AISAFETY 2021

IS UNCERTAINTY QUANTIFICATION IN DEEP LEARNING SUFFICIENT FOR OUT-OF-DISTRIBUTION DETECTION?

ADRIAN SCHWAIGER, POULAMI SINHAMAHAPATRA, JENS GANSLOSER, KARSTEN ROSCHER

MOTIVATION

- **Uncertainty estimation is important for a dependable perception of the environment**
- **Benchmarks to assess quality of uncertainty estimation exist almost exclusively for in-distribution data**
  - In-distribution (**ID**) data: Data that is conceptually similar to the training data
  - Out-of-distribution (**OoD**) data: Data that conceptually differs (strongly) from the training data
- **We asses the quality of SotA uncertainty estimation methods for detecting OoD samples on the task of image classification**
  - Extension of our work on benchmarking uncertainty estimation methods

ID VS OOD EXAMPLES

CIFAR-10 (ID)

Frog

Truck

GTSRB (ID)

Poor road surface

Priority road ahead

NWPU (ID)

Intersection

Residential area

CIFAR-100 (OoD)

Lizard

Pickup truck

BTSRB (OoD)

Speed bump

Priority road ends

NWPU OoD split (OoD)

Harbor

River

METHODS FOR UNCERTAINTY ESTIMATION

- **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 & DATASETS

- **Uncertainty & Correctness:** 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
  - Correct classification of OoD samples not possible ➔ Should always be labelled as uncertain

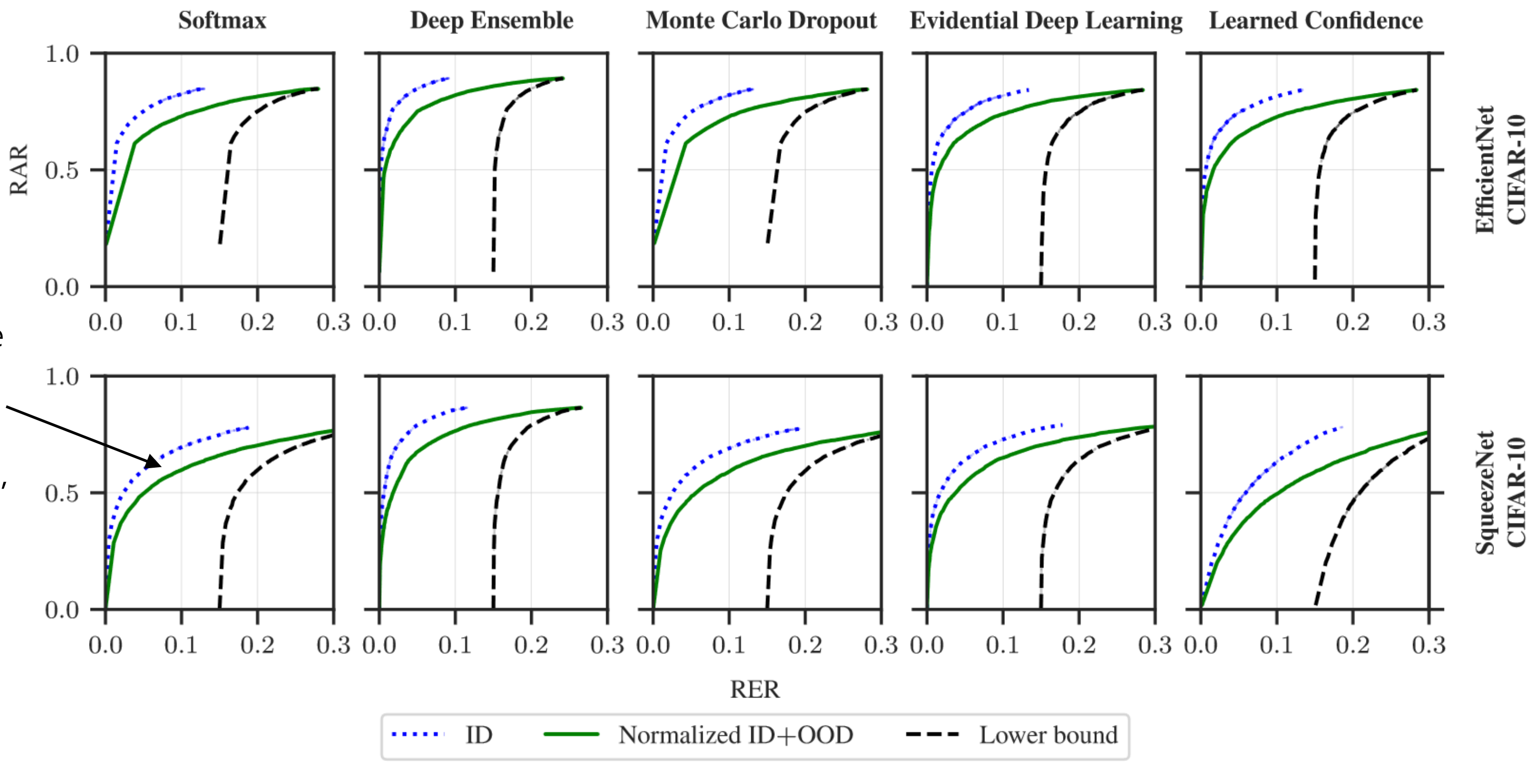
- **Remaining Error Rate**
  - $RER = \frac{CF}{N}$, Error ratio when discarding uncertain predictions

