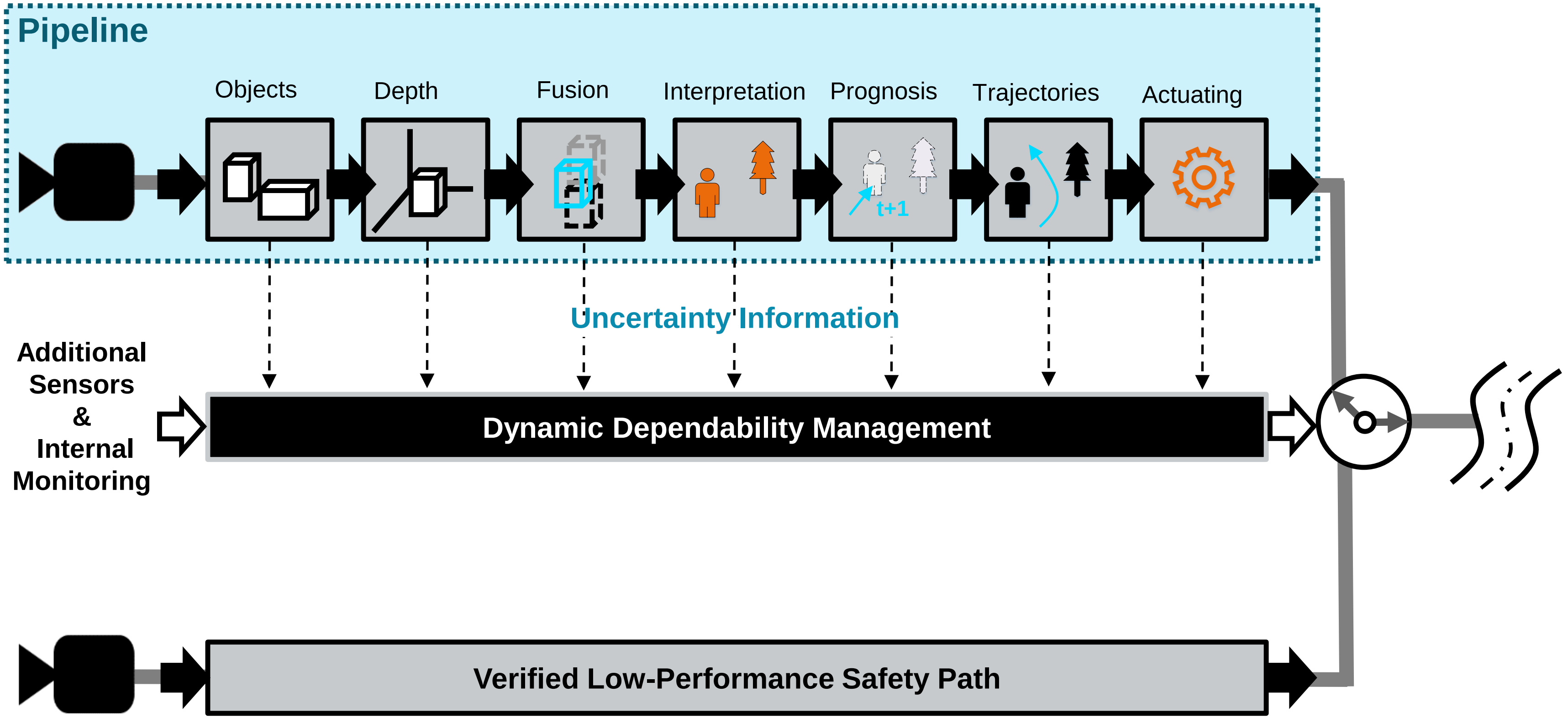
SAFEAI 2020

BENCHMARKING UNCERTAINTY ESTIMATION METHODS FOR DEEP LEARNING WITH SAFETY-RELATED METRICS

MAXIMILIAN HENNE, **ADRIAN SCHWAIGER**, KARSTEN ROSCHER, GEREON WEISS

DYNAMIC DEPENDABILITY MANAGEMENT



UNCERTAINTY QUANTIFICATION

Softmax:	Default network output
Monte-Carlo Dropout (MCDO):	Sample over same network with different dropout masks
Deep Ensembles (DE):	Sample over multiple, differently initialized networks
Evidential Deep Learning (EDL):	Learn parameters of a predictive Dirichlet distribution
Learned Confidence (LC):	Additional confidence head

