How can temporal information be applied to improve trust in AI techniques for object detection and tracking in drone video streams.

Temporal information plays a crucial role in improving the reliability and trustworthiness of AI-based object detection and tracking in drone video streams. By leveraging temporal cues—patterns and consistencies across video frames—AI systems can achieve more robust, accurate, and explainable results, especially in challenging aerial environments.

Mechanisms for Leveraging Temporal Information

- Temporal Embedding and Association: Integrating temporal embedding structures and temporal-association networks allows models to better represent object trajectories over time, reducing errors from frame-to-frame inconsistencies and improving both detection and tracking accuracy (Lin et al., 2022; Wu et al., 2025; Zhou et al., 2024; Yao et al., 2021).
- Temporal-Spatial Feature Interaction: Combining temporal and spatial features, such as through feedback loops or transformer-based architectures, enables the system to use information from previous frames and multiple viewpoints, enhancing robustness against occlusion, appearance changes, and background clutter (Wu et al., 2025; Zhou et al., 2024; Chen et al., 2023).
- Long-Term Temporal Context: Utilizing long-term temporal information, rather than just adjacent frames, helps the system recover from transient failures and maintain reliable tracking even when short-term cues are unreliable (Chen et al., 2020; Zhou et al., 2024; Fujitake & Sugimoto, 2022).

Trust and Reliability Improvements

- Consistency and Smoothness: Temporal modeling enforces smoother object trajectories and more consistent detections, reducing false positives and tracking drift, which are critical for user trust (Lin et al., 2022; Zhou et al., 2024; Yao et al., 2021).
- Handling Challenging Scenarios: Temporal information helps maintain performance during occlusions, rapid motion, and visual similarity between objects, which are common in drone footage (Wu et al., 2025; Zhou et al., 2024; Yao et al., 2021; Chen et al., 2023).
- Explainability: Temporal attention and memory mechanisms can provide interpretable cues about why the system made certain decisions, further supporting trust (Yadav et al., 2022; Fujitake & Sugimoto, 2022; Chen et al., 2023).

Performance Gains

Method/Approach	Temporal Info Used	Key Trust-Related Gains	Citations
Temporal-Association Network	Temporal embedding	+4.9% MOTA, improved tracking	(Lin et al., 2022)
Temporal-Spatial Feature Interaction	Tracklet feedback loops	+2.76–4.66% MOTA, robust ID	(Wu et al., 2025)
Recurrent Motion Attention	Motion features, optical flow	+6.99% mAP, stable detection	(Zhou et al., 2024)
Cross-Drone Transformer	Multi-drone temporal fusion	Enhanced recovery from lost targets	(Chen et al., 2023)

Conclusion

Applying temporal information in AI models for drone video object detection and tracking significantly enhances trust by improving accuracy, consistency, and robustness, especially in complex aerial scenarios. Temporal modeling not only boosts performance but also provides more reliable and interpretable results, which are essential for building user confidence in AI-driven drone applications.

References

Lin, Y., Wang, M., Chen, W., Gao, W., Li, L., & Liu, Y. (2022). Multiple Object Tracking of Drone Videos by a Temporal-Association Network with Separated-Tasks Structure. Remote. Sens., 14, 3862. https://doi.org/10.3390/rs14163862

Wu, H., Sun, H., Ji, K., & Kuang, G. (2025). Temporal-Spatial Feature Interaction Network for Multi-Drone Multi-Object Tracking. IEEE Transactions on Circuits and Systems for Video Technology, 35, 1165-1179. https://doi.org/10.1109/TCSVT.2024.3478758

Chen, C., Wang, G., Peng, C., Zhang, X., & Qin, H. (2020). Improved Robust Video Saliency Detection Based on Long-Term Spatial-Temporal Information. IEEE Transactions on Image Processing, 29, 1090-1100. https://doi.org/10.1109/TIP.2019.2934350

Zhou, Z., Yu, X., & Wang, X. (2024). Object detection in drone video based on recurrent motion attention. Pattern Recognit. Lett., 183, 56-63. https://doi.org/10.1016/j.patrec.2024.04.029

Yadav, S., Luthra, A., Pahwa, E., Tiwari, K., Rathore, H., Pandey, H., & Corcoran, P. (2022). DroneAttention: Sparse Weighted Temporal Attention for Drone-Camera Based Activity Recognition. Neural networks : the official journal of the International Neural Network Society, 159, 57-69. https://doi.org/10.1016/j.neunet.2022.12.005

Fujitake, M., & Sugimoto, A. (2022). Temporal feature enhancement network with external memory for live-stream video object detection. Pattern Recognit., 131, 108847. https://doi.org/10.1016/j.patcog.2022.108847

Yao, S., Zhang, H., Ren, W., , C., Han, X., & Cao, X. (2021). Robust Online Tracking via Contrastive Spatio-Temporal Aware Network. IEEE Transactions on Image Processing, 30, 1989-2002. https://doi.org/10.1109/TIP.2021.3050314

Chen, G., Zhu, P., Cao, B., Wang, X., & Hu, Q. (2023). Cross-Drone Transformer Network for Robust Single Object Tracking. IEEE Transactions on Circuits and Systems

for Video Technology, 33, 4552-4563. https://doi.org/10.1109/TCSVT.2023.3281557