

# Uncertainty-sensitive Activity Recognition: a Reliability Benchmark and the CARING Models

*International Conference on Pattern Recognition (ICPR)*

Computer Vision for Human-Computer Interaction Lab (cv:hci)  
Institute for Anthropomatics and Robotics (IAR)

**Alina Roitberg**



**Monica Haurilet**



**Manuel Martinez**

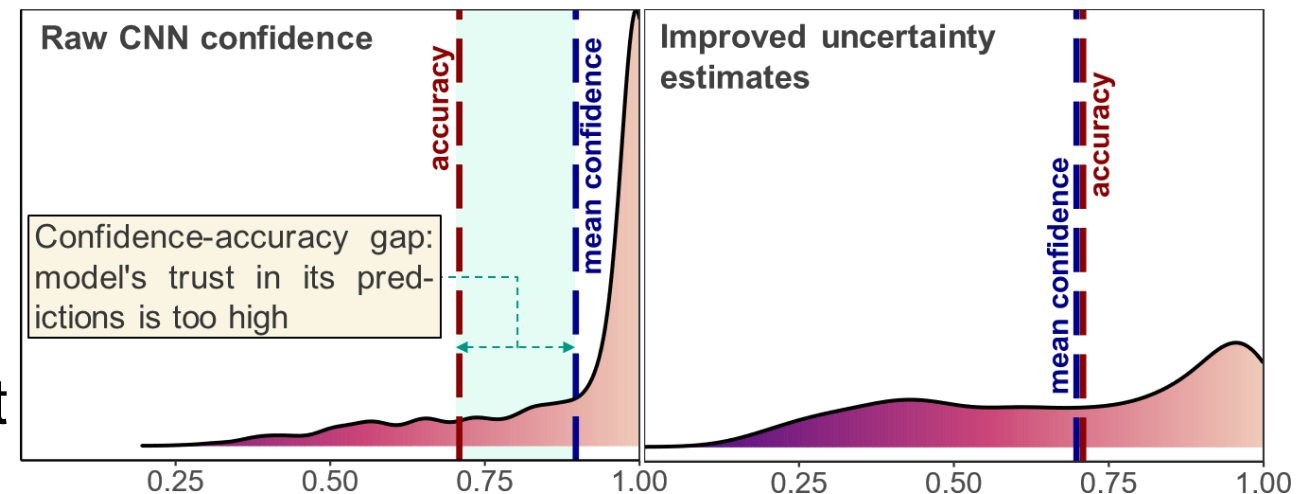


**Rainer Stiefelhagen**



# Highlights

- Goal: Action Recognition (AR) with *reliable confidence estimates*
- First study of *reliability of confidence values* in modern AR architectures
- Improving the reliability of model confidence in action recognition:
  - Enhancing the AR models with temperature scaling (*Guo et al.*)
  - Calibrated Action Recognition with Input Guidance (CARING): a new learning technique to scale the logits depending on the video input

