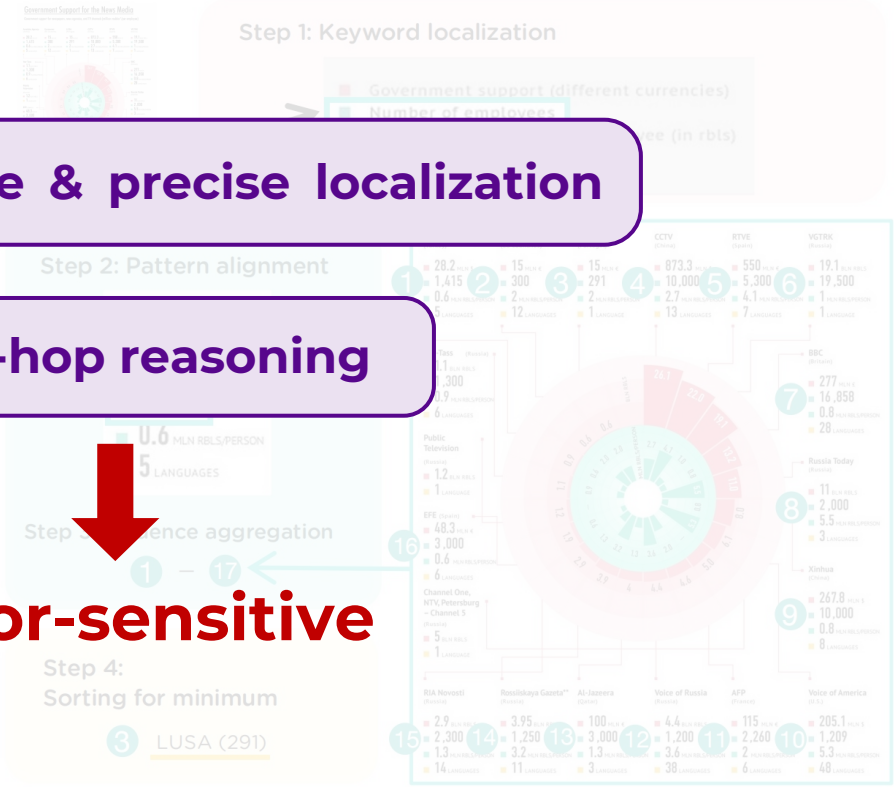


Comprehensive & precise localization

Multi-hop reasoning

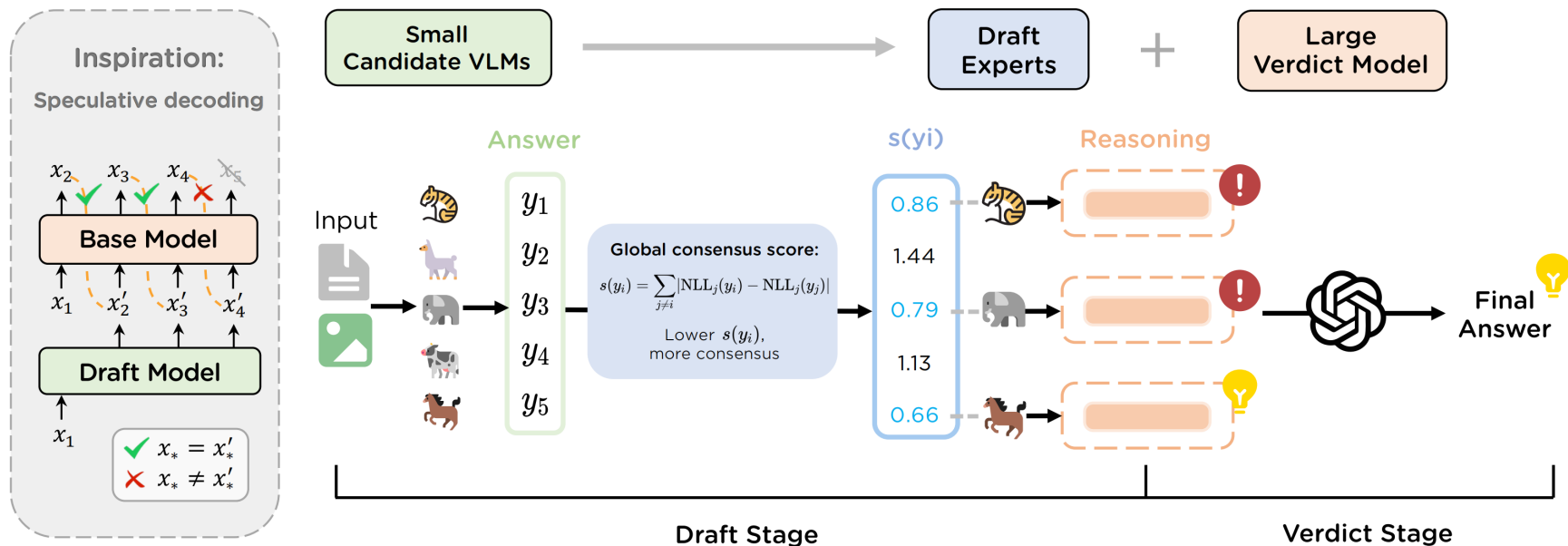


