AxiomVision: Accuracy-Guaranteed Adaptive Visual Model Selection for Perspective-Aware Video Analytics

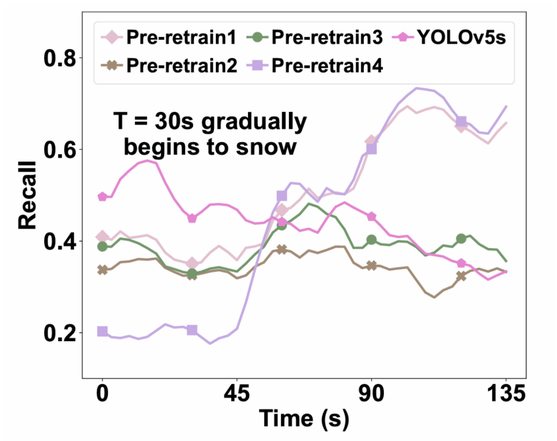
Xiangxiang DAI, Zeyu Zhang, Peng Yang, Yuedong Xu, Xutong Liu, John C.S. Lui

Introduction (Challenges)

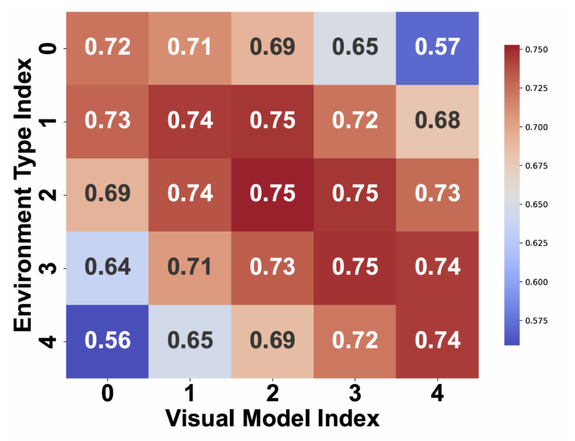
- Which visual model to apply that caters to application-specific requirements under a dynamic environment.
- The video analytic system transitions from single-camera sources to multi-camera feeds.
- The application-specific requirements and adjustable perspective require moving beyond traditional offline learning methods.

Introduction (Contributions)

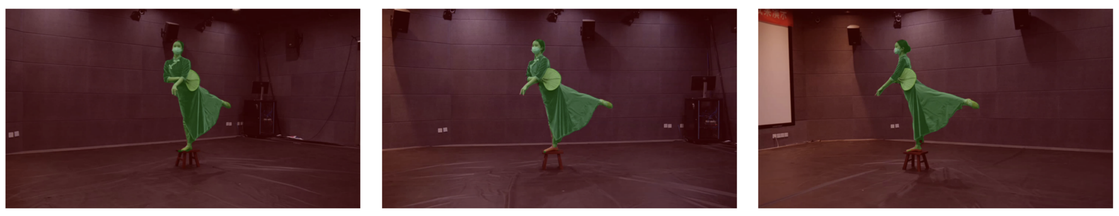
- **Dynamic Visual Model Selection**: Enhance task-specific performance via online learning on an edge-cloud architecture.
- **Camera Network Topology Utilization**: Leverage grouped camera networks to speed up model selection.
- **Camera Perspective Consideration**: Develop a perspective-aware learning method through online sensing.



I). (c) Dynamic performance shifts



(d) Environment and model correlation



II). (a) Perspective 1 (0°) (b) Perspective 2 (30°) (c) Perspective 3 (60°)

Fig. 1: I). Visual model performance across different environmental conditions; II). Semantic segmentation across diverse camera perspectives for the same dancer.

