A METHOD OF DATA LABEL CHECKING AND THE WRONG LABELS IN NIST SD19V2 AND MNIST

A PROOF OF ‘BAD’ HANDWRITING IN MNIST TRAINING DATASET MAKING CNN TO PREDICT ‘GOOD’ HANDWRITING WRONG

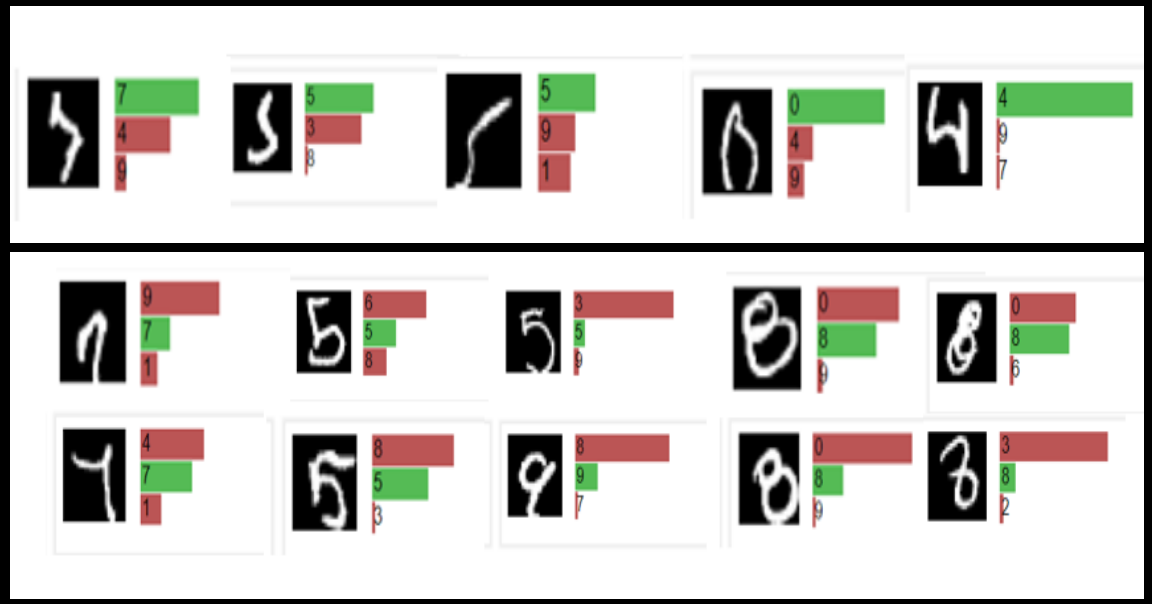
DeepAI

Abner.Zhang Steven Lu

WHY 99% ACCURACY CNN ARE WRONG AT GOOD HANDWRITING

Bad handwriting

Good handwriting



convnetjs

Karpathy

Possible Reason: ‘bad’ handwriting in training dataset

- SD-3 Census Bureau employees
- SD-1 high-school students
- SD-3 is much cleaner and easier to recognize than SD-1
- The MNIST training set is composed of 30,000 patterns from SD-3 and 30,000 patterns from SD-1

