Model Refinement Report — Hangong Chen

The baseline model is a single-head transformer: three encoder layers, width 192 with six heads, dropout 0.1, and one regression output. It is trained only to predict the rain amount, with an extreme-weighted MSE to emphasize heavy rain.

The current improved model changes both shape and objective. Architecturally, it's a bit shallower and narrower (four encoder layers, width 160 with five heads, higher dropout 0.12, weight decay) and adds a second head for rain/no-rain classification. In training, it optimizes both heads: regression in real units with log1p compression plus penalties for over-forecasting and clipping, and a BCE loss on event occurrence with a higher weight. At evaluation, it uses the classifier to gate the regression output, zeroing low-confidence rain to cut false alarms.

Performance trade-off: the improved model boosts event detection and average error (MAE ~0.082 vs 0.128; CSI ~0.60 vs 0.57; POD ~0.81 vs 0.69; extreme POD ~0.82 vs ~0.74), but it still has higher RMSE (0.42 vs 0.38) and a higher false-alarm rate (0.30 vs 0.24). In short, the new architecture and loss make it better at catching rain events with lower average error, at the cost of more false positives and slightly worse RMSE.

The performance comes from the changes we made to the model and loss. We added a rain/no-rain head and trained it jointly with the regressor, then gated the regression output at inference. That makes the model more willing to call rain when the classifier is confident, which boosts POD and CSI but can raise false alarms. The regression loss now works in real units with log1p compression, an overprediction penalty, and clipping penalties; this pulls the average error (MAE) down and keeps outputs reasonable, but the extra emphasis on catching events can spread residuals and slightly lift RMSE. Architecturally, the model is shallower and narrower with heavier dropout and weight decay, which reduces overfitting and stabilizes training, but some of the baseline's variance-fitting capacity was traded for better event detection. The net result is better hit rates and lower MAE, at the cost of more false positives and a modest RMSE increase.