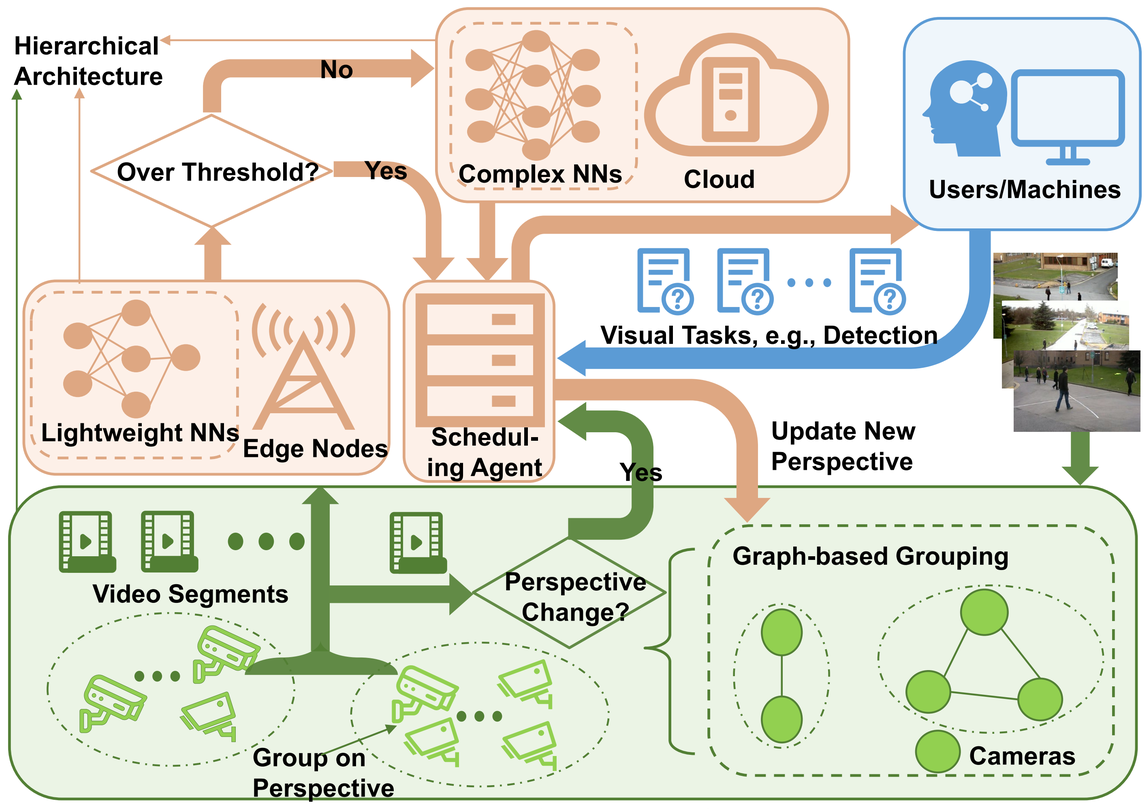


Fig. 2: Overview of AxiomVision framework.

Motivation Observation

- A single universal model faces significant challenges when performing consistently in dynamic environments.
- Perspective v.s. model selection: challenging perspectives, like a distant blurred view in object recognition, require sophisticated models, while simpler perspectives can use basic visual models.

Model & Problem Formulation

- **Hierarchical Setup:** Combines cloud and edge processing.
- **Round-Based Analytics:** Adapts model choice per task.
- **Feedback System:** Adjusts models based on performance.
- **Model Choices:** Selects from multiple models for accuracy.
- **Camera Grouping:** Group by perspective similarity.
- **Goal:** Maximizes performance across all tasks and rounds.

Continual Learning of AxiomVision

Algorithm Design

1. Assigning Inferred Groups for Processed Cameras.
2. Perspective-aware Weight Estimation.
3. Selecting Visual Model with Optimistic Approach.
4. Optimizing Selection for Adaptive Accuracy.
5. Updating Dynamic Graph for Grouping.
6. Adaptive Graph Reconstruction Strategy.

