Spatio-Temporal LSTM

Motivation: (1) LSTM network is suitable for modeling dependence in temporal domain. (2) There is also high dependence among joints in spatial domain.

Spatio-temporal LSTM with Trust Gates for 3D Human Action Recognition

Jun Liu¹    Amir Shahroudy¹    Dong Xu²    Gang Wang¹
¹Nanyang Technological University   ²University of Sydney

1 Spatio-Temporal LSTM

Motivation: (1) LSTM network is suitable for modeling dependence in temporal domain. (2) There is also high dependence among joints in spatial domain.

Spatio-temporal LSTM.

- Spatial direction: body joints are fed in a sequence.
- Temporal direction: locations of corresponding joints are fed frame by frame.

2 Tree-Structure based Traversal

It is beneficial to model the spatial dependency of the joints based on their adjacency tree structure (see figure c).

(a) Simple joint chain: joints are fed to ST-LSTM in a naive order of 1-2-3-...-16. (b) Skeleton is transformed to a tree structure. (c) Proposed tree traversal over spatial steps: 1-2-3-2-4-5-6-5-4-2-...-16.

It is beneficial to model the spatial dependency of the joints based on their adjacency tree structure (see figure c).

3 Spatio-Temporal LSTM with Trust Gate

Unreliable input data (noisy joint coordinates) restricts classification accuracy.

Learn to predict the input using contextual information:

\[ p_{j,t} = \tanh \left( M_p \left( h_{j,t-1}, h_{j,t} \right) \right) \]

Prediction of input

Contextual information

Assess reliability of actual input by comparing it with the prediction:

\[ x_{j,t} = \tanh \left( M_x \left( x_{j,t} \right) \right) \]

\[ \tau_{j,t} = G(x_{j,t} - p_{j,t}) \]

\[ G(z) = \exp(-\lambda z^2) \]

Updating cell state using trust gate \( \tau_{j,t} \):

\[ c_{j,t} = \tau_{j,t} \odot \tilde{c}_{j,t} \odot u_{j,t} \]

\[ + (1 - \tau_{j,t}) \odot f^S_{j,t} \odot c_{j-1,t} \]

\[ + (1 - \tau_{j,t}) \odot f^T_{j,t} \odot c_{j,t-1} \]

4 Action Recognition Results

(a) Simple joint chain: joints are fed to ST-LSTM in a naive order of 1-2-3-...-16. (b) Skeleton is transformed to a tree structure. (c) Proposed tree traversal over spatial steps: 1-2-3-2-4-5-6-5-4-2-...-16.

Contributions

- Spatio-temporal design of LSTM network for 3D action recognition.
- Skeleton-based tree traversal technique to feed the structure of the skeleton data into a sequential LSTM.
- Improving the design of spatio-temporal LSTM by adding the trust gate to deal with noisy input.
- Achieving state-of-the-art performance on NTU RGB+D, UT-Kinect, SBU-Interaction, and MHAD datasets.