Generating Visual Explanations

**نویسنده اول: Lisa Anne Hendricks**

**سال انتشار: 2016**

**محل چاپ: ECCV**

**لینک به صفحه دانلود مقاله: https://arxiv.org/abs/1603.08507**

## در یک جمله بنویسید مقاله چه کار کرده

یک مدل جدید طراحی کردن که بتونه علاوه بر اینکه classification رو انجام میده، بتونه توضیح بده چرا اون شئ رو در اون دسته قرار داده.

## چکیده مقاله

Abstract. Clearly explaining a rationale for a classification decision to an end-user can be as important as the decision itself. Existing approaches for deep visual recognition are generally opaque and do not output any justification text; contemporary vision-language models can describe image content but fail to take into account class-discriminative image aspects which justify visual predictions. We propose a new model that focuses on the discriminating properties of the visible object, jointly predicts a class label, and explains why the predicted label is appropriate for the image. We propose a novel loss function based on sampling and reinforcement learning that learns to generate sentences that realize a global sentence property, such as class specificity. Our results on a fine-grained bird species classification dataset show that our model is able to generate explanations which are not only consistent with an image but also more discriminative than descriptions produced by existing captioning methods

## کارای قبلی که مقاله ذکر کرده چی بودند (background literature, previous work)

این مقاله از اولین مقاله هایی است که روی تولید توضیحات برای مسئله classification کار میکند و قبل از آن به مسئله classification به صورت جدا پرداخته شده بود. همچنین روی image captioning هم کار شده بود ولی باز هم توضیحات راجع به عکس بود نه راجع به فرایند تصمیم گیری مدل.

## به طور خلاصه در یک پاراگراف بگید که مقاله، چالش، ایراد یا ضعف کارهای قبلی رو چگونه ارزیابی کرده

کارهای قبلی اکثرا فقط عمل classification رو انجام میدادن و نهایتا یه متنی هم تولید میکردن که خود عکس رو توضیح میداد مثلا میگفت این عکس حاوی یک سیب است که روی درخت آویزان است (combining vision and language for image captioning) ولی توضیحی راجع به اینکه چرا این عکس یک سیب است نمیداد.

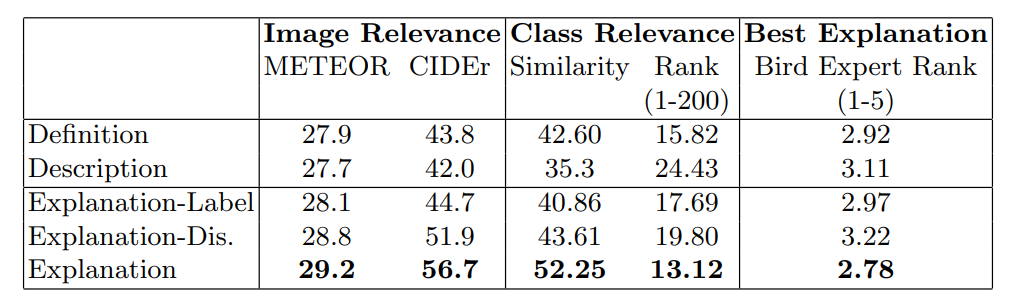
## ایده مقاله برای حل چالشها، ایرادها یا ضعف های پاراگراف قبل چی بوده (خلاصه)

طراحی یک مدل جدید که روی ویژگیهای متمایز کننده ی شئ تمرکز کنه و همچنین طراحی یک تابع ضرر جدید بر اساس تکنیکهای sampling و reinforcement learning.

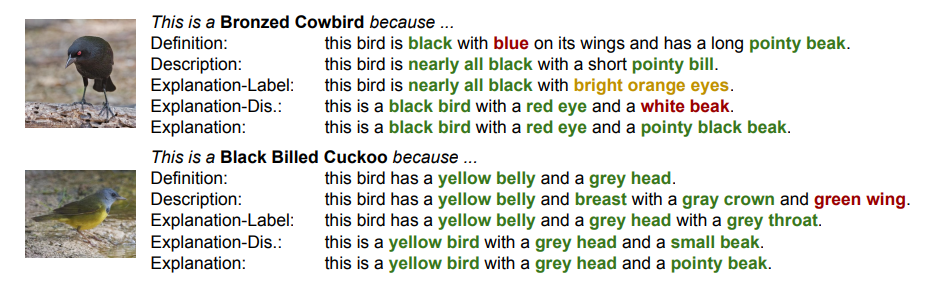
We propose a new model that focuses on the discriminating properties of the visible object, jointly predicts a class label, and explains why the predicted label is appropriate for the image. We propose a novel loss function based on sampling and reinforcement learning that learns to generate sentences that realize a global sentence property, such as class specificity.

## مقاله تا چه حد تونسته با روش پیشنهادی خودش به اون چالشها، ایرادها و ضعف ها پاسخ بده (metrics,evaluations,results)

اومده از معیارهای METEOR و CIDEr برای شباهت جمله ها از لحاظ مفهومی استفاده کرده. مدل خودش رو با مدل Definition و Description مقایسه کرده. اولی صرفا تعریف هر کلاس رو خروجی میده و دومی صرفا یه توصیفی برای اون عکس میاره بدون اینکه به ویژگیهای کلاس پیش بینی شده توجهی بکنه. همچنین مدل خودش رو به دو شکل ناقص کرده تا نشون بده ایده هایی که برای اون مدل استفاده کرده تاثیر گذار بوده اند. نسخه نهایی مدلشان همان ردیف مربوط به Explanation است که بیشترین دقت را بدست آورده است هم از لحاظ مرتبط بودن توضیحات به تصاویر و هم مرتبط بودن آن با کلاس پیشبینی شده. همچنین از متخصصانی هم برای امتیازدهی استفاده کردن که توی ستون Best Explanation اومده و مقدارش هرچی کمتر باشه بهتره.



این هم دو نمونه از خروجیهای هر مدل:



## شما برای حل چالشهای گفته شده چه ایده ای دارین؟ (خلاصه)

استفاده از مدلهای زبانی بزرگ برای یادگیری توضیحات

## شما برای بهبود این مقاله چه ایده ای دارین؟ (خلاصه)

استفاده از ترکیب چندین تابع ضرر و تست کردن آن تا به دقت بهتری شاید برسد.