Mass Estimation From Images Using Deep Neural Network and Sparse Ground Truth

Muhammad Hamdan
Department of Electrical and Computer Engineering
Iowa State University

Dec 19, 2019

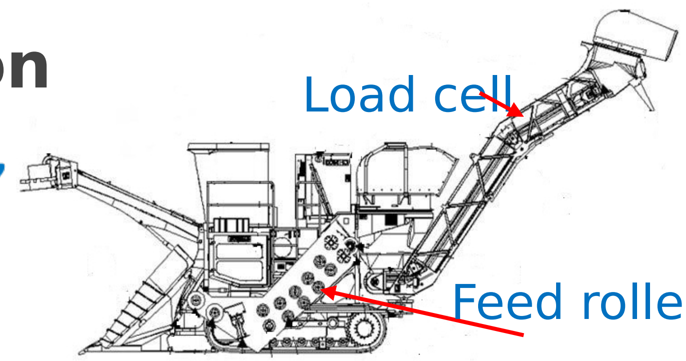
Motivation

Literature Review

Sugarcane mass flow estimation methods

- Mass measurement through **load cell** **[1, 2, 3, 4, 5, 6]**
- Volume measurement through roller displacement **[7]**
- Volume measurement via optical sensor **[8, 9]**
  - Inexpensive, simple, and relatively accurate ( )
  - Depends on ambient light (night time and early morning)
  - Requires calibration and highly affected by changes in material density

Sugarcane mass flow estimation methods
- Mass measurement through load cell [1, 2, 3, 4, 5, 6]
- Volume measurement through roller displacement [7]
- Volume measurement via optical sensor [8, 9]
  - Inexpensive, simple, and relatively accurate ($\mu = 7.5\%, \sigma = 6.3\%$)
  - Depends on ambient light (night time and early morning)
  - Requires calibration and highly affected by changes in material density



- **Mass measurement through images from stereo camera**

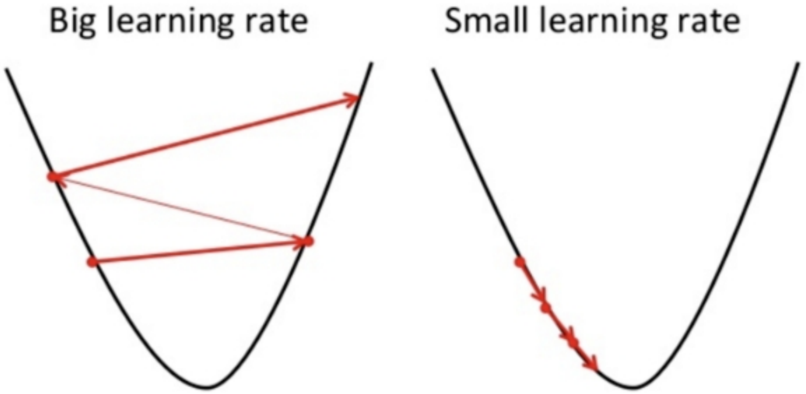
Problem Complexity

- **Factors**
  - Angle of capture
  - Mass flow rate
  - Frame overlap
  - Variable elevator spe
  - Different run sizes
  - Different lighting con
  - Sparse ground truth

Deep Learning Basics

What to consider when deciding on using a DNN?

- CNN architecture
  - AlexNet, VGG, GoogleNet, ResNet, Your own?
- Activation function
  - Sigmoid, Tanh, ReLU, ELU
- Choice of hyper-parameters:
  - Learning rate
- Loss function
  - Classification: Softmax
  - Regression: MSE



$$\mathbf{MSE} = \quad L(y;\hat{y}) = \sum_{i=1}^{k} \frac{1}{n}\Big(y_i - \hat{y}_i\Big)^2$$

Loss Function

$$L_i(x, y; w) = \frac{1}{n_i}\left\{y_i - \boxed{\sum_{j=1}^{n_i}(f(x_{ij}; w) \times v_{ij} \times t})\right\}^2$$

$$L_i(x, y; w) = \frac{1}{n_i}\left\{y_i - \boxed{\sum_{j=1}^{n_i} \hat{y}_{ij}}\right\}^2$$

that we handled frame overlap, we need to figure out to obtain correct predictions per frame

