

DNN-based Photography Rule Prediction using Photo Tags

Steve Göring, Rasmus Merten, Alexander Raake

Audiovisual Technology Group; Technische Universität Ilmenau, Germany
Email: [steve.goering, rasmus-leo-lukas.merten, alexander.raake]@tu-ilmenau.de

Code & Data: http://git.avt-imt.de/sophoappeal_rule_prediction_extension



Figure 1: Example images (CC0 licensed) for all tags ("framing_framed", "ruleofthirds", "patterns", "symmetric", "leadinglines", "texture", "simplicity", and "repetition").

Introduction

- ▶ many images are uploaded to Instagram or Flickr
- ▶ not all images of high appeal → "rules of thumb" for photography
 - rule of thirds, simplicity, leading lines, framing, symmetry, and patterns [5]
- ▶ related work: prediction of individual rules [8, 7, 4]
- ▶ our target: **prediction as multi-class**, with dataset based on 8 Flickr tags
- ▶ training and evaluation of several DNNs for prediction
- ▶ rules can be used to improve appeal, thus quality, compare Fig. 2

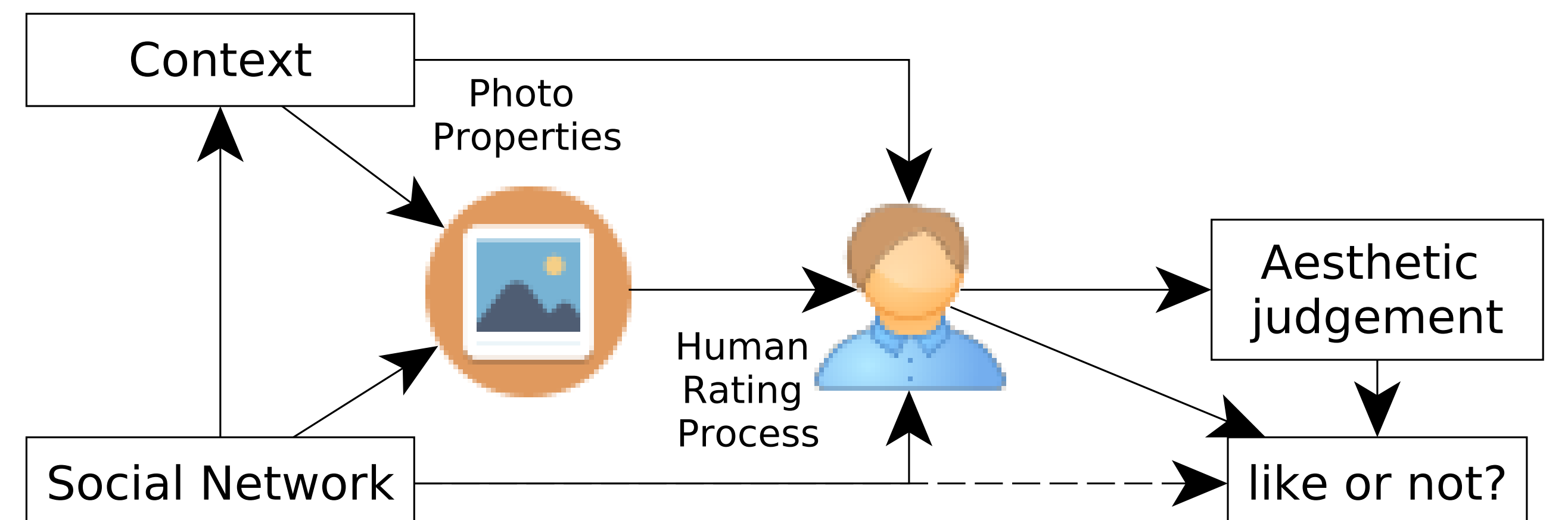


Figure 2: How humans rate aesthetic and decide liking [3], based on Leder et al.'s model [6].

Dataset

- ▶ Flickr tags, matching photo rules:
 - "leadinglines", "ruleofthirds", "simplicity", "symmetric", "texture", "framing", "framed", "patterns", and "repetitions"
- ▶ downloaded highest resolutions, removing duplicates (phash [10])
- ▶ merged "framed", "framing" to "framing_framed"
- ▶ distribution of images per class, see Fig. 3
 - min 1698 for "patterns",
 - max 3226 for "leadinglines"
- ▶ example images, see Fig 1
- ▶ total 20.449 images; URLs of all images shared

