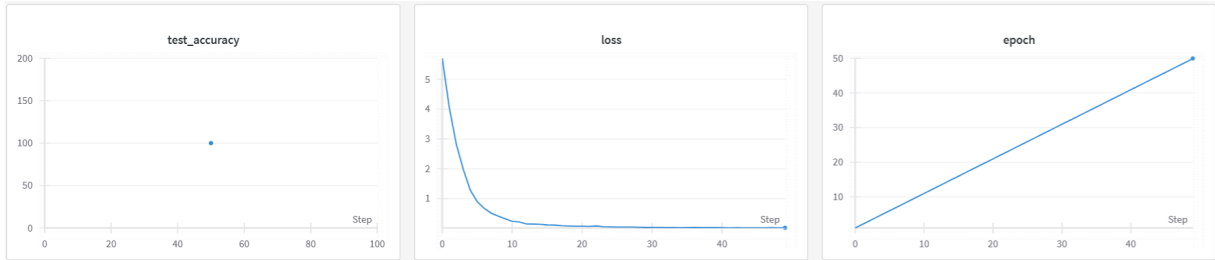


Week01

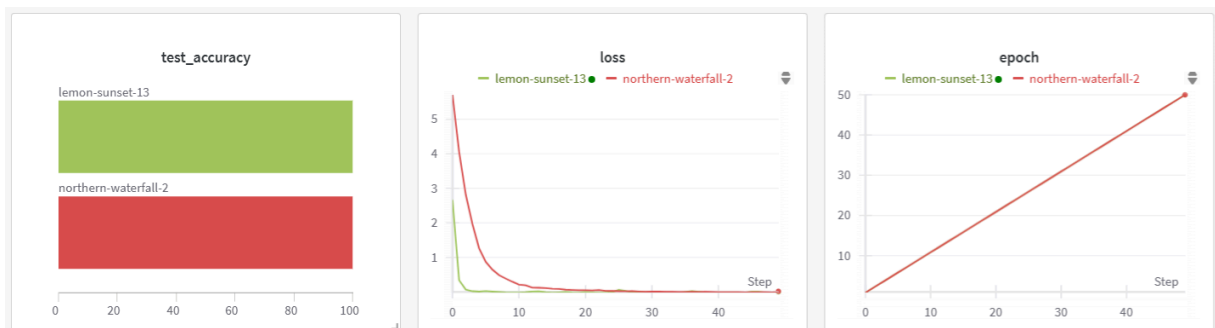
▼ 1st trial



```
src_vocab_size = 100
tgt_vocab_size = 200
d_model = 32
n_heads = 4
d_ff = 64
num_encoder_layers = 6
num_decoder_layers = 6
dropout = 0.1
batch_size = 32
num_epochs = 50
learning_rate = 0.0001
```

▼ 2nd trial

- 과적합을 방지하고자 encoder와 decoder layer를 각각 3으로 줄임
- 정확도를 높이하고자 batch size는 절반으로 줄임
- train 속도를 높이하고자 learning rate을 10배해 0.001로 실행



```
src_vocab_size = 100
tgt_vocab_size = 200
d_model = 32
```

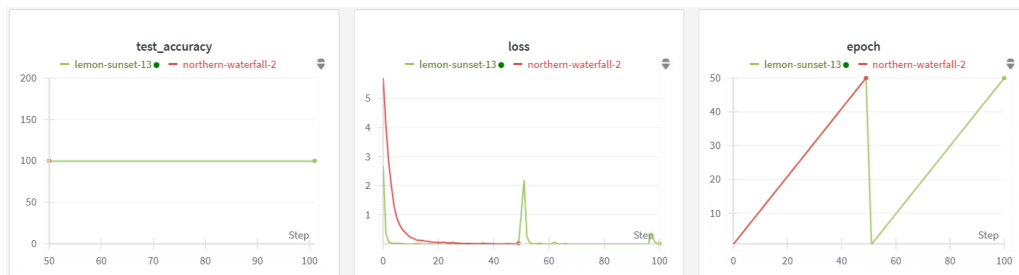
```

n_heads = 4
d_ff = 64
num_encoder_layers = 3
num_decoder_layers = 3
dropout = 0.1
batch_size = 16
num_epochs = 50
learning_rate = 0.001

```

▼ 3rd trial

- 2nd trial과 모든 것을 동일하게 설정한 상태에서 head의 개수만 8로 늘렸더니 중간에 보이는 그래프처럼 train loss가 튀는 현상이 관찰되었다. batch size가 작거나 learning rate이 크면 이러한 문제가 발생할 수 있다고 한다. → 2nd trial에 사용한 하이퍼파라미터를 사용하는 게 가장 적합할 듯!



```

src_vocab_size = 100
tgt_vocab_size = 200
d_model = 32
n_heads = 8
d_ff = 64
num_encoder_layers = 3
num_decoder_layers = 3
dropout = 0.1
batch_size = 16
num_epochs = 50
learning_rate = 0.001

```