EVALUATION METRICS

- Incorporate uncertainty in addition to the correctness of a prediction
  - **CT:** Certain True, **CF:** Certain False, **UT:** Uncertain True, **UF:** Uncertain False
  - Depends on a threshold for the certainty

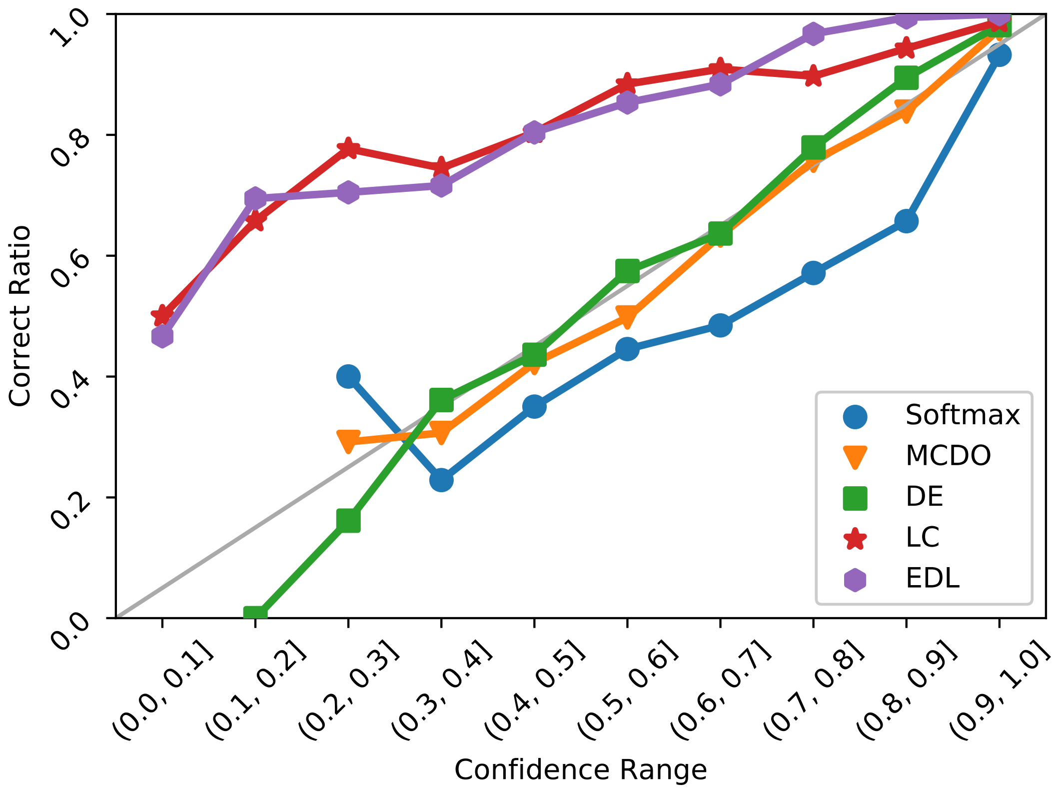
- **Remaining Error Rate**
  - $RER = \frac{CF}{N}$, Error ratio when discarding uncertain predictions

- **Remaining Accuracy Rate**
  - $RAR = \frac{CT}{N}$, Accuracy ratio when discarding uncertain predictions