Yann LeCun

DATA: NIST SD19V2 WITH ID AND ORIGINAL FORM
METHOD: 360-DEGREE PERFORMANCE EVALUATION

CNN 1-5 training data
From high school

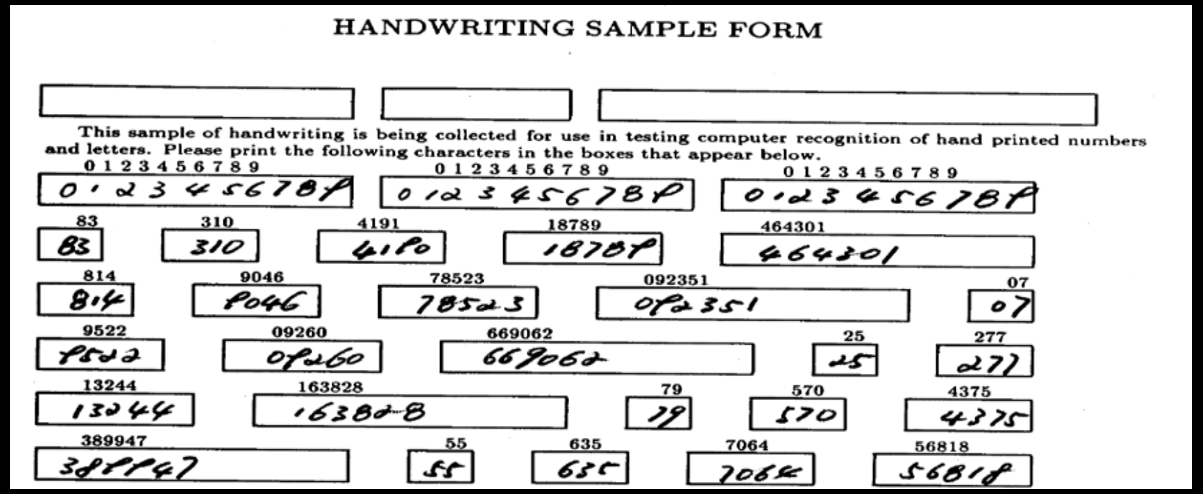
CNN 6-10 training data
From employee

Mixed group

15 CNNs predict
100,000 digits of
905 writers

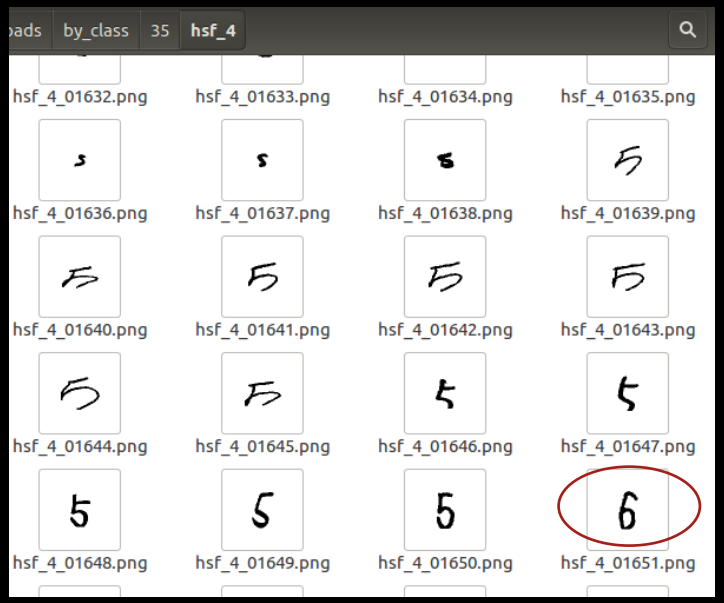
RESULT OF DIGITS OF WRITER F2564

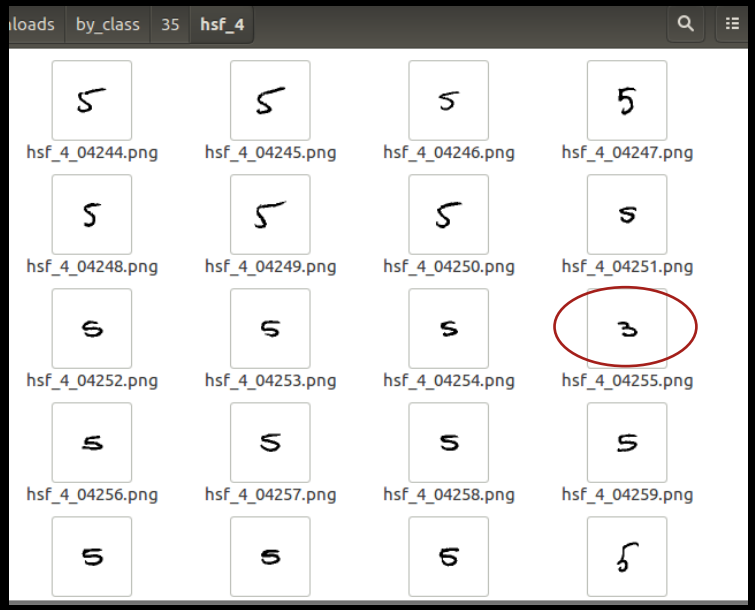
	digits	predby 1	predby 2	predby 3	predby 4	predby 5	predby 6	predby 7	predby 8	predby 9	predby 10	predby 16	predby 27	predby 38	predby 49	predby 510
d2564_59_00004.png	1	1	8	8	8	8	8	9	9	9	8	1	8	8	8	8



