Protocol 20: RNN Training

24.05.2019

Recurrent Neural Networks (RNN)

1. Introduction

After trying different types of CNNs for lameness detection, RNNs are applied in this protocol. More specifically, a hierarchical recurrent neural network is used to train the data of cow skeleton. The concept of the network is from [**?**]: the skeleton is divided into several parts, each of which is fed into a subnet; as the number of layers increases, the representations extracted from the subnets are hierarchically fused based on the correlation between the parts. The architecture of the network is illustrated in Figure 1.

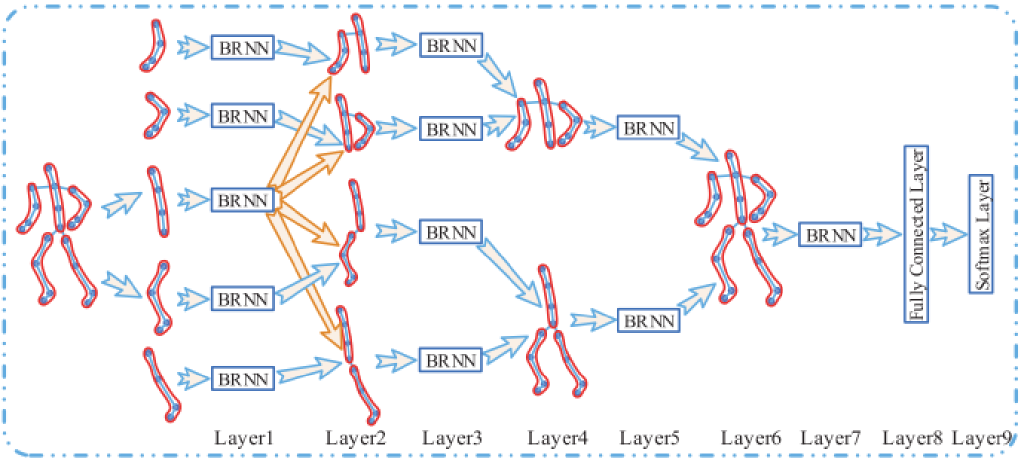

Figure 1: Hierarchical recurrent neural network. The human skeleton is divided into five parts, which are fed into five bidirectional recurrent neural networks (BRNNs) [**?**].

As cows are quadrupedal rather than bipedal like humans, the division of body parts is different. The 25 skeletal joints of cows (Figure 2) are divided into six parts:

- head and neck (joints 0-2)
- trunk (joints 3-8)
- front-left leg (joints 9-12)
- front-right leg (joints 13-16)
- rear-left leg (joints 17-20)
- rear-right leg (joints 21-24)

2. Experiment

- **Dataset:** The original dataset contains 501 samples, each of which contains the coordinates of 25 skeletal joints from more than 100 video frames and a locomotion score as the label. As in the previous protocols, the locomotion scores fall into four classes. The dataset is augmented by mirroring and stretching, such that the amount of data is tripled. The streching is only applied to x coordinate based on the assumption that the cow is standing on level ground in the frames.

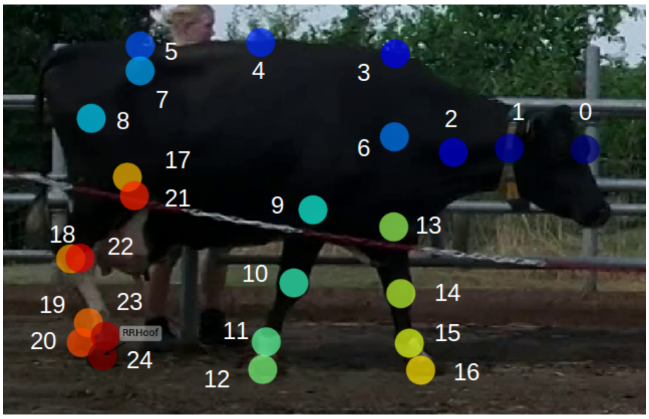

Figure 2: The 25 keypoints of cow skeleton.

- **Keypoints:** Figure 2 shows the 25 keypoints (skeletal joints) from by pose estimation. Note that the coordinate of all the points are calculated from the point 5 instead of the absolute position in the frame. Besides, the values of x and y coordinates are divided by the frame width and height, respectively.
- **Tuning:** In order to improve the performance, different hyperparameters and functions have been applied:
  - Different optimization methods: Adam, Adagrad, and SGD
  - The activation function in the RNN: tanh, ReLU
  - Batch size: 4, 16, 32
- **Result:** The training and validation loss and accuracy curves of different image sizes are displayed in Figure 6 and 7.
- **Notes:**
  - A single LSTM without hierarchical architecture has been used, but the result look the same as that using the complex one.
  - The training does not seem to converge and the loss is stuck in a local minimum. The network always predicts class 1 (locomotion score 1-2).
  - There may be an issue with network initialization since the initial network predicts the same class (class 1 or 4) for the whole dataset rather than random guess.

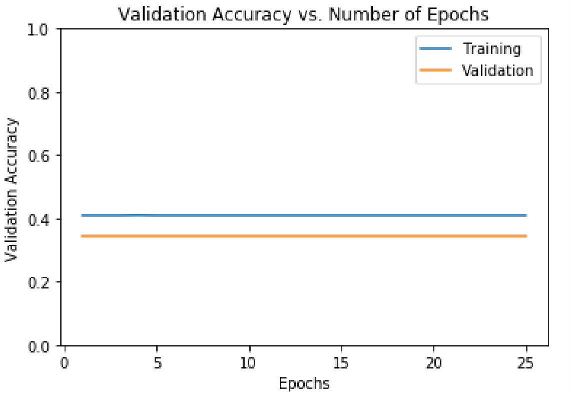

Figure 3: Training/validation accuracy.

3. Discussion

An RNN-based network has been applied to the lameness detection. Unlike the previous trials with different types of CNNs, it is expected that RNNs can learn the temporal information from the skeleton from sequence of video frames. The stagnation of training curve can have various causes:

- Data: poor data quality (noisy coordinates), unnormalized data between layers
- Data distibution: imbalanced data (prdominant class 1)
- Unsuitable haperparameters and weight initialization
- Bugs in network or training codes
- Inappropriate network architecture

These probable causes will be taken into account, such that the network training can jump out of the local minimum and thus the result can be improved.

Reference

[1] Y. Du, W. Wang, and L. Wang, Hierarchical recurrent neural network for skeleton based action recognition, in Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 11101118, 2015.