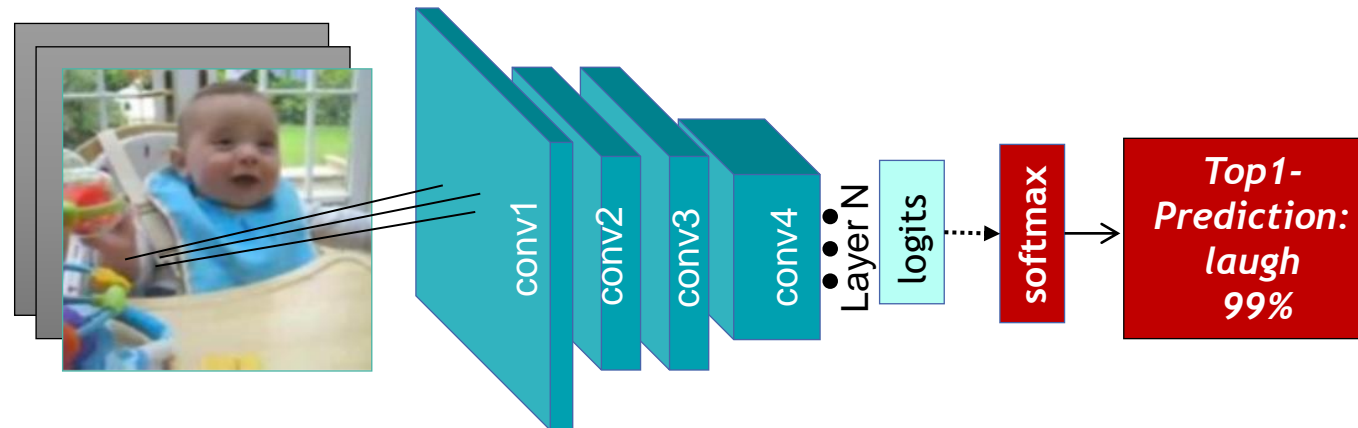


# A note on Softmax

- Normalizes the output of the last fully-connected layer
  - The outputs sum up to one (often inaccurately called „probabilities“)
  - Input to Cross-Entropy loss, usually to maximize the Top-1 accuracy
- Excellent Top-1 action recognition results, but does it reflect the true confidence?

$$\sigma(\mathbf{z})_i = \frac{e^{z_i}}{\sum_{j=1}^K e^{z_j}}$$

$$H(p, q) = - \sum_{x \in \mathcal{X}} p(x) \log q(x)$$

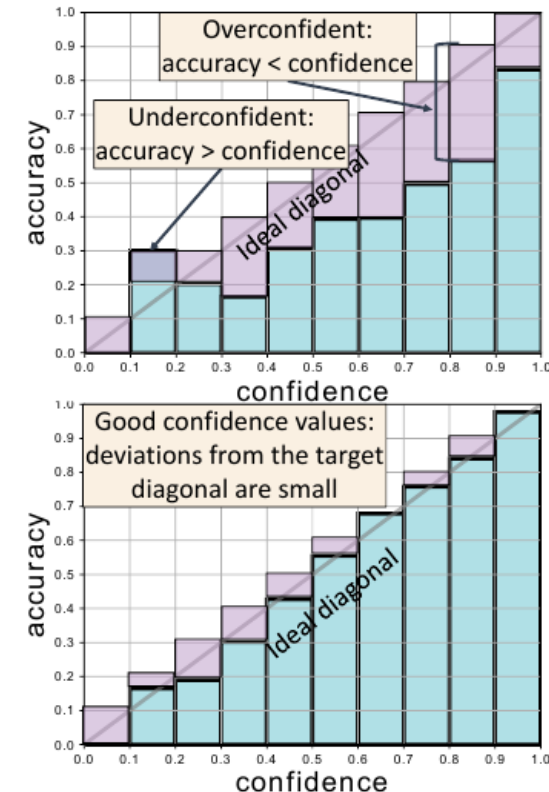
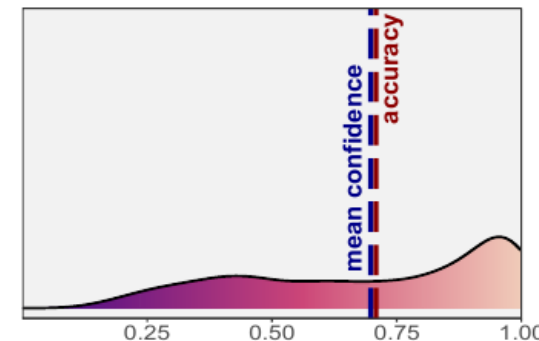
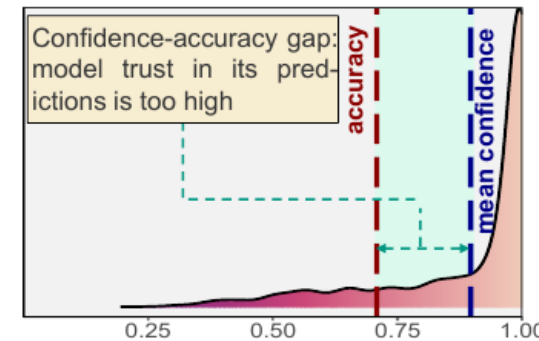


# Expected Calibration Error

- We integrate *reliability of model confidence* in the activity recognition evaluation
- We present a learning-based approach for improving it

