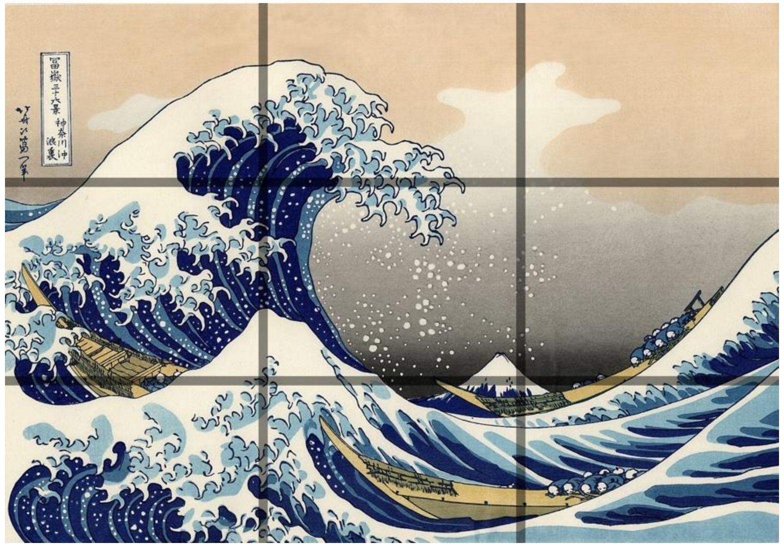
Rule-of-Thirds Detection with Interpretable Geometric Features

