Interpretable and Accurate Fine-grained Recognition via Region Grouping
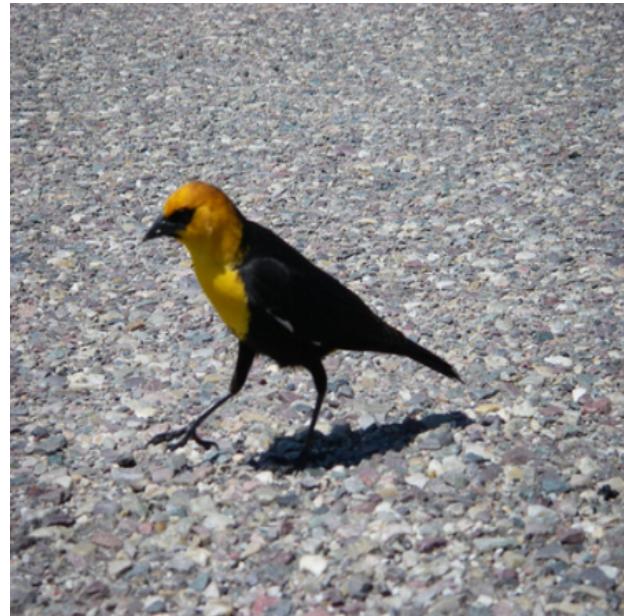
Zixuan Huang [1], *Yin Li* [2,1]

1. Department of Computer Sciences
2. Department of Biostatistics & Medical Informatics
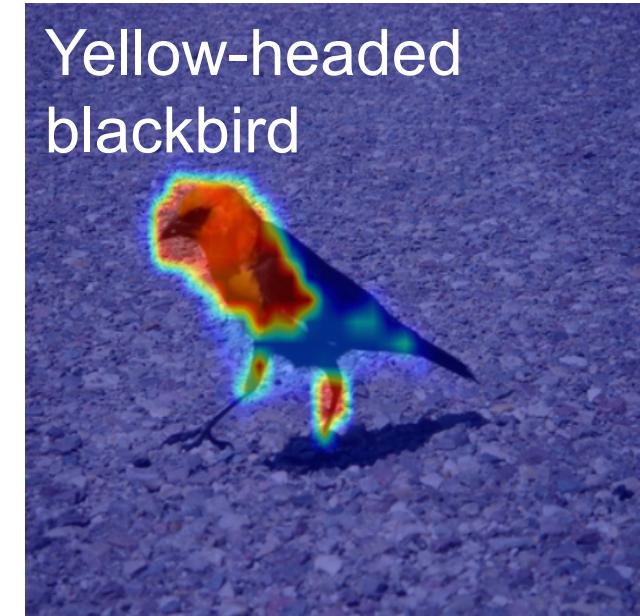
Interpretation = part segmentation + part attribution

