1. Description of data usage:

80% for train, 20% for validation. One hot encoded. No augmentation

2. Description of architecture:

We used LSTM (2 layers) followed by Attention with 5 heads.

3. Training procedure:

Loss: RMSE

Learning rate: Adam

Epochs=10

4. Contributions:

Yuanfang Guan implemented the algorithm