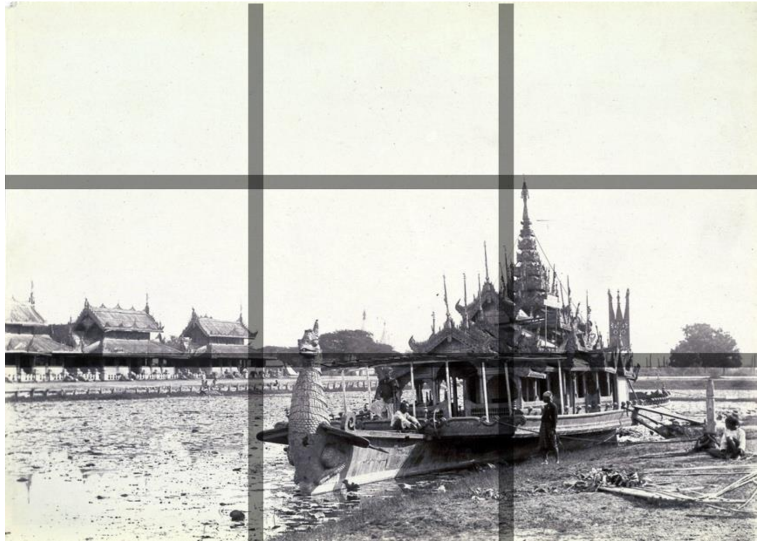
SUMAC @ ACM Multimedia 2025

Armin Dadras
Tingyu Lin
Robert Sablatnig
Markus Seidl

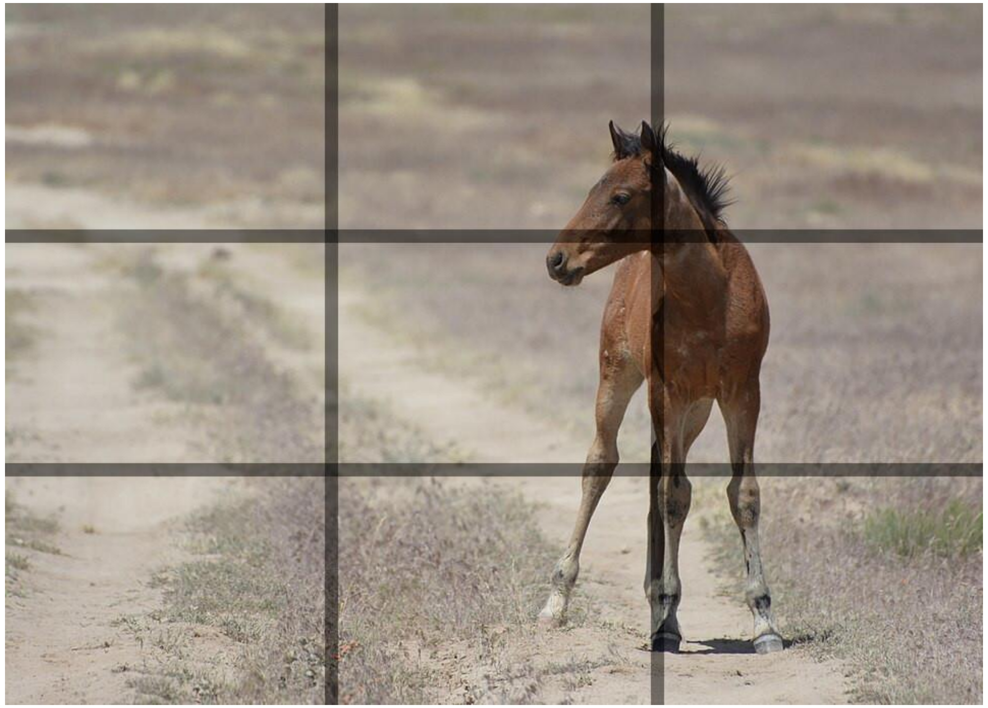
armin.dadras@{fhstp.ac.at,tuwien.ac.at}

The Rule-of-Thirds





The Rule-of-Thirds



Purpose

- Popular heuristic for visual media (e.g. Photography, Art History, Film studies)
- Related tasks: Aesthetic Scoring, Image Cropping, Image Composition

Difficulty

- Inherently Subjective
- Vaguely defined

Existing Research for Classification

Traditional Methods [23,24,2]	Deep Learning [10,11,35]
• Saliency Algorithms • Low-Level Features • Heuristics	• End-to-End • Saliency Features • Extended to more classes

Idea: combine deep models with traditional methods

(-) No High-Level Features (Semantics)

