

MNIST-Augmentaion

<https://github.com/zhangabner>

MNIST-Augmentaion

- Why this new dataset similar to MNIST?
- Please quit caring about MNIST

(Goodfellow 2017)

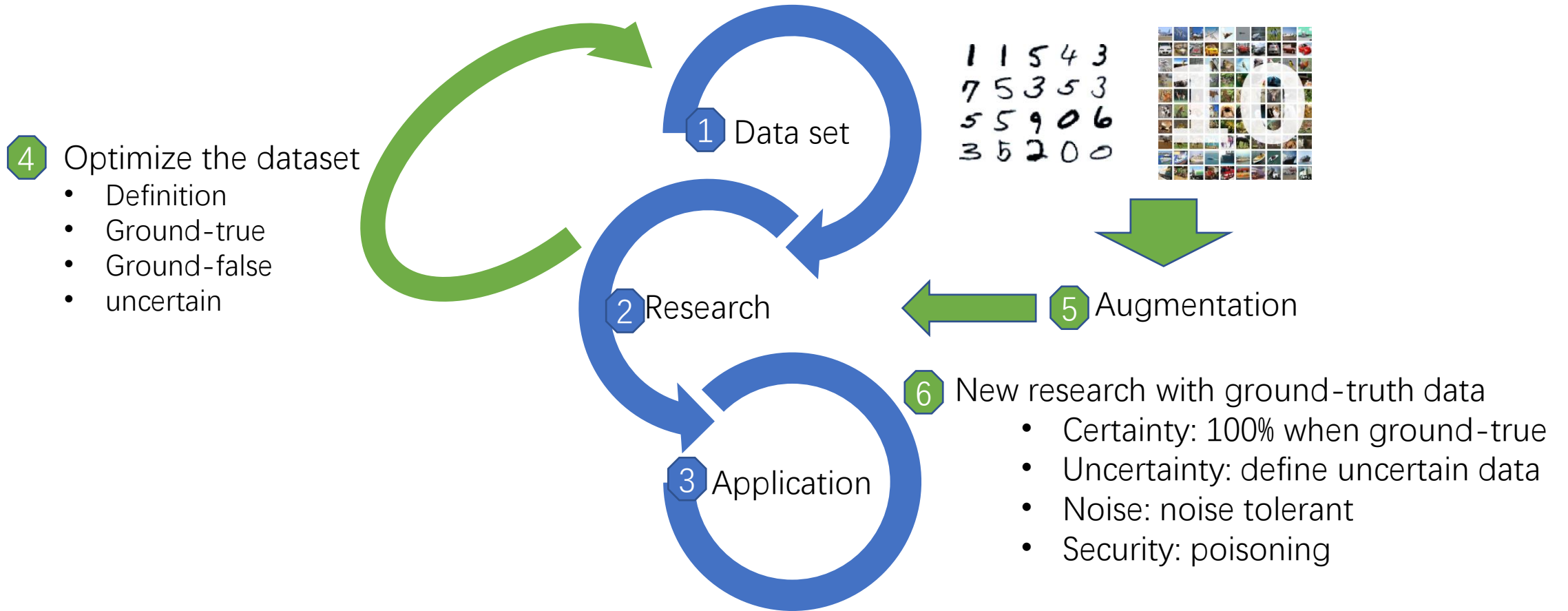
To Serious Machine Learning Researchers

Seriously, we are talking about replacing MNIST. Here are some good reasons:

- **MNIST is too easy.** Convolutional nets can achieve 99.7% on MNIST. Classic machine learning algorithms can also achieve 97% easily. Check out [our side-by-side benchmark for Fashion-MNIST vs. MNIST](#), and read "[Most pairs of MNIST digits can be distinguished pretty well by just one pixel.](#)"
- **MNIST is overused.** In [this April 2017 Twitter thread](#), Google Brain research scientist and deep learning expert Ian Goodfellow calls for people to move away from MNIST.
- **MNIST can not represent modern CV tasks**, as noted in [this April 2017 Twitter thread](#), deep learning expert/Keras author François Chollet.