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

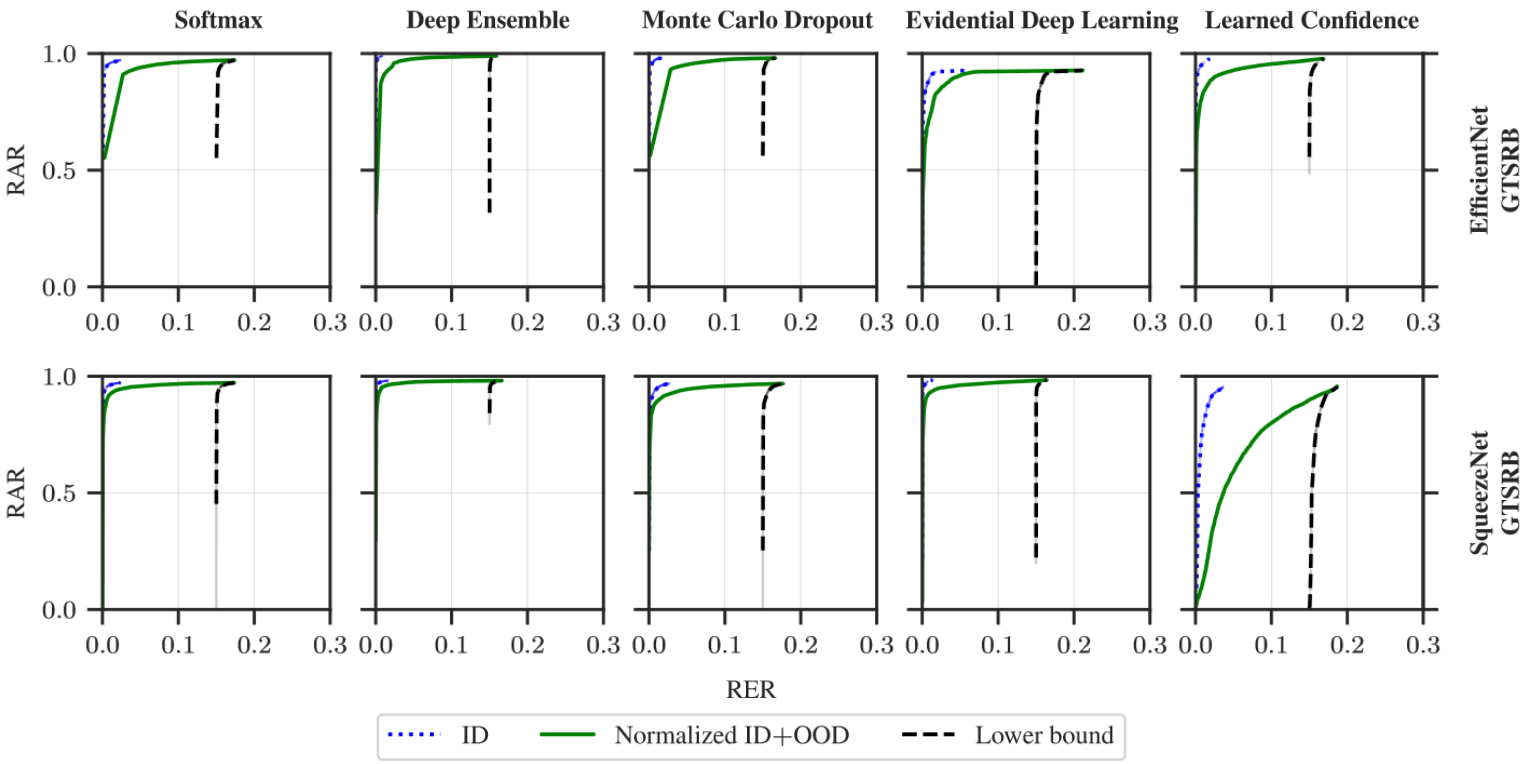
- **Datsets**
  - CIFAR-10 (OoD: CIFAR-100), GTSRB (OoD: BTSRB), NWPU-RESISC45 (OoD: Split from training data)
  - Two evaluation datasets
    - First one contains 100% ID data
    - Second one consists of 85% ID and 15% OoD data