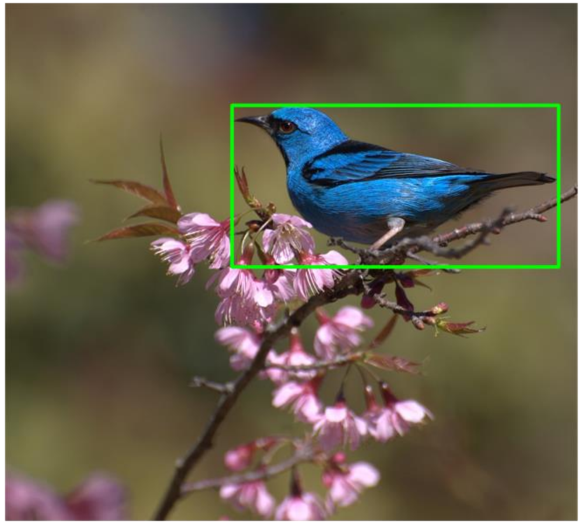
(-) Semantic Bias
(-) Opaque

Our Method

Input Image

Saliency

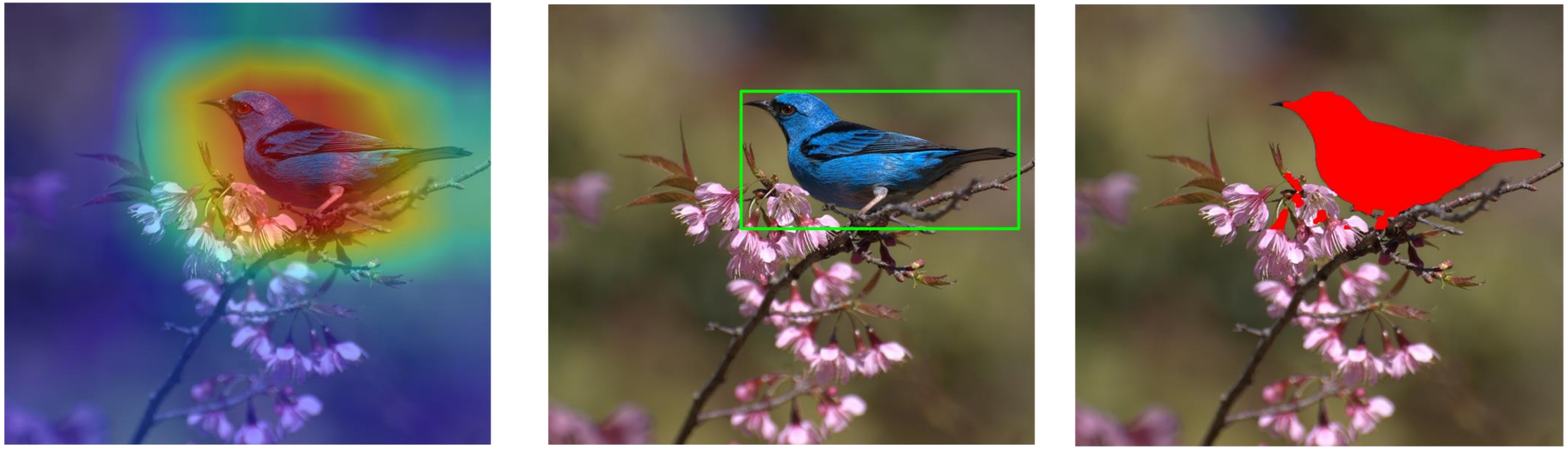
Bounding Box



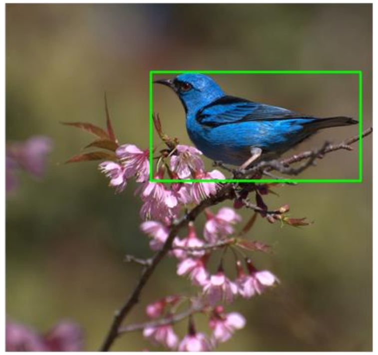
Segmentation Mask

Computer Vision Lab

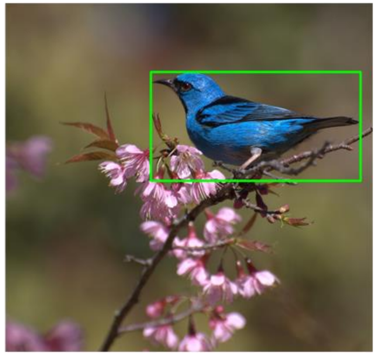
Our Method



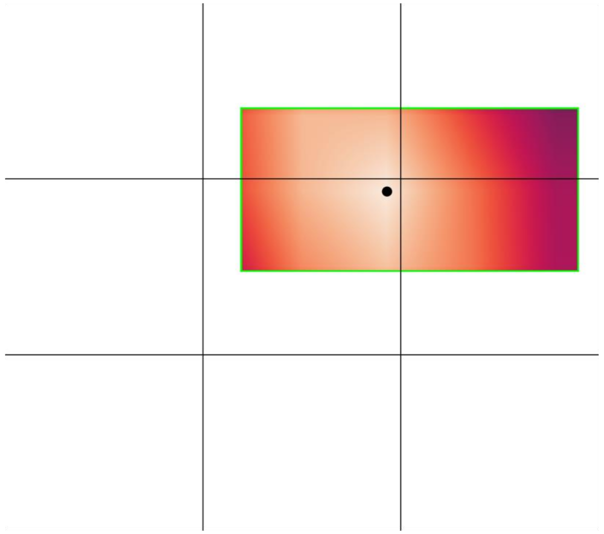
Our Method

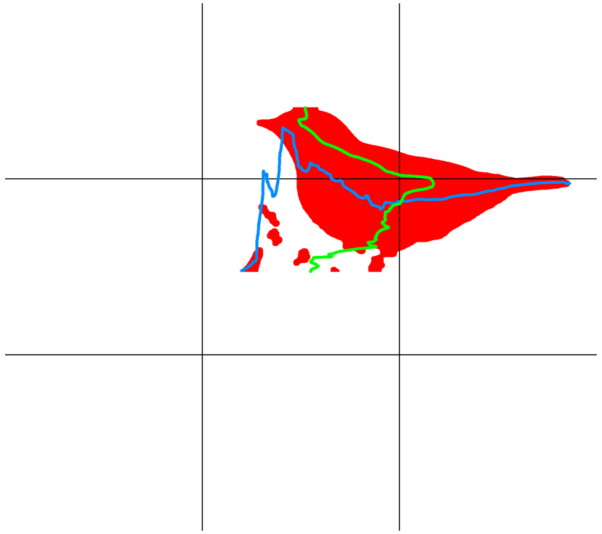


Our Method

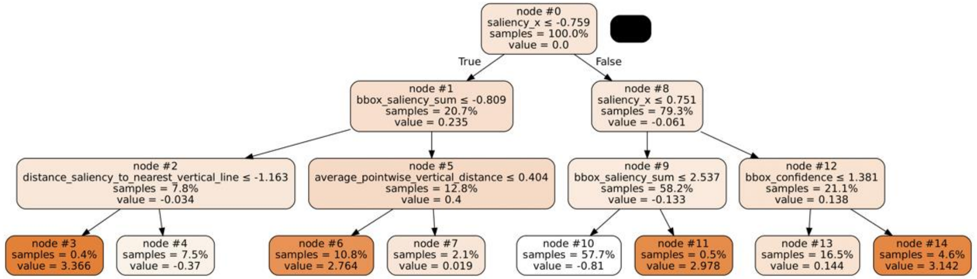


Feature Extraction

