# Generating Visual Explanations

Lisa Anne Hendricks[1], Zeynep Akata[2], Marcus Rohrbach[1],
Jeff Donahue[1], Bernt Schiele[2], Trevor Darrell[1]
[1]University of California Berkeley, [2]Max Planck Institute for Informatics

## Motivation

- Explanations are important for understanding and interacting with intelligent systems

## Visual Explanations

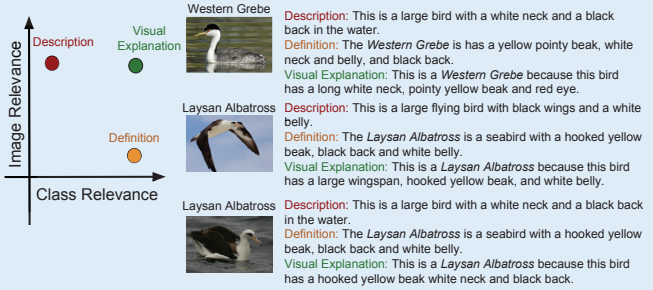
- Visual explanations must be image relevant and class consistent

## Justification vs. Introspection

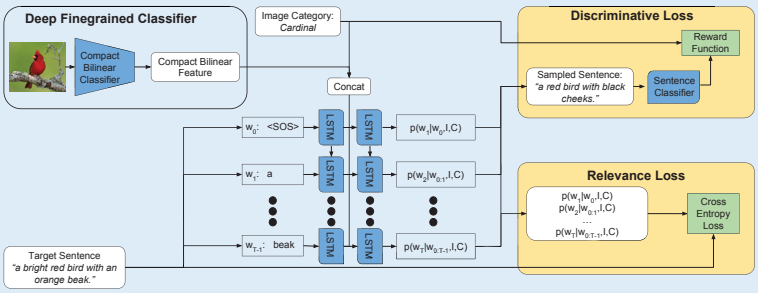
- *Justification* explanations detail why a prediction is compatible with visual evidence
- *Introspective* explanations detail the internal mechanisms of a model
- Here, concentrate on justifications as they are useful for non-experts

## Method:

- Introduce novel discriminative loss based on REINFORCE [1]




# Model




## Relevance Loss

*Encourages sentences to be **image relevant** by minimizing cross entropy between ground truth and predicted words.*

$$L_R = \sum_{t=0}^{T-1} \log p(w_{t+1}|w_{0:t}, I, C)$$

where $I$ is an image feature, $w_t$ is a ground truth word, and $C$ is a class label.

## Discriminative Loss

*Encourages sentences to be **class discriminative** by maximizing a reward, $R_D$, which measures class discriminativeness of a sampled sentence $\widetilde{w}$.*

$$L_D = -E_{\widetilde{w}\sim p(w)}[R_D(\widetilde{w})]$$

Estimate $L_D$ with Monte Carlo sampling because computing an expectation over descriptions is intractable.

Use REINFORCE [1] to estimate the expected gradient of $L_D$:

$$\nabla E_{\widetilde{w}\sim p(\widetilde{w})}[R_D(\widetilde{w})] = E_{\widetilde{w}\sim p(\widetilde{w})}[R_D(\widetilde{w})\nabla_W \log p(\widetilde{w})]$$

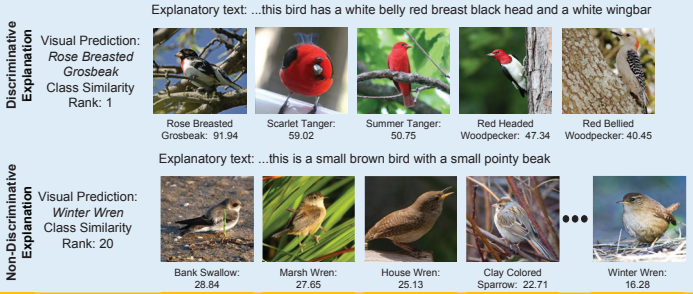
# Data

- Use finegrained descriptions collected in [2]
- 200 bird classes, > 11k images, 5 captions/image
- Finegrained descriptions more class informative than attributes.

*Ground truth descriptions:*
...there is a **black bird** with a **black beak** and **red eyes**.
...this bird has a **large, black, curved bill**, **black tarsuses and feet**, and a **black throat**, **breast**, and **belly**.
...**angular beak** and **red eye**, this otherwise nondescript black bird has **alarmingly large feet**.