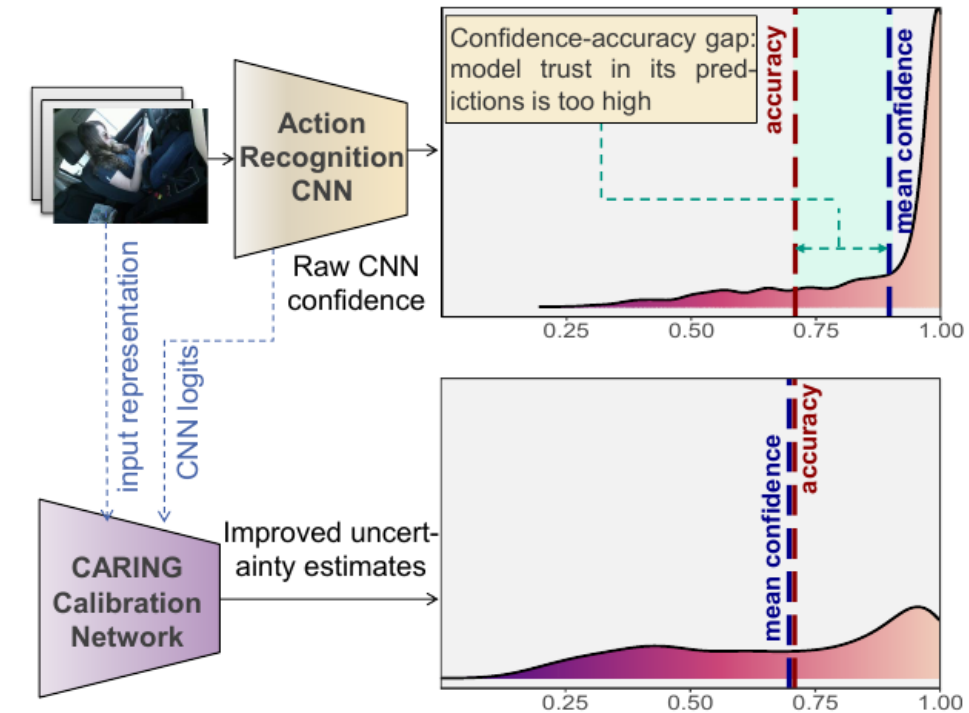
- What is a good confidence value?
  - Metrics: Expected Calibration Error (ECE)
  - Visualized with reliability diagrams

$$ECE = \sum_{i=1}^K \frac{N_{bin_i}}{N_{total}} |acc(bin_i) - conf(bin_i)|$$



# CARING: Calibrated Action Recognition with Input Guidance

- We present a learning-based approach for reliable confidence estimates
- Calibrated Action Recognition with Input Guidance (CARING)
- Coarse idea: additional network learning to scale the logits depending on the input
- Compared to native action recognition architectures and *temperature scaling*, where a *single* scaling factor is learned (Guo et al., 2017)

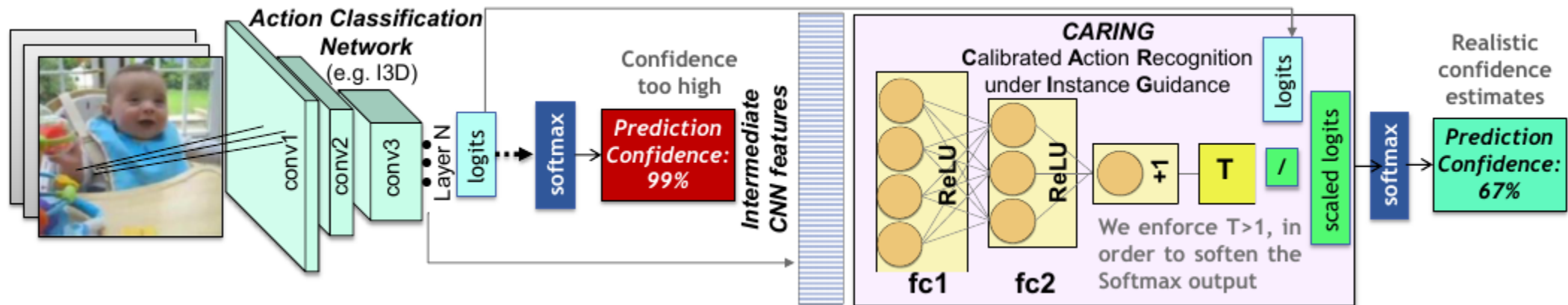


# Learning Proper Confidence Estimates

- The CARING network produces different temperature values  $T(z)$  depending on the input  $z$
- The logits are scaled by  $T(z)$ , the final confidence becomes:

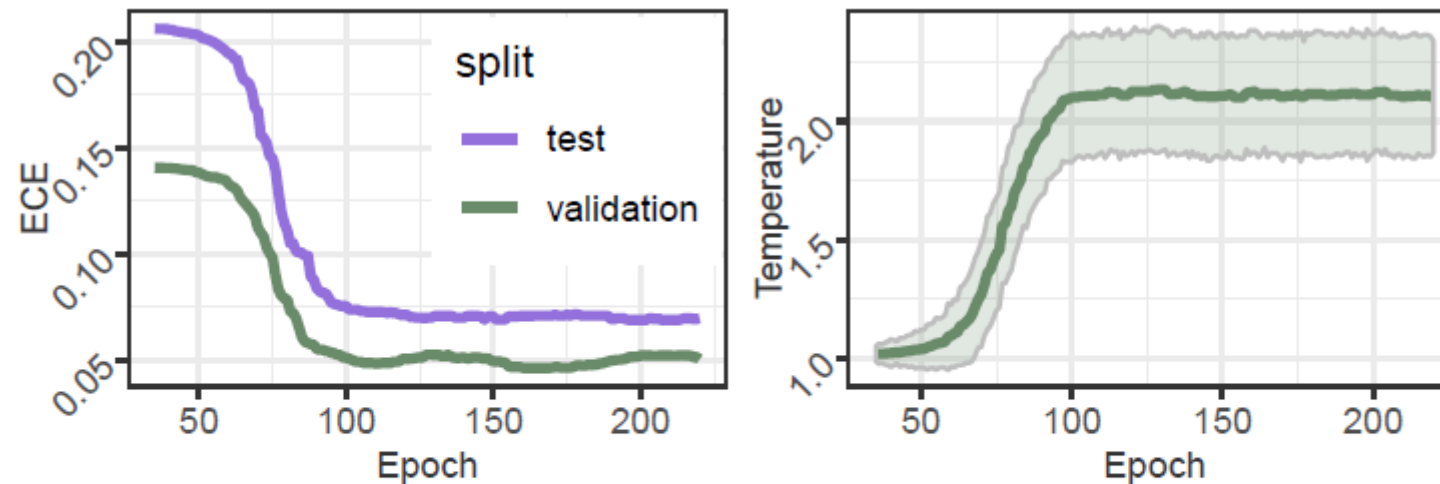
$$conf(a_{pred}) = \max_{a \in \mathcal{A}} \frac{\exp(\frac{y_a}{T(\vec{z})})}{\sum_{\hat{a} \in \mathcal{A}} \exp(\frac{y_{\hat{a}}}{T(\vec{z})})}$$