(fashion-mnist)

A good dataset will motivate a thousands good papers



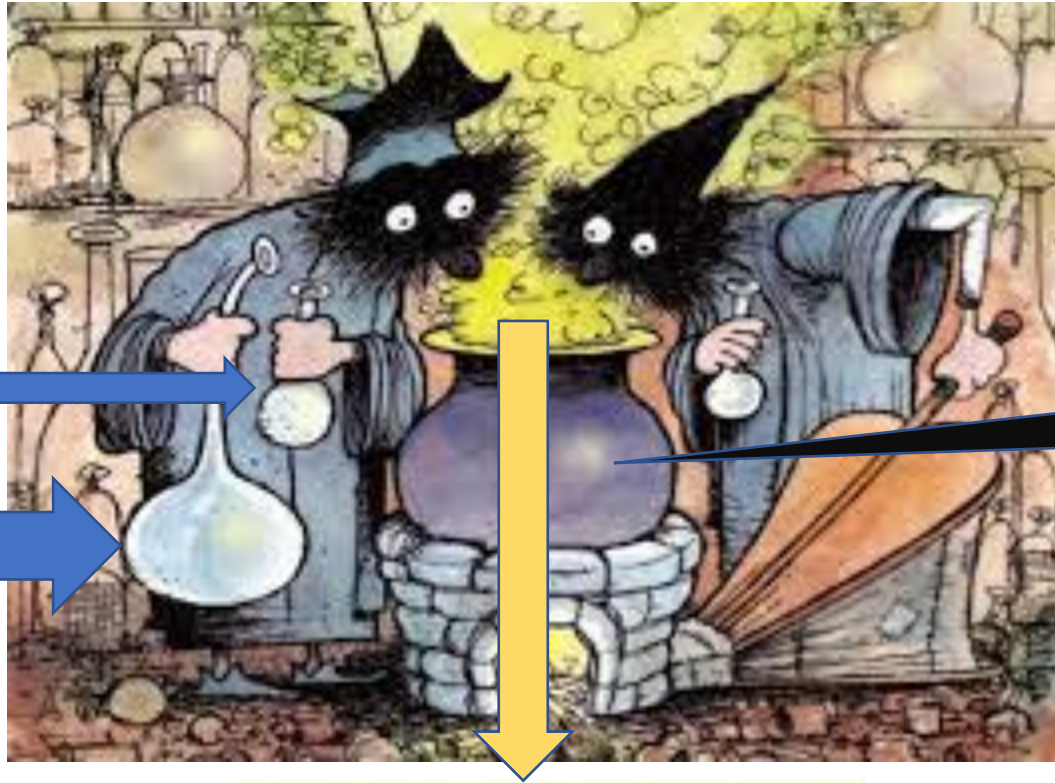
Current state of research based on MNIST

MNIST test

- Main part: ground true
- Some part: uncertain
- Few: ground false

MNIST training

- Main part: ground true
- Some part: uncertain
- Few: ground false

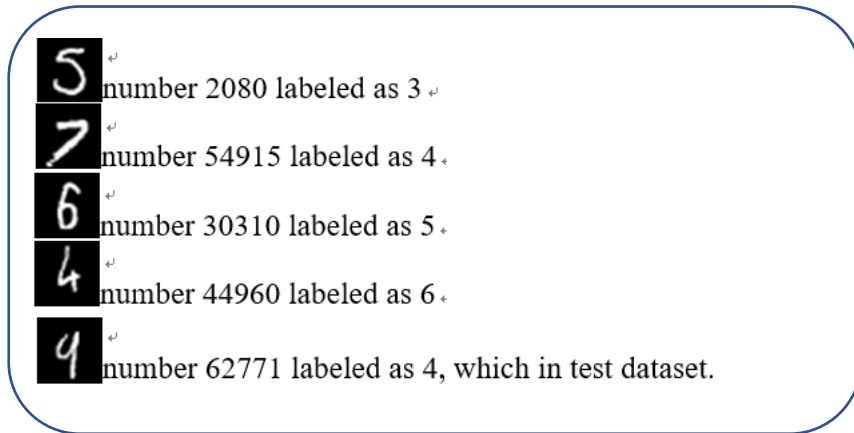


What's really happened there?