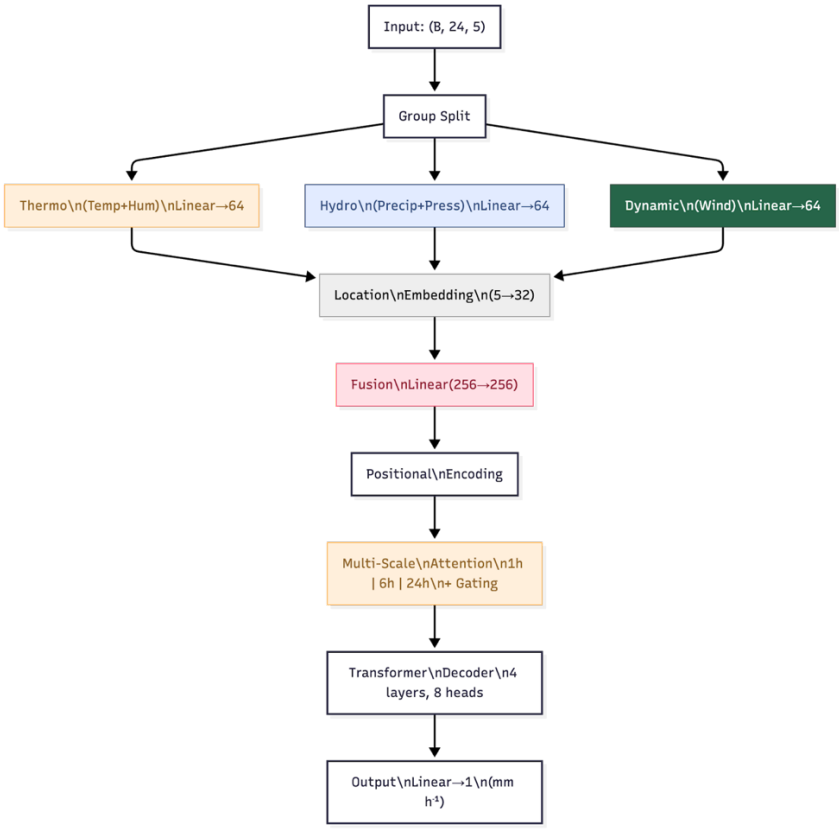

Fig.1 Original model architecture

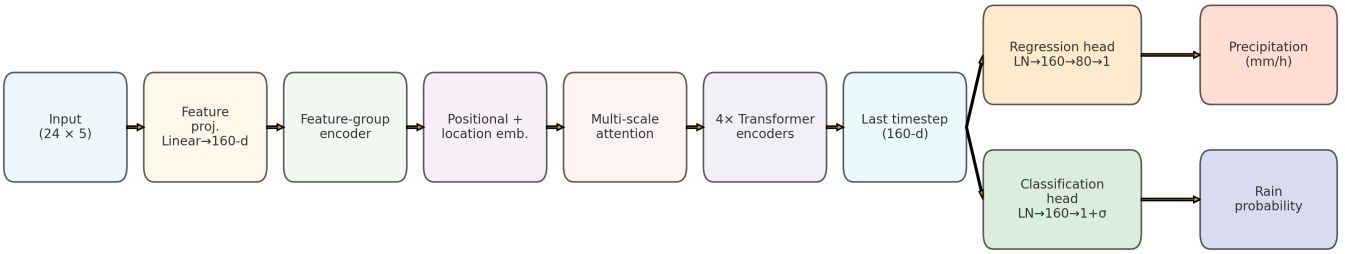

Fig.2 Improved model architecture

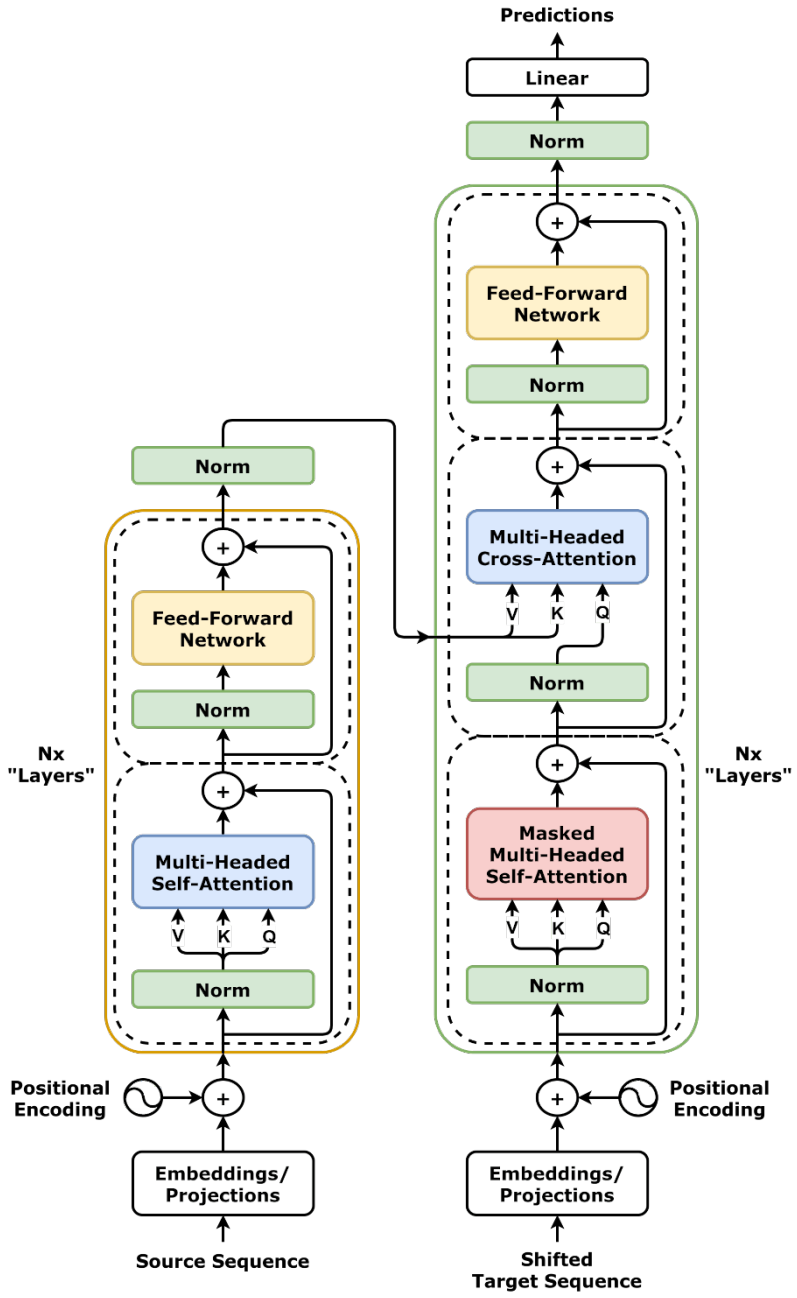

Fig.3 Original transformer model architecture