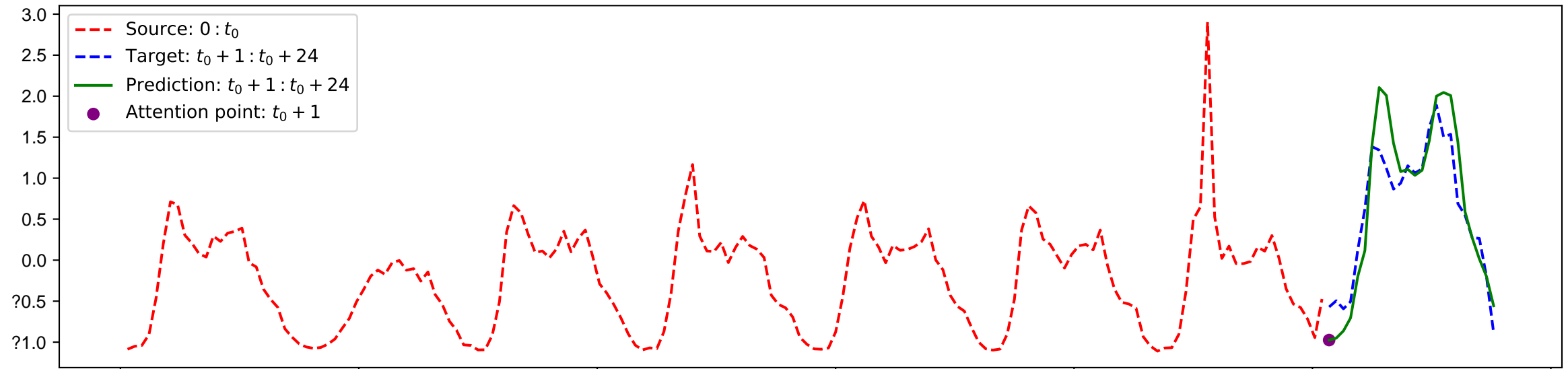Traffic dataset, linear_kernel, 24-steps non-overlapping, 12-th time series, total epochs: 300

Legend:
- Source: $0 : t_0$
- Target: $t_0 + 1 : t_0 + 24$
- Prediction: $t_0 + 1 : t_0 + 24$
- Attention point: $t_0 + 1$

Attention pattens for $t_0 + 1$

- attn_layer_1
- attn_layer_2
- attn_layer_3