Performance Analysis

- Regret: difference in payoff between the theoretically ideal visual model (not known beforehand) and the visual model chosen.
- Regret Upper Bound : $Reg(T_q) \leq O\left(\sqrt{T_q}\log T_q\right)$, where T_q denotes the total rounds for visual task q.

Performance Evaluation

Setting: Public 360° VR camera feeds from [1, 2, 3, 4]. Rectilinear images are extracted from panorama to function as adjustable perspectives [5]. Visual tasks include *Classification*, *Counting*, *Detection*, and *Aggregation*.

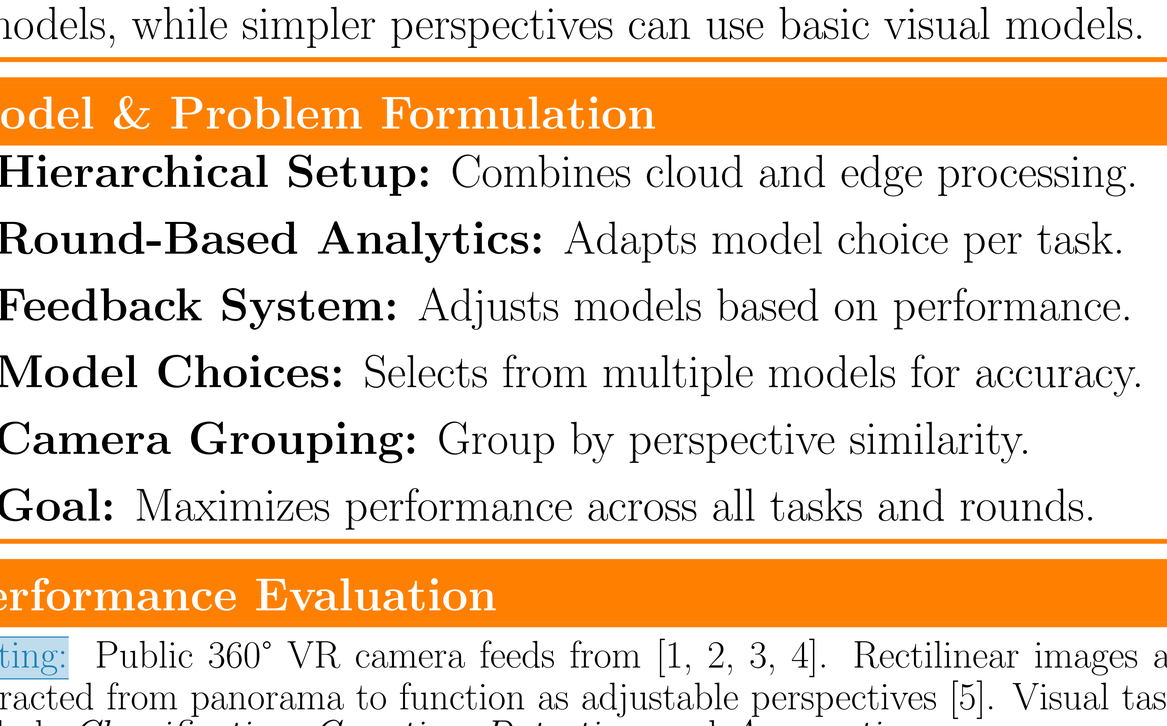


Fig. 3: Accuracy on fixed and adjustable perspectives.

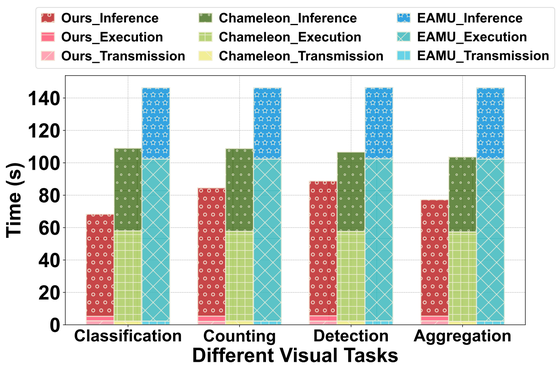


Fig. 4: Decomposition of total time overhead.

