1. Dataset Structure

- Downloaded and extracted
  - Raw: data/CCPD2019/
  - Subsets:

Sub-Dataset	Description	Image Number
CCPD-Base	Ordinary license plate picture	200 k
CCPD-FN	License plate is relatively close or far from the camera's shooting position	20 k
CCPD-DB	Brighter, darker or unevenly lit license plate areas	20 k
CCPD-Rotate	License plate tilted 20 to 50 degrees horizontally, −10 to 10 degrees vertically	10 k
CCPD-Tilt	License plate tilted 15 to 45 degrees horizontally and 15 to 45 degrees vertically	10 k
CCPD-Weather	License plate photographed in rain, snow and fog	10 k
CCPD-Challenge	The more challenging pictures in the plate detection recognition task	10 k
CCPD-Blur	Blurred plate images due to camera lens shake	5 k
CCPD-NP	Picture of a new car without plates fitted	5 k

2. Yolov5 setup and train

- YOLOv5 repo cloned and configured
- YAML file created with train/val paths
- Generate YOLO-style proto-labels
- YOLO training script uses CCPD2019 raw images
- Labels generated using filename metadata
- Outputs YOLO-style .txt files per image
- Trained weights saved for cropping step

3. Cropping

- Crops license plates using YOLOv5 bounding boxes
- Output images are 144×48 RGB crops
- Used as input for recognition models (baseline, PDLPR)

4. Baseline Pipeline

Input

- **Source**: Cropped license plate images (RGB) 48 × 144 pixels

Tokenizer

- Converts CCPD-encoded filenames (e.g., `"0_0_22_27_27_33_16"`) to license plates (e.g., `"皖A04025"`)
- No `<SOS>` or `<EOS>` tokens are used in CTC; blank index (0) is reserved

Feature Extraction: CNN+ Bi-GRU

- 3× convolutional blocks with BatchNorm + ReLU + MaxPooling
- Final output reshaped into sequence for temporal modeling

- **Output**: Tensor of shape (batch_size, width/8, channels × height/8)

Sequence Modeling

Bidirectional GRU (3 layers):

- Captures left–right temporal dependencies
- Input: flattened feature maps
- Output: sequence of features for classification

Decoder

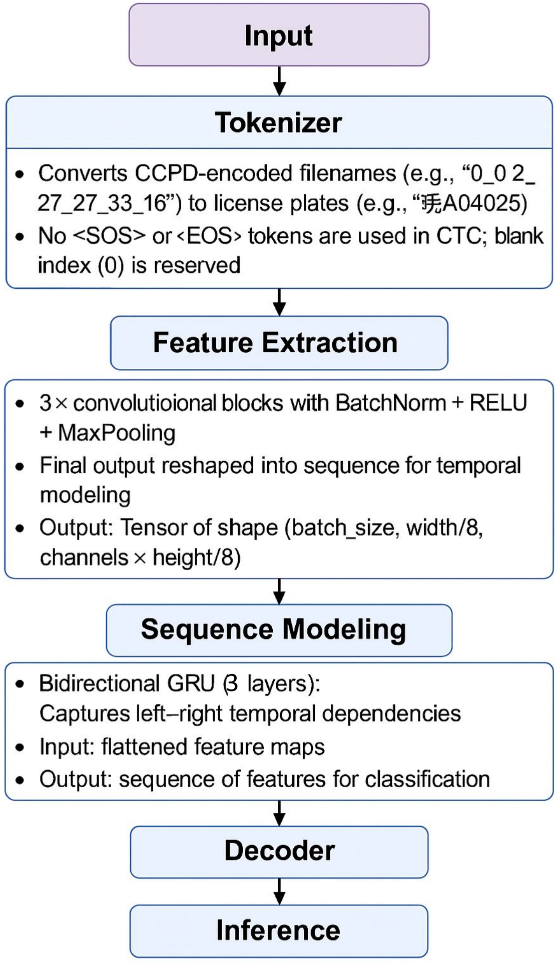
- Fully-connected linear layer maps each timestep to logits over character classes (including blank)

Loss Function

- **CTC Loss (Connectionist Temporal Classification):** Allows flexible alignment between predicted sequence and ground truth and does not require character-level alignment or segmentation

Inference

- Greedy decoding with collapsing repeated characters and removing blanks
- Ground truth derived directly from filename structure (CCPD encoding)



5. PDLPR Pipeline

Input

- **Source**: Cropped license plate images (RGB) 48 × 144 pixels

Tokenizer

- Converts plate strings (e.g., "京A12345") to integer sequences, including special tokens like <SOS> and <EOS>.

