

Practical Implementation Tips



STABILITY

Gradient Clipping

Prevent exploding gradients during training



NORMALIZATION

Batch/Layer Normalization

Stabilize training and improve convergence



REGULARIZATION

Dropout

Apply to non-recurrent connections only



TRAINING

Learning Rate Scheduling

Use warmup + decay strategy



AUGMENTATION

Data Augmentation

Time warping, noise injection



MONITORING

Validation Metrics

Track to prevent overfitting