RESULTS ON CIFAR-10/CIFAR-100

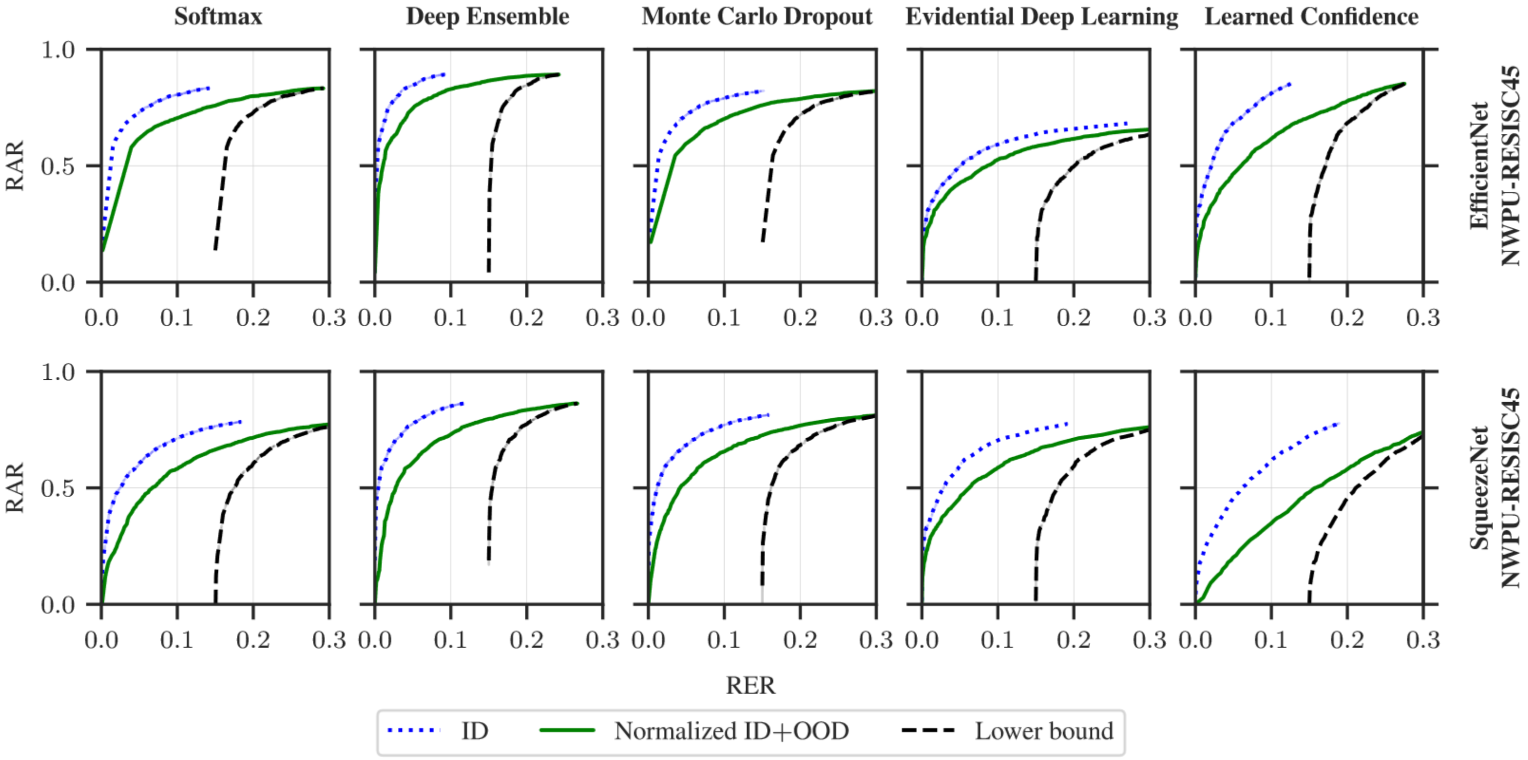
Corridor between blue and green curve should be minimized, ideally non-existent



