AI Model Performance for Drone-Based Person Tracking

Project Kestrel

Prepared by the Model Team

July 25th, 2025

---

1. Objective

To develop a lightweight, high-performance AI model capable of reliably re-identifying individuals over at least 15 minutes of drone footage, even after periods of occlusion.

---

2. Our Methodology: State-of-the-Art & Efficient

We are fine-tuning a **MobileNetV3** architecture, a neural network optimized for high performance on edge devices like our **NVIDIA Jetson Nano**. The model is trained using **Triplet Loss**, a technique that teaches it to create a unique "appearance vector" for each individual, allowing for robust matching.

The model was fine-tuned on the large-scale **AG-VPReID dataset**, which contains over **1.47 million images** across ~7,000 tracklets, providing a challenging and diverse training environment.

---

3. The Journey to a Robust Evaluation

A key part of our work was developing a fair and realistic testing protocol.

- **Initial Test (Intra-Camera)**

Our first evaluation, which involved matching near-identical video frames, showed a misleading **100% Rank-1 Accuracy**. While successful, this didn't test the model's ability to handle significant appearance changes, which is crucial for long-term tracking.

- **Rigorous Test (Cross-Camera Evaluation)**

We implemented the industry-standard **cross-camera evaluation**. This tests the model's core ability to re-identify a person when they are viewed from entirely different camera angles and locations—a much more challenging and realistic scenario.

---

4. Final Performance Metrics (Cross-Camera)

After fine-tuning on our dataset, the MobileNetV3 model established the following quantitative baseline on the challenging cross-camera test:

- **mAP (Mean Average Precision): 23.64%**
    - *This measures the overall quality of the ranking for all correct matches.*
- **Rank-1 Accuracy: 35.34%**
    - *The model correctly identifies the person as the top match over one-third of the time.*
- **Rank-5 Accuracy: 48.99%**
    - *The correct person is found within the top 5 results nearly 50% of the time.*

These metrics represent a strong development milestone and a solid foundation for achieving our final performance goals.

---

5. Key Takeaways & Next Steps

1. **Promising Initial Results:** Fine-tuning the MobileNetV3 architecture has proven to be a successful strategy, yielding a significant performance improvement over our initial model.
2. **Validated for Edge Hardware:** The chosen model is lightweight and efficient, confirming its suitability for our target NVIDIA Jetson Nano hardware.
3. **Next Steps:** Our immediate next step is to continue strengthening the model through further hyperparameter tuning and data augmentation experiments. Concurrently, we will begin integrating the current model's embeddings as a similarity cost within the full DeepSORT tracking system to test its real-world effectiveness.