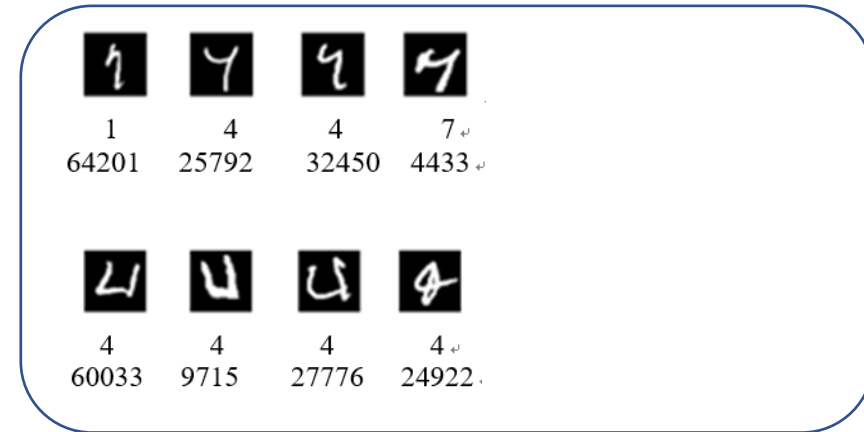
Result	Method	Venue	Details
0.21%	Regularization of Neural Networks using DropConnect	ICML 2013	
0.23%	Multi-column Deep Neural Networks for Image Classification	CVPR 2012	
0.23%	APAC: Augmented Pattern Classification with Neural Networks	arXiv 2015	
0.24%	Batch-normalized Maxout Network in Network	arXiv 2015	Details
0.29%	Generalizing Pooling Functions in Convolutional Neural Networks: Mixed, Gated, and Tree	AISTATS 2016	Details
0.31%	Recurrent Convolutional Neural Network for Object Recognition	CVPR 2015	
0.31%	On the Importance of Normalisation Layers in Deep Learning with Piecewise Linear Activation Units	arXiv 2015	
0.32%	Fractional Max-Pooling	arXiv 2015	Details
0.33%	Competitive Multi-scale Convolution	arXiv 2015	

The state of the art on MNIST

Ground-false and uncertain data in MNIST



Ground-false instances by applying majority voting on MNIST



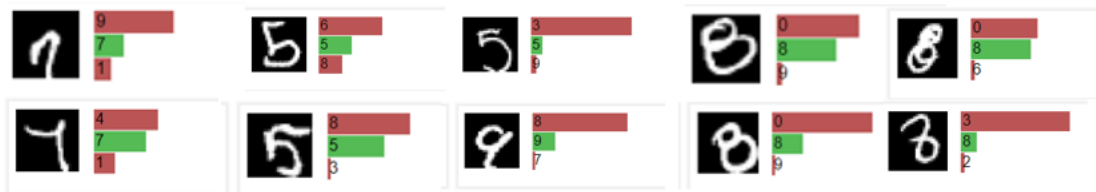
Uncertain instances by analyzing the process of NIST SD19v2

Do we understand the prediction of CNN on MNIST?