Input

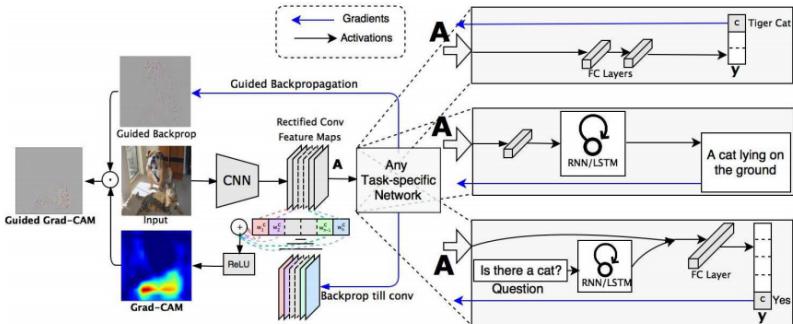Part segmentation

Yellow-headed blackbird

Part attribution

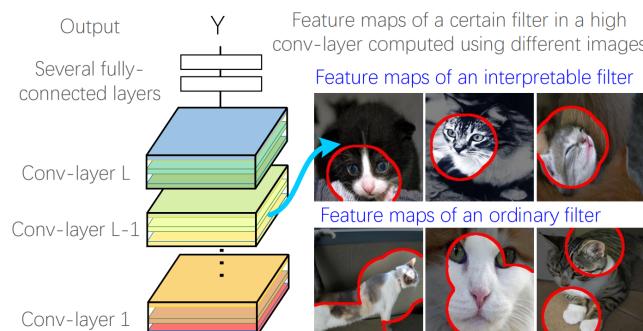Only image-level label required!
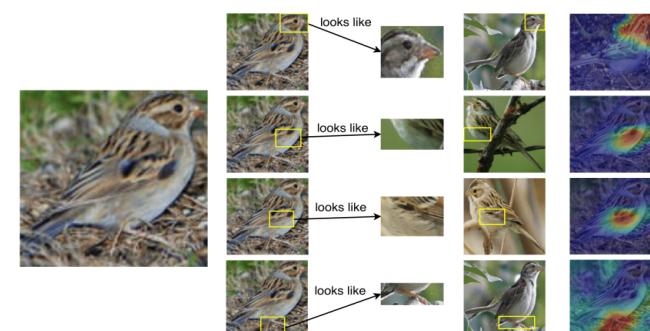
Related work

Zhou et al., CVPR'16

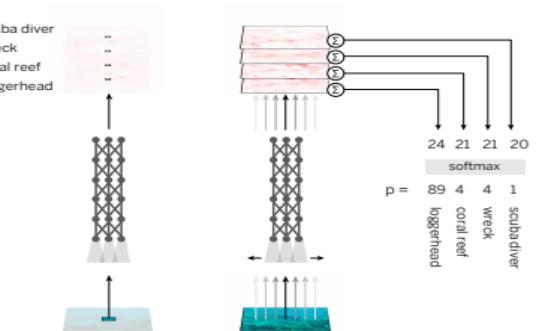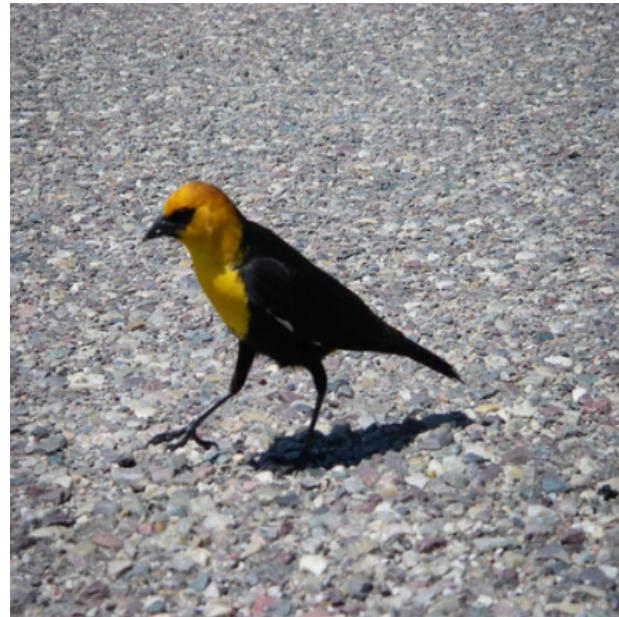Selvaraju et al., ICCV'17