Error-sensitive



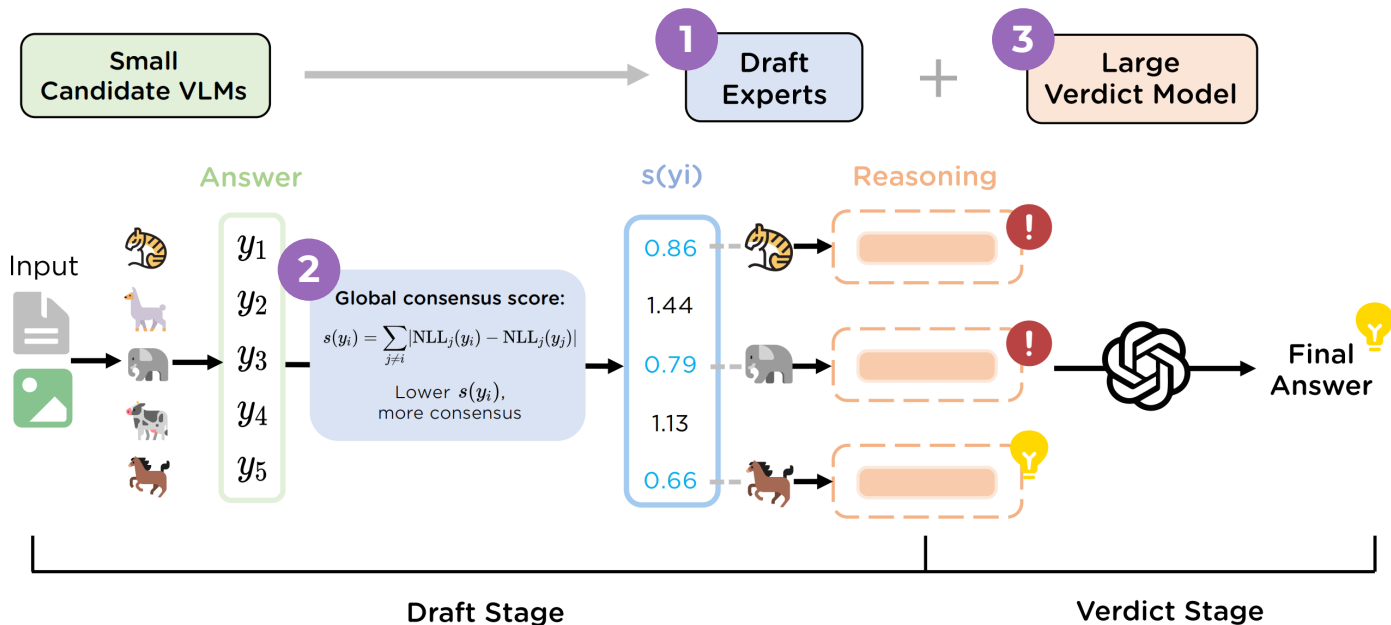
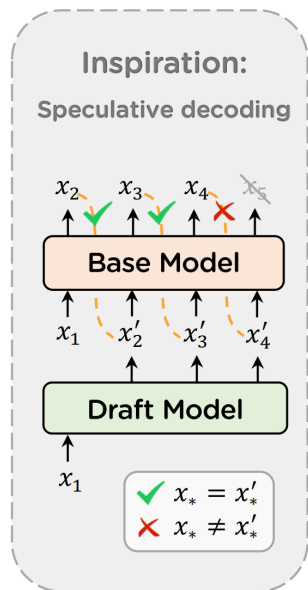
Speculative Verdict (SV)