# Metrics

- Measure **image relevance** using CIDEr.
- Measure **class relevance** using class similarity rank.




# Results

## Explanations are more **class relevant** than descriptions.

*This is a **White Necked Raven** because...*

*Description:* ...this bird is black in color with a black beak and black eye rings.
*Explanation:* ...this is a black bird with a **white nape** and a **large black beak**.

*This is a **Common Raven** because...*

*Description:* ...this bird is black in color with a black beak and black eye rings.
*Explanation:* ...this is a black bird with a long black tail feather and **pointy black beak**.

*This is a **Brown Pelican** because...*

*Description:* ...this bird has a long neck and long bill.
*Explanation:* ...this is a large **brown bird** with a long white neck and a long bill.

*This is a **White Pelican** because...*

*Description:* ...this is a large white bird with a long neck and a long orange beak.
*Explanation:* ...this bird has **a white body**, long neck, and large orange bill.

*This is a **Scarlet Tanger** because...*

*Description:* ...this bird has a red crown, a short bill, and a red belly.
*Explanation:* ...this is a red bird with **black wings** and a small beak.

*This is a **Cardinal** because...*

*Description:* ...this bird has a red crown, a short bill, and a red breast.
*Explanation:* ...this is a red bird with a **black cheek patch** and **a red beak**.

## Explanations are more **image relevant** than definitions.

*This is a **Crested Auklet** because...*

*Definition:* ...this bird is black with a bright orange bill and white eyes.
*Explanation:* ...this is a black bird with a white eye and an orange beak.

*This is a **Crested Auklet** because...*

*Definition:* ...this bird is black with a bright orange bill and white eyes.
*Explanation:* ...this is a grey bird with a white eye and an orange beak.

*This is a **Red Headed Woodpecker** because...*

*Definition:* ...this bird has a red crown, a white belly, and a white wing bar.
*Explanation:* ...this bird has a white belly and breast, red crown, and black back and back.

*This is a **Red Headed Woodpecker** because...*

*Definition:* ...this bird has a red crown, a white belly, and a white wing bar.
*Explanation:* ...this is a black and white bird with a red head and a pointy beak.

*This is a **Ruby Throated Hummingbird** because...*

*Definition:* ...this bird is white and brown in color with a long skinny beak, and brown eye rings.
*Explanation:* ...this is a small bird with a long thin beak and a white breast and belly.

*This is a **Ruby Throated Hummingbird** because...*

*Definition:* ...this bird is white and brown in color with a long skinny beak, and brown eye rings.
*Explanation:* ...this is a small bird with a red throat and a long thin bill.

## Explanations for incorrect classification decisions.

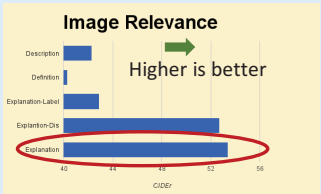
**Correct**: Anna Hummingbird
**Predicted:** Green Violetear

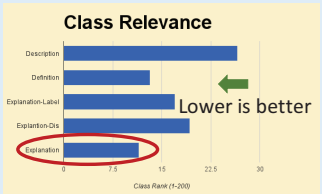
*Explanation:* ...this is a green and green bird with a long pointy black beak
*Green Violetear Definition:* ...this bird is green with blue on its head and has a long pointy beak.

**Correct**: Anna Hummingbird
**Predicted:** Anna Hummingbird

*Explanation:* ...this bird has a white belly, a long, thin bill, a white throat, and a green head.
*Anna Hummingbird Definition:* ...this bird is white and brown in color with a long skinny beak, and brown eye rings.

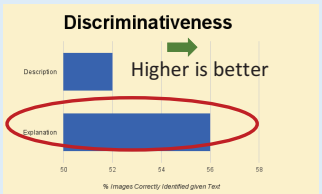
## Quantitative Evaluations




## Human Evaluations


Asked two experienced bird watchers to rank quality of explanations.




Asked non-experts to choose between two similar bird classes given either a description or an explanation.

[1] Williams, Ronald J. "Simple statistical gradient-following algorithms for connectionist reinforcement learning." *Machine learning* 8.3-4 1992.
[2] Reed, Scott, et al. "Learning Deep Representations of Fine-Grained Visual Descriptions." *CVPR 2016.*