- Trained on a held-out validation set



# Learning Proper Confidence Estimates

## ■ CARING model evolution during training

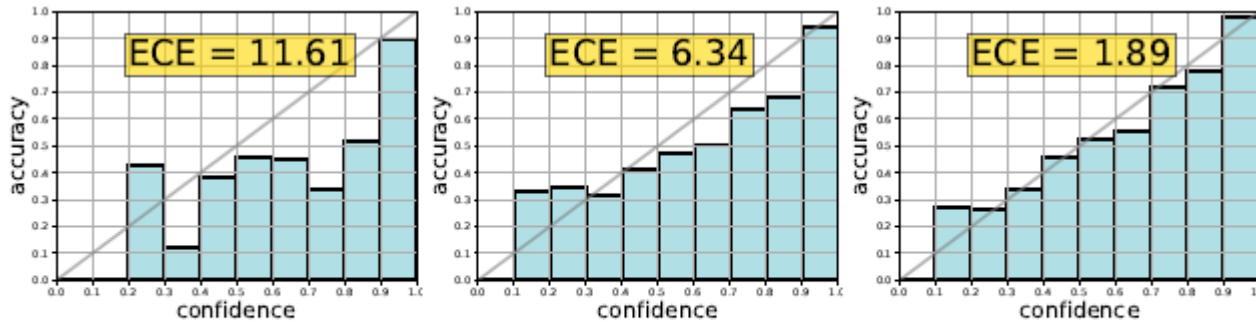


(a) Expected Calibration Error improvement during the training procedure for validation and test data. (b) Average temperature and its standard deviation estimated by our model during training.



# Confidence Estimates Study

- Original models - poor confidence estimates
- Clearly improved by Temperature Scaling
- Our CARING model leads to the best results
- Still: a lot of room for improvement for underrepresented classes!



(a) I3D (original), all action classes      (b) I3D + temp. scaling, all action classes      (c) CARING-I3D, all action classes

| Model                                 | ECE          |              | NLL         |             |
|---------------------------------------|--------------|--------------|-------------|-------------|
|                                       | validation   | test         | validation  | test        |
| <b>Drive&amp;Act - Common Classes</b> |              |              |             |             |
| P3D [61] \$                           | 16.9         | 19.39        | 1.63        | 1.85        |
| I3D [6] \$                            | 10.22        | 13.38        | 0.90        | 1.27        |
| P3D + Temperature Scaling [20] U      | 5.65         | 5.7          | 1.28        | 1.48        |
| I3D + Temperature Scaling [20] U      | 5.31         | 6.99         | 0.57        | 0.83        |
| CARING - P3D (ours) U                 | 4.81         | <b>4.27</b>  | 1.19        | 1.42        |
| CARING - I3D (ours) U                 | <b>2.57</b>  | 5.26         | <b>0.50</b> | <b>0.78</b> |
| <b>Drive&amp;Act - Rare Classes</b>   |              |              |             |             |
| P3D [61] \$                           | 31.49        | 37.25        | 3.43        | 4.68        |
| I3D [6] \$                            | 31.48        | 43.32        | 3.41        | 4.54        |
| P3D + Temperature Scaling [20] U      | 17.83        | 21.09        | 2.26        | 2.99        |
| I3D + Temperature Scaling [20] U      | 24.97        | 32.38        | 1.96        | 2.62        |
| CARING - P3D (ours) U                 | <b>13.73</b> | <b>19.92</b> | 2.12        | 2.93        |
| CARING - I3D (ours) U                 | 18.34        | 23.6         | <b>1.55</b> | <b>2.17</b> |
| <b>Drive&amp;Act - All Classes</b>    |              |              |             |             |
| P3D [61] \$                           | 17.89        | 21.09        | 1.77        | 2.12        |
| I3D [6] \$                            | 11.72        | 15.97        | 1.10        | 1.56        |
| P3D + Temperature Scaling [20] U      | 5.89         | 6.41         | 1.35        | 1.63        |
| I3D + Temperature Scaling [20] U      | 6.59         | 8.55         | 0.68        | 0.99        |
| CARING - P3D (ours) U                 | 4.58         | <b>5.26</b>  | 1.26        | 1.57        |
| CARING - I3D (ours) U                 | <b>3.03</b>  | 6.02         | <b>0.58</b> | <b>0.9</b>  |
| <b>HMDB-51</b>                        |              |              |             |             |
| I3D [6] \$                            | 10.29        | 20.11        | 0.98        | 1.97        |
| I3D + Temperature Scaling [20] U      | 4.00         | 7.75         | <b>0.81</b> | 1.57        |
| CARING - I3D (ours) U                 | <b>3.38</b>  | <b>5.98</b>  | <b>0.81</b> | <b>1.54</b> |

\$ Standard activity recognition models

U Uncertainty-aware models



*Thank you for your attention!*

Check out our paper😊

**“Uncertainty-sensitive Activity Recognition: a Reliability Benchmark and the CARING Models”**

*Alina Roitberg, Monica Haurilet, Manuel Martinez, Rainer Stiefelhagen*