|  |  |
| --- | --- |
| نام و نام خانوادگی | **پویا شیخ الاسلامی** |
| شماره دانشجویی | **810100394** |
| تاریخ ارسال گزارش | **۱۴۰2.۰3.۰3** |

|  |  |  |
| --- | --- | --- |
|  | **به نام خدا**  **دانشگاه تهران**  **دانشکده‌ مهندسی برق و کامپیوتر** |  |
| **درس شبکه‌های عصبی و یادگیری عمیق**  **تمرین چهارم** | | |

**فهرست**

[**پاسخ 1**. **عنوان پرسش اول به فارسی** 1](#_Toc115453335)

[۱-۱. عنوان بخش اول 1](#_Toc115453336)

[**پاسخ ۲** **- عنوان پرسش دوم به فارسی** 2](#_Toc115453337)

[۱-۲. **عنوان بخش اول** 2](#_Toc115453338)

[**پاسخ ۳** **– عنوان پرسش سوم به فارسی** 3](#_Toc115453339)

[۱-۳. **عنوان بخش اول** 3](#_Toc115453340)

[**پاسخ ۴** **– عنوان پرسش چهارم به فارسی** 4](#_Toc115453341)

[۱-۴. **عنوان بخش اول** 4](#_Toc115453342)

**شکل‌ها**

شکل 1: تصویر پوشه دادگان 1

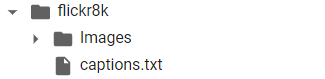
**جدول‌ها**

[جدول 1. عنوان جدول نمونه 1](#_Toc115453324)

# **پاسخ 1**. **توصیف عکس**

۱-۱. مجموعه دادگان و پیش پردازش آن ها

با توجه به محدودیت های سخت افزاری و امکانات موجود از محیط google colab برای آموزش و اجرای کد ها استفاده شده است. برای دانلود کردن دیتاست از روی سایت Kaggle به جای دانلود کردن دیتاست و آپلود آن به صورت دستی، این کار به صورت اتوماتیک در کد برنامه قرار داده شد. بدین صورت که یک فایل json از مشخصات حساب kaggle به گوگل داده شده و با توجه به آن و استفاده از API سایت، دیتاست دانلود و در پوشه مناسب خود قرار داده می‌شود. در نهایت فایل های دانلود شده به شکل زیر خواهد بود:



شکل 1: تصویر پوشه دادگان

در این پوشه یک فایل به نام captions.txt موجود است که در آن 5 جمله به ازای هر یک از عکس های موجود در دادگان موجود است. در پوشه Images نیز تمامی عکس ها موجود هستند.

در کلاس Vocabulary که برای تبدیل جملات به توکن تولید شده روند طی شده به این صورت است:

برای تبدیل جملات به توکن های مختلف از کتابخانه SPACY استفاده شده است و جمله در آن به توکن های مختلف تبدیل شده و در نهایت به حروف کوچک تبدیل شده و پس از آن دیکشنری های تبدیل کلمه به ایندکس و ایندکس به کلمه ساخته می‌شوند. در نهایت تابع numericalize برای تبدیل یک جمله به عدد ها، در آن تعریف شده است.

در کلاس FlickrDataset که برای خواندن و پردازش دادگان تعریف شده است، ابتدا داده ها خوانده، سپس دیتافریم آن ساخته شده و برای هرکدام از عکس ها، جمله متناظر آن ذخیره می‌شود.

در تابع get\_loader داده های مورد نیاز برای آموزش دادن شبکه عصبی ساخته و اماده می‌شوند.

2-۱. مدل شبکه

مدل شبکه عصبی همانطور که در داکیومنت توضیح داده شد از دو قسمت شبکه ResNet18 و LSTM تشکیل شده است که بخش اول برای استخراج ویژگی از تصویر استفاده می‌شود و بخش دوم برای ترکیب ویژگی های تصویر با ویژگی های متن توصیف تصویر و یادگیری آن استفاده می‌شود.

پس از آموزش دادن شبکه با استفاده این روش دقت کسب شده برای تصویر به اندازه کافی مناسب نبود و پس از 30 ایپاک آموزش دادن شبکه، متن تولید شده همچنان دارای دقت بالایی نبود. مقدار loss شبکه در این حالت در 10 ایپاک آخر به شکل زیر کاهش پیدا کرد(شروع خطای شبکه از 3.14 بود):

شکل 2: میزان خطای شبکه در 10 ایپاک اخر

و در نهایت نمونه خروجی آن به صورت زیر بود:



شکل 3: نمونه خروجی مدل اول

برای بهبود دقت شبکه طبق مطالعات انجام شده از شبکه های attention در بین لایه های مدل استفاده شد تا دقت را بهبود دهد.

در کد پیاده سازی شده در قسمت EncoderCNN یک مدل ResNet18 بدون در نظر گرفتن لایه های تماما متصل انتهایی آن برای انکود کردن ویژگی های تصویر استفاده شده است و لایه های آن در دو حالت freeze و unfreeze شده اند.

