A Perspective on Deep Vision Performance with Standard Image and Video Codecs

Christoph Reich[1,2,3,5], **Oliver Hahn**[1], **Daniel Cremers**[2,5], **Stefan Roth**[1,4], **and Biplob Debnath**[3]

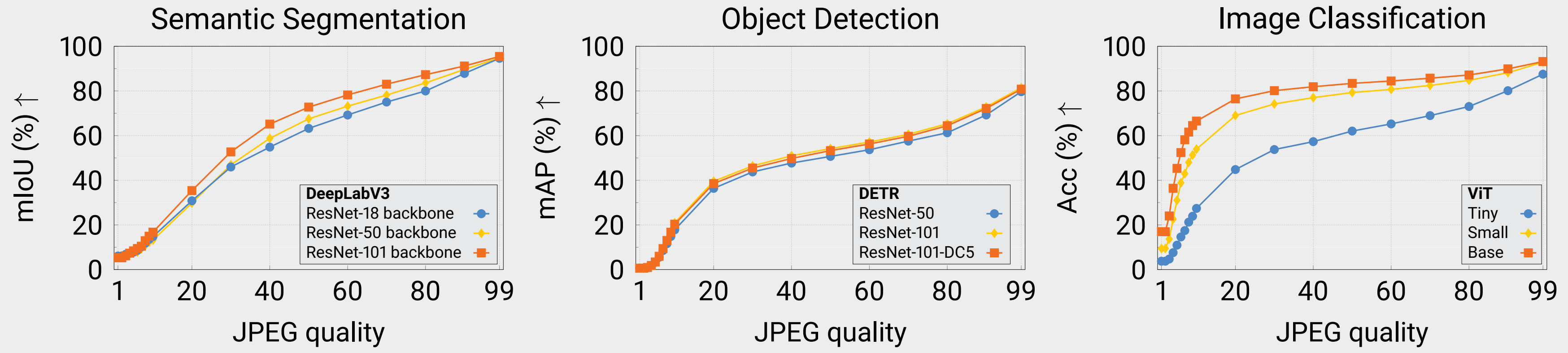
[1] TU Darmstadt, [2] TU Munich, [3] NEC Laboratories America, Inc., [4] Hessian Center for AI (hessian.AI), [5] Munich Center for Machine Learning (MCML)

Introduction

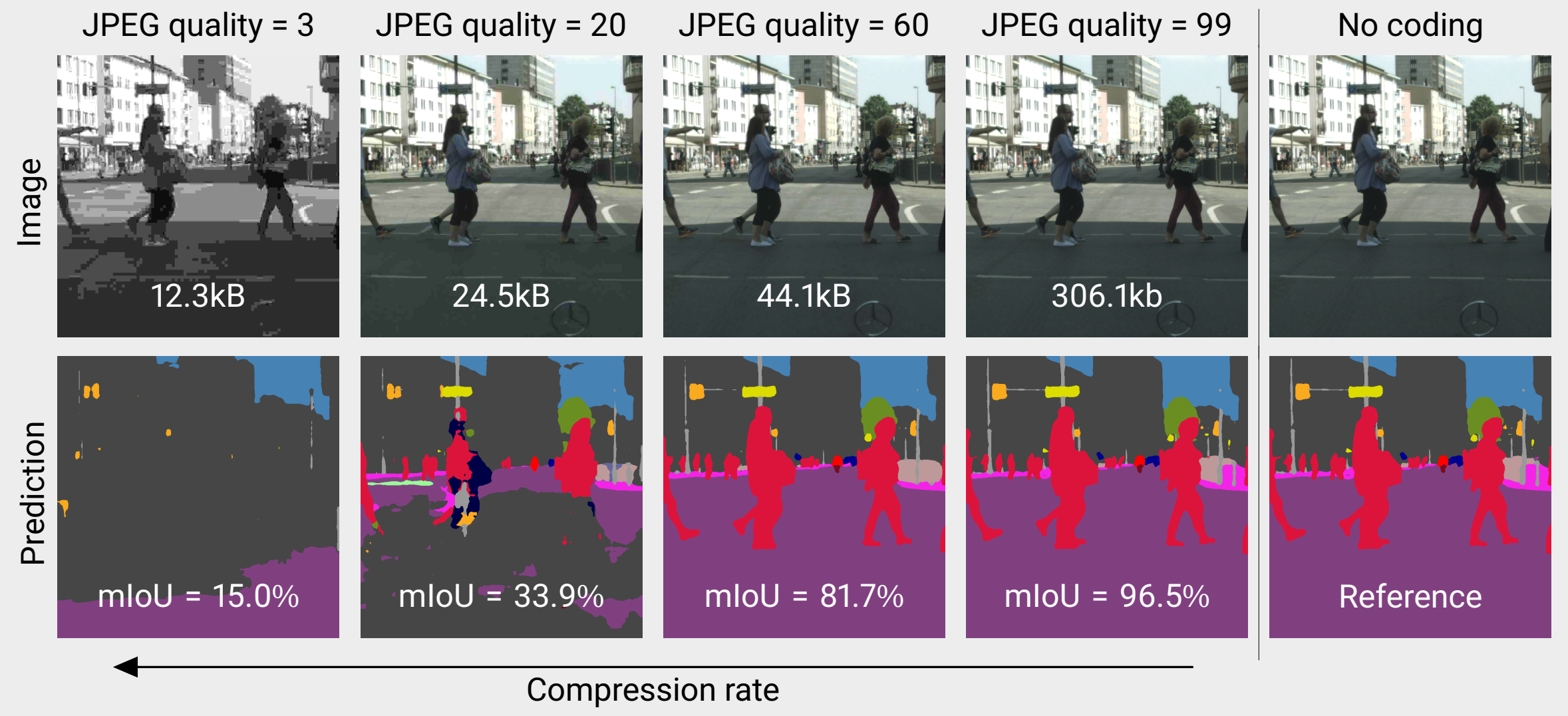
- Standard image & video codecs are the *de facto* standard in real-world image & video processing pipelines
- The use of standard codecs facilitates *low costs* and *interoperability*
- Standard codecs have been naïvely incorporated into deep vision pipelines
- Rate-distortion has been studied through the lens of Shannon's rate-distortion theory and via perceptual quality [1, 2]
- **We analyze the implications of standard codecs on the performance of deep vision models across downstream tasks**

