Unsupervised Object Detection and Tracking with Informed ML

Final Presentation

Mohamad Akbari

University Bonn

06.03.2023

Lab Development and Application of Data Mining and Learning Systems: Data Science and Big Data
Supervisor: Laura von Rueden

Outline

Problem Definition

problem

In traditional supervised learning setting, the labeling process is time-consuming and the result of the process is only specific to one data set.

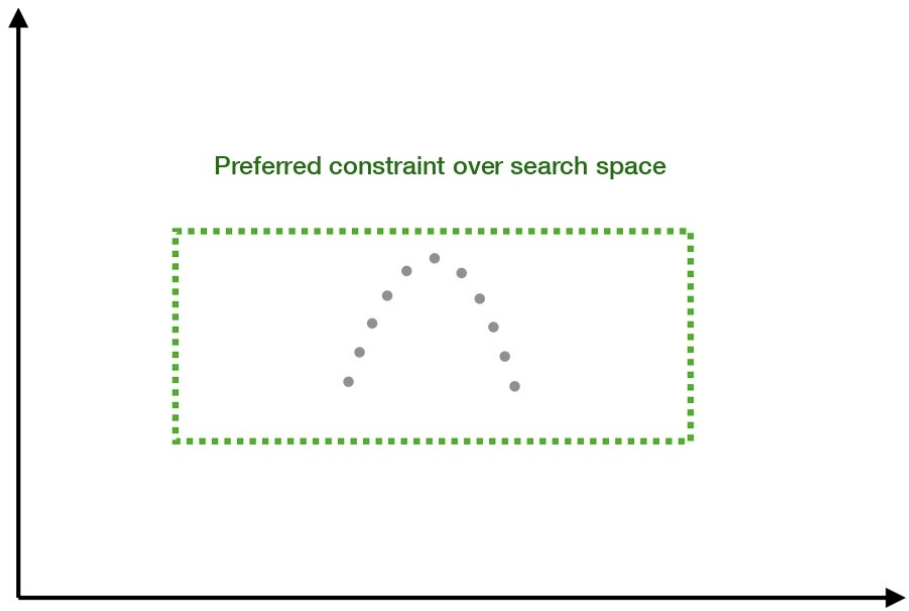
Problem Definition

problem

In traditional supervised learning setting, the labeling process is time-consuming and the result of the process is only specific to one data set.

proposed solution

If prior knowledge informs us that outputs of f^* have unique properties, we may use them for training instead of direct labels.



Experiment

Tracking an object in free fall

1. Record Data Set
2. Automatic Labeling
3. Supervised Training
    - 3.1. Supervised Training on single Image
    - 3.2. Supervised Training on Image Sequence
4. Unsupervised Training with Informed ML

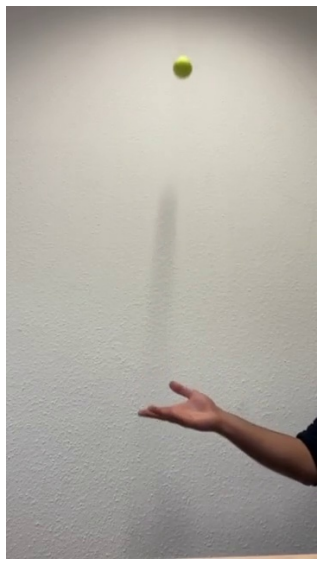
1.Record Data

Data Preparation

1. record slow-motion(120fps) HD videos of free fall of tennis ball in front of the white wall,
Number of Videos: 59

2. extract image sequence with $\Delta t = 0.1s$ from the video,
Image size : $360x640$
Number of Images : 610