The Karpathy's convnetjs can predict these correctly: ↵



but make errors at some 'good' handwriting such as: ↵



Turn the mixed dataset into ground-truth dataset



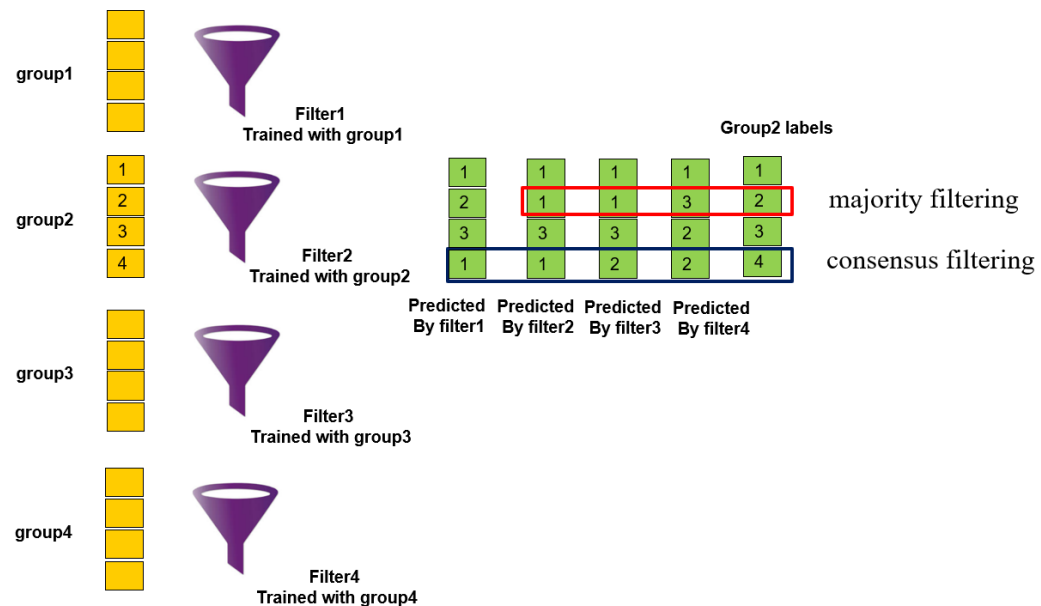
Ground-true

Ground-false

Uncertain



Majority voting to split MNIST



Majority voting process

count of wrong predicted times	count of digits	
0	62161	Ground true
1	2059	Uncertain
2	1008	
3	733	
4 (Majority Vote)	622	
5 (Majority Vote)	593	Ground false
6 (Majority Vote)	722	
7 (Consensus Filters)	2023	

Majority voting result

Majority voting by
[Brodley *et al.*, 1999]

Why that good handwriting being predicted wrong?

Training

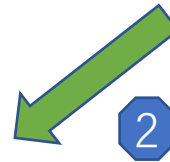
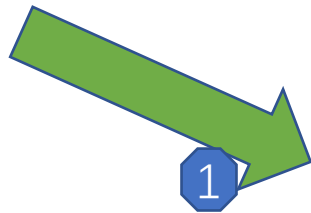


Test



Ground true Ground false
Uncertain

Ground true Uncertain



2 Accuracy = 100%
Ensemble method, num_networks = 5

Why that good handwriting being predicted wrong?

Training

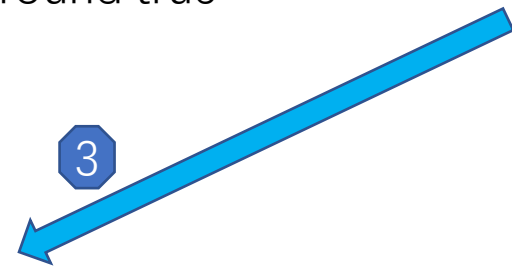
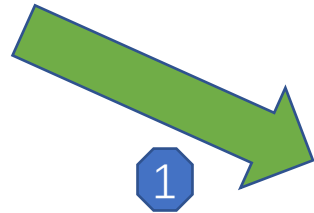


Test



Ground true Ground false Uncertain

Ground true Uncertain



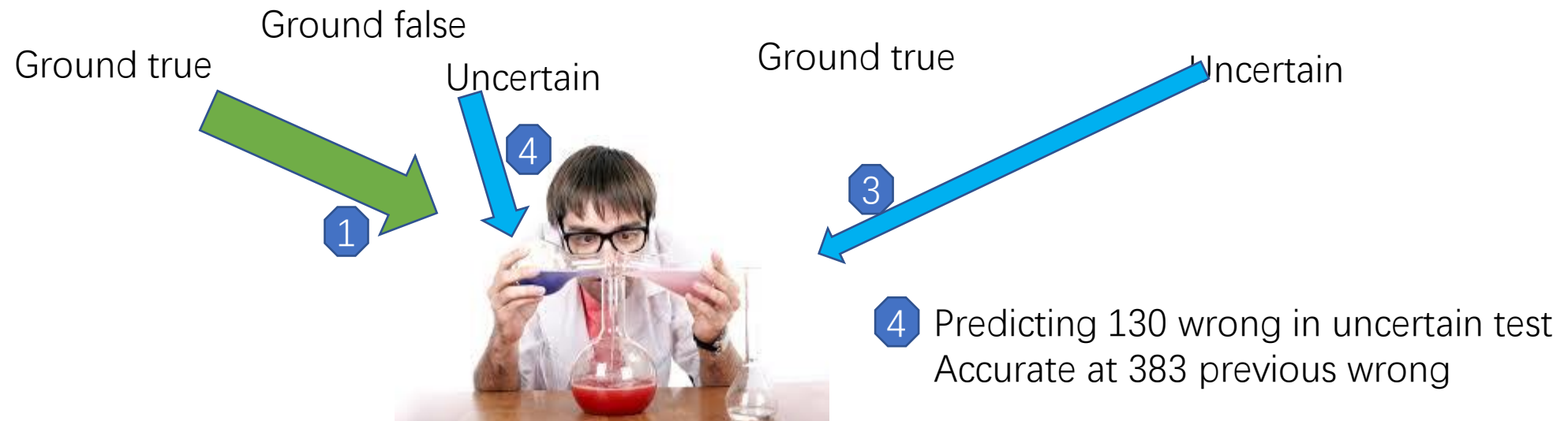
3 Predicting 513 wrong in uncertain test

Why that good handwriting being predicted wrong?