Performance of Deep Vision Models on JPEG-coded Images

- We evaluated the accuracy of deep vision models on JPEG-coded images (w.r.t. the prediction on the original images)
- We evaluated 20 different vision models and three vision tasks (from classification to dense prediction)



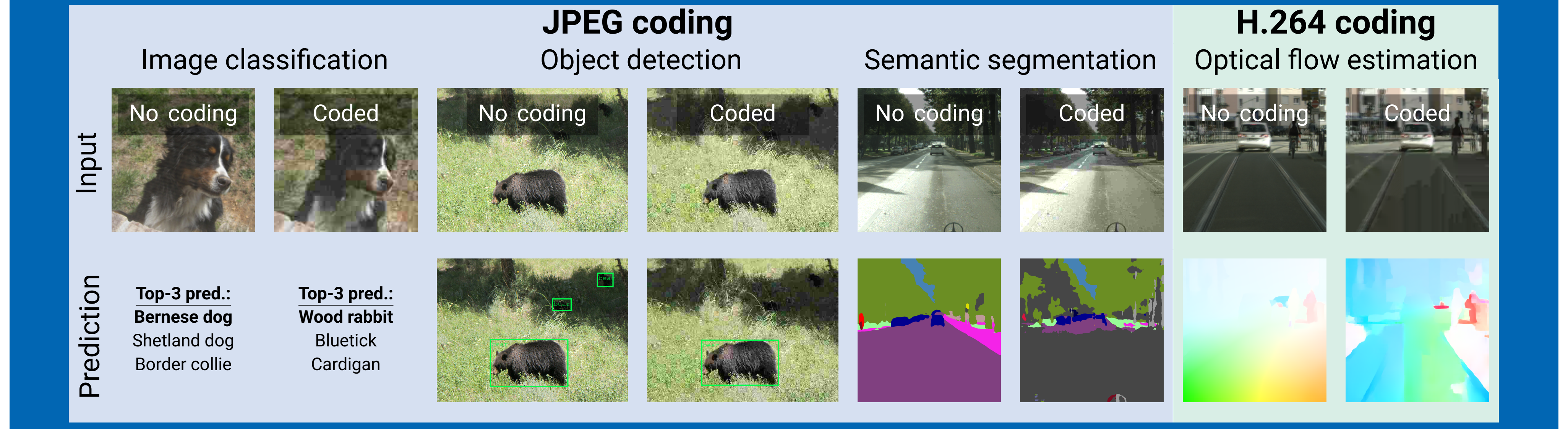
- **All deep vision models tested significantly suffer from JPEG coding**
- Dense prediction tasks suffer more from JPEG coding than image classification
- Larger capacity models offer more robustness against JPEG coding



- Weak compression rates can lead to wrong predictions – strong coding leads to a collapse in segmentation accuracy

tl;dr:

We examine the implications of employing standardized codecs within deep vision pipelines.



