Attribute Prototype Network for Zero-Shot Learning

**نویسنده اول: Wenjia Xu**

**سال انتشار: 2020**

**محل چاپ: NeurIPS**

**لینک به صفحه دانلود مقاله: https://arxiv.org/abs/2008.08290**

## در یک جمله بنویسید مقاله چه کار کرده

روشی جدید برای attribute localization در تصویر ارائه داده و از آن برای بهبود مسئله zero-shot learning استفاده کرده است.

## چکیده مقاله

From the beginning of zero-shot learning research, visual attributes have been shown to play an important role. In order to better transfer attribute-based knowledge from known to unknown classes, we argue that an image representation with integrated attribute localization ability would be beneficial for zero-shot learning. To this end, we propose a novel zero-shot representation learning framework that jointly learns discriminative global and local features using only class-level attributes. While a visual-semantic embedding layer learns global features, local features are learned through an attribute prototype network that simultaneously regresses and decorrelates attributes from intermediate features. We show that our locality augmented image representations achieve a new state-of-the-art on three zero-shot learning benchmarks. As an additional benefit, our model points to the visual evidence of the attributes in an image, e.g. for the CUB dataset, confirming the improved attribute localization ability of our image representation.

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

قبلا برای بهبود مسئله zero-shot learning روشهایی مثل visual attention برای مشخص کردن محل رخداد یک attribute وجود داشت.

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

دارای بعضی ایرادها بود، به عنوان مثال خود را با داده های آموزشی خیلی تطبیق میداد و برای صفت هایی که معمولا در داده های آموزشی با هم رخ میدادند، در داده های تست، اگر با هم نبودند، دچار مشکل میشد. در مقاله به صورت زیر به آن اشاره شده است:

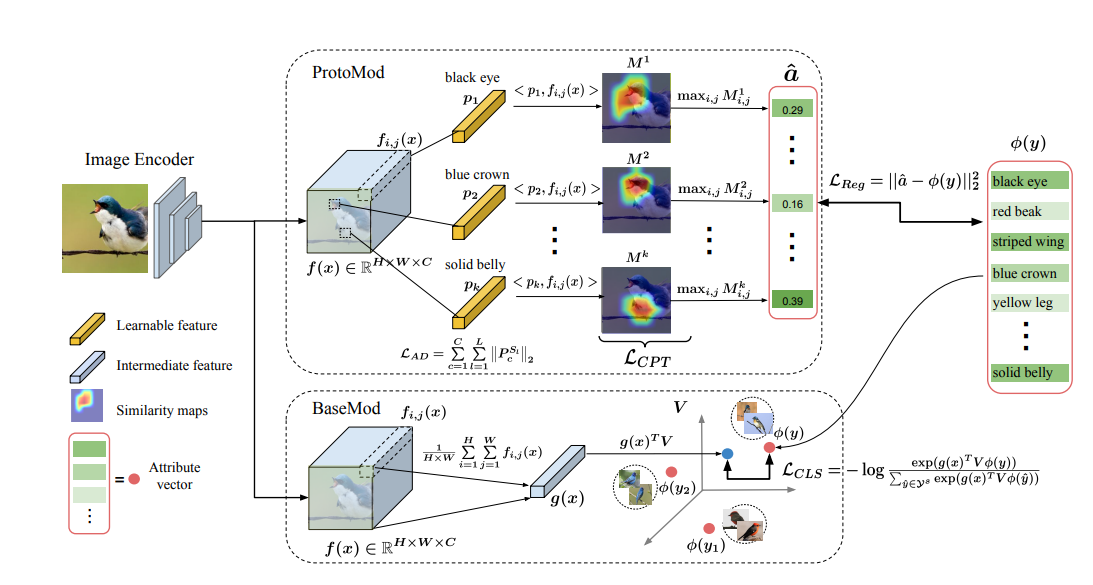
Although visual attention accurately focuses on some object parts, often the discovered parts and attributes are biased towards training classes due to the learned correlations. For instance, the attributes yellow crown and yellow belly co-occur frequently (e.g. for Yellow Warbler). Such correlations may be learned as a shortcut to maximize the likelihood of training data and therefore fail to deal with unknown configurations of attributes in novel classes such as black crown and yellow belly (e.g. for Scott Oriole) as this attribute combination has not been observed before.

