Speech Draft Template for UAV Object Detection Presentation

"Good morning, everyone. We are the UAV Object Detection Group. My name is Hu Kaiwei, and I'm here to deliver a presentation representing my team."

"Today, we will go through our key learnings and progress. Our presentation is divided into four main parts:"

1. **Experiment comparison based on existing research**
2. **Training the model according to paper**
3. **Performance analysis**
4. **Our Confusion**

---

1. Experiment Comparison

---

*"Last week, we trained our models from the YOLO series. We used the YOLO framework, setting the image size to **640**, training for **100 epochs**, and using an initial learning rate of **0.01**. After testing, we obtained the following results:"*

- *The best-performing model was **YOLOv11-Extra**, which achieved:*
    - **mAP50 = 39%**
    - **mAP = 23.7%**

---

*"We also tried training YOLOv11-Extra with **200 epochs**, but the performance did not improve."*

---

*"At first, we followed the instructions from our professor and referred to some research papers. We analyzed the **VisDrone challenge papers** from **2018, 2019, and 2021**. Based on these references, we compared our results with the given methods"*

- **VisDrone 2018:** HAL-RetinaNet and DPNet achieved over **30% mAP**, while our best was **23.7%**.

- **VisDrone 2019:** DPNet-Ensemble achieved **29.62% mAP**, slightly worse than 2018.

- **VisDrone 2021:** DBNet reached nearly **40% mAP**, much higher than ours.

"After this comparison, we realized we had done something wrong in our training setup."

_"We found another research paper analyzing the impact of different YOLO versions on the VisDrone dataset.

This paper highlighted that **image size is a key performance factor**. And it uses four hyperparameters which are image size, learning rate, model version and optimizer. Based on this, we made some adjustments:"_

2. Training Adjustments & Improvements

Increased image size to 1920 while keeping other settings the same (epochs = 100, learning rate = 0.01).
- *New results:* ***mAP50 = 49.9%, mAP = 30.9%*** *(significant improvement).*

Trained YOLOv11-Extra for 150 epochs instead of 100.
- *New results:* ***mAP50 = 50.3%, mAP = 31.3%.***

Reduced learning rate to 0.001 while keeping image size at 1920 and training for 100 epochs.

- *New results: **mAP50 = 53.9%**, **mAP = 33.3%** (our best so far).*

"These adjustments confirmed that increasing image size and fine-tuning the learning rate significantly impact performance especially the image size."

The following is our testing outputs,
This is the normalized confusion matrix
These are F1 confidence curve, Precision-Confidence curve, Precision-Recall Curve, and Recall-Confidence Curve
These are three batch prediction and each of them contains 8 images because we set the batch size to eight.

---

3. Model Analysis

"We notice that image size really matters in drone-based object detection due to three main aspects"

1. **Large-scale variation**
2. **Occlusion**
3. **Class imbalance of the VisDrone dataset**

---

*"We also analyzed feature fusion in YOLOv11. The **neck structure** is designed to extract as much information as possible from the backbone feature maps. However, if the localization and semantic feature quality are poor, the neck will aggregate weak features, leading to poor detection performance."*

*"This is why **increasing image size improves detection accuracy**—it enhances both spatial and semantic feature extraction."*

---

4. Is our Confusion and we need help

*"Even after optimizing our hyperparameters, our best result (**mAP50 = 53.9%, mAP = 33.3%**) is still lower than the results reported in research papers, which reach nearly **mAP50 = 60% and mAP = 40%**."*

So I am quite confused about what possible reasons cause this situation and is our fine-tuned model reasonable?

That's all, thank you so much for listening.