



Context-aware Captions from Context-agnostic Supervision Devi Parikh³ **Gal Chechik² Kevin Murphy² Samy Bengio²** Rama Vedantam¹ ¹Virginia Tech ²Google Research ³Georgia Tech

Highlights

Image Captioning **Task 1: Discriminative Image Captioning** Model Context: Distractor Image **Task 2: Discriminative Justification** Justify: Why do you think this image is a prairie warbler, and not a mourning warbler? Justification:





Discriminative Image Captioning Results Qualitative Results Quantitative Evaluation

Human accuracy at picking target Image from distractor images

Approach	Easy Confusion	Hard Confusio
neuraltalk2	74.6	52.5
neuraltalk2+IS λ =0.3	89.0	74.1

Conclusion

Novel inference procedure to generate context-aware discriminative language from context agnostic supervision

- 3161 image, target distractor classes
- Results on discriminative image captioning as well as justification show improvement over generative and speaker-listener approaches.





crossing the street

Two novel tasks: discriminative image captioning and justification New dataset CUB-Justify with ground truth justifications for evaluation with