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

MAXIMILIAN HENNE, ADRIAN SCHWAIGER, KARSTEN ROSCHER, GEREON WEISS
DYNAMIC DEPENDABILITY MANAGEMENT

Pipeline

Objects → Depth → Fusion → Interpretation → Prognosis → Trajectories → Actuating

Additional Sensors & Internal Monitoring → Uncertainty Information → Dynamic Dependability Management

Verified Low-Performance Safety Path
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

![Graph showing the correct ratio across different confidence ranges for various methods.]

- Softmax
- MCDO
- DE
- LC
- EDL
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!
REMAINING ERROR RATE VS REMAINING ACCURACY RATE

(CIFAR-10)

\[
RER = \frac{CF}{N} \quad RAR = \frac{CT}{N}
\]
REMAINING ERROR RATE VS REMAINING ACCURACY RATE (MNIST)

\[ RER = \frac{CF}{N} \]

\[ RAR = \frac{CT}{N} \]
## SIMPLECNN ARCHITECTURE

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