Example

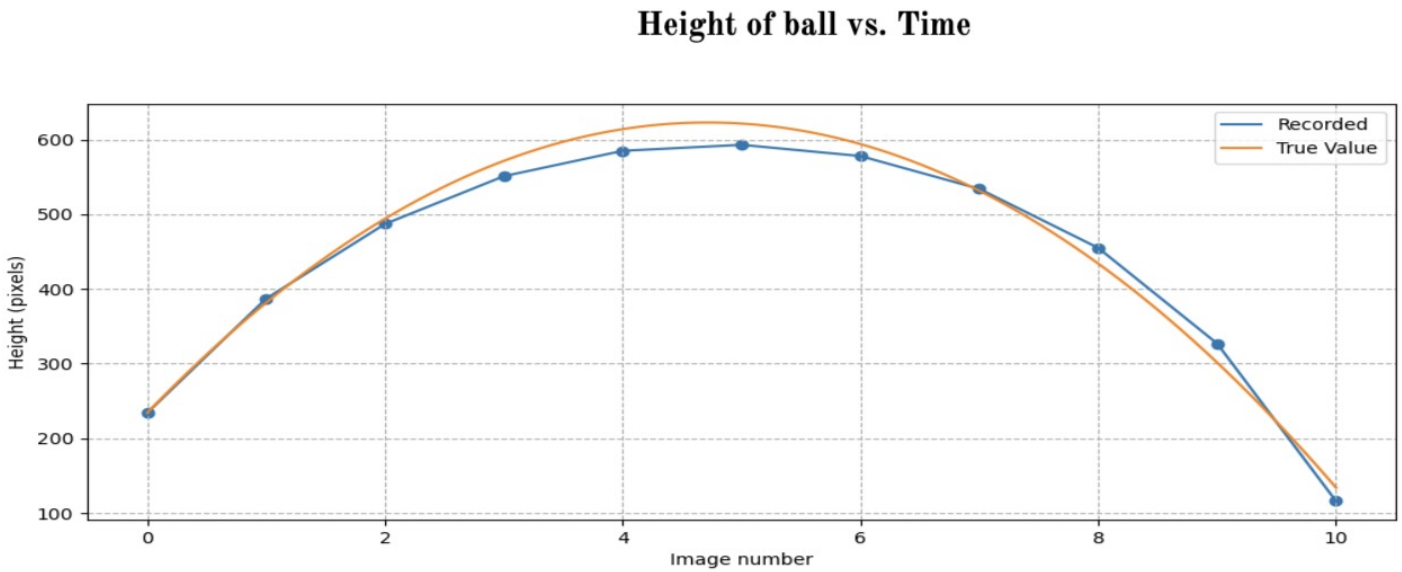


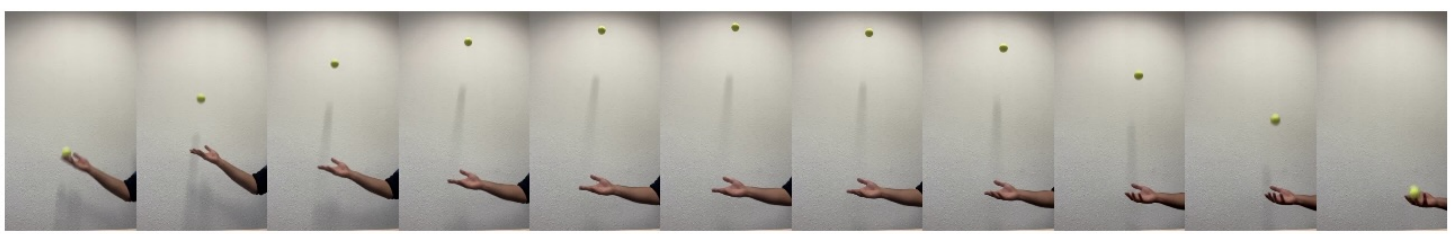
2. Auto Labeling

Creating a true label data set of heights
for step 3 and for measuring the accuracy of models later on

$$y_i = y_0 + v_0(i\Delta t) + a(i\Delta t)^2$$

$$\Delta y = v_0(0.1) + a(0.1)^2 \rightarrow v_0 = \frac{\Delta y - a(0.1)^2}{0.1} (1)$$





