Case Study 1. Multimodal Driver Behavior Monitoring in Autonomous Vehicles

Problem Statement: Design a deep learning system to monitor and classify driver behavior (e.g., alert, distracted, drowsy, aggressive) in real-time using in-cabin video data. Optionally, incorporate biometric signals (e.g., heart rate) and steering telemetry.

Tasks:

- Propose a multimodal architecture that fuses video, sensor data, and physiological signals.
- Justify how temporal information will be modeled (e.g., LSTM, 3D CNN).
- Design a suitable loss function (e.g., multi-class cross-entropy, contrastive loss).
- Specify required data modalities and collection parameters.
- Define the accuracy assessment method (e.g., F1 score, ROC-AUC, balanced accuracy).

Case Study 2. Cross-Domain Medical Image Segmentation with Scarce Labels

Problem Statement: Given a few labeled MR brain scans from one hospital and many unlabeled scans from others (with domain shifts), develop a method to segment tumor regions robustly across domains under label scarcity.

Tasks:

- Propose a model architecture that supports domain generalization or domain adaptation.
- Explore data normalization, style transfer, or adversarial techniques to handle domain shift.
- Formulate a suitable semi-supervised or domain-adaptive loss function.
- Specify the imaging modalities used (e.g., T1, T2, FLAIR).
- Define accuracy metrics such as Dice coefficient, Hausdorff distance, or average surface distance.

Case Study 3. Fine-Grained Action Recognition in Sports Videos

Problem Statement: Given high-resolution tennis match videos, build a system that recognizes fine-grained player actions (e.g., serve, forehand, volley), localizes them temporally, and handles camera and player variability.

Tasks:

- Propose a spatiotemporal model (e.g., SlowFast network, 3D CNN + RNN).
- Optionally use pose estimation or object tracking as auxiliary signals.
- Design a multi-task loss for action classification and boundary localization.
- Specify data requirements (e.g., frame rate, annotations).
- Evaluate performance using mAP, temporal IoU, or action-wise F1 score.

Case Study 4. Visual Grounding for Remote Drone Navigation in Disaster Zones

Problem Statement: Design a system that enables drones to interpret natural language navigation commands (e.g., "fly to the red building past the collapsed tower") using real-time video, GPS, and IMU data in disaster zones.

Tasks:

- Propose a vision-language-spatial model (e.g., CNN-LSTM-based grounding + policy head).
- Design appropriate loss functions (e.g., contrastive loss, alignment loss).
- Specify required modalities: onboard video, textual instructions, GPS/IMU data.
- Define evaluation metrics: grounding accuracy (IoU), navigation success rate, and goal-reaching time.