Transformer, RNN (GRU), CNN and DNN models loading time Transformer: model load time RNN (GRU): model load time 3.0 CNN: model load time DNN: model load time 2.5 Model load time (seconds) 2.0 1.0 0.5 1000 2000 5000 10000 20000 25000 30000 35000 3000 4000 Training step