Feng & Vedaldi, ICCV'17

Zhang et al., CVPR'18

Chen et al., NeurIPS'19

Brendel et al., ICLR'19
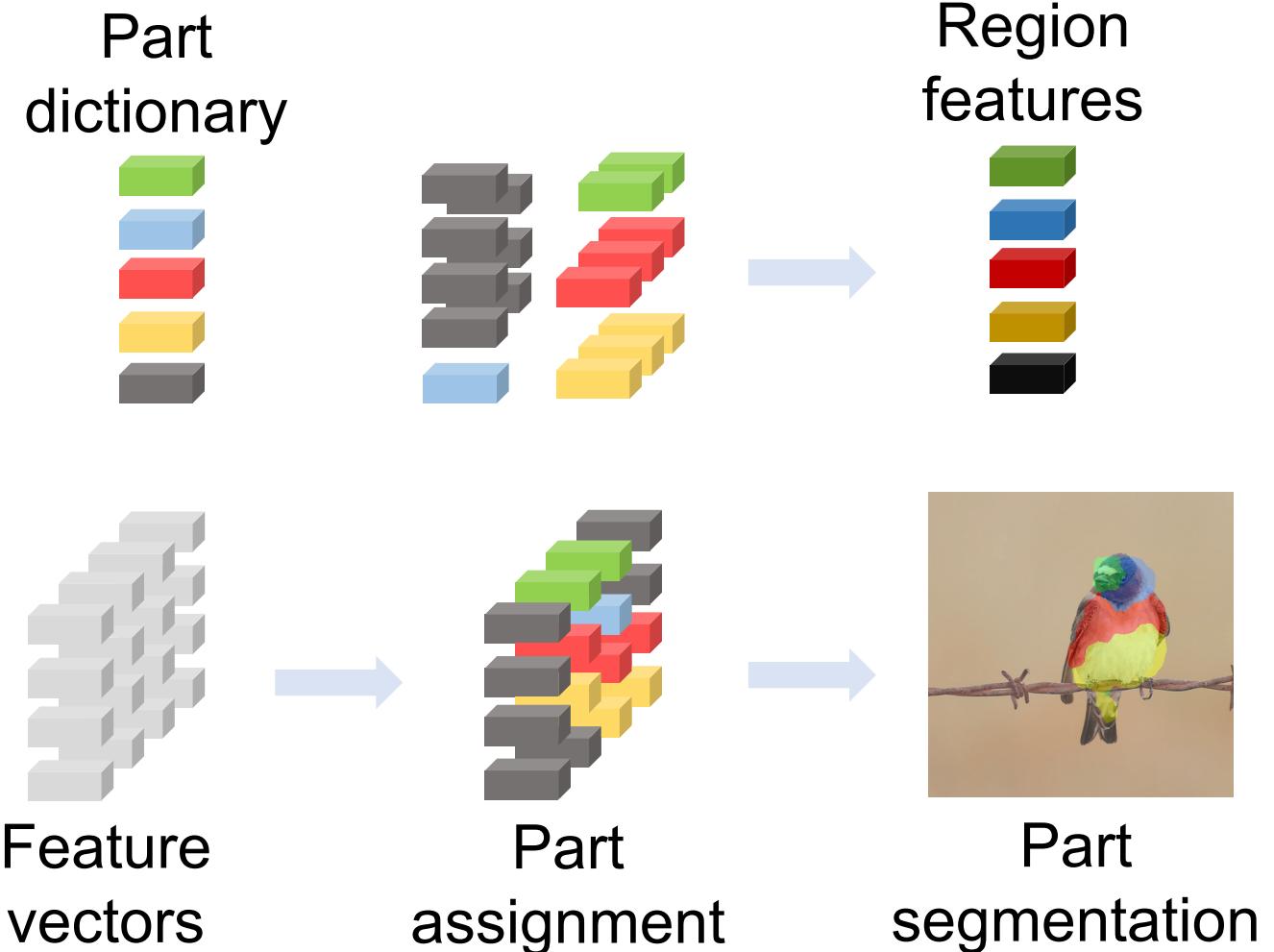
Interpretation = part segmentation + part attribution