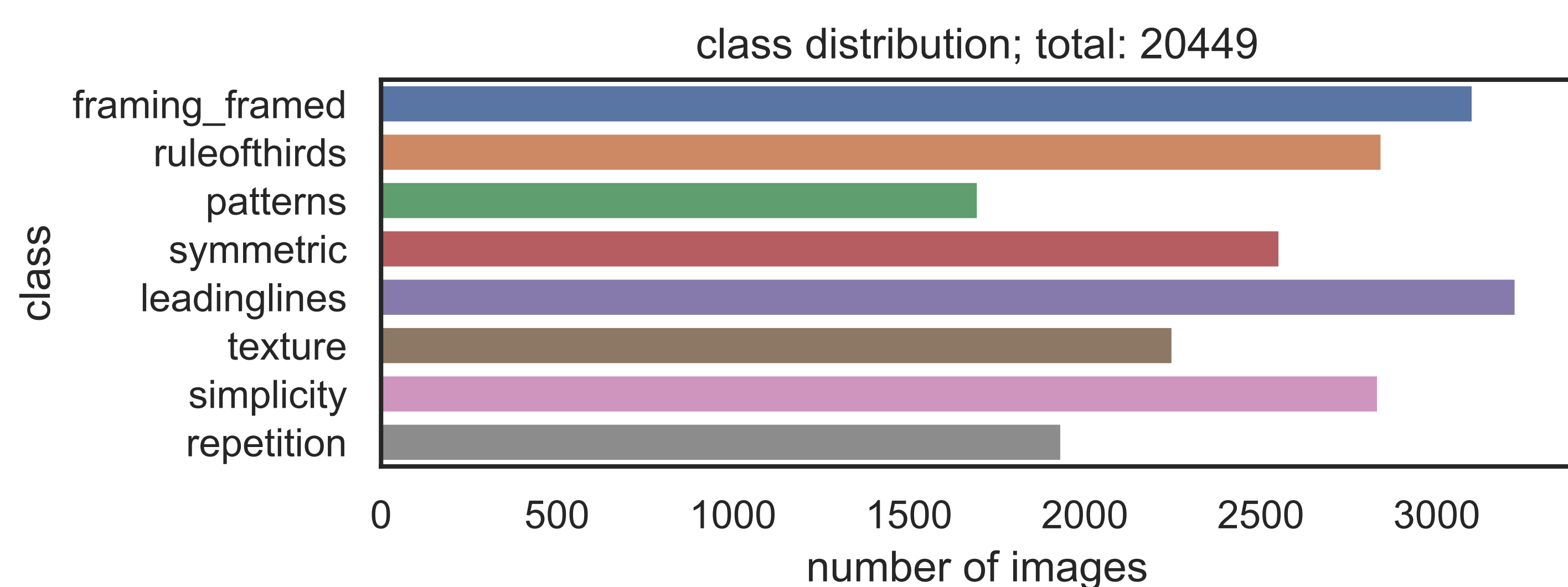


Figure 3: Class distribution for each of the tags.

Evaluation

- ▶ transfer-learning approach [9], similar to [4, 2]
- ▶ input layer 224x224 → image pre-processing → baseline DNN
- ▶ baseline DNN:
 - removed last layer
 - added flattening layer
 - dropout layer (rate 0.2)
 - dense fully connected layer, $n = 8$ output signals, softmax
- ▶ training only the changed parameters
- ▶ in total 13 baseline DNNs (from Keras [1]), variants of: Xception, DenseNet, VGG, MobileNet, ResNet, and Inception
- ▶ best model: ResNet50, accuracy ≈ 0.563 , see Fig. 4
- ▶ worst model: InceptionV3, accuracy ≈ 0.488