Gradient Update

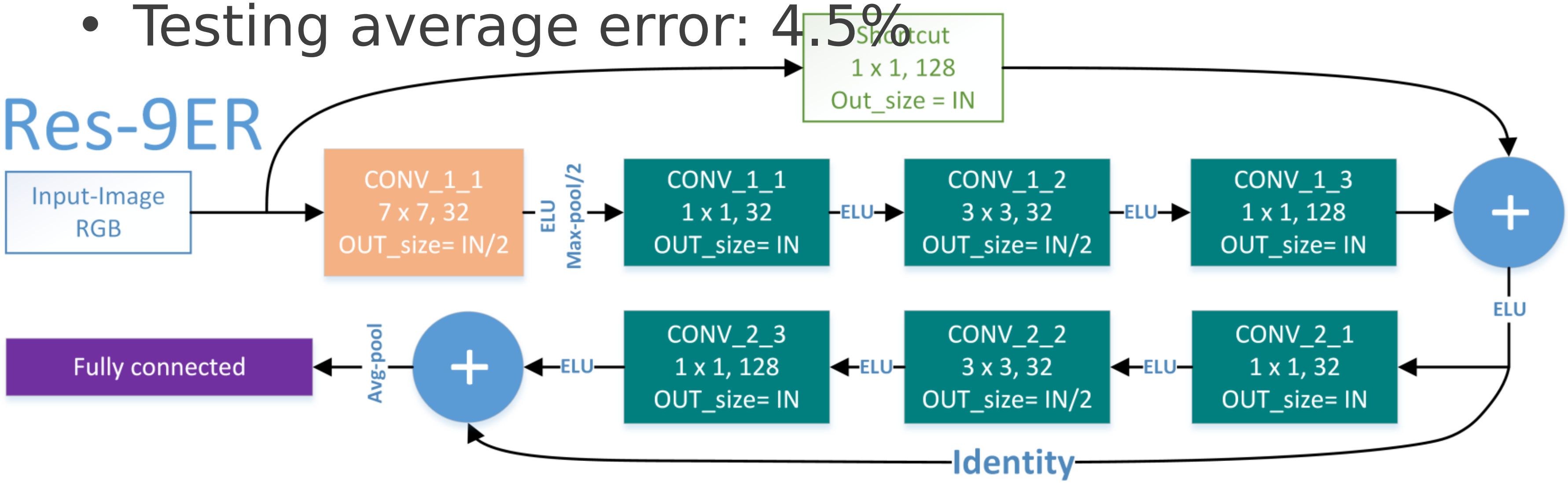
- Our loss function
- Gradient update occurs at every end of a run
- We keep a running sum of gradients and predictions
- Compute the derivative of the loss function to apply loss

$$\frac{\partial L_i}{\partial w} \leftarrow -\frac{2}{n_i}\left[y_i - \sum_{j=1}^{n_i} \hat{y}_{ij}\right] \times \sum_{j=1}^{n_i} \frac{\partial \hat{y}_{ij}}{\partial w}$$