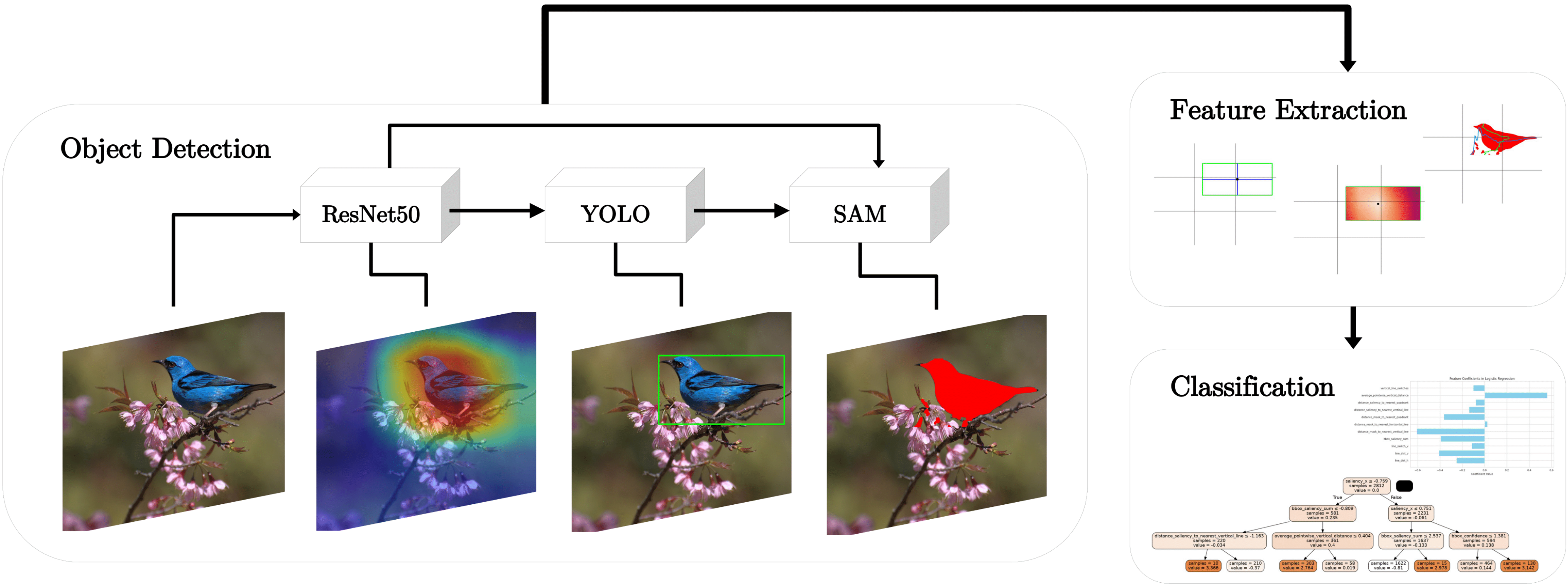




Classification



Our Method



Object Detection
ResNet50
YOLO
SAM
Feature Extraction
Classification

Data

3 Datasets

RoT-DS [Mai 2011]

- Binary Baseline
- Object Centric

KU-PCP [Lee 2018]

- Includes more classes
- Many Landscapes

AVT [Göring 2021]

- Multiple Labeler
- Noisy

Long Mai, Hoang Le, Yuzhen Niu, and Feng Liu. 2011. Rule of thirds detection from photograph. In 2011 IEEE international symposium on multimedia. IEEE, 91–96.