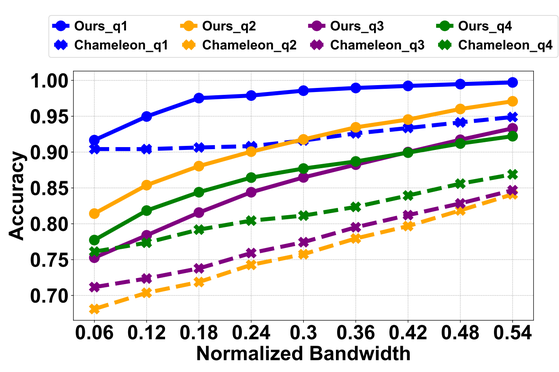


Fig. 5: Accuracy with varying bandwidth.

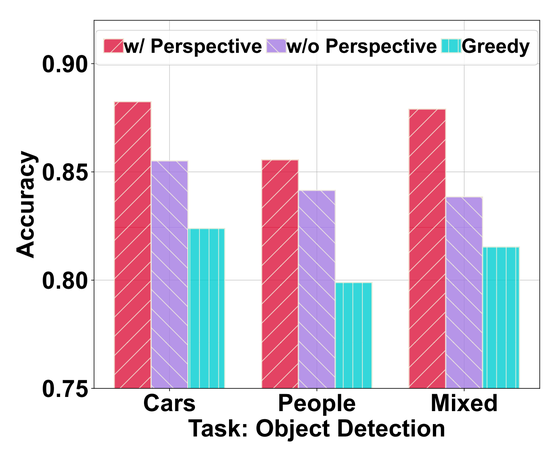


Fig. 6: Perspective impact.

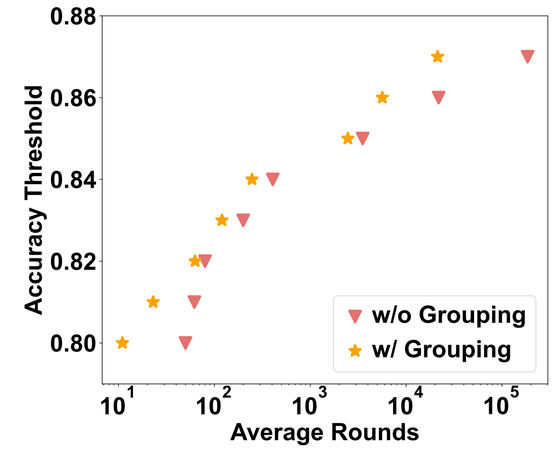


Fig. 7: Grouping cameras for acceleration.

References

[1] 4K Urban Life. 5k vancouver downtown 360° vr video - vancouver harbourfront, canada - 1 hr. [Online] https://www.youtube.com/watch?v=oeKr9O6z4IU, Accessed: March 17, 2024.

[2] 4K Urban Life. Seattle 5k 360° vr video - seattle city north downtown. part #2. [Online] https://www.youtube.com/watch?v=sAMF5BkmO5O, Accessed: March 17, 2024.

[3] 4K Urban Life. Seattle downtown - city tour 360 vr - 4k video. part 1 - 1 hr. [Online] https://www.youtube.com/watch?v=Zy2ihEV-ooI, Accessed: March 17, 2024.

[4] 4K Urban Life. Seattle traffic in 5k 360° vr video - seattle highways & stadiums. [Online] https://www.youtube.com/watch?v=znSzP4R_1a8, Accessed: March 17, 2024.

[5] Ali Lenjani, Chul Min Yeum, Shirley Dyke, and Ilias Bilionis. Automated building image extraction from 360 panoramas for postdisaster evaluation. *Computer-Aided Civil and Infrastructure Engineering*, 35(3):241–257, 2020.