Training



Test

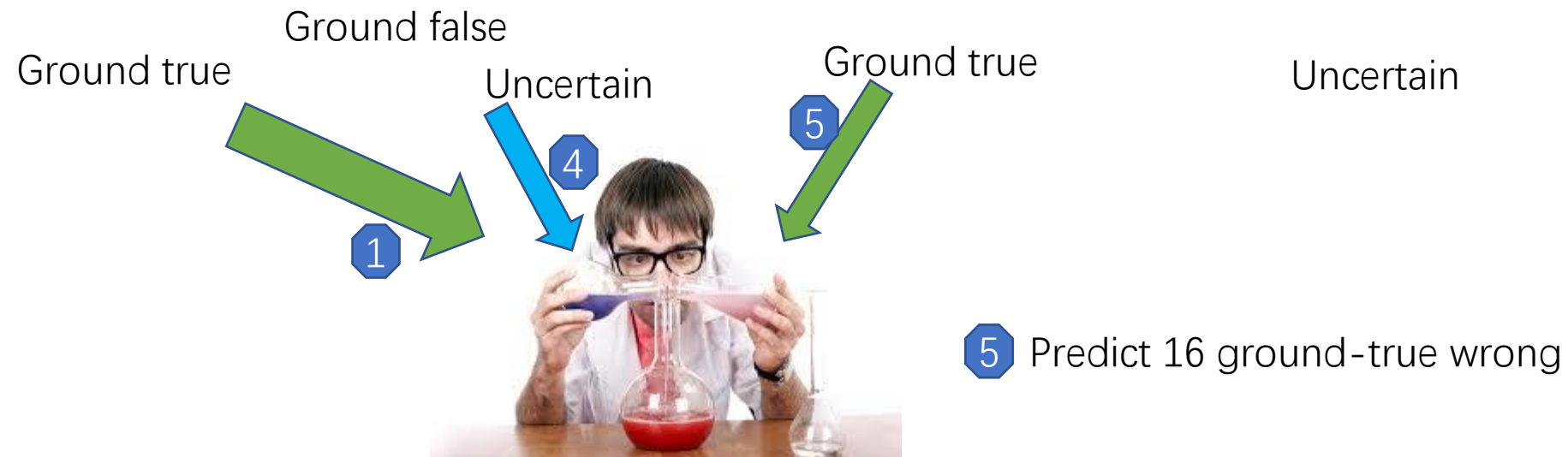


Why that good handwriting being predicted wrong?

Training



Test



Adding more uncertain data in training

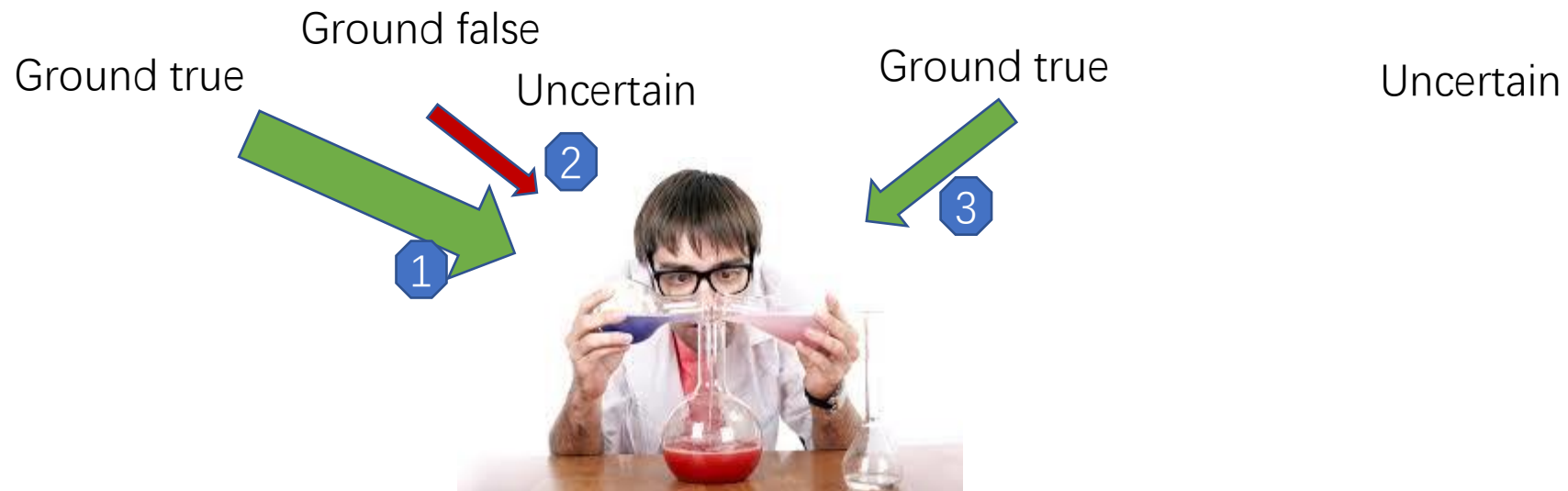
	Dn0'	Dn1'	Dn2'	Dn3'	Dn4'	Dn5'	Dn6'	Dn7'
Dn0	0	2	18	37	53	61	96	276
Dn0+n1	0	1	5	18	46	46	87	271
Dn0+n2	0	1	6	15	31	39	80	269
Dn0+n3	1	2	3	18	38	33	67	251
Dn0+n4	1	4	6	16	21	32	52	253
Dn0+n5	3	2	7	20	19	18	50	232
Dn0+n6	5	4	6	21	17	17	33	211
Dn0+n7	15	4	9	8	6	5	13	93
Dtraining	16	4	7	11	5	5	11	70