DNN Architecture Summary

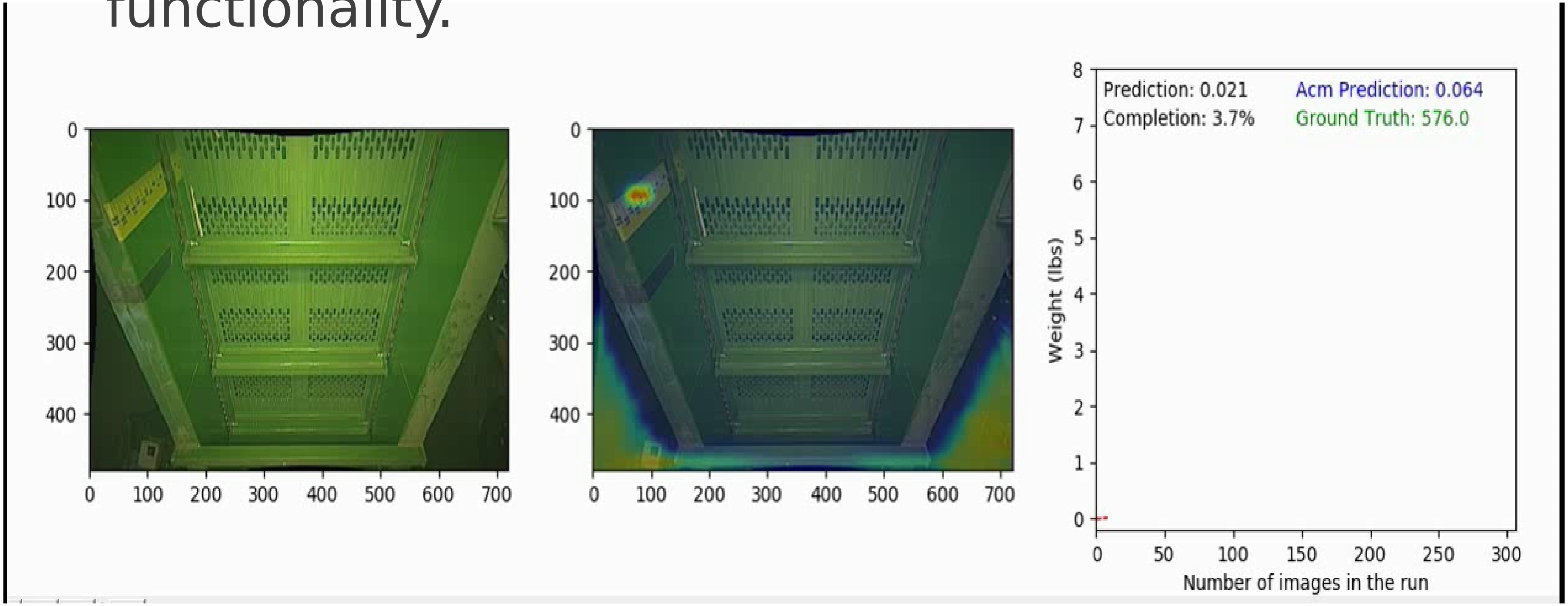
- DNN Architecture
  - Input image size: 96 x 144 (5th original size)
  - Parameters: 4K and Size of parameters: 0.17 MB
  - Training time: ~11 hours
  - Testing average error: 4.5%

  - Input image size: 96 x 144 (5th original size)
  - Parameters: K and Size of parameters: 0.17 MB
  - Training time: ~hours
  - Testing average error: 4.5%