A training-free framework for information-intensive visual reasoning



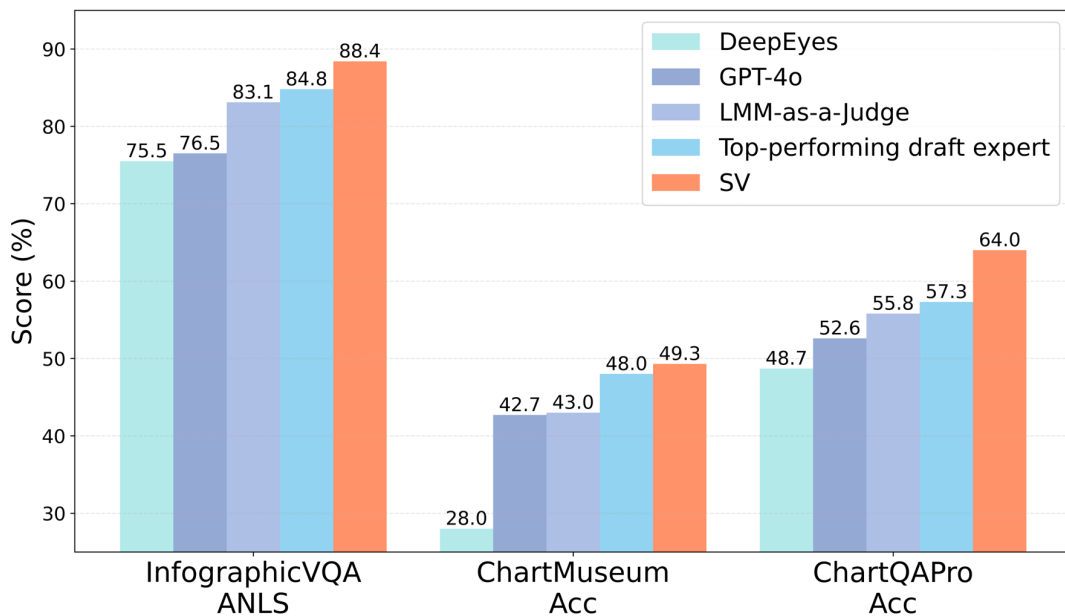
Speculative Verdict (SV)