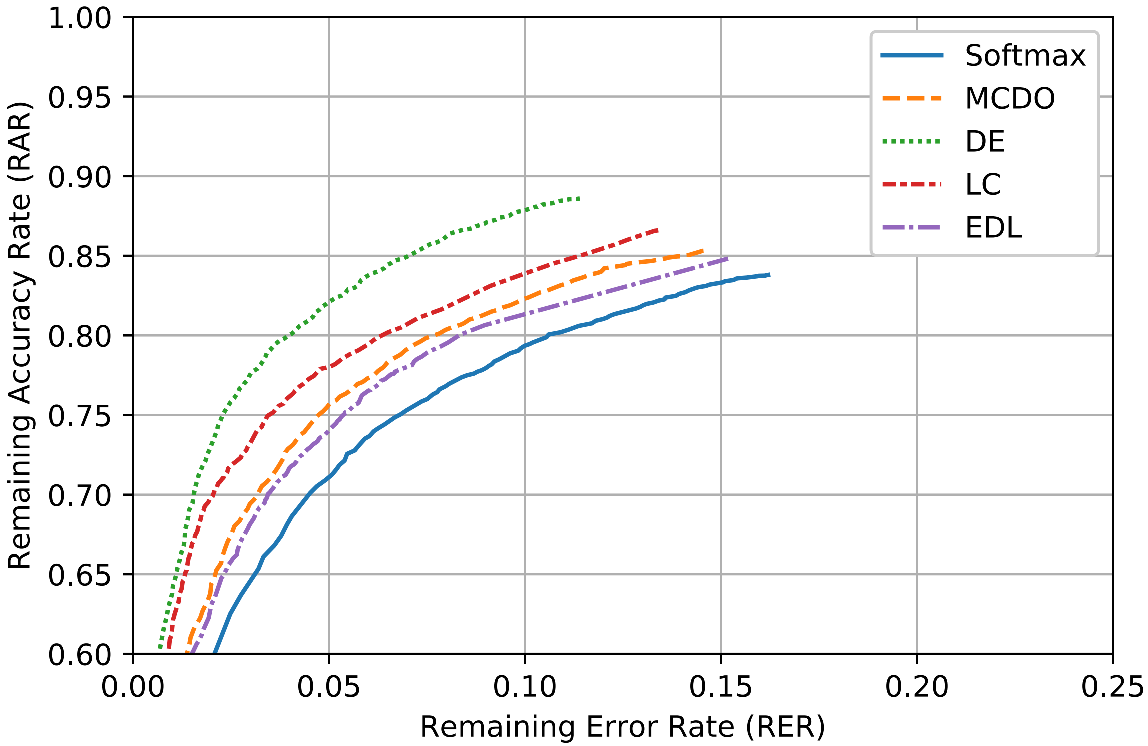
EXPERIMENTS SETUP

- **Task:** Image classification
- **Network Architectures**
  - VGG16 and a simple 6-Layer CNN (SimpleCNN)
  - Both perform very similar wrt. accuracy
  - SimpleCNN used for most of the evaluation, except when using learned confidences
- **Datasets**
  - **CIFAR-10**
  - MNIST
  - German Traffic Sign Recognition Benchmark (GTSRB)

CALIBRATION ON CIFAR-10



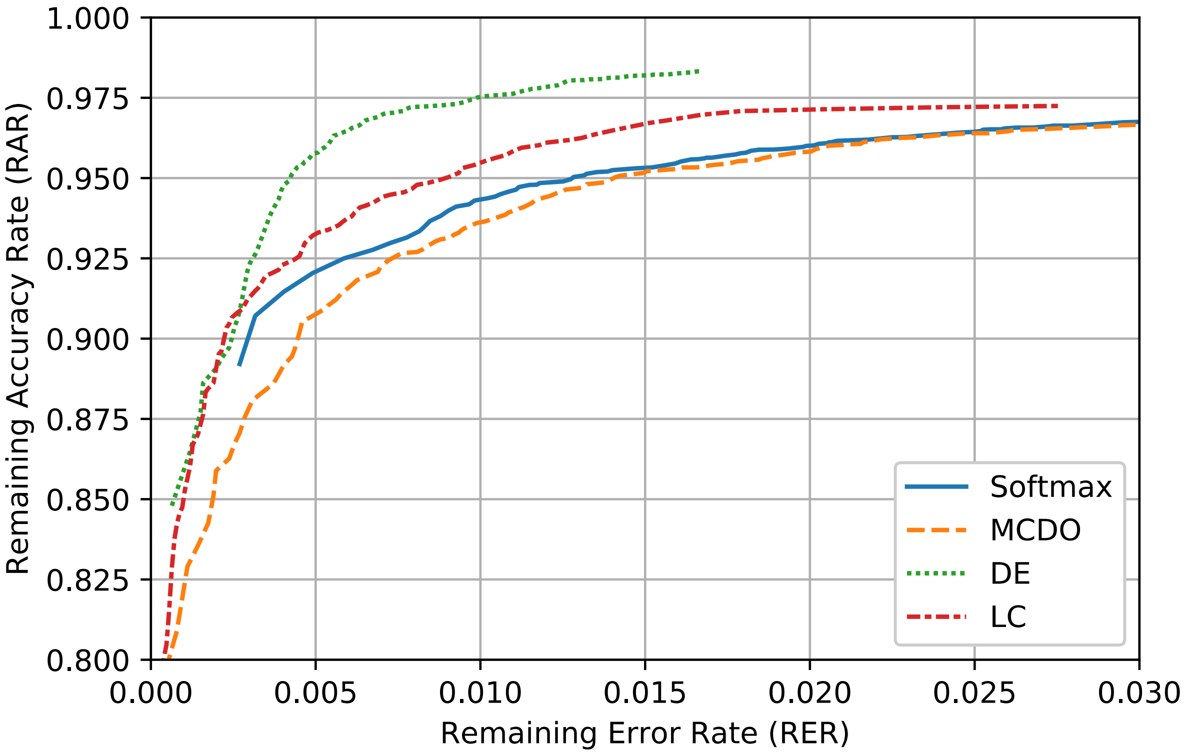
REMAINING ERROR RATE VS REMAINING ACCURACY RATE (CIFAR-10)