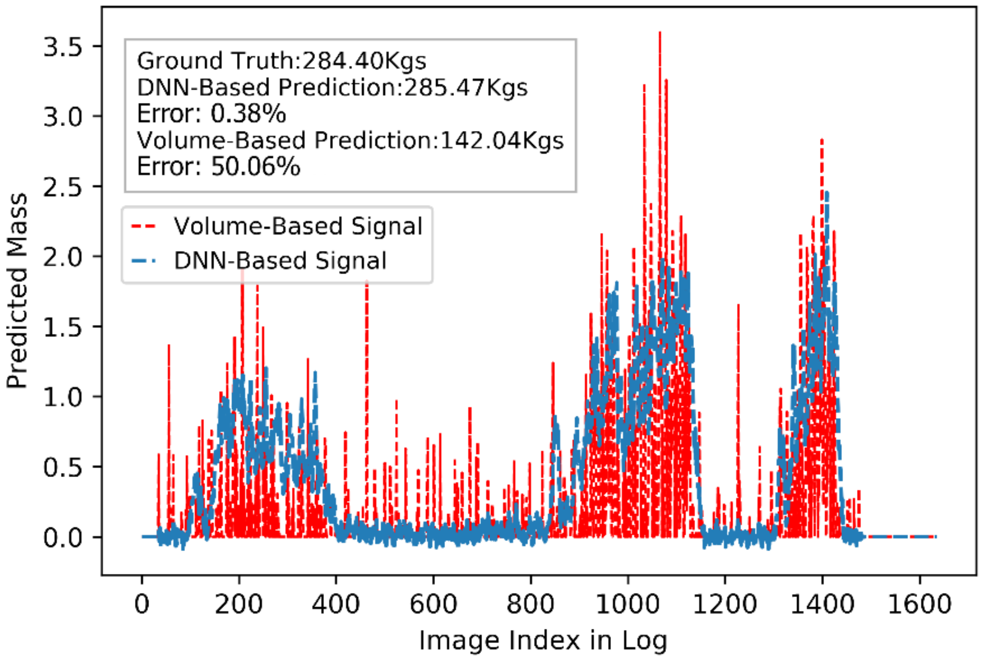


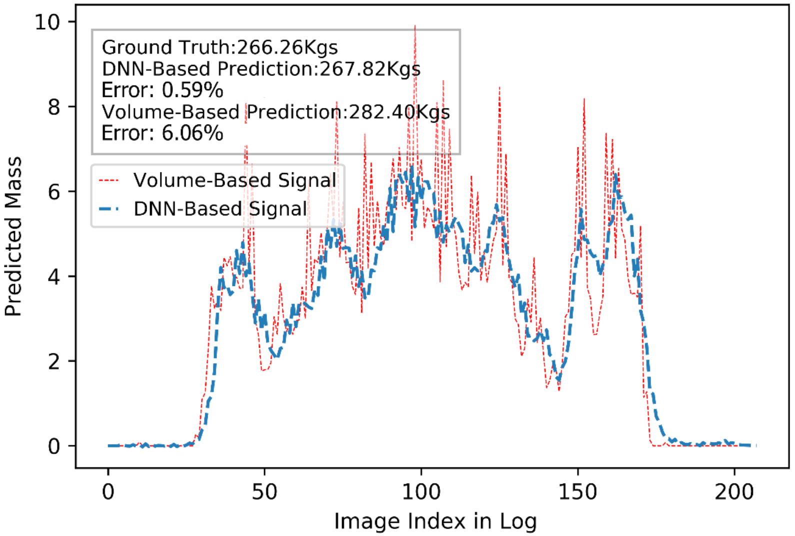
What is Going on Behind the Scenes?

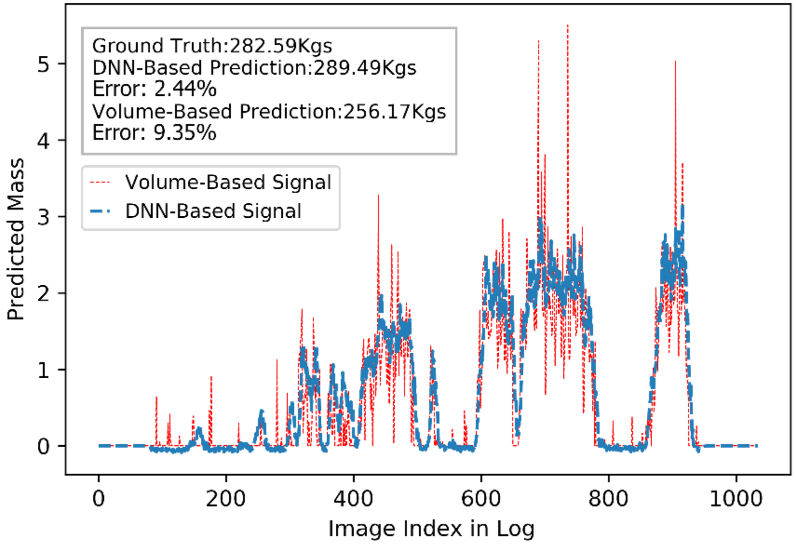
- Proper visualization techniques can support the investigation of DNN functionality.