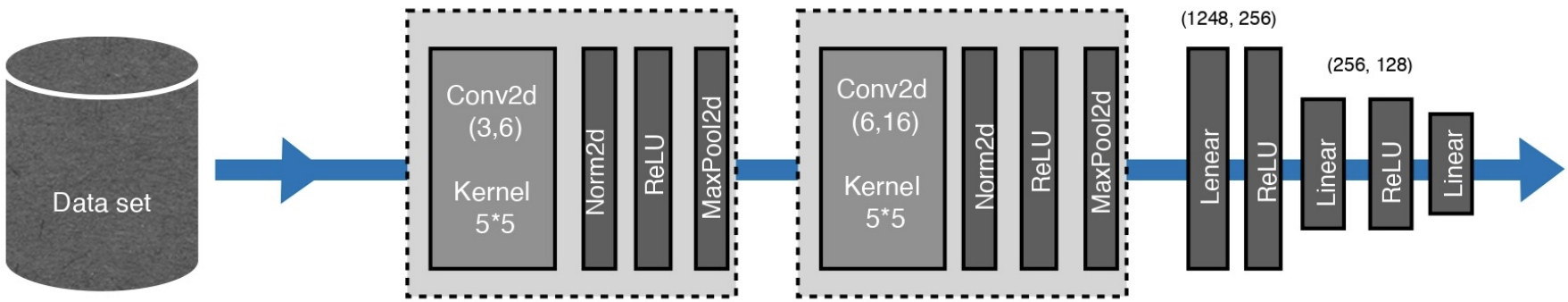
3.1. Supervised Training on single Image

1. Create Data Set & Data Loader

Including : 1- true label acquired in the last step and 2- resized images (3∗64∗36). Batch size 16, implemented in Pycharm.

2. Model Architecture

Model consists of two Blocks of Convolution Network/ RELU/ MaxPool followed by 2 Fully connected Linear/ RELU layers and finally a linear layer



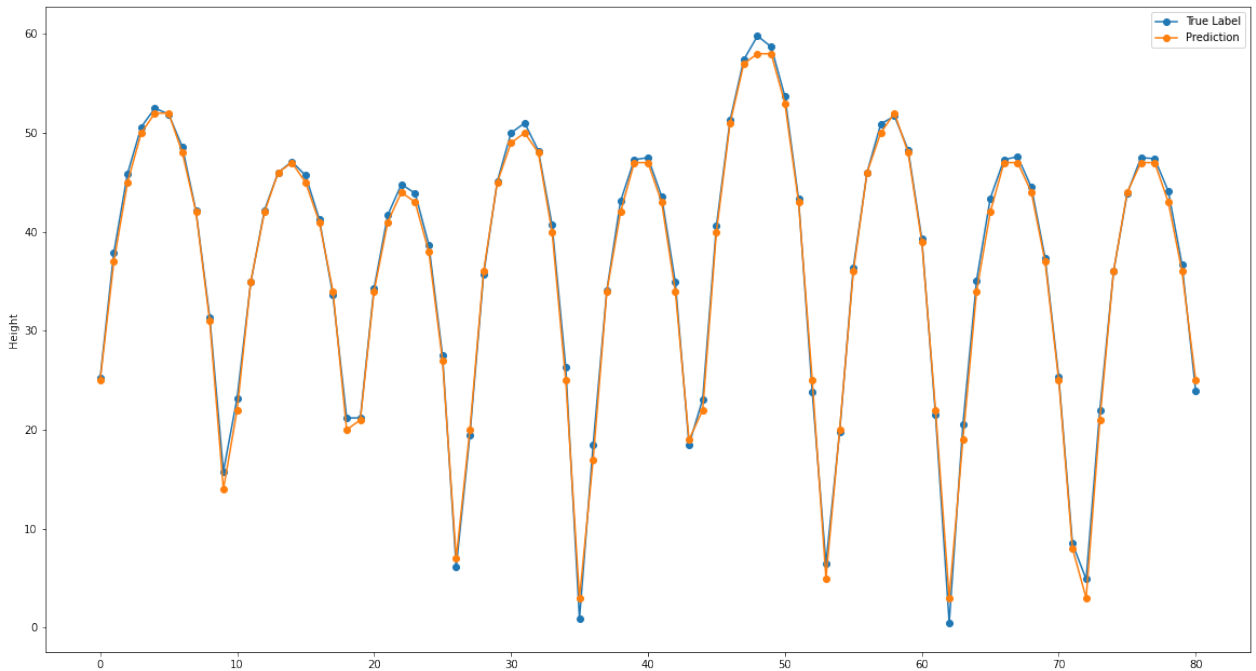
3.1. Supervised Training on single Image

