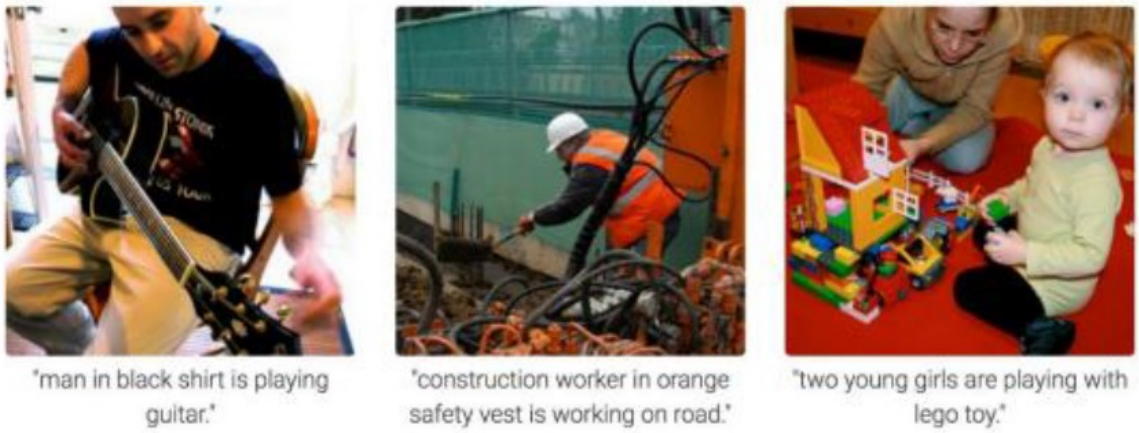
دانشکدگان علوم
دانشکده ریاضی، آمار و علوم کامپیوتر

پروژه حسابگری زیستی مهلت تحویل: ۲۰ تیر

توصیف عکس

هدف ما در این پروژه ایجاد و آموزش مدلی است که بتواند یک تصویر را به عنوان ورودی بگیرد و در نهایت یک جمله در توصیف آن عکس در خروجی خود تولید کند. تصویر زیر نمونه‌ای از خروجی این شبکه را نشان می‌دهد.



شکل ۱: خروجی یک مدل آموزش دیده برای Image Captioning

ساختار کلی این مدل‌ها به این صورت است که یک شبکه CNN جهت تولید ویژگی‌های تصاویر وجود دارد و در کنار آن روش‌های مختلفی برای Embedding جملات موجود است که در نهایت بردار ویژگی تصاویرو متن در کنار هم قرار گرفته و به عنوان ورودی یک شبکه بازگشتی اعمال می‌شود تا در نهایت جمله نهایی را تولید نماید. برای آشنایی بیشتر این مقاله را مطالعه کنید.

مجموعه دادگان

مجموعه دادگان را دریافت کنید. این مجموعه از دو بخش به نام Image و Caption.txt تشکیل شده است که پوشه

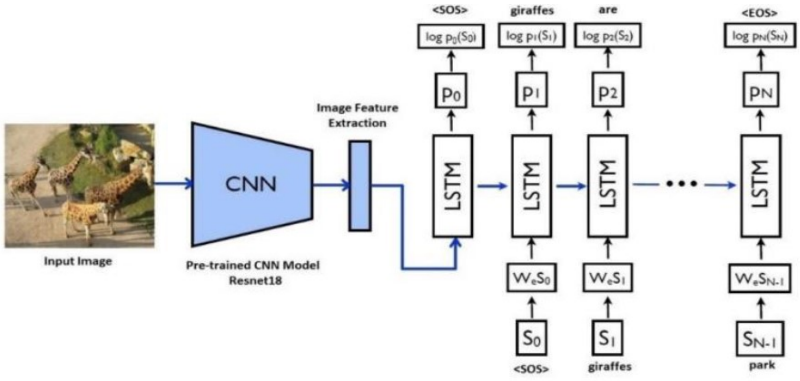
Image شامل ۸۰۹۱ تصویر و Caption.txt شامل ۴۰۴۵۵ جمله است که برای هر تصویر ۵ جمله مختلف توسط افراد مختلف جمع آوری شده است. در کنار هر جمله نام تصویر مورد نظر نیز آورده شده است.

پیش‌پردازش داده‌ها

با آماده سازی تصاویر برای اعمال به شبکه‌های کانولوشنی پیشتر آشنا شدید. در اینجا جملات نیز باید پیش‌پردازش شوند تا به بردارهایی از اعداد تبدیل شوند.(برای سادگی پیشنهاد می‌شود که از لایه Embedding در پایتورچ استفاده کنید) برای هر کلمه یک بردار عددی با طول ۳۰۰ در نظر بگیرید. نکته که مهمی که در پیش‌پردازش داده‌ها باید توجه نمایید، این است که باید برای هر جمله از توکن‌های شروع و پایان جمله <SOS>و <EOS> استفاده نماییم. که توکن‌های خاصی می‌باشد که توسط خود شما تعریف می‌شوند. همچنین باید مجموعه لغات موجود در مجموعه دادگان خود را پردازش و به هر کدام از آنها یک Index نسبت دهید. بهتر است علامت‌های نگارشی از جملات حذف شوند. همچنین از آنجایی که جملات Caption ها طول‌های متفاوتی دارند باید طول آن‌ها باهم یکسان شوند، که این کار را با Padding مناسب می‌توانید انجام دهید که می‌توان یک طول مشخص ثابت را در نظر گرفت یا یکسان‌سازی را در هر mini batch انجام داد.