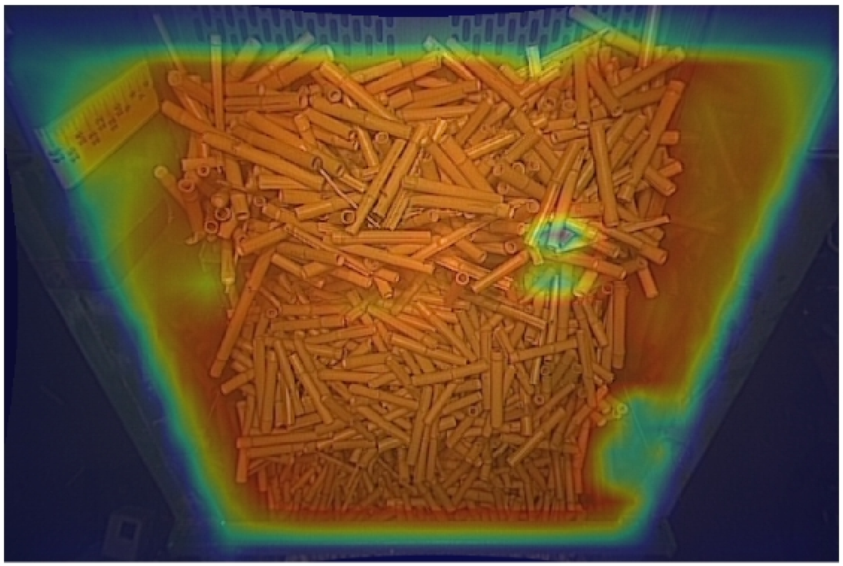
Robustness

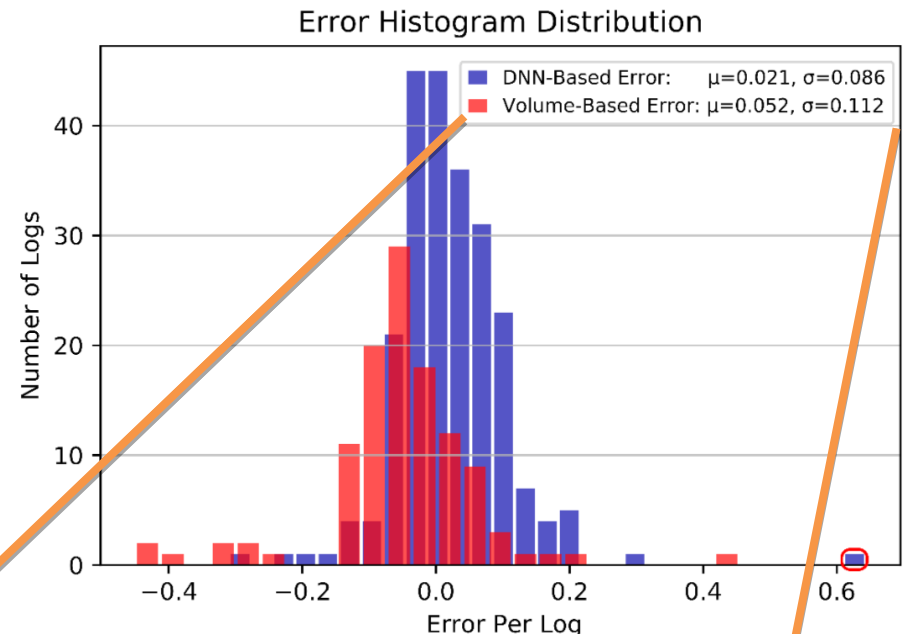

Ground Truth:284.40Kgs
DNN-Based Prediction:285.47Kgs
Error: 0.38%
Volume-Based Prediction:142.04Kgs
Error: 50.06%
Volume-Based Signal
DNN-Based Signal
Predicted Mass
Image Index in Log
0.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
0
200
400
600
800
1000
1200
1400
1600

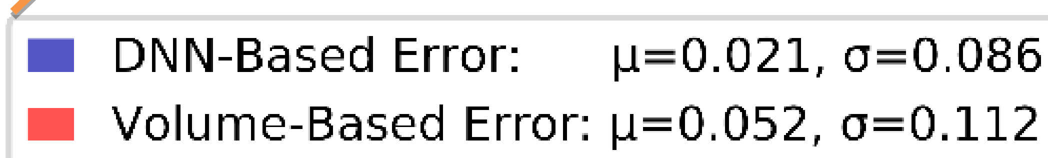

Ground Truth:266.26Kgs
DNN-Based Prediction:267.82Kgs
Error: 0.59%
Volume-Based Prediction:282.40Kgs
Error: 6.06%
Volume-Based Signal
DNN-Based Signal
Predicted Mass
Image Index in Log
0
2
4
6
8
10
0
50
100
150
200


Ground Truth:282.59Kgs
DNN-Based Prediction:289.49Kgs
Error: 2.44%
Volume-Based Prediction:256.17Kgs
Error: 9.35%
Volume-Based Signal
DNN-Based Signal
Predicted Mass
Image Index in Log
0
1
2
3
4
5
0
200
400
600
800
1000

Histogram Distribution of Error and Outliers







Questions

Questions?