$$RER = \frac{CF}{N}$$

$$RAR = \frac{CT}{N}$$

REMAINING ERROR RATE VS REMAINING ACCURACY RATE (GTSRB)



$$RER = \frac{CF}{N}$$

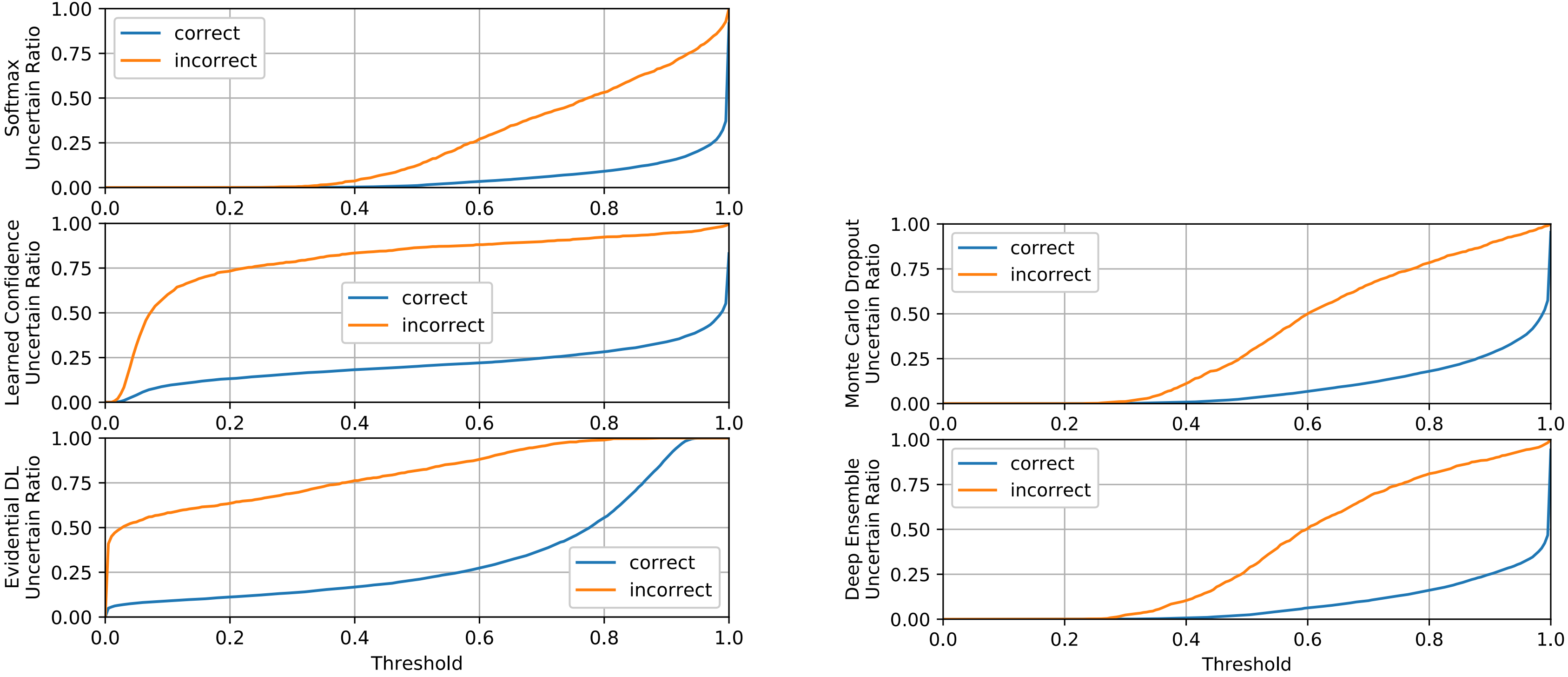
$$RAR = \frac{CT}{N}$$

SUMMARY AND OUTLOOK

- **Conclusions**
  - No single best method
  - Tested sampling-free approaches generally more cautious