3. Result:

The model performed poorly in regression task (1 output) with 62.18% accuracy on the test set.

However, accuracy had noticeable improvement for the classification task (64 outputs): 98.41%

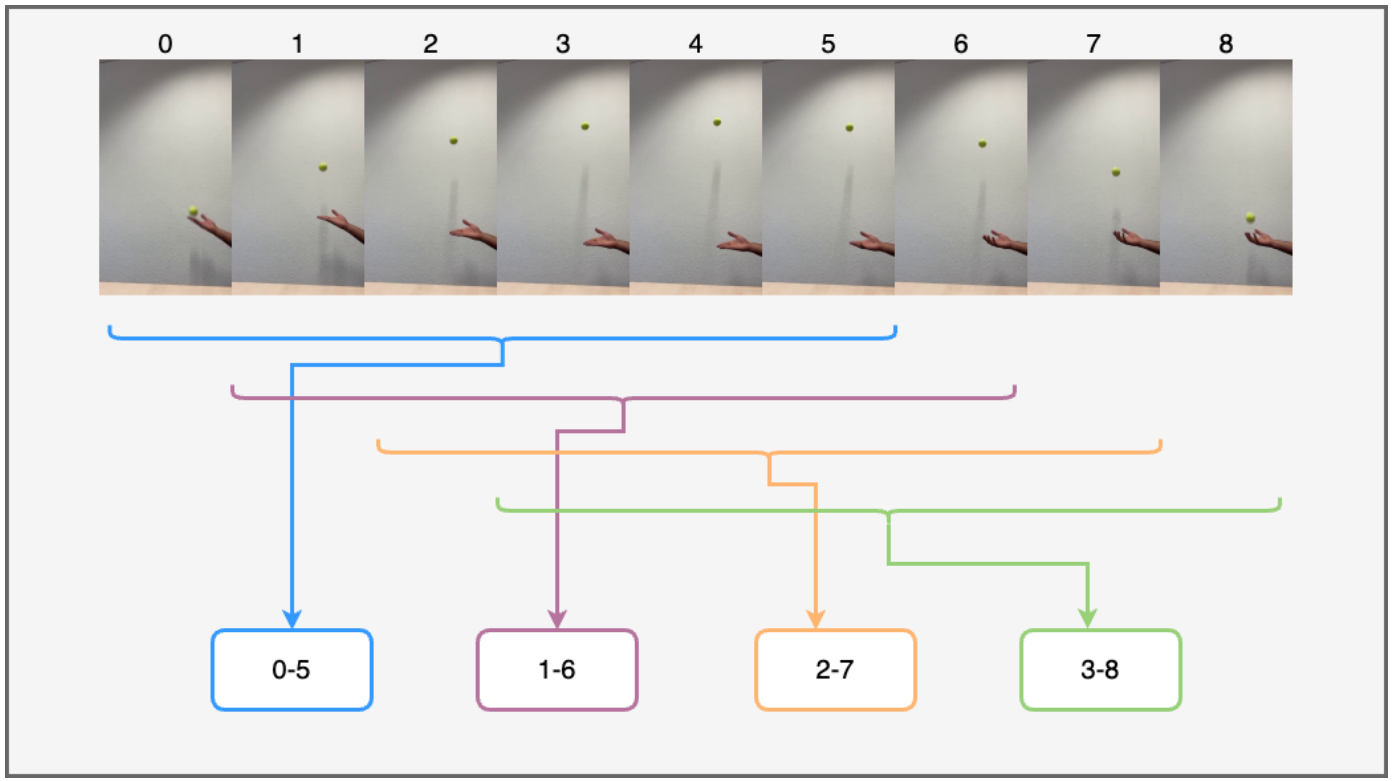
$$\text{Accuracy} = 1 - (MAE(targets, prediction))/(Mean(targets))$$