References

[1] Graeme Cox, H Harris, R Pax, and R Dick. Monitoring cane yield by measuring mass flow rate through the harvester. In PROCEEDINGS-AUSTRALIAN SOCIETY OF SUGAR CANE TECHNOLOGISTS, pages 152{157. WATSON FERGUSON AND COMPANY, 1996
[2] G Cox, H Harris, and R Pax. Development and testing of a prototype yield mapping system. In Proceedings-Australian Society of Sugar Cane Technologists, pages 38{43. WATSON FERGUSON AND COMPANY, 1997.
[3] NB Pagnano and PG Magalhaes. Sugarcane yield measurement. faculdade de engenharia agricola unicamp campinas sp, brazil 13083-970. In Proceeding of 3rd European Conference on Precision Agriculture, pages 839{844, 2001.
[4] JP Molin and LAA Menegatti. Field-testing of a sugar cane yield monitor in brazil. In 2004 ASAE Annual Meeting, page 1. American Society of Agricultural and Biological Engineers, 2004.
[5] Domingos GP Cerri and Paulo Graziano Magalh~aes. Sugar cane yield monitor. In 2005 ASAE Annual Meeting, page 1. American Society of Agricultural and Biological Engineers, 2005.
[6] Mike Mailander, Caryn Benjamin, Randy Price, and Steven Hall. Sugar cane yield monitoring system. Applied engineering in agriculture, 26(6):965–969, 2010.
[7] Cox, Graeme J. "A yield mapping system for sugar cane chopper harvesters." PhD diss., University of Southern Queensland, 2002.
[8] Mike Mailander, Caryn Benjamin, Randy Price, and Steven Hall. Sugar cane yield monitoring system. Applied engineering in agriculture, 26(6):965{969, 2010
[9] RR Price, RM Johnson, RP Viator, J Larsen, and A Peters. Fiber optic yield monitor for a sugarcane harvester. Transactions of the ASABE, 54(1):31–39, 2011

Empty Slide

Empty Slide

Volumetric-Based Approach to Mass estimation

- Instant volume measurement is available
- Ground truth (true mass) is only available by run

- Instant volume measurement is available
- Ground truth (true mass) is only available by run

$$\frac{\text{MASS}}{\text{sec}} = \frac{\text{Volume}}{\text{sec}} \times \text{DENSITY}$$

$$Mass = f(max(V-\beta,0);\theta) \times max(V-\beta,0) \times v_{elev} \times t$$

Where "f" is a 2-layer neural network parameterized by "θ" that outputs a prediction of density based on the volume (V), scaled by elevator speed (V_{elev}) and capture time (t), with tanh activation

Where "f" is a 2-layer neural network parameterized by "" that outputs a prediction of density based on the volume (V), scaled by elevator speed () and capture time (t), with tanh activation.

$Percent\ error_{avg}$ using neural network **including** low ligh runs: $\mathbf{12.58}\%$

$Percent\ error_{avg}$ using neural network **without** low ligh runs: $\mathbf{8.65}\%$

Data Summary

Laboratory data summary

Location	Runs	Samples	Material	Representation	Environment
ISU	239	>120K	Bamboo	Images and point cloud	Controlled

Laboratory etup

Temporal Smoothness

- Images near in time should have more similarity in mass than images further away in time
- Hyper-parameter λ (chosen empirically 0.05)
- This term is added to the loss function

$$L_i(x, y; w) = \frac{1}{n_i}\left\{y_i - \sum_{j=1}^{n_i}(f(x_{ij}; w) \times v_{ij} \times t)\right\}^2 +$$

$$\frac{\lambda}{n_i}\sum_{j=1}^{n_i}\left\{f(x_{ij}; w) - f(x_{i(j-1)}; w)\right\}^2$$