  - No guarantees can be given for any of the considered uncertainty quantification methods
- **Future Work**
  - Combination of approaches
  - Embedding in a safety concept
  - More complicated datasets, out-of-distribution examples and other perception tasks

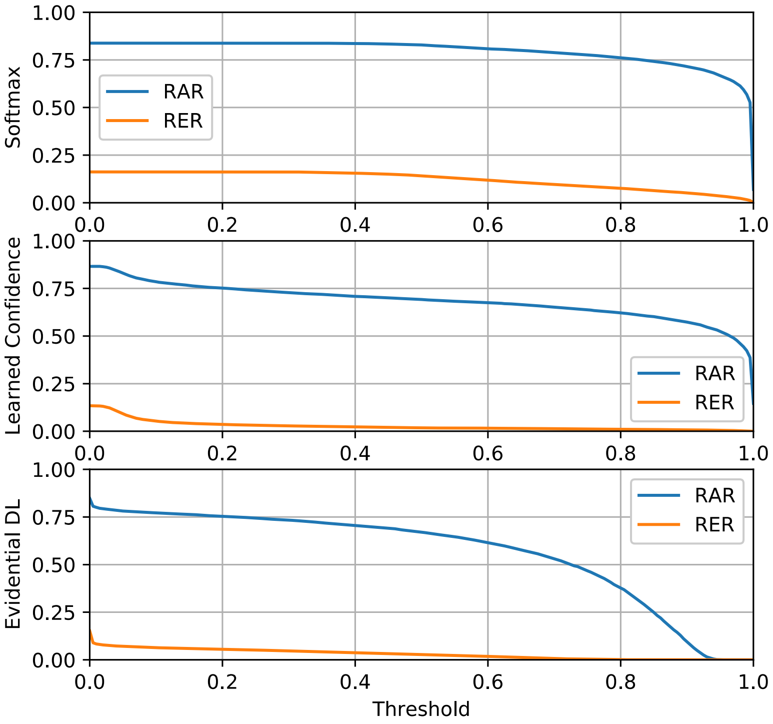
THANK YOU FOR YOUR ATTENTION!

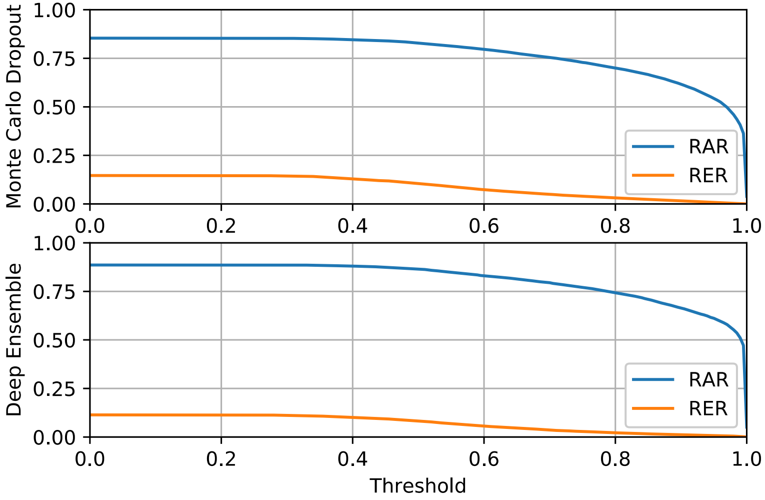
UNCERTAINTY RATIOS (CIFAR-10)