Feature Extraction: IGFE (Improved Global Feature Extractor)

- Based on CNN with:
  - **Focus structure** (YOLOv5-inspired) for spatial slicing without information loss
  - **ConvDownSampling** instead of pooling to preserve semantic content
  - **ResBlocks** to increase feature richness and stability
- **Output**: Feature map of shape `(batch_size, 512, 6, 18)`

Encoder

- Flattens feature map to sequence: (batch_size, 27, 512)
- Adds **fixed sinusoidal positional encodings**
- Passes through a stack of **Transformer encoder blocks** with:
  - Multi-Head Self-Attention
  - Feedforward layers
  - Residual connections + LayerNorm

Decoder: Parallel Transformer Decoder

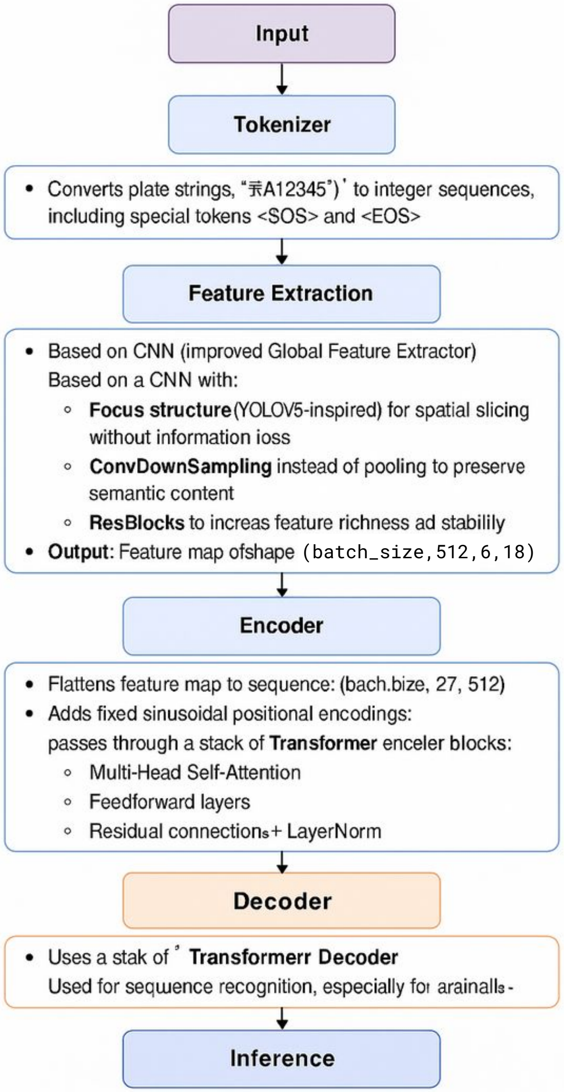
- Uses a stack of **Transformer decoder blocks**
- During training learns to attend to encoder features and previously decoded tokens. It uses masked Multi-Head Self-Attention to prevent attending to future tokens, enabling parallel computation during training.

Loss Function

- **CTC Loss (Connectionist Temporal Classification):** Used for sequence recognition

Inference

- The model iteratively predicts the next token.



6. Results – Baseline

Base:

- **Plate similarity (Leveshtein)** : 99.6 %
- **Throughput**   : 3972.7 FPS (plates/s)

Averages on all sub-datasets:

- **Plate similarity (Leveshtein)**: 77.7%
- **Throughput**: 4128.2FPS (Plates/s)

7. Evaluation PDLPR

Base:

- **Plate accuracy** : 99.60 %
- **Char accuracy** : 99.83 %
- **Throughput** : 2968.5 FPS (plates/s)

Averages on all sub-datasets:

- **Plate Accuracy**: 76.47%
- **Char Accuracy**: 92.81%
- **Throughput**: 2845.7 FPS (Plates/s)