3.2. Supervised Training on Image Sequence

1. Transform Data Set

transformed the data from a single image to a sequence of images from the same throw with a specific length: 6
list of labels stored for each image sequence

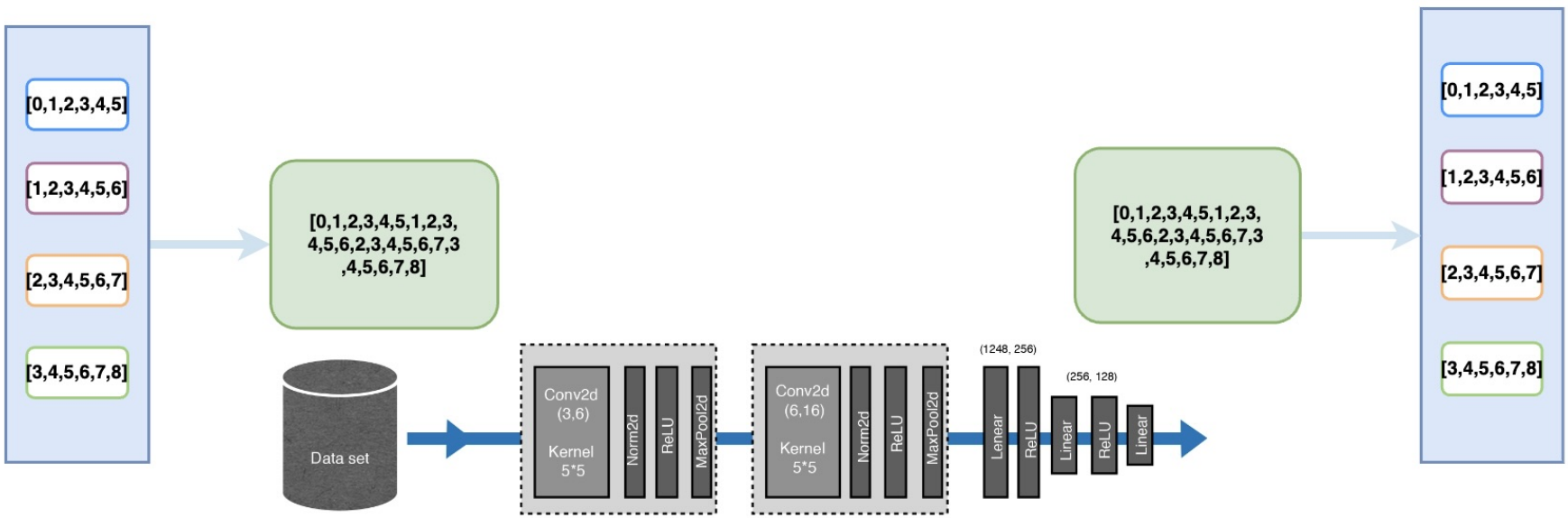


3.2. Supervised Training on Image Sequence

2. adjustments

Model architecture is based on the paper instruction as before

Implementation modified to work on image sequence as data

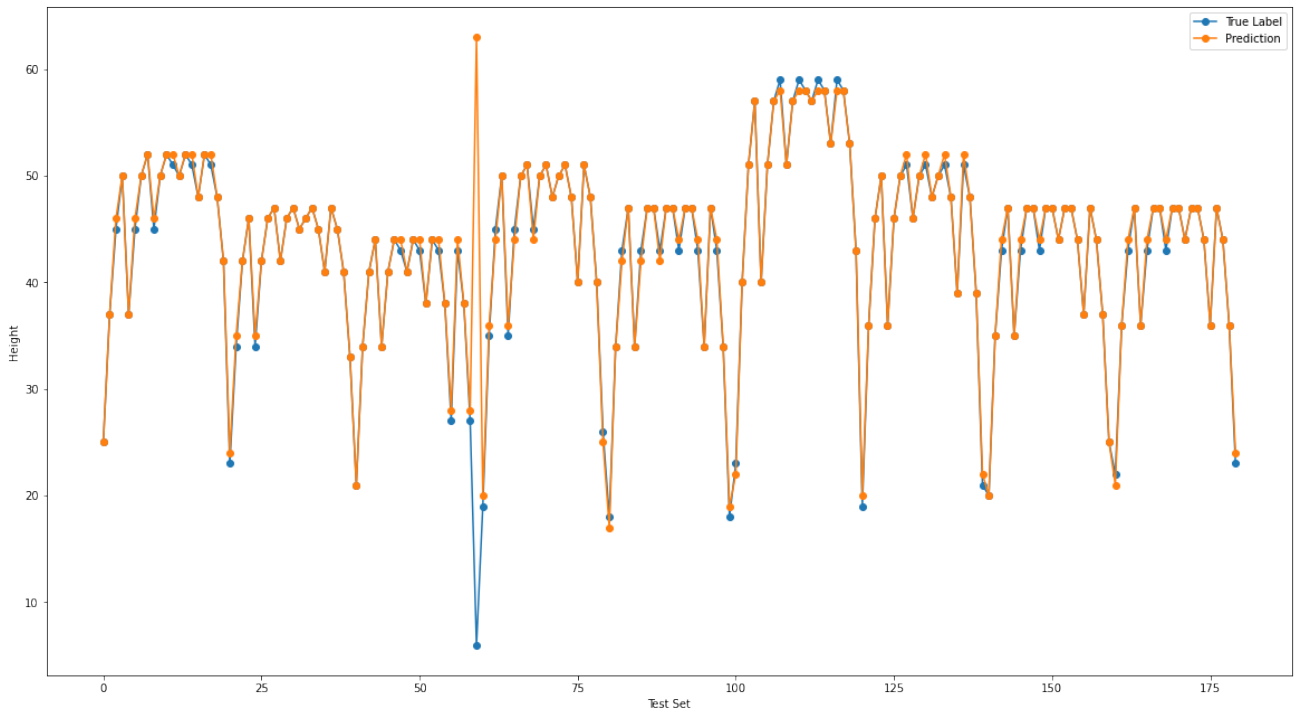


3.2. Supervised Training on Image Sequence

3. Results

Supervised sequence training shows accuracy of 96.27%

1. Accuracy $= 1 - \frac{MAE(targets, prediction)}{Mean(targets)}$
2. Mean absolute error $= 1.5$



4. Unsupervised Training with Informed ML

1. Loss Function

The loss function is based on the regression formula of free fall parabola

$$\mathbf{y}_i = y_0 + v_0(i\Delta t) + a(i\Delta t)^2$$

$$\hat{\mathbf{y}} = \mathbf{a} + \mathbf{A}(\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T(f(\mathbf{x}) - \mathbf{a})$$

$$\mathbf{A} = \begin{bmatrix} \Delta t & 1 \\ 2\Delta t & 1 \\ 3\Delta t & 1 \\ \vdots & \vdots \\ N\Delta t & 1 \end{bmatrix}$$

$$\mathbf{a} = [a\Delta t^2, a(2\Delta t)^2, \ldots, \bar{a}(N\Delta t)^2]$$

$$g(\mathbf{x}, f(\mathbf{x})) = g(f(\mathbf{x})) = \sum_{i=1}^{N} |\hat{\mathbf{y}}_i - f(\mathbf{x})_i|$$

