Protocol 17: Experiment with Different Data

03.05.2019

Spatial Temporal Graph Convolutional Networks (ST-GCN)

Since last few weeks, ST-GCN has been applied to lameness detection. It seemed that the pose estimation affected the training result since the input data were the coordinates of the skeleton joints from pose estimation. Some experiments were carried out in order to find out the causes of poor performance and to improve the result. As in the previous experiments, the lameness detection is treated as a four-class problem.

1. Experiment 1: With less occlusion

As the pose estimation was affected by occlusion, only the first few frames of the video data were considered. In this experiment, 60 frames (3 seconds) were used since the cows in most of the videos were occluded by a fence between three to six second.

- **Dataset:** Each data sample contains the coordinates of 18 skeleton joints from 60 video frames and a locomotion score as the label. The dataset has 501 samples, divided into training and test sets with a 70/30 ratio.
- **Keypoints:** Figure 1 shows the 18 keypoints (skeleton joints) from by pose estimation. Since the hook joint (keypoint 5) is comparatively stable, the coordinate of all the points are calculated from the point instead of the absolute position in the frame.
- **Result:** The overall accuracy is 41.06%, and the confusion matrix is shown in Table 1.
- **Notes:**
  - Another experiment with 100 frames was executed, but the accuracy is lower, which may somewhat explain that occlusion indirectly has an impact on lameness detection.
  - The experiment with 100 frames shows an improvement in class 2.

Table 1: Confusion matrix: 60 frames

	Predicted Class 1	Predicted Class 2	Predicted Class 3	Predicted Class 4
Actual Class 1	31	20	9	7
Actual Class 2	13	14	7	8
Actual Class 3	3	6	7	6
Actual Class 4	4	4	2	10

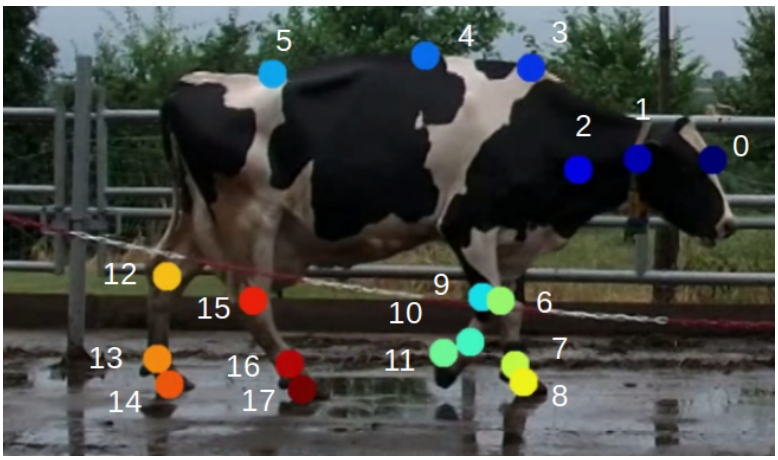

Figure 1: The 18 keypoints extracted from videos for lameness detection.

2. Experiment 2: Use only true labels

This experiment only used the data with true labels. Unlike the previous experiments, the data with interpolated labels were not included, and the size of dataset was reduced.

- **Dataset:** Each data sample contains the coordinates of 18 skeleton joints from 300 video frames and a true locomotion score as the label. The dataset has only 174 samples, divided into training and test sets with a 70/30 ratio.
- **Keypoints:** The 18 keypoints shown in Figure 1 were considered.
- **Result:** The overall accuracy is 46.94%, and the confusion matrix is shown in Table 2.
- **Notes:**
  - Compared to the other experiments, this one shows the highest accuracy of class 1, which may be due to the unbalanced class distribution. The effect is more obvious with less data.
  - With some repetition, the accuracy jumps between 40 to 55%. The interpolation of labels may be feasible, but can lead to wrong labeling for some samples.

Table 2: Confusion matrix: true locomotion score

	Predicted Class 1	Predicted Class 2	Predicted Class 3	Predicted Class 4
Actual Class 1	11	2	0	1
Actual Class 2	5	6	2	1
Actual Class 3	2	2	3	3
Actual Class 4	2	1	5	3

3. Experiment 3: More joints

The previous experiments only used 18 keypoints. As the amount of data is small, it my help by including more skeletal joints to capture more information from the cow's pose.

- **Dataset:** Each data sample contains the coordinates of 25 skeleton joints from 300 video frames and a true locomotion score as the label. The dataset has only 174 samples, divided into training and test sets with a 70/30 ratio.
- **Keypoints:** The 25 keypoints shown in Figure 2 were considered. The hook joint (keypoint 5) is comparatively stable, the coordinate of all the points are calculated from the point instead of the absolute position in the frame.
- **Result:** The overall accuracy is 41.72%, and the confusion matrix is shown in Table 3.
- **Notes:**
  - Even with more jonits, the network does not seem to capture more spatio-temporal information about cow's lameness behaviors.
  - The ordering and connection of the joints may be another factor to be tested.

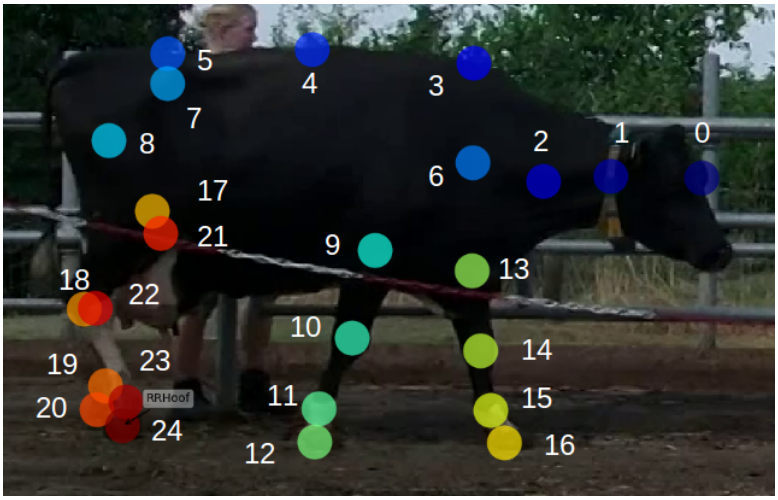

Figure 2: The 25 keypoints extracted from videos for lameness detection.

Table 3: Confusion matrix: 25 keypoints

	Predicted Class 1	Predicted Class 2	Predicted Class 3	Predicted Class 4
Actual Class 1	38	14	7	5
Actual Class 2	13	17	6	8
Actual Class 3	6	6	3	10
Actual Class 4	0	5	8	5

3. Comments and Next Step

After trying different experiments, it seems that the skeletal data only provide limited information for lameness detection. This can be inferred from the experiments with more accurate pose estimation or more skeletal joints. The training and test accuracy are respectively around 60% and 50%, so the overfitting is currently not of great concern. The ST-GCN already has multiple layers, and thus the structure cannot be expanded to improve the training result with limited amount of data. In order to improve the result of lameness detection, more vital and effective spatio-temporal information needs to be extracted.