در قسمت دوم پیاده سازی مدل، بخش شبکه attention اضافه شده است. این قسمت به ترتیب شامل لایه های زیر است:

* انکودر
* دیکدر
* خطی ساز
* ReLU
* Softmax

و در نهایت قسمت نهایی LSTM که با دریافت پارامتر های ورودی توسط کاربر به شکل زیر ساخته شده است:

attention\_dim = 700

embed\_dim = 700

decoder\_dim = 700

dropout = 0.5

vocab\_size = len(dataset.vocab)

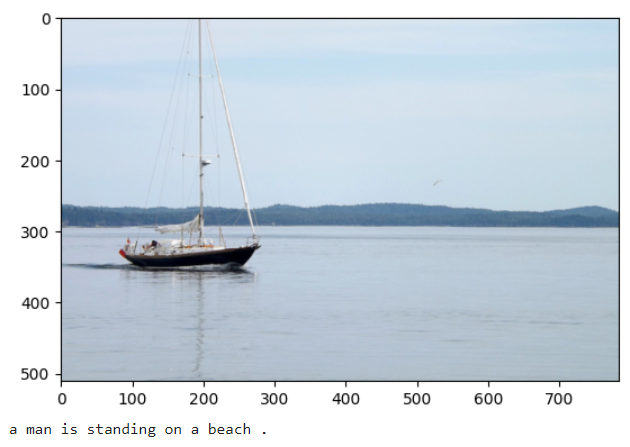
learning\_rate = 1e-03

num\_epochs = 2

3-۱. آموزش شبکه

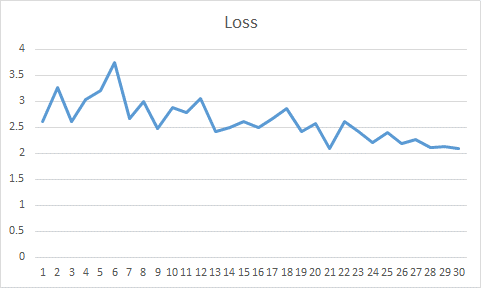
برای آموزش شبکه در google colab به دلیل محدودیت زمانی امکان آموزش دادن شبکه بیش از 30 ایپاک موجود نبود. آموزش به صورت دوره هایی با 2 ایپاک انجام می‌گرفت و هربار خروجی شبکه بررسی می‌شد. ولی شبکه تا 30 ایپاک خروجی هایی به شکل زیر داشت:

شکل 4: خطای شبکه

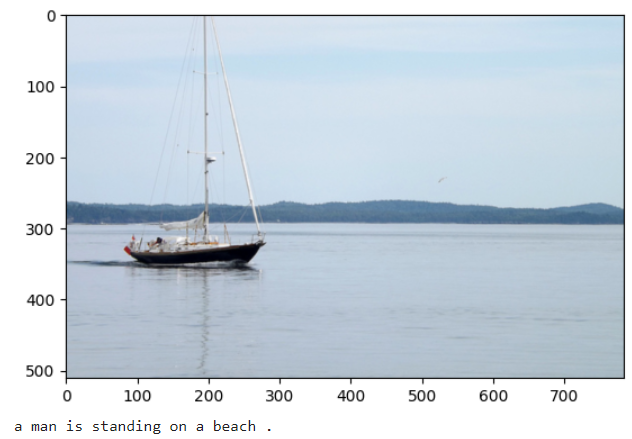


شکل 5: نمونه تصاویر خروجی شبکه و توضیحات آن ها

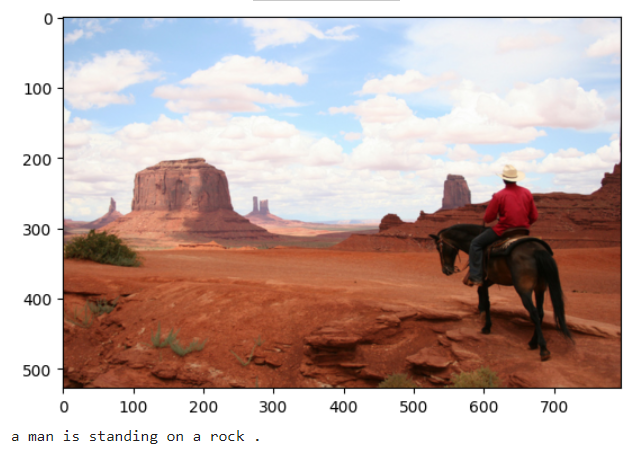
خروجی های بالا در صورتی بودند که وزن های شبکه ResNet18 ثابت باشند. در صورت اضافه شدن پارامتر های این شبکه به پارامتر های قابل آموزش، میزان خطای شبکه و خروجی به شکل زیر خواهد بود:



شکل 6: خطای شبکه با آموزش دادن شبکه resnet







شکل 7: نمونه خروجی های شبکه وقتی که resnet آموزش داده می‌شود

.

# **پاسخ ۲** **– تشخیص اندیشه**

## ۱-۲. **عنوان بخش اول**

متن نمونه

.

# **پاسخ ۳** **– عنوان پرسش سوم به فارسی**

## ۱-۳. **عنوان بخش اول**

متن نمونه

.

# **پاسخ ۴** **– عنوان پرسش چهارم به فارسی**

## ۱-۴. **پیش پردازش**

متن نمونه

## . **۲**-۴. **MLP**

لش