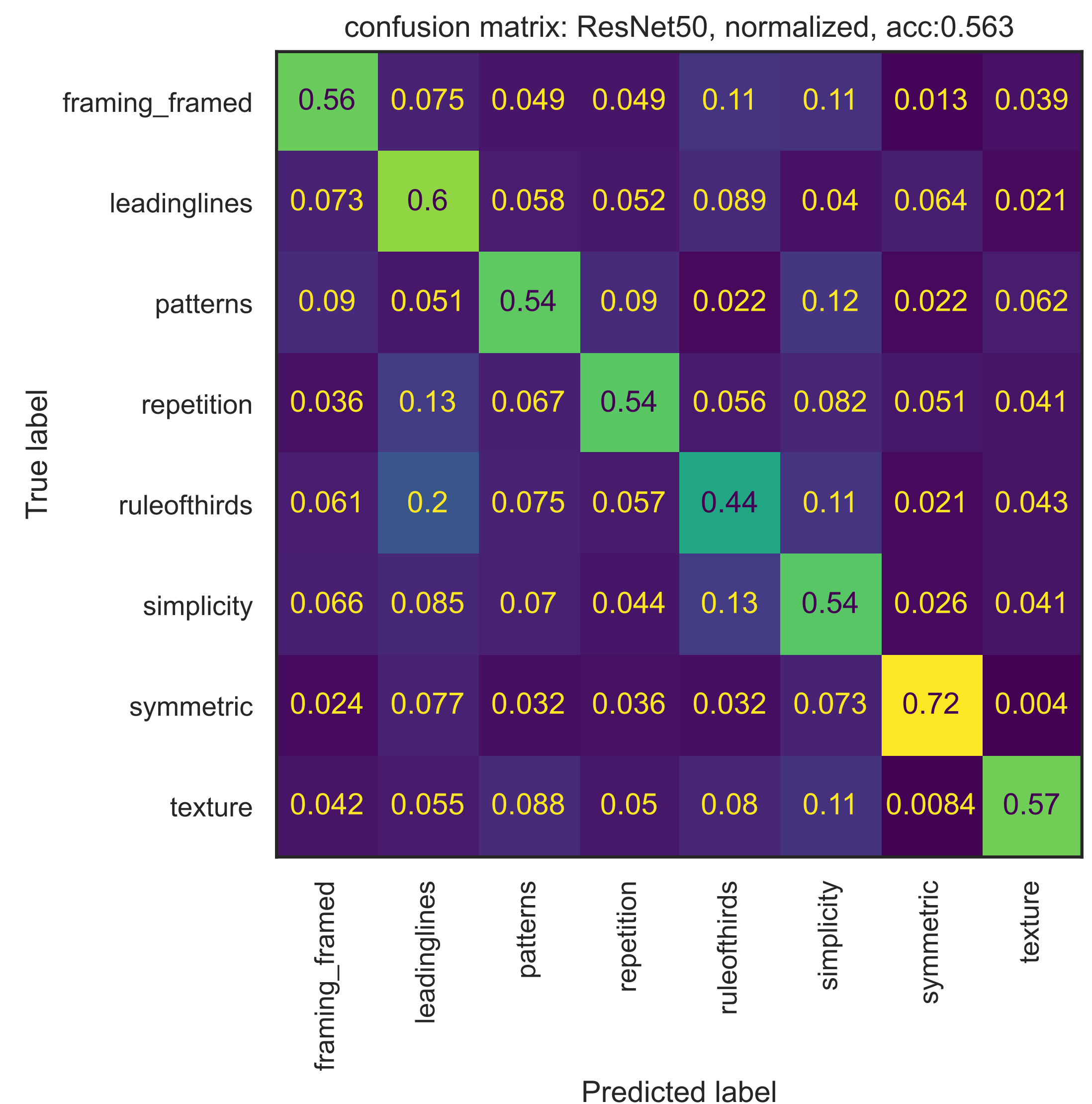


Figure 4: Confusion matrix of the best performing model (ResNet50) for photo rule prediction.

Conclusion

- ▶ presented extension of photo rule prediction
- ▶ dataset based on Flickr image tags
- ▶ trained 13 DNNs for prediction; similar performance, best ResNet50

Future Work

- ▶ evaluation with human annotations
- ▶ include in prediction systems for image appeal
- ▶ or as guidances system

References

- [1] F. Chollet et al. *Keras*. <https://keras.io>. 2015.
- [2] S. Göring et al. "Analysis of Appeal for realistic AI-generated Photos". In: vol. 11. 2023, pp. 38999–39012.
- [3] S. Göring et al. "Extended Features using Machine Learning Techniques for Photo Liking Prediction". In: *QoMEX*. Sardinia, Italy, May 2018.
- [4] S. Göring et al. "Rule of Thirds and Simplicity for Image Aesthetics using Deep Neural Networks". In: *23st International Workshop on Multimedia Signal Processing (MMSP)*. IEEE. 2021, pp. 1–6.
- [5] B. Krages. *Photography: The Art of Composition*. Allworth, 2012.
- [6] H. Leder et al. "A model of aesthetic appreciation and aesthetic judgments". In: *British journal of psychology* 95.4 (2004), pp. 489–508.
- [7] L. Mai et al. "Detecting rule of simplicity from photos". In: *ACM int. conf. on Multimedia*. 2012, pp. 1149–1152.
- [8] L. Mai et al. "Rule of thirds detection from photograph". In: *Int. Symp. on Multimedia*. IEEE. 2011, pp. 91–96.
- [9] L. Torrey et al. "Transfer learning". In: *Handbook of research on machine learning applications and trends: algorithms, methods, and techniques*. IGI global, 2010, pp. 242–264.
- [10] C. Zauner. *Implementation and benchmarking of perceptual image hash functions*. 2010.

Acknowledgment

This work is funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – DFG-437543412. Furthermore, we want to thank the "AG Wissenschaftliches Rechnen" of the TU Ilmenau for computing resources.