Using standard coding significantly deteriorates the accuracy across vision tasks and models. For dense prediction tasks, moderate coding already leads to a significant loss of performance.

References

[1] Y. Blau *et al.*, "Rethinking lossy compression: The rate-distortion-perception tradeoff," in *ICML*, 2019.

[2] C. E. Shannon, "Communication in the Presence of Noise," *Proceedings of the IRE*, 1949.

[3] C. Reich *et al.*, "Deep video codec control for vision models," in *CVPRW*, 2024.

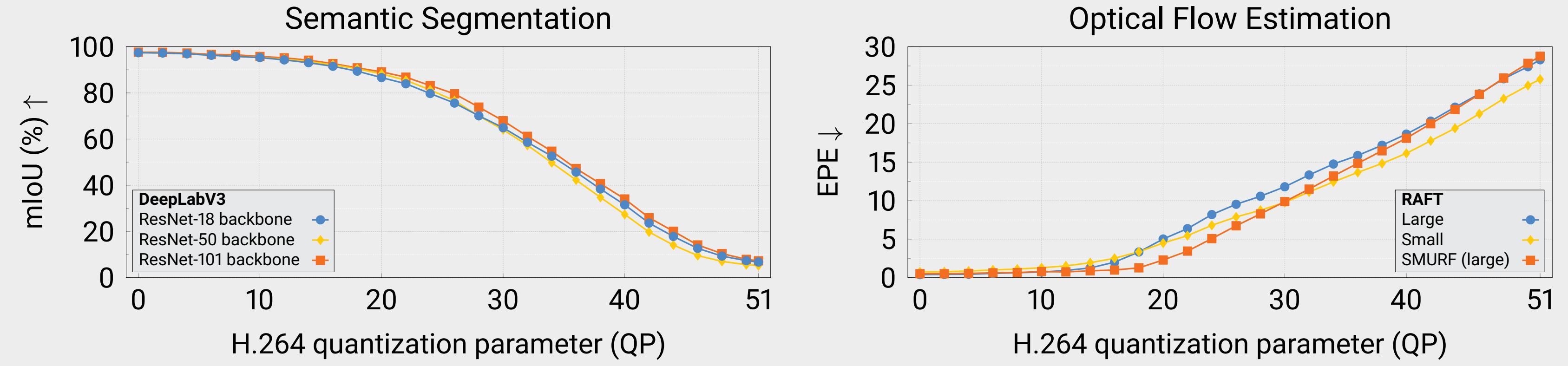
[4] Y.-H. Chen *et al.*, "TransTIC: Transferring transformer-based image compression from human perception to machine perception," in *ICCV*, 2023.

[5] A. Otani *et al.*, "Performance evaluation of action recognition models on low quality videos," *IEEE Access*, 2022.

