Traffic dataset, scaled\_dot\_product, 24-steps non-overlapping, 29-th time series, total epochs: 300 --- Source:  $0:t_0$ --- Target:  $t_0 + 1: t_0 + 24$ Prediction:  $t_0 + 1 : t_0 + 24$ Attention point:  $t_0 + 1$ ?1 Attention layer 1 for  $t_0 + 1$ 0.125 0.100 0.075 0.050 0.025 Attention layer 2 for  $t_0 + 1$ 0.25 0.20 0.15 0.10 0.05 Attention layer 3 for  $t_0 + 1$ 0.08 0.07 0.06 0.05

48000

50000

52000

0.04

42000

44000

46000