REMAINING ERROR RATE VS REMAINING ACCURACY RATE (CI

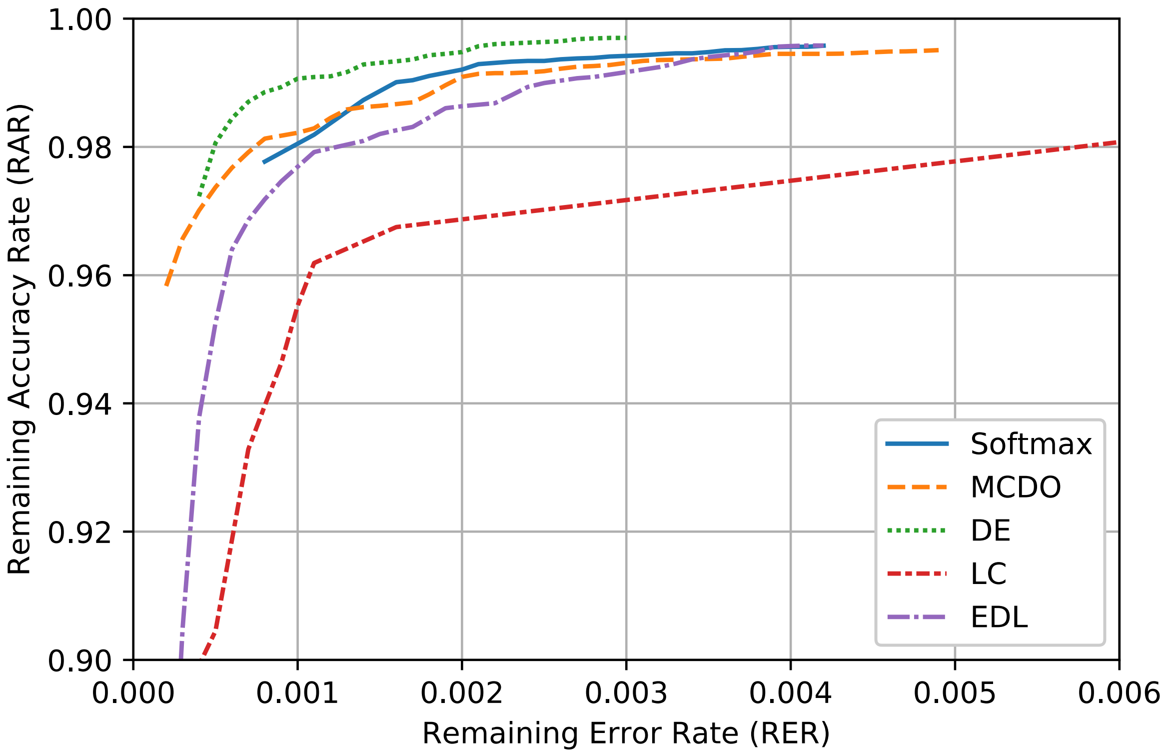
$$RER = \frac{CF}{N} \qquad RAR = \frac{CT}{N}$$





Fraunhofer IKS

REMAINING ERROR RATE VS REMAINING ACCURACY RATE (MNIST)



$$RER = \frac{CF}{N}$$

$$RAR = \frac{CT}{N}$$

SIMPLECNN ARCHITECTURE

Layer (type)	Output Shape	Param #
inputLayer	[(None, 32, 32, 3)]	0
Conv2D	(None, 32, 32, 32)	896
BatchNormalization	(None, 32, 32, 32)	128
Conv2D	(None, 32, 32, 32)	9248
BatchNormalization	(None, 32, 32, 32)	128
MaxPooling2D	(None, 16, 16, 32)	0
Dropout	(None, 16, 16, 32)	0
Conv2D	(None, 16, 16, 64)	18496
BatchNormalization	(None, 16, 16, 64)	256
Conv2D	(None, 16, 16, 64)	36928
BatchNormalization	(None, 16, 16, 64)	256
MaxPooling2D	(None, 8, 8, 64)	0
Dropout	(None, 8, 8, 64)	0
Conv2D	(None, 8, 8, 128)	73856
BatchNormalization	(None, 8, 8, 128)	512
Conv2D	(None, 8, 8, 128)	147584
BatchNormalization	(None, 8, 8, 128)	512
MaxPooling2D	(None, 4, 4, 128)	0
Dropout	(None, 4, 4, 128)	0
Flatten	(None, 2048)	0
Dense	(None, 10)	20490