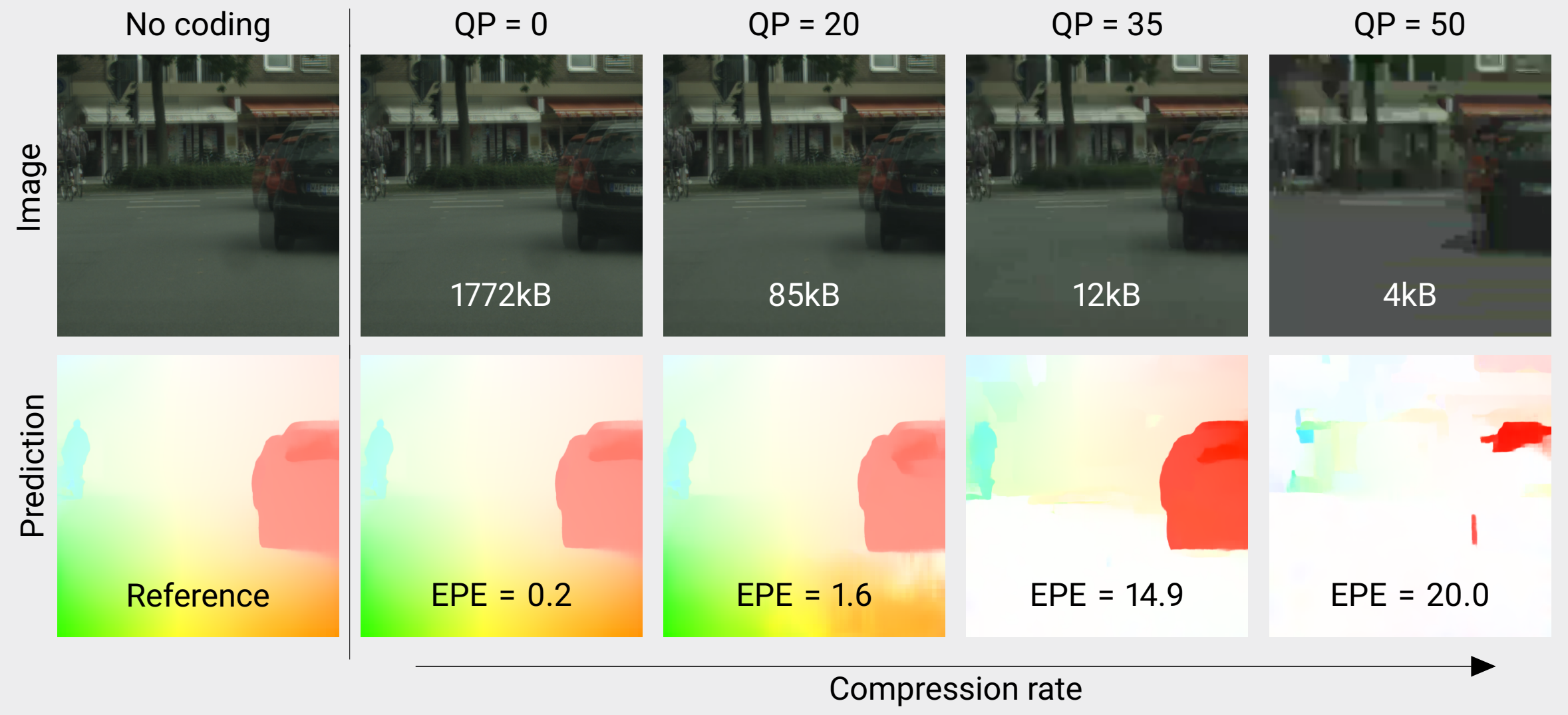
[6] J. Park and J. Johnson, "RGB no more: Minimally-decoded jpeg vision transformers," in *CVPR*, 2023.

Performance of Deep Vision Models on H.264-coded Videos

- We evaluated the accuracy of deep vision models on H.264-coded video clips (w.r.t. the prediction on the original clip)
- We evaluated 6 different vision models and two vision tasks



- **All deep vision models tested significantly suffer from H.264 coding**
- Surprisingly, larger models do not introduce more robustness (different from JPEG coding)



- Strong H.264 coding leads to a complete breakdown in optical flow estimation

Conclusion & Discussion

- **All 23 models tested significantly suffered from standard coding**
- For strong compression rates downstream deep vision performance can completely break down

How to overcome the deterioration in downstream deep vision performance?

- **Optimizing standard codecs** (see our other poster [3])
- Deep codecs for deep vision models (*e.g.*, [4])
- Data augmentation (*e.g.*, [5] & [3])
- Adapting deep vision models for coded data (*e.g.*, [6])

Paper: `https://arxiv.org/abs/2404.12330`
Correspondence: `c.reich@tum.de`

Acknowledgement. This project was supported by the European Research Council (ERC) Advanced Grant SIMULACRON. This project has also received funding from the ERC under the European Union's Horizon 2020 research and innovation programme (grant agreement No. 866008). In part, the project has been supported further by the State of Hesse through the cluster project "The Third Wave of Artificial Intelligence (3AI)". Moreover, we thank NEC Laboratories America, Inc., for the support. Finally, we thank Deep Patel and Srimat Chakradhar for the insightful discussions.

Funded by the European Union
European Research Council
Established by the European Commission

CVPR
JUNE 17-21, 2024
SEATTLE, WA