Input

Part segmentation

Interpretable and Accurate Fine-grained Recognition via Region Grouping

Part segmentation via region grouping
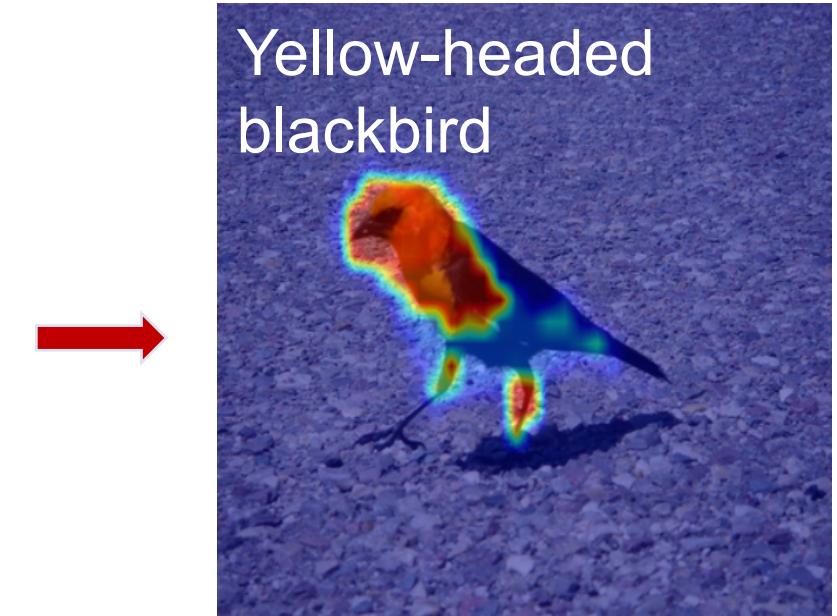
- Assign feature vectors to different centers

- Encode each part into one vector

Part dictionary

Region features

Input image → Feature map → Feature vectors → Part assignment → Part segmentation

Interpretation = part segmentation + part attribution
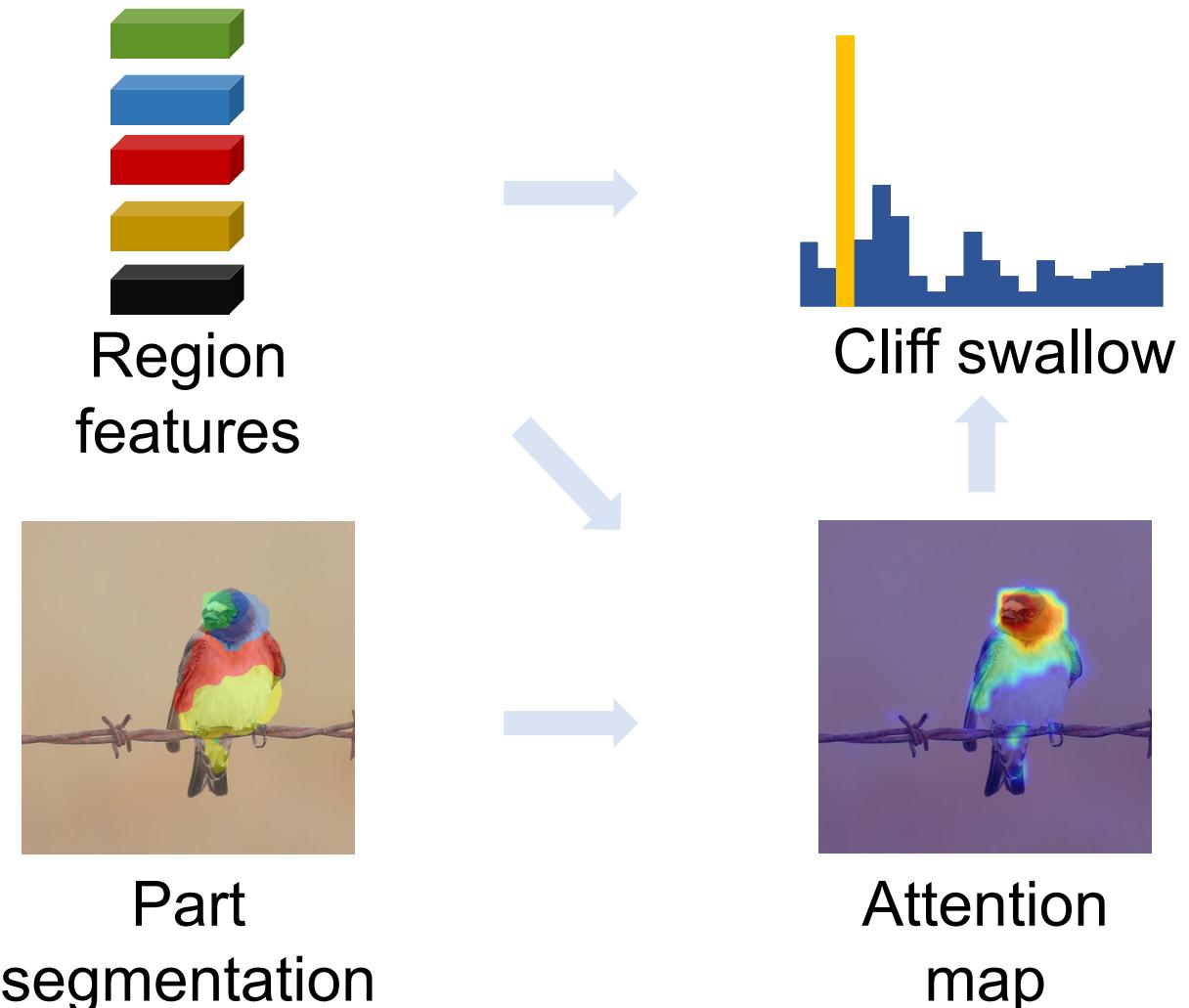