Jun-Tae Lee, Han-Ul Kim, Chul Lee, and Chang-Su Kim. 2018. Photographic composition classification and dominant geometric element detection for outdoor scenes. Journal of Visual Communication and Image Representation 55 (2018), 91–105

Steve Göring and Alexander Raake. 2021. Rule of thirds and simplicity for image aesthetics using deep neural networks. In 2021 IEEE 23rd International Workshop on Multimedia Signal Processing (MMSP). IEEE, 1–6.

Results: Baseline vs. Ours on RoT-DS

- 9.8% lower accuracy than end-to-end

Method	Accuracy
Mai et al.	80.5%
Males et al.	77.7%
Göring et al. (SOTA)	**84.1%**
Proposed Method	**74.3%**

Results: Ablation Studies

- 9.8% lower accuracy than end-to-end
- Combination of all features best
  - Segmentation based features better on landscape focused dataset (KU-PCP)
  - Bounding Box based features better on object focused dataset (RoT-DS)

F1 Scores

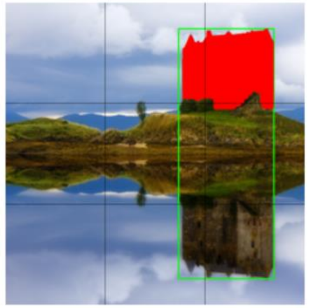
Configuration	RoT-DS	KUPCP
All Features	**0.748**	0.835
BBox + Saliency	0.751	0.657
Segmentation Only	0.683	**0.851**
BBox Only	0.745	0.657

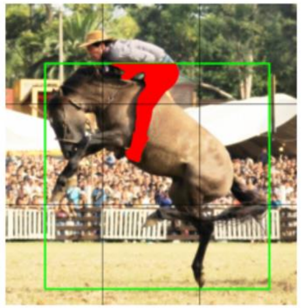
Results: Noisy Dataset (AVT)

- 9.8% lower accuracy than end-to-end
- Combination of all features best
  - Segmentation based features better on landscape focused dataset (KU-PCP)
  - Bounding Box based features better on object focused dataset (RoT-DS)
- On par with SOTA for noisy data (AVT)

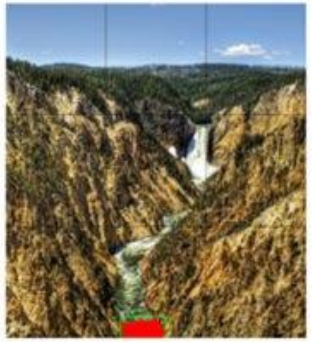
Method	Accuracy	F1-Score
Göring et al.	67.2%	0.381
Proposed Method	**75.6%**	**0.340**

Limitations





- Difficult Part-Object Relations



- Failing Segmentation/Saliency

- Insufficient Features

Conclusion & Future Research

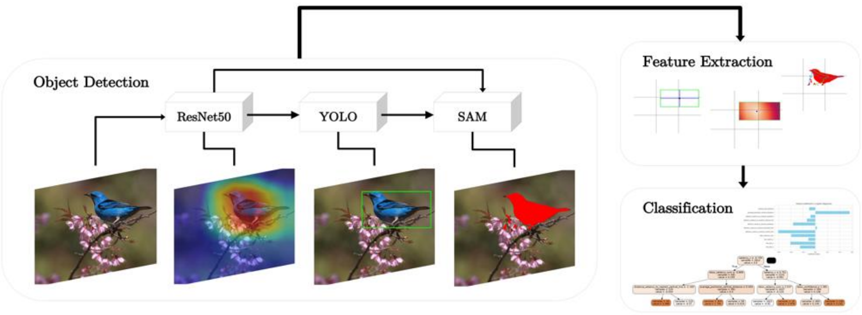
Conclusion

- Slightly below SOTA despite no end-to-end training
- Combination of features performs best
- On par with SOTA on noisy data

Limitations due to rigidness

We suggest:

- Improved part-object modelling
- Other Saliency Algorithms
- Better Integration with end-to-end
- Adaptability and transparency instead of pure accuracy



https://github.com/ADadras/Rule-of-Thirds-Detection-with-Interpretable-Geometric-Features

Thank you

Armin Dadras , armin.dadras@{fhstp.ac.at,tuwien.ac.at}