Poisoning Attack

Training



Test



Example: poisoning attack

$F(M2, Dn0, x61466) = 5$

$F(M2, Dn0 + d2080, x61466) = 3$

$F(M2, Dn0, x68553) = 5$

$F(M2, Dn0 + d2080, x68553) = 3$

$F(M2, D0, x61941) = 7$

$F(M2, D0 + d54915, x61941) = 4$

The poisoning effect

5	number 2080 labeled as 3
7	number 54915 labeled as 4
6	number 30310 labeled as 5
4	number 44960 labeled as 6

$F(M2, Dn0, x61466) = 5$

$F(M2, Dn0 + d2080, x61466) = 3$

$F(M2, Dn0 + n7, x61466) = 5$

$F(M2, Dn0, x68553) = 5$

$F(M2, Dn0 + d2080, x68553) = 3$

$F(M2, Dn0 + n7, x68553) = 5$

The neutralized effect

5	x61466, y61466 is 5
5	x68553, y68553 is 5
7	x61941, y61941 is 7

$F(M2, Dn0 + n1, x60115) = y60115$

$F(M2, Dn0 + n2, x60115) = y60115$

$F(M2, Dn0 + n3, x60115) = y60115$

$F(M2, Dn0 + n4, x60115) = y60115$

$F(M2, Dn0 + n5, x60115) = y60115$

$F(M2, Dn0 + n6, x60115) = y60115$

$F(M2, Dn0 + n7, x60115) = y60115$

$F(M2, Dn0 + n1-7, x60115) \neq y60115$

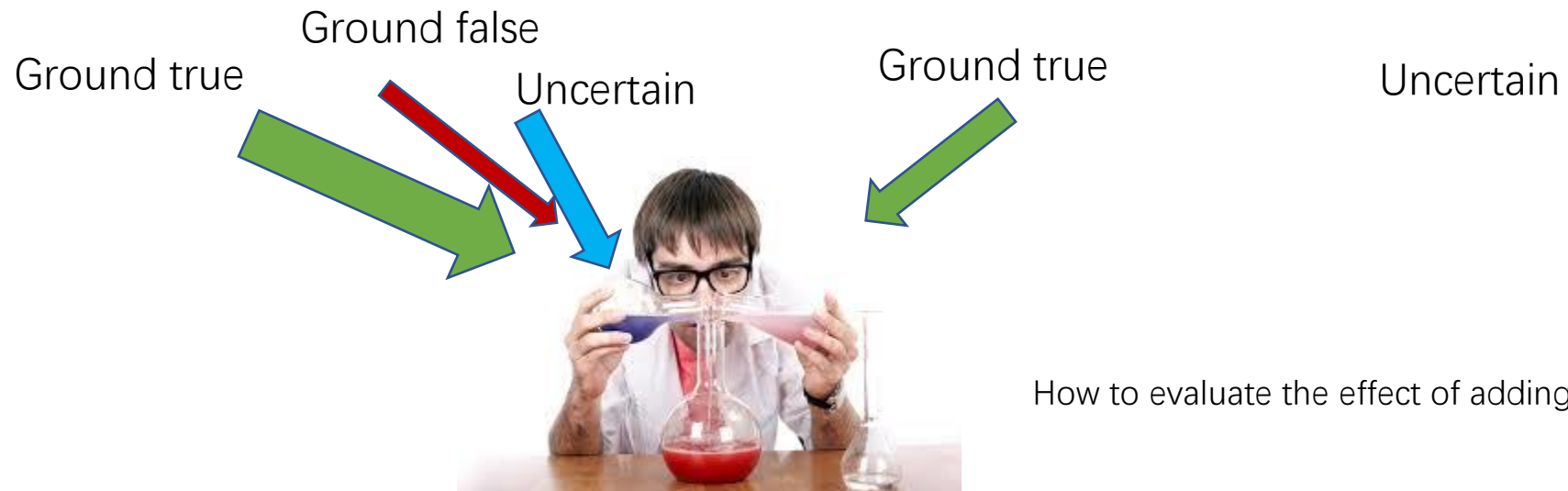
The compounded effect

Poisoning Attack

Training



Test



How to evaluate the effect of adding more Ground false?

Pseudo label in MNIST



12958 in the context of form is 9, but in a large context of 0-9 and a-z, it should be 'g', which is unknown in 0-9.

