ESTd 2020

Acme AI Ltd. Technical Test

Core task is to develop, train, and benchmark a lightweight, temporal computer vision model that recognizes jersey numbers from image (player bounding box) sequences. Crucially, the model must be architected so that it can be generalised later to all two-digit numbers (00-99), despite the limited training data.

Context, Data Structure, and Requirements

Objective: Reliable, temporal stable recognition of jersey numbers.

Constraint: The model must be rather lightweight (avoid massive LLM kind of networks).

Data Structure:

- The dataset here is organized into folders based on jersey number classes. URL to ZIP-file: https://drive.google.com/file/d/183CXUk99IyCRgfAPeRnpo5xBYCER1-an/view?usp=sharing
- Inside the class folders, there are 10 subfolders representing **temporal sequences** (bounding box of a player across several frames).
    - **IMPORTANT:** we only included 10 folders with limited data to enable you to train a model without needing a lot of resources.
- Each sequence folder contains:
    - Multiple .jpg files: The individual bounding box crops for the player over time.
    - anchor.jpg: The single frame image with potentially the clearest view of the jersey number **(only for use in training)**.

Extra:

- **Temporal Modeling Requirement:** The model must use the sequence of images (not just the anchor.jpg) to produce a final, stable prediction.
    - *Why this is crucial:* Jersey numbers can be partially occluded, blurry, or rapidly changing between frames. Using only a single-frame prediction leads to unstable output. A temporal model (like a recurrent or attention-based network) can use information from *previous and subsequent frames* to produce a more stable and accurate prediction for a specific player's tracking ID.

Generalization Requirement: Although trained on only the given jersey number data, the model structure must be able to be adapted to support the prediction of any number from 00 to 99. Example: Now it predicts digits and then classifies the number. Soon we need to be able to predict numbers that weren't in the train classes (training data has '4', '9' and '49', so later we need to also be able to detect '94' even though we don't have this class in the training data).

Deliverables and Expectations

The candidate must submit the following:

Technical Implementation (Code)

- *Codebase:* A complete, runnable solution for data loading, training, and evaluation.
- *Input:* The model needs to be trained and tested with the provided data
- *Output:* A single, stable, number prediction.
- *Temporal Architecture:* Implementation of a temporal component

Acme AI Ltd.
Address
H #385, R #6, Mirpur DOHS, Dhaka 1216, Bangladesh
E: info@acmeai.tech
W: www.acmeai.tech

- **Generalization Strategy:** The solution must be designed to recognize numbers from 00 to 99, despite training data only having 10 classes. Candidates must demonstrate how their model structure (e.g., predicting two independent digits instead of a single 100-class label) achieves this generalizability.
- Provide a summary of trained weights

Comprehensive Report (PDF or Markdown)

The report should act as a technical communication document, covering:

- Model Architecture & Strategy:
  - *Network Diagram:* A clear, detailed diagram of the proposed network, showing the flow from the input image sequence to the final two-digit output.
  - *Generalization Explanation:* In-depth justification for using the two-digit prediction method and how it allows the model to correctly identify numbers it hasn't seen during training (e.g., '10', '23', '77', etc.).
- Training & Optimization:
  - *Data Loading*: Description of how the temporal sequence and the anchor.jpg (if used) are handled.
  - *Optimization:* Detailed summary of techniques used to keep the model lightweight
- Results & Analysis:
  - *Graphs:* Training/validation loss/accuracy curves.
  - *Quantitative Results:* Report on the test set performance using the provided test folder.
- Conclusion & Next Steps:
  - *Interpretation:* Analysis of temporal stability. Did the model perform better on the sequence than on the anchor.jpg alone? Why?
  - *Actionable Next Steps:* Specific recommendations for improving the solution, focusing on scaling to the full 0-99 dataset and further reducing latency.

AI Usage Disclaimer

- Use of AI tools (e.g., Copilot, ChatGPT, Gemini, etc.) is allowed by upto 60%.
- If you use any AI-generated content (e.g., for mock responses or code snippets), you must cite it clearly in your submission.
- We use AI detection tools to verify originality so any AI-generated imprint without disclosure will result in negative markings and/or disqualification.

Submission guideline: All tasks, code files, and written responses must be compiled into a single PDF document. Please include any external references as hyperlinks within the document.

File name format: Name your submission file as: **SCVE-AAI-[Your Full Name]**

Submission link: Once completed, upload the PDF to the Google Form linked **here.**

Deadline: Submissions must be lodged by 13 December 2025, by 11.59 PM

Acme AI Ltd.

Address
H #385, R #6, Mirpur DOHS, Dhaka 1216, Bangladesh

E: info@acmeai.tech
W: www.acmeai.tech