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

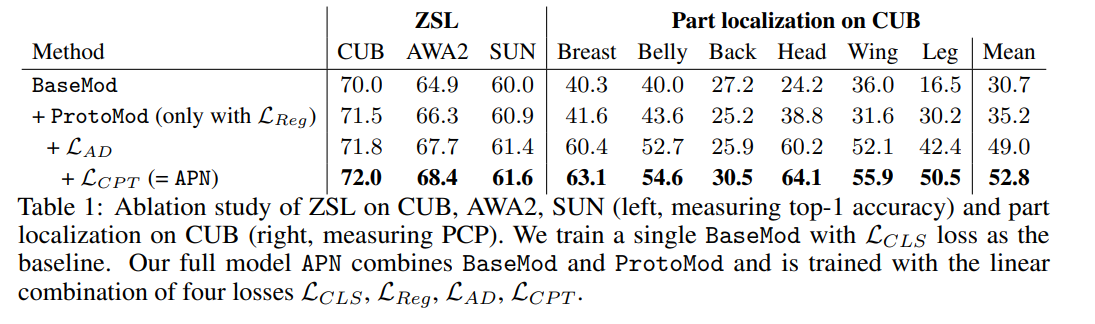
اومده یه فریم ورک برای representation learning ایجاد کرده که با نظارت ضعیف هست (weakly supervised) که میاد صفات بصری رو localize و decorrelate میکنه. به طور ویژه تر، میاد توابع ضرری رو در لایه های ابتدایی شبکه کانولوشنی اعمال میکنه تا فیچرهای این لایه ها یاد بگیرن ویژگی های محلی رو استخراج کنن. در مقاله به صورت زیر آمده است:

To improve locality and mitigate the above weaknesses of image representations, we develop a weakly supervised representation learning framework that localizes and decorrelates visual attributes. More specifically, we learn local features by injecting losses on intermediate layers of CNNs and enforce these features to encode visual attributes defining visual characteristics of objects. It is worth noting that we use only class-level attributes and semantic relatedness of them as the supervisory signal, in other words, no human annotated association between the local features and visual attributes is given during training. In addition, we propose to alleviate the impact of incidentally correlated attributes by leveraging their semantic relatedness while learning these local features

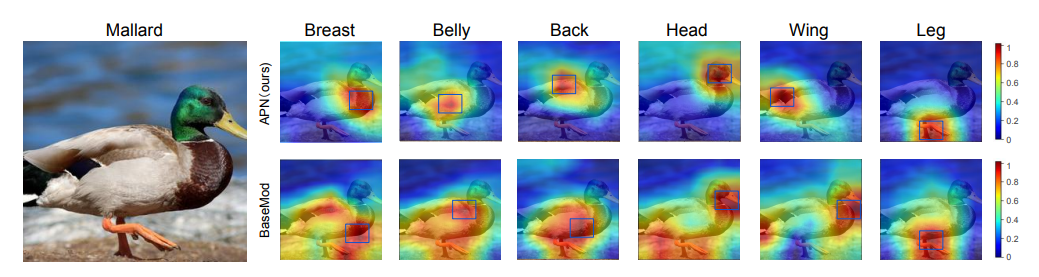
شبکه ای که طراحی کردن و اسمش رو هم APN(Attribute Prototype Network) گذاشتن:

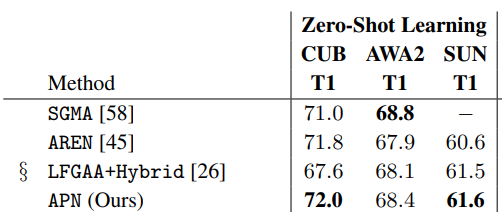


## مقاله تا چه حد تونسته با روش پیشنهادی خودش به اون چالشها، ایرادها و ضعف ها پاسخ بده (metrics,evaluations,results)



وقتی همه‌ی ایده های این مقاله روی هم جمع میشود، در مسئله zero-shot learning روی دیتاست CUB به دقت 72 میرسد. همچنین در مشخص کردن محل رخداد یک صفت، عملکرد بهتری نسبت به مدلهای قبلی دارد و این مسئله در عکسهای پایین هم مشخص است:





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

فکر کنم بشه در هنگام آموزش، یه سری Hint هم به مدل بدیم، مثلا صفاتی که به هم ربطی ندارن، اینجوری مدل میتونه یه معیاری در بیاره از چیزایی که به هم ربط خاصی ندارن و صرفا مثلا توی داده آموزشی زیاد با هم رخ دادن. در این صورت دیگه اون مشکل بوجود نمیاد.

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

میتونن توابع ضرر متنوعی رو در شبکه شون امتحان کنن یا از بعضی تکنیکها در شبکه شون بهره ببرن تا شاید به درصد دقت بهتری هم برسند.