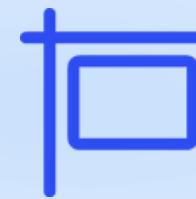


Hand Gesture Recognition

Developing robust deep learning systems for interactive applications.

Revolutionising Interaction: Key Applications

Our advanced gesture recognition system opens doors to transformative applications, enhancing how humans interact with technology and each other.



Sign Language Translation

Bridging communication gaps by accurately interpreting sign language in real-time.



Human-Computer Interaction

Enabling intuitive control of devices and systems through natural hand movements.



AR/VR Gesture Control

Providing immersive and seamless control within augmented and virtual reality environments.

Dataset Overview and Preprocessing Pipeline

A meticulously curated dataset underpins our model's performance, ensuring robustness and generalisability across various gesture classes.



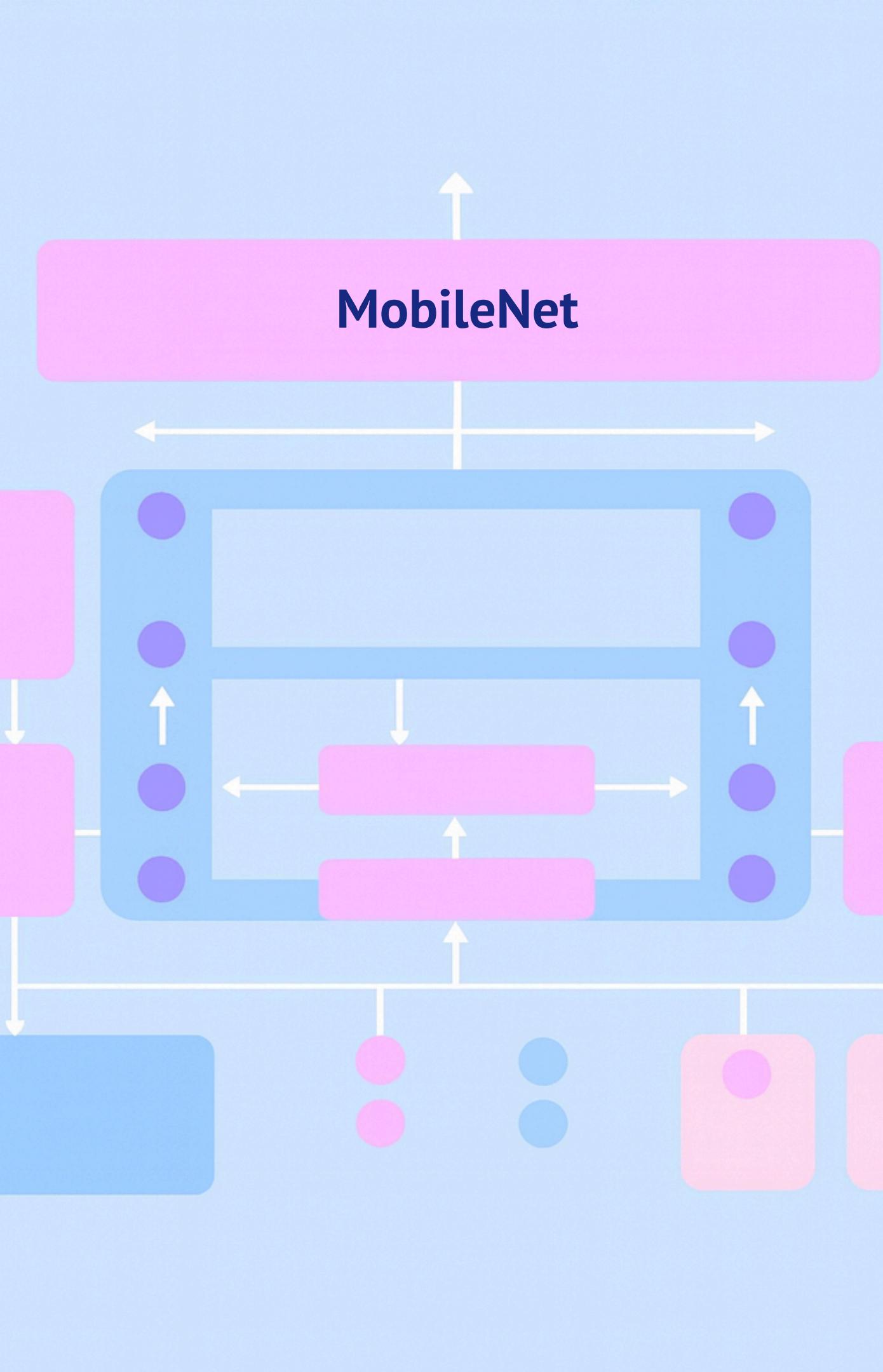
- **Structure:** Organised into distinct training, validation, and testing splits for rigorous evaluation.
- **Classes:** Approximately 27 unique gesture classes, including the English alphabet (A-Z), 'space', 'delete', and 'nothing' gestures.
- **Preprocessing:** Images underwent resizing to a uniform dimension, normalisation of pixel values, and extensive augmentation to enhance model robustness.