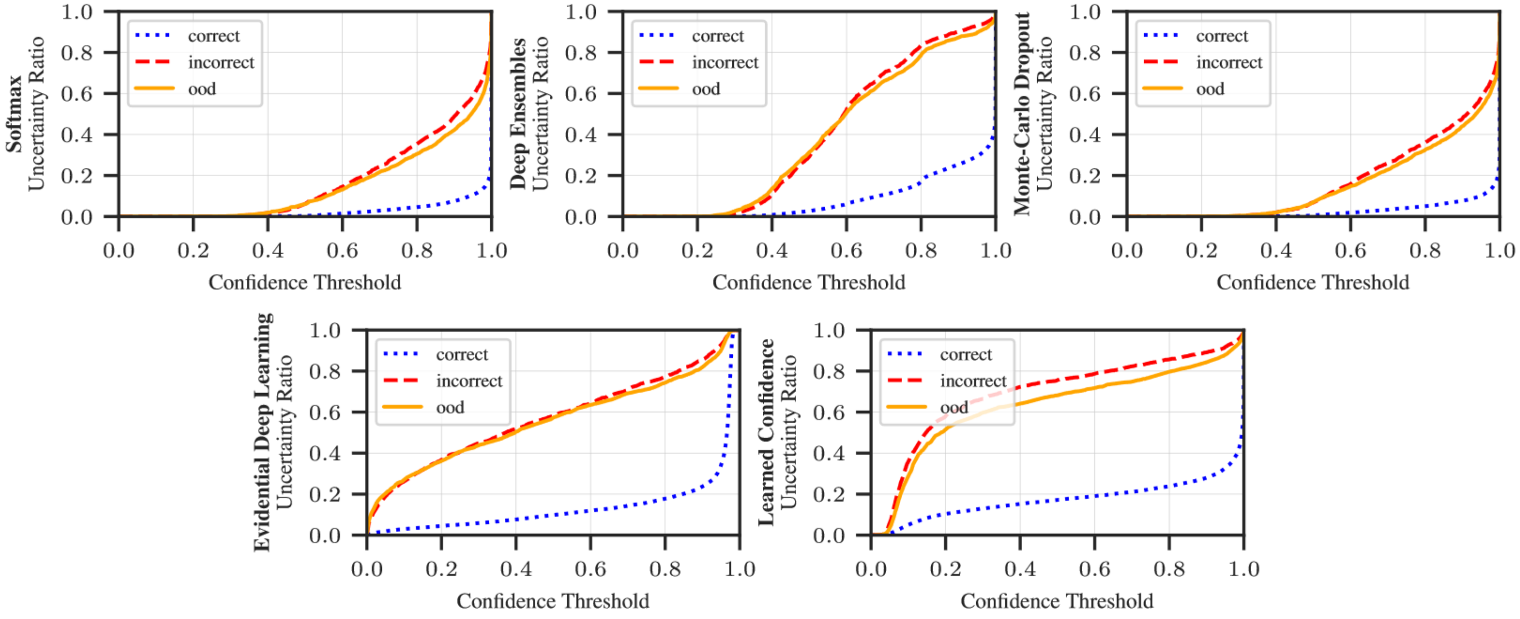
RESULTS ON GTSRB/BTSRB



RESULTS ON NWPU-RESISC45

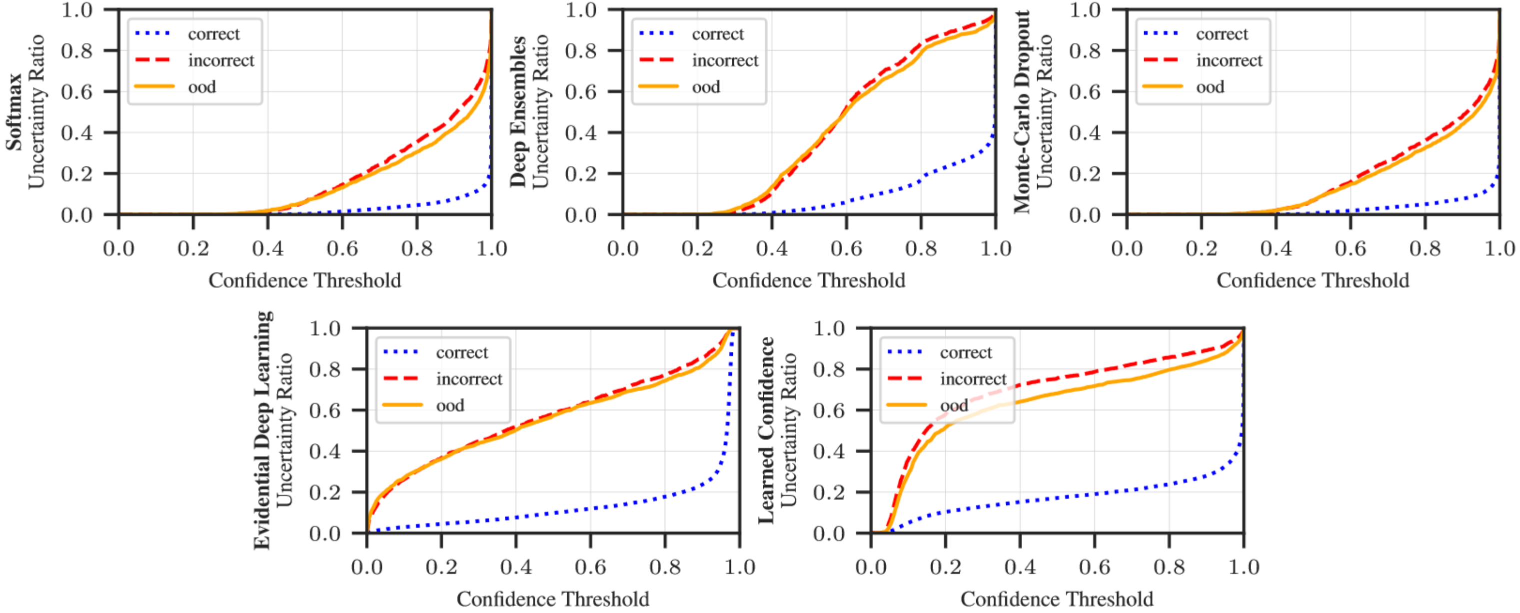


RATIOS OF INPUTS MARKED AS UNCERTAIN (EFFICIENTNET / CIFAR-10)



Interestingly, the uncertainty ratios for incorrect and OoD samples match very closely ➔ Raises the question about correlation of both classes and if better uncertainty estimation is able to reliable detect novel inputs

RATIOS OF INPUTS MARKED AS UNCERTAIN (EFFICIENTNET / CIFAR-10)



Interestingly, the uncertainty ratios for incorrect and OoD samples match very closely ➔ Raises the question about correlation of both classes and if better uncertainty estimation is able to reliable detect novel inputs

Findings did not always hold in later experiments!

CONCLUSIONS

- **SotA uncertainty estimation techniques for image classification are not sufficient enough for OoD detection**
  - In reality even more difficult OoD examples could occur
- **Hints that some OoD inputs are conceptually harder to identify**
  - Possibly additional insights into the difficulty of OoD detection in general when analyzed
- **Future directions**
  - Comparison with SotA OoD detectors
  - Transfer to other perception tasks