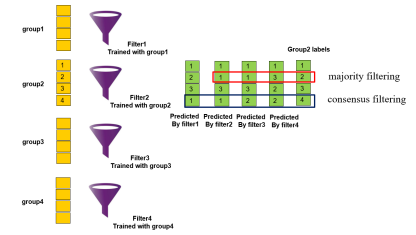


23652 is labeled as 4 and some readers may think it is closer to 9, but in a large context of 0-9 and a-z, it should be 'q', which is unknown in 0-9.

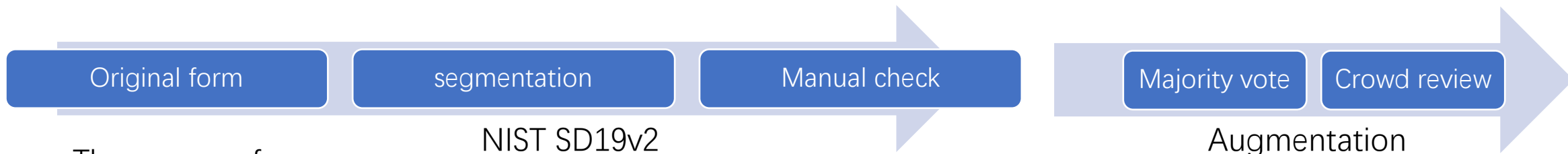
Reason:

- Lack of ground-truth source
- Definition:
- It must be one of digits
- Bias:
- Different personal views

How to be ground-truth



HSF_4 and HSF_6
images
groundtrueV01.csv



The source of Ground-truth

NIST SD19v2

Augmentation

Definition:

- The print version
- Distinguishable

CIFAR-10

By web crawlers from internet

Tiny Images dataset 80 million 32×32

CIFAR-10 by students

Lack the source of ground-truth clear images



Image 16925 and 22490 are labeled as automobile.



Image 55416 and 5013 are labeled as truck.

- “The only criteria for including an image were that the image contain one dominant instance of a CIFAR-10 class, and that the object in the image be easily identifiable as belonging to the class indicated by the image label.”
- “There is no overlap between automobiles and trucks. “Automobile” includes sedans, SUVs, things of that sort. “Truck” includes only big trucks. Neither includes pickup trucks”