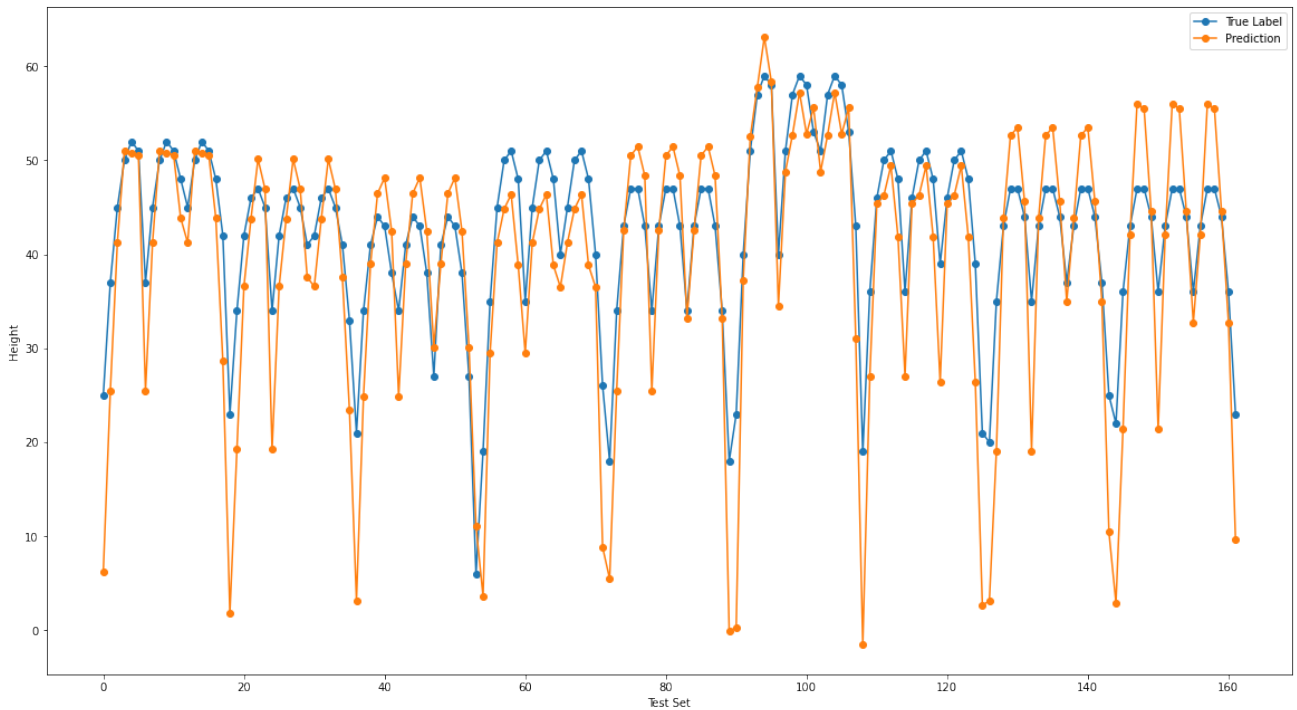
4. Unsupervised Training with Informed ML