Yellow-headed blackbird

Part segmentation

Part attribution

Part attribution via region attention

Attention selects important regions for classification

- Generate region-based attention

- Attention-guided classification

Region features

Part segmentation

Cliff swallow

Attention map

Learning with image-level labels

How does an object part occur in natural images?

U-shaped distribution

Regularization by part occurrence
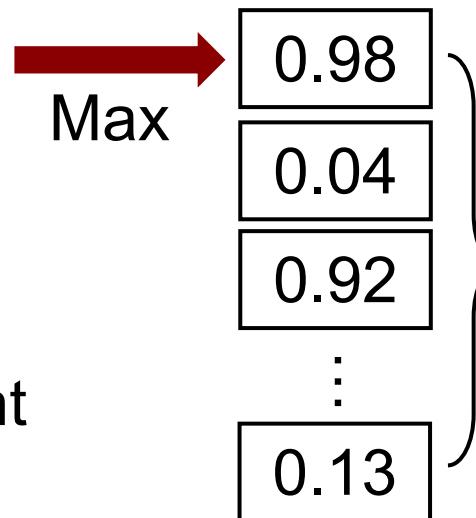
Max-pooling of part assignment as a part detector

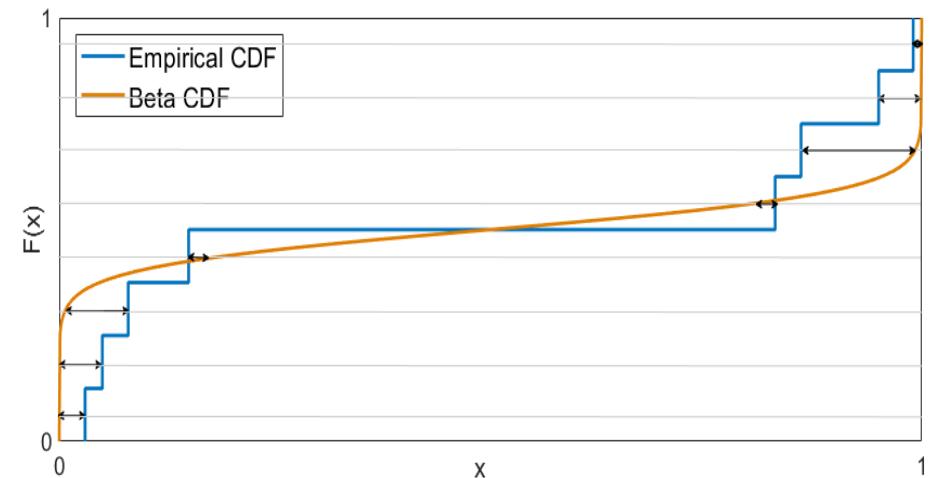Match the empirical distribution to prior using Earth-Mover distance