4 Unknown labels:



1569 labeled as a deer



18310 labeled as truck



5074 labeled as a deer



52226 labeled as a bird

4 Wrong labels:



52405 labeled as cat



52804 labeled as cat



21347 labeled as cat



17455 labeled as a cat

4 Multi-objects labels



15696: a bird on a car



56859 is a cat on a car

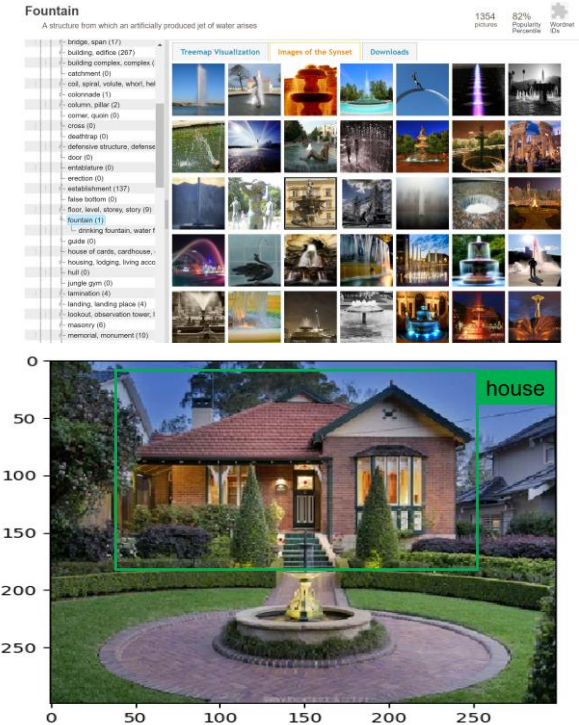
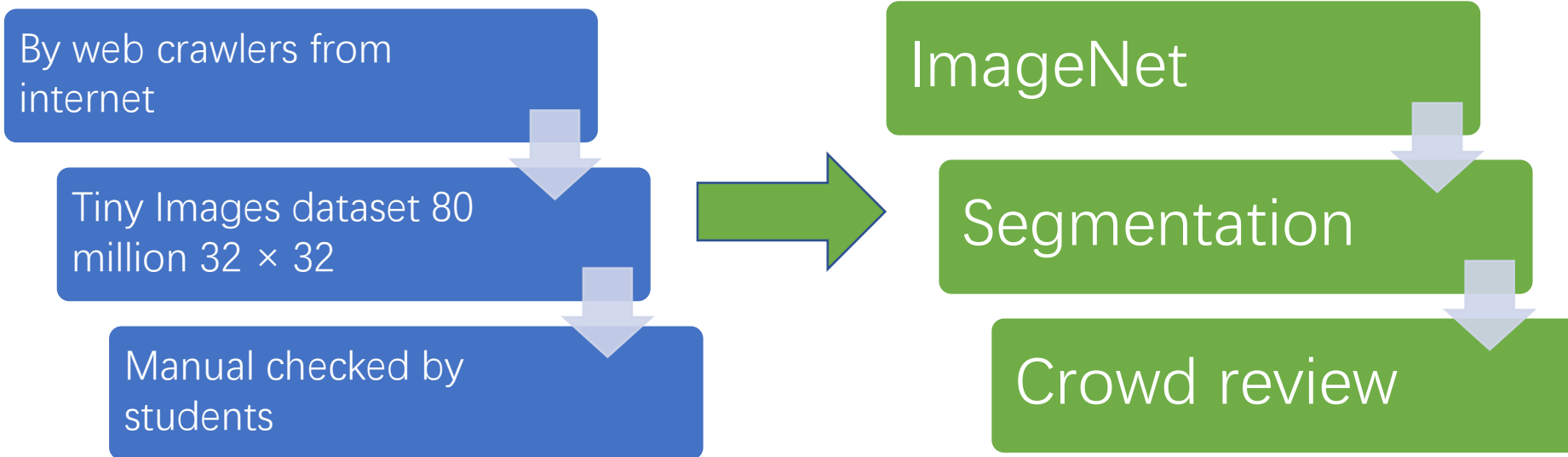


8208: a deer in front of a car



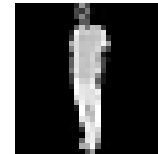
35829: a truck beside a car

CIFAR-10 augmentation



fashion-mnist

- Unlike CIFAR-10, Fashion-MNIST doesn't provide the process of labeling and the mutually exclusive definition.
- There is no definition of female t-shirt and female shirt, and the definition of male t-shirt and male shirt is not mutually exclusive.



- Image 2693 19545,34248,27110 in class trousers are a model.
- Image 37183,30191 in class pullover is a model.
- Image 7388, 14274 in class dress is a model.
- Image 21701 in class coat is a model.
- Image 65249 in class shirt is a model.

How to use MNIST-augmentation

- <https://github.com/zhangabner/ML>
 - hsf_4.tar.gz and hsf_6.tar.gz
 - groundtrueV1.csv
 - Groundtrue2.py
 - demo how to input new instances and replace MNIST instances
 - <https://www.kaggle.com/abnerzhang/mnistaugmentationv1>
- How to change instances (when found errors)
 - Change Groundtrue2.py to replace errors
- Using it as a new benchmark after it stable

- Thanks