

مبانی بینایی کامپیوتر

تمرین یازدهم

نیکی نزاکتی

98522094

1.

(الف)

اینکه لایه های کانولوشنی پارامترهای کمتری نسبت به لایه های fully connected دارند به این دلیل است که در لایه کانولوشنی، هنگامی که هسته کانولوشن در سراسر تصویر اسلاید می شود، از پارامترها مجدداً استفاده می شود. Convolution ها به طور متراکم به هم متصل نیستند، همه node های ورودی بر همه node های خروجی تأثیر نمی گذارند. این به لایه های کانولوشن انعطاف بیشتری در یادگیری می دهد. علاوه بر این، تعداد وزن ها در هر لایه بسیار کمتر است، که به ورودی های با ابعاد بالا مانند داده های تصویر کمک زیادی می کند. در نتیجه لایه های کانولوشنی عملکرد بهتری دارند.

(ب)

$16 \times 16 \times 5 \rightarrow 16 \text{ filter } 5 \times 5 \rightarrow 12 \times 12 \times 16$

اگر بخواهیم طول و عرض خروجی این لایه تفاوتی نکند مقدار گسترش مرز باید ۴ باشد.

$20 \times 20 \times 5 \rightarrow 16 \text{ filter } 5 \times 5 \rightarrow 16 \times 16 \times 16$

تعداد پارامترهای این لایه:

$$((5 \times 5 \times 5) + 1) \times 8 = 1008$$

(پ)

$32 \times 32 \times 3 \rightarrow 3 \text{ filter } 5 \times 5 \rightarrow \text{output shape: } (28, 28, 3)$

$32 \times 32 \times 3 \rightarrow 9 \text{ filter } 3 \times 3 \rightarrow 30 \times 30 \times 9 \rightarrow 9 \text{ filter } 3 \times 3 \rightarrow \text{output shape: } (28, 28, 9)$

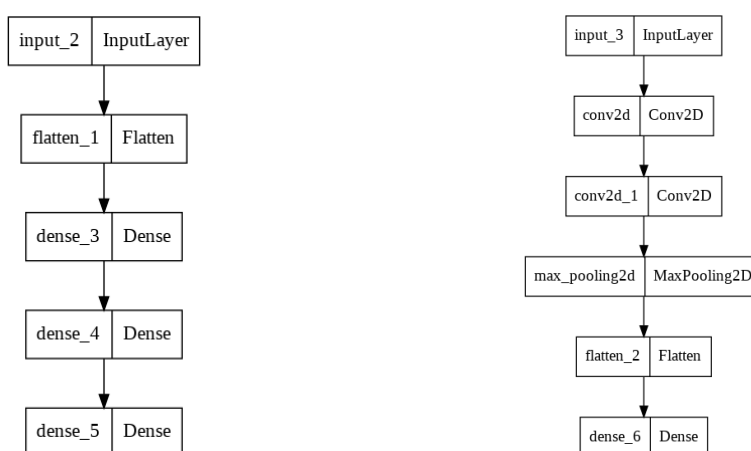
(ت)

Max Pooling پیکسل های روشن تر را از تصویر انتخاب می کند و این زمانی مفید است که پس زمینه تصویر تاریک است و ما فقط به پیکسل های روشن تر تصویر علاقه مندیم. مثلاً در مجموعه داده MNIST، ارقام به رنگ سفید و پس زمینه سیاه نشان داده می شوند. بنابراین از Max Pooling استفاده می کنیم. به طور مشابه، در موارد برعکس از Min Pooling استفاده می شود. روش Average pooling تصویر را صاف می کند و از این رو ممکن است هنگام استفاده از این روش ادغام، ویژگی های واضح شناسایی نشود. علاوه بر این از Global Average pooling میتوان برای کوچک کردن شبکه و جلوگیری از overfitting استفاده کرد.

ث) فیلترهای کانولوشنال در Resnet به آرامی ایجاد می شوند و در مقایسه با VGG از هسته های کمتری استفاده می کنند، اما تعداد بیشتری از آنها به طور متناوب بین عملیات کانولوشن و توابع فعال سازی غیرخطی روی هم چیده شده اند. VGG از ایده استفاده از شبکه های عمیق تر استفاده کرد و در مقابل، Resnet از ایده لایه کانولوشن 1x1 استفاده کرد تا اجازه دهد عمق کانال را قبل از اعمال کانولوشن 3x3 کاهش داده و بعد از آن دوباره آن را افزایش دهد.

2.

مدل های ما به صورت زیر می باشند:



(الف)

میزان خطا و دقت روی داده های Test برای مدل fully connected :

loss: 1.6915 - accuracy: 0.3905

میزان خطا و دقت روی داده های Test برای مدل Convolutional :

loss: 1.2823 - accuracy: 0.6227

در این مورد میبینیم که خطای کمتر به معنای دقت بیشتر است.

ب) مدت زمان اجرای هر اپیک در مدل fully connected حدودا 5 ثانیه و در مدل Convolutional حدودا 6 ثانیه است.

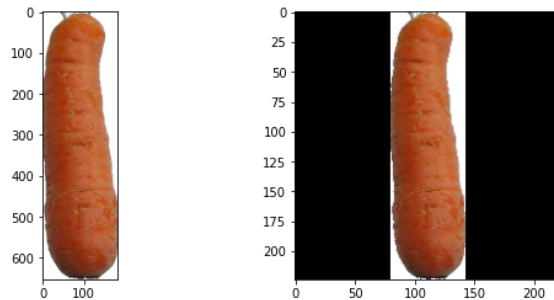
پ) همانطور که میبینیم با اینکه تعداد پارامترهای مدل fully connected تقریباً ۲ برابر مدل Convolutional است، مدت زمان انجام هر اپیاک در مدل fully connected کمتر است. پس ارتباط مستقیم بین زمان و پارامترهای مدل ها وجود ندارد و این زمان بیشتر میتوان به محاسبات بیشتر لایه های کانولوشنی مرتبط باشد.

3.

(الف)

برای resize کردن تصویر بدون اینکه ساختار آن عوض شود به صورت زیر عمل می کنیم:

```
new_img=tf.image.resize_with_pad(img,desired_size,desired_size,method=tf.image.ResizeMethod.BILINEAR,antialias=False)
new_img /= 255
new_img = np.array(new_img,np.float64)
```



(ب)

عملکرد آموزش مدل Resnet بدون فریز کردن لایه ها و وزنه های رندوم به روی دیتاست میوه ها:

loss: 11.1204 - acc: 0.0489

پ) عملکرد آموزش مدل Resnet آموزش دیده به روی Imagenet :

loss: 3.7032 - acc: 0.0522

ت) نتایج حاصل از evaluate کردن این دو مدل:

resnet: loss: 27655.7559 - acc: 0.0260

fine_tune_resnet : loss: 3.1712 - acc: 0.0502

همانطور که مشاهده می شود عملکرد مدلی که از Resnet همراه با وزن های آن استفاده می کند هم در آموزش و هم در evaluate بهتر است. این به این دلیل است که وزن های آموزش دیده Resnet بر روی Imagenet ، از قبل دارای نمایش هایی برای شناسایی ویژگی های مهم مورد نیاز برای تشخیص تصویر است و به نقطه ای بهتری به نسبت زمانی که از وزن های تصادفی استفاده می کردیم همگرا می شود.

Resources:

<https://www.geeksforgeeks.org/>

<https://stackoverflow.com/>

<https://machinelearningmastery.com/>

<https://towardsdatascience.com/>