- 1 Expanded evidence coverage
- 2 Dynamic expert selection
- 3 Error correction

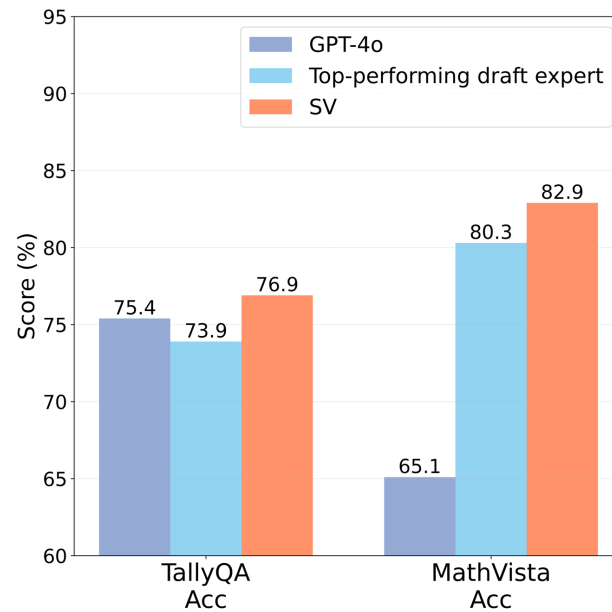


SV significantly outperforms strong baselines

Information-intensive VQA benchmarks

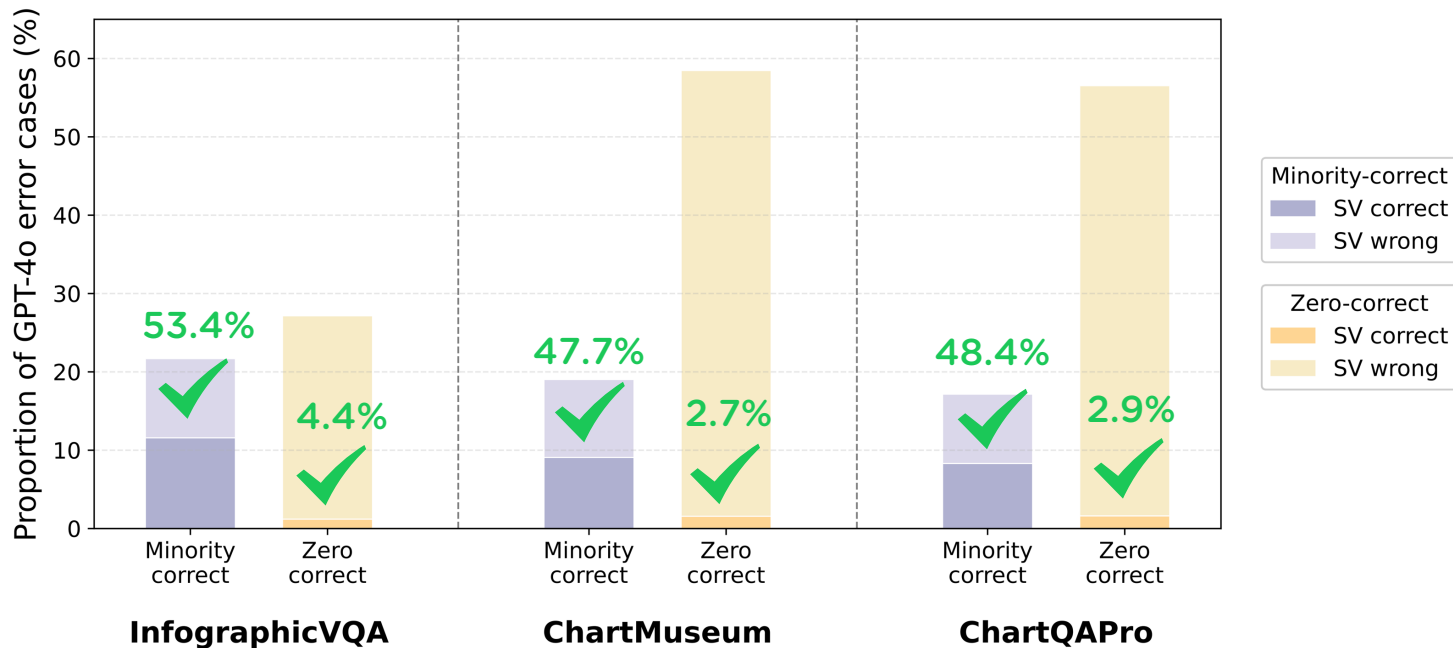


Generalization



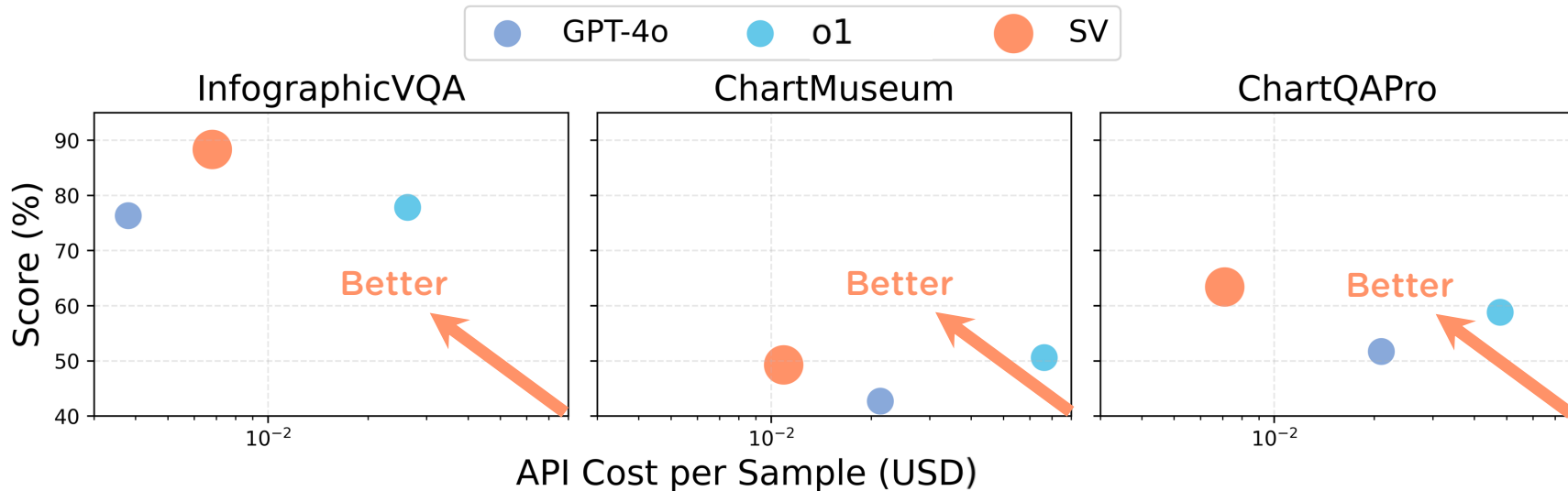
Why does SV work? Error correction ability

Complementary experts → effective synthesis



SV achieves superior cost-efficiency

Matches/outperforms o1 at only 15-26% of its cost



Takeaways

- 1 **Information-intensive visual reasoning** requires:
(1) comprehensive & precise localization, (2) multi-hop reasoning
- 2 SV is a training-free framework that effectively reframes visual reasoning as a **draft-then-verify** process
- 3 **Complementarity** is central to error correction in SV
- 4 **Cost-efficient** paradigm



Paper



Code