Efficient Model Architecture: MobileNetV2 Backbone

We leverage the MobileNetV2 architecture for its optimal balance of efficiency and accuracy, critical for real-time gesture recognition.

Custom Classifier Details:

- **Dropout Layers:** Strategically incorporated to mitigate overfitting and improve generalisation.
- **Fully Connected Layers:** Maps extracted features to the distinct gesture classes.
- **Pre-training:** Initialised with weights from ImageNet, then fine-tuned on our gesture dataset for domain-specific performance.



Training Setup

Our training regimen employs state-of-the-art tools and hardware to achieve optimal model performance.

Framework

PyTorch (`torch==2.3.1`,
`torchvision==0.18.1`)

Loss Function

`CrossEntropyLoss` for multi-class classification.

Optimizer

`Adam`, with a learning rate (LR) of $1e-3$.

Batch & Image Size

Batch size: 32, Image dimensions:
224x224 pixels.

Hardware

High-Performance Computing (HPC) GPU nodes, featuring NVIDIA A100 (40GB) GPUs.

Achieved Performance: High Accuracy & Robustness

Our model demonstrates exceptional accuracy on both training and validation datasets, indicating strong learning and generalisation capabilities.



- **Training Accuracy:** Achieved approximately 94% accuracy during the training phase.
- **Validation Accuracy:** Reached an impressive 98% on the unseen validation set.
- **Confusion Matrix:** Detailed analysis via the confusion matrix confirms robust per-class performance, with minimal misclassifications.



The best performing model checkpoint is saved at:
`outputs/mobilenetv2_gesture.pth`

Comprehensive Evaluation Metrics

We conducted a thorough evaluation using standard classification metrics to provide a complete understanding of the model's efficacy.

Key Metrics

Accuracy, Precision, Recall, and F1-score were calculated for each gesture class.

Detailed Reports

A comprehensive classification report and visual confusion matrix are generated.

Visualisation

Loss and accuracy curves provide insight into the training dynamics.

All evaluation outputs are stored in the `outputs/` directory, including:

- `classification_report.csv`
- `confusion_matrix.png`
- `loss_curve.png, accuracy_curve.png`

Flexible Deployment Options

Our gesture recognition system is engineered for versatile deployment, capable of integration into various real-world scenarios.



Static Image Analysis

The model can process individual images for gesture prediction using `predict_single.py`.



Real-time Webcam Stream

Enables live gesture recognition from a webcam feed via `predict_realtime.py`.



Core Dependencies

Requires PyTorch (version 2.3.1 or newer), OpenCV, and torchvision for full functionality.



Gesture Labels

Gesture class mappings are conveniently stored in `labels.json`.



Conclusion – Key Highlights

- High Performance: The model consistently achieved strong accuracy across training and validation.
- Generalization: Robust recognition across 29 hand gesture classes.
- Efficiency: Lightweight MobileNetV2 backbone enables fast inference.
- Deployment Ready: Supports single-image testing and real-time webcam prediction.
- Transparency: Training history, evaluation reports, and confusion matrix provided for reliability.



Thank You!

Any Question

Team:

Faris Abouagour

Mohamed Ghonem

Ahmed Ammar

Mohamed Yehia

Youssef Fady