ساخت مدل

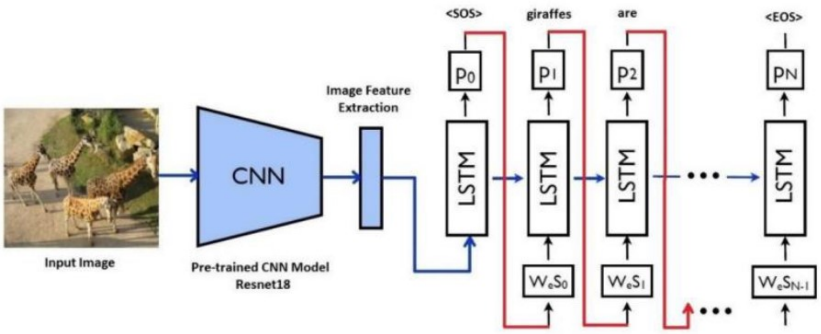
در تصویر زیر مدل کلی مد نظر را مشاهده می‌کنید. بخشی از مدل جهت استخراج ویژگی تصاویر مورد استفاده قرار می‌گیرد. در این مسئله ما قصد داریم از یک مدل از پیش‌آموزش دیده Resnet18 استفاده نماییم.از آخرین لایه شبکه کانولوشنی آن ویژگی‌های تصویر استخراج می‌شود که در نهایت نیاز است به یک لایه خطی جهت استخراج ویژگی‌های مورد نظر با ابعاد مناسب جهت ورود به شبکه بازگشتی، استفاده نمود.



در این قسمت از یک لایه شبکه LSTM با تعداد ۲۵۶ لایه پنهان استفاده می‌نماییم و بردارهای Embed شده جملات در کنار بردار تصویر به آن داده شده و خروجی آن به یک لایه‌خطی به سایز ورودی Hidden State و سایز خروجی تعداد کلمات موجود در مجموعه دادگان اعمال میشود و به این ترتیب به محاسبه خطا و پیشبینی مدل می‌پردازیم.

پیش‌بینی شبکه

بعد از آموزش شبکه، نیاز دارید تا شبکه را ارزیابی نمایید. جهت ارزیابی شبکه باید به صورتی که در تصویر زیرنشان داده شده از شبکه استفاده نماییم.



همانطور که می‌دانیم در زمان تست شبکه آموزش داده شده، Caption وجود ندارد و ما باید برای یک تصویر Caption تولید نماییم. برای این منظور روش‌های مختلفی وجود دارد ولی ما در اینجا مدل بالا را پیشنهاد می‌دهیم. در یک تابع به عنوان ورودی، تصویر تست و مدل آموزش داده شده را جهت پیش‌بینی کلمات اعمال می‌کنیم. قطعه کد زیر الگوریتم این شبکه را نمایش دادهاست.

```
x = self.encoderCNN(image).unsqueeze(0)
states = None

for _ in range(max_length):
    hiddens, states = self.decoderRNN.lstm(x, states)
    output = self.decoderRNN.linear(hiddens.squeeze(0))
    predicted = output.argmax(1)
    result_caption.append(predicted.item())
    x = self.decoderRNN.embed(predicted).unsqueeze(0)

    if vocabulary.itos[predicted.item()] == "<EOS>":
        break
```

در نهایت caption-prediction مجموعه index های کلمات می‌باشد که در نهایت به کمک دایره لغات موجود در مجموعه دادگان قابل تبدیل به کلمات می‌باشد. توجه داشته باشید که الگوریتم فوق فقط مراحل کار را نشان داده است و نیاز به بازنویسی درست، رعایت ابعاد تنسورها و غیره دارد که بر عهده شما میباشد.

پرسش‌ها

۱. از یک مدل از پیش‌آموزش دیده Resnet18 به عنوان شبکه CNN استفاده نمایید و به جز لایه خطی آخر تمامی لایه‌های آن را Freeze نمایید تا در عملیات بروزرسانی وزن‌ها شرکت نداشته باشند. سپس خروجی آن را در کنار

بردارهای Embed شده جملات به یک لایه شبکه LSTM یک طرفه اعمال کرده و نمودار خطای آموزش و تست را در طول یادگیری گزارش نمایید. از تابع خطای CrossEntropy و تابع بهینه‌ساز Adam می‌توانید استفاده نمایید. بعد از فرآیند آموزش،۳عدد عکس از دادگان تست را جهت پیشبینی مدل، به آن اعمال کرده و خروجی آن را در گزارش کار خود ذکر نمایید.

۲. با حفظ موارد گفته شده سؤال قبل تمامی لایه‌های شبکه Resnet18 را Unfreeze نمایید و مجددا موارد خواسته شده در سوال قبل را بررسی نمایید و نتایج بدست آمده را با سؤال قبل مقایسه کنید.