original form

‘BAD’ HANDWRITING: WRONG LABELED





SD19 V2

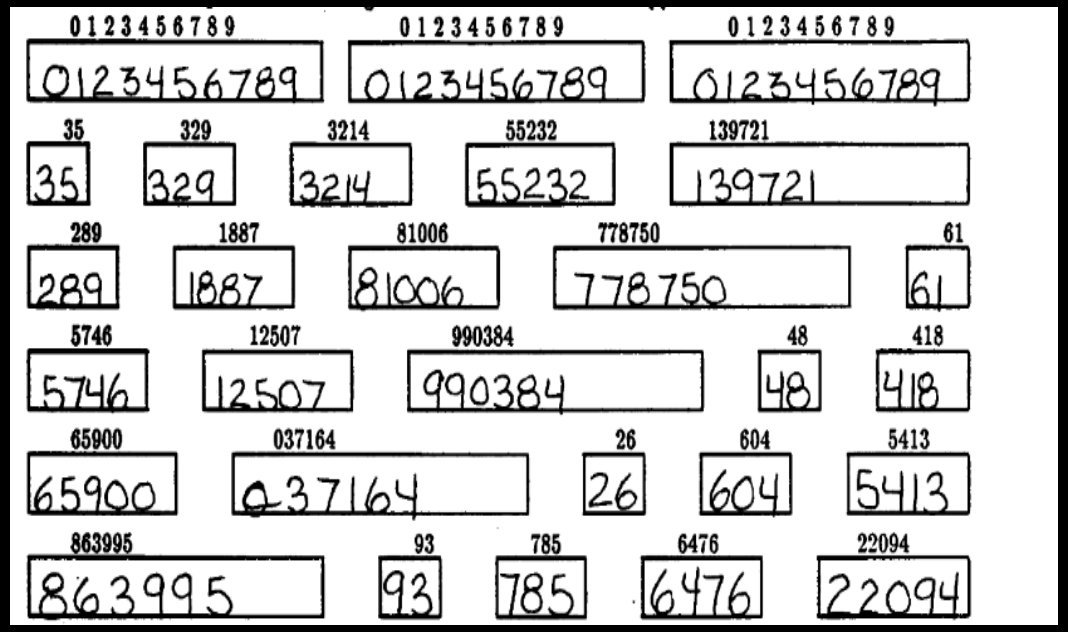
54915 4

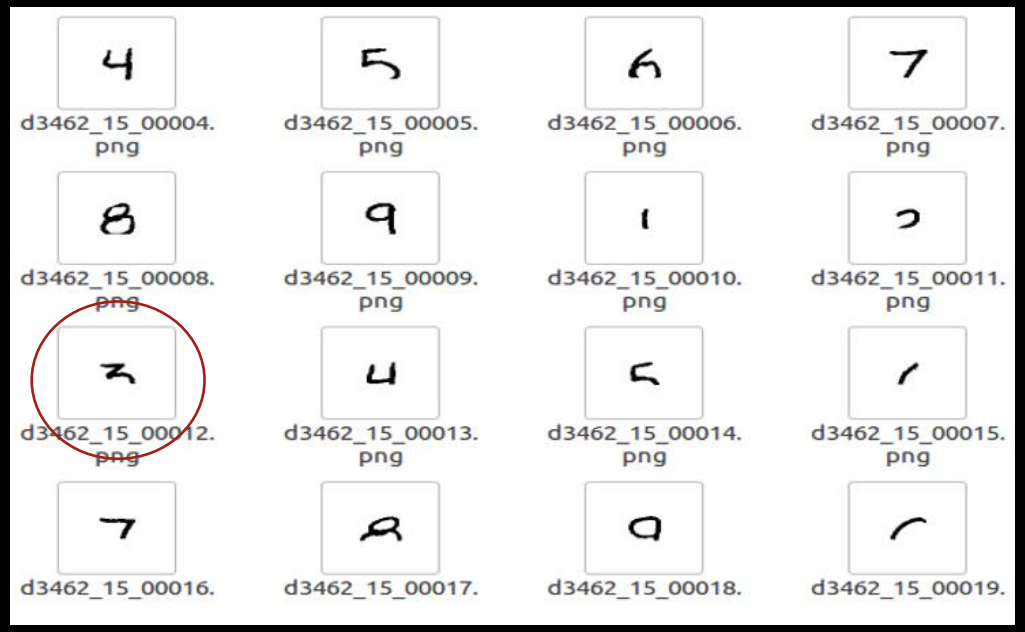
45340 3

44960 6

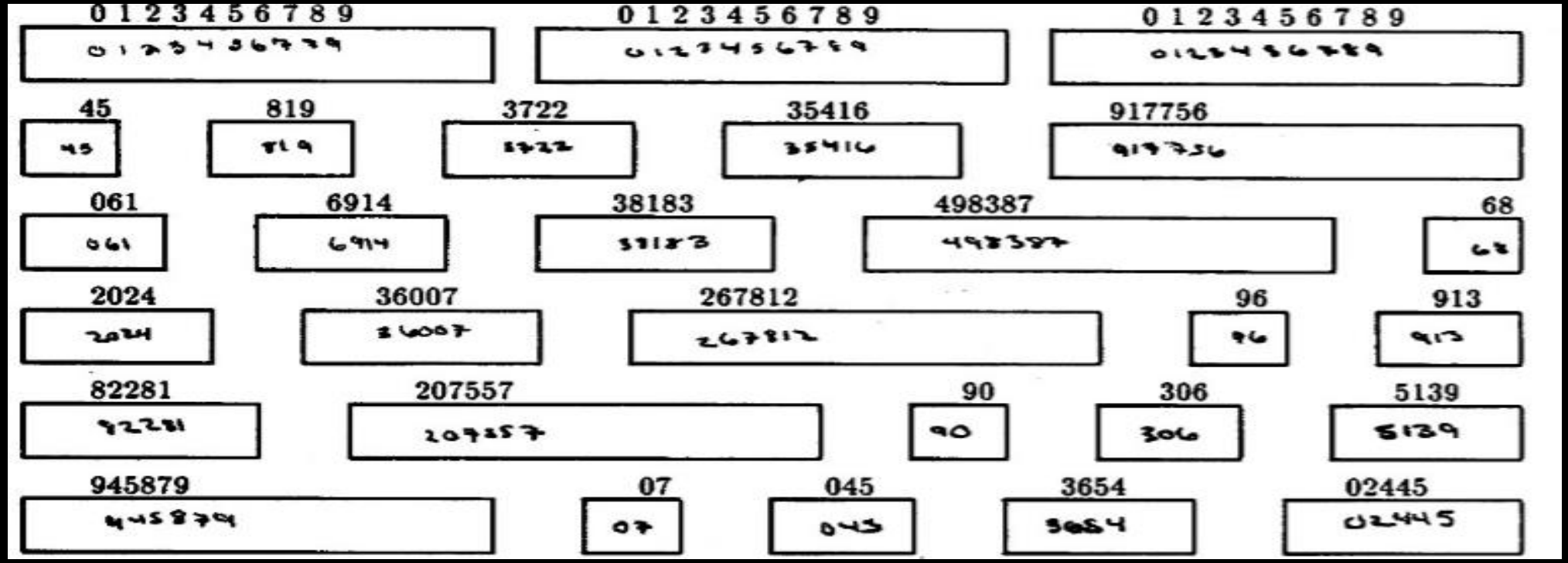
MNIST

‘BAD’ HANDWRITING: WRONG CUT

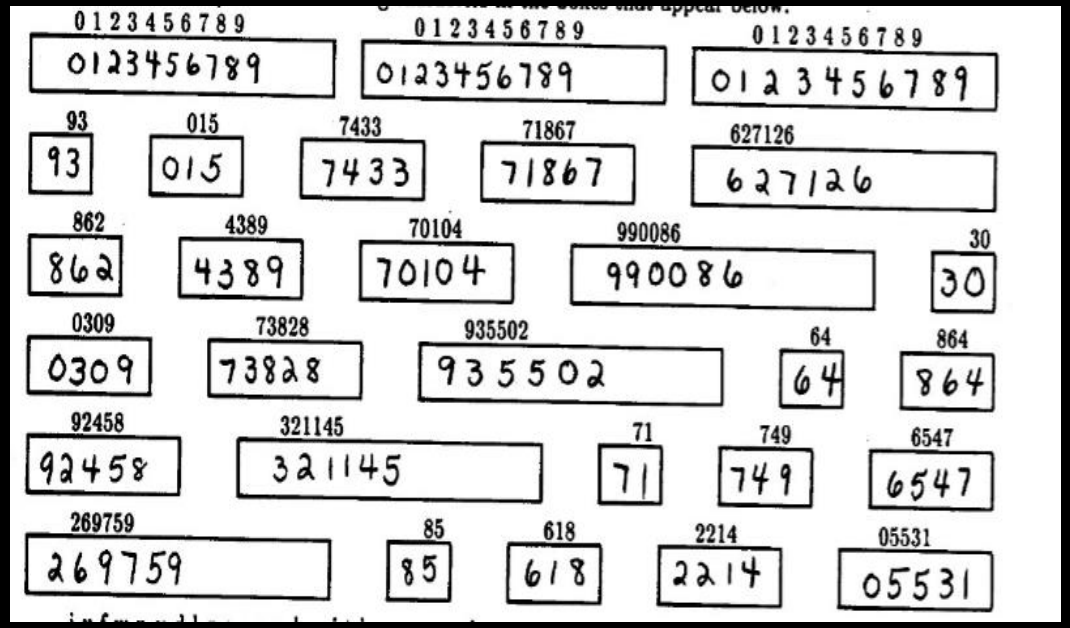


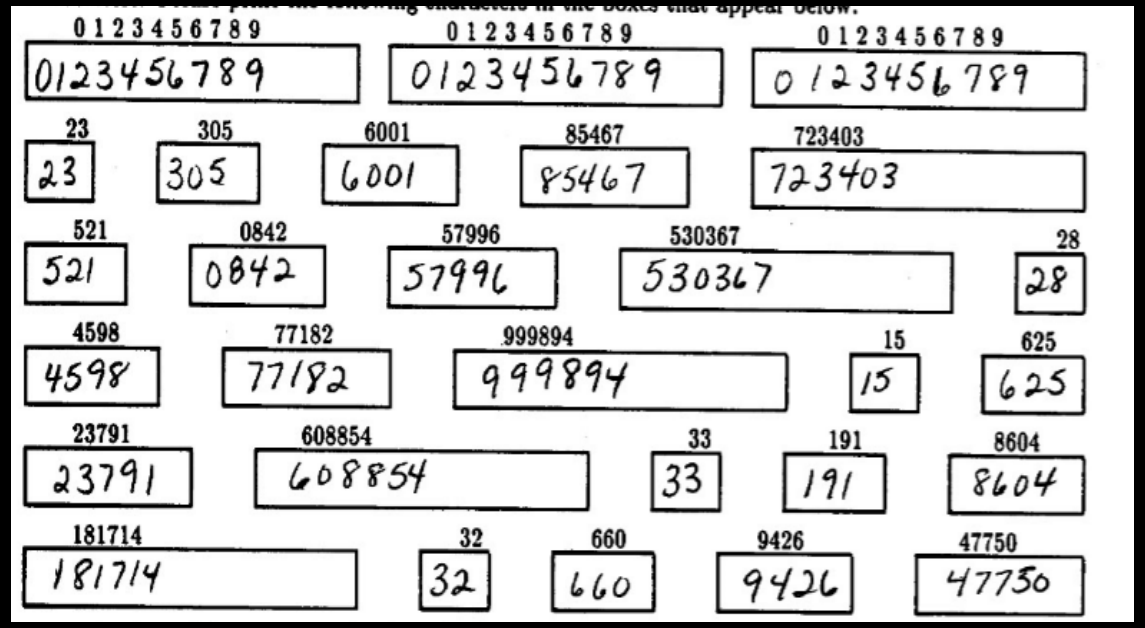


‘BAD’ HANDWRITING: MESSED



‘GOOD’ HANDWRITING DATA





PROOF

	CNN learn only ‘good’	CNN also learn ‘bad’
‘good’ test data	5 (99.9%)	22(99.5%)

It approves that after the CNN has learnt more 31570 ‘bad’ training data, it make 17 more errors at ‘good’ test data than the CNN only learnt ‘good’ handwriting.

http://ssrn.com/abstract=3056117

http://ssrn.com/abstract=3049684

WHAT'S NEXT?

- To build a 360-degree evaluation method in Keras to relabel data
- To check wrong labeled or cut digits pictures in SD19v2
- To find out why there is still 5 errors
- 99.9% accuracy is better than 100%,

- Thanks