Part assignment for bird head

Max

| 0.98 |
| 0.04 |
| 0.92 |
| ⋮ |
| 0.13 |

Empirical distribution

Minibatch of N samples

Calculate Earth-Mover distance
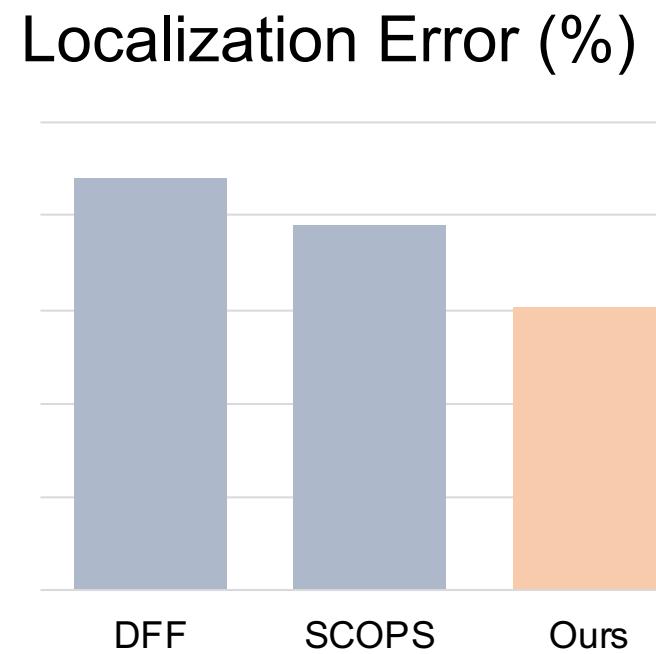
Results - CUB200

Bird species classification (accuracy)

Accuracy (%)

Bird landmark localization (interpretability)

Localization Error (%)

Qualitative results

Input　　　　　　Assignment　　　　　　Attention
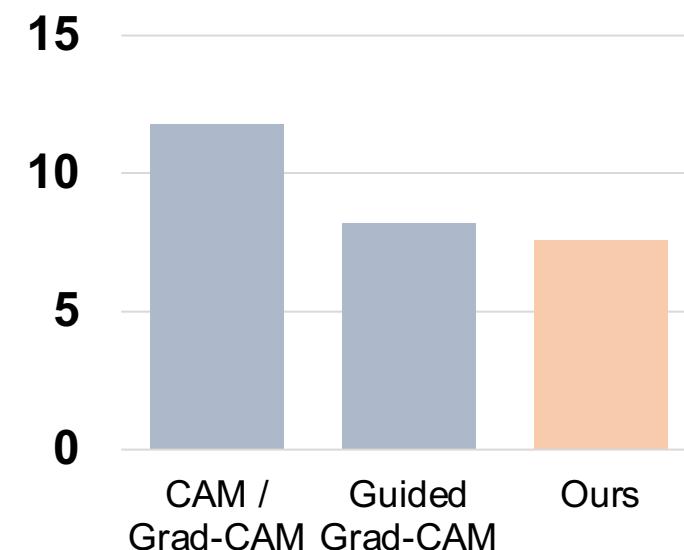
Results - iNaturalist2017

Species classification (accuracy)

Accuracy (%)

Pointing game using attention (interpretability)
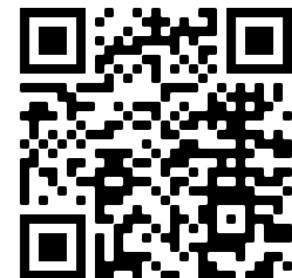
Pointing Error (%)

See our paper for more results on iNaturalist and CelebA datasets

Conclusion

- An interpretable and accurate model for fine-grained classification
- Region grouping + attention = interpretability
- A novel prior as regularization
- Strong performance over challenging datasets

Thank you!

Project website: https://www.biostat.wisc.edu/~yli/cvpr2020-interp/