2. Results

Unsupervised sequence training model is more unstable than before

1. Best accuracy is 88.6% but over multiple runs the average is 80-82%

2. Best mean absolute error = 5.6

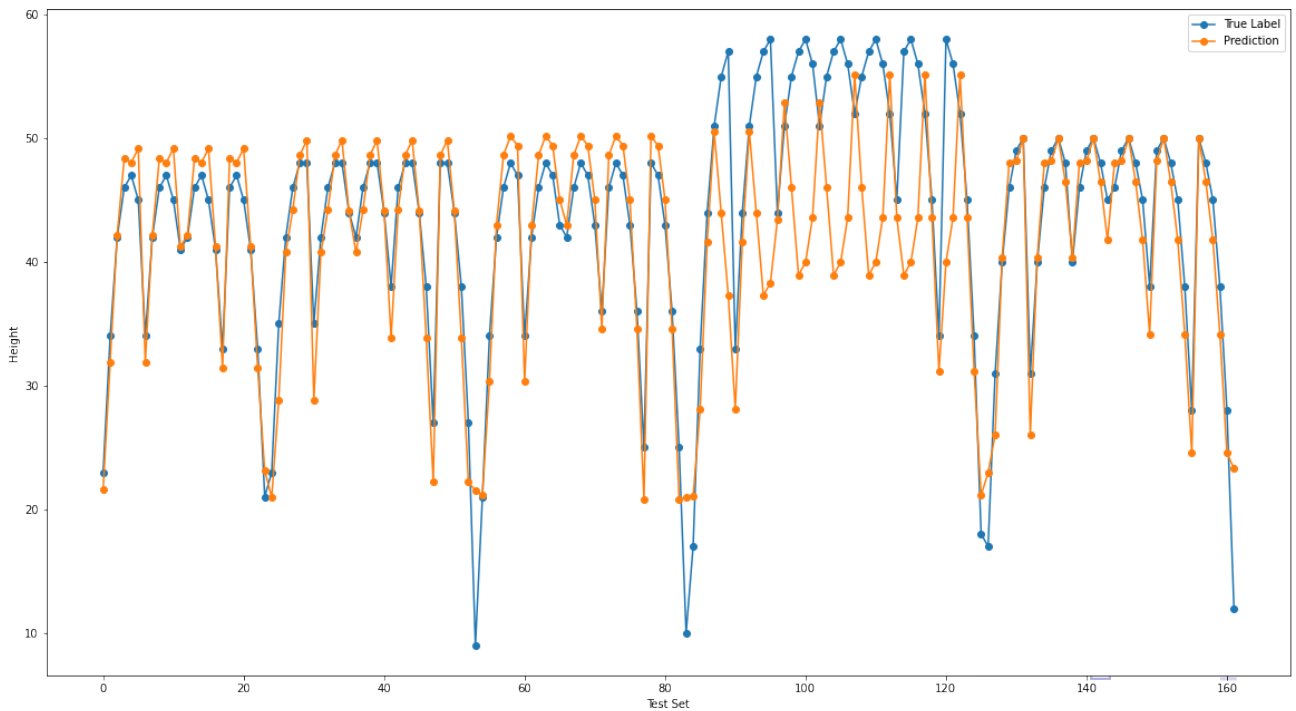


4. Unsupervised Training with Informed ML

2. Results

After adding constant C (height of initial image in a throw) based on the instructions of the paper result improved

1. The best accuracy is 90.74% on the test set
2. Best mean absolute error = 4.02



Lessons Learned

1. Supervised object tracking classification models have better performance than regression (single or sequence input)

2. The larger the sequence length the higher the information model gets from each instance. However, it means fewer data as data size from each throw is equal to: $len(throw) - len(sequence) + 1$

3. When we record our data the distance from the camera is essential to be fixed because it affects the gravitational pull (meter-to-pixel ratio)

4. Model performance improves when provided with a piece of information(height of the initial image in a throw) because the information our loss function provides is the relation of points in a sequence and not actual height, so adjusting $y_0 = 0$ make information of the loss function more relevant.

Future Work

1. Adding noise to the existing data base

2. Follow the same protocol on images with objects in the background instead of white wall

3. Investigate the effectiveness of this approach on other forms of movement rather than free fall

References

1 Russell Stewart , Stefano Ermon, Label-Free Supervision of Neural Networks with Physics and Domain Knowledge,Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence .2576–2582. (AAAI-17)