

سهیل صالحی

بہمن 1401

—

فهرست سوالات

3 مقدمه
3 بررسی شبکه MobileNet
5 Visualize
8 Conclusion
9 مراجع

مقدمه

در این پروژه قصد داریم تا feature map یک شبکه CNN را به صورت عکس مشاهده کنیم. در واقع میخواهیم خروجی لایه های این شبکه را مشاهده کنیم که چگونه به پردازش عکس داده شده می پردازد. Feature map، خروجی هر لایه در یک شبکه عصبی عمیق هستند و ویژگی های آموخته شده داده های ورودی را نشان می دهند. با مشاهده feature map یک مدل، می توانیم بینشی در مورد اینکه مدل چگونه داده های ورودی را پردازش می کند و چه ویژگی هایی را یاد می گیرد، به دست آوریم. مدل MobileNet یک شبکه عصبی CNN است که برای دستگاه های تلفن همراه طراحی شده است. این مدل تعداد پارامترهای کمی دارد و عمق شبکه آن نسبت به سایر شبکه ها (به جز VGG) کمتر می باشد و برای برنامه های کم مصرف بهینه شده است.

بررسی شبکه MobileNet

MobileNet نوعی از شبکه های CNN می باشد که مختص تلفن های همراه و دستگاه هایی که منابع محاسبات ضعیف و یا محدودی دارند طراحی شده است.

این مدل برای برنامه های کم مصرف بهینه سازی شده است و این کار را با استفاده از نوع خاصی از لایه های Convolutional به نام depthwise convolutional انجام می دهد. این لایه ها، عملیات convolutional را به دو بخش تقسیم می کند: بخش depthwise که یک فیلتر واحد را برای هر کانال ورودی اعمال می کند و بخش pointwise که خروجی بخش depthwise را با استفاده از یک convolutional 1×1 ترکیب می کند.

این جداسازی باعث می شود تا تعداد پارامترهای مدل کاهش یابد و محاسبات مورد نیاز نیز کاهش میابد اما عملکرد آن بدون تغییر باقی می ماند. در کنار این موضوع، این مدل با استفاده از یک convolutional 1×1 بعد از بخش depthwise، باعث می شود تا تعداد کانال ها افزایش یابد که این کار باعث حفظ عملکرد شبکه حین کاهش پارامترهای آن می شود.

نسخه های متعددی از این نوع شبکه وجود دارد که در عمق، عرض و وضوح ورودی با یکدیگر متفاوتند. نسخه استفاده شده در این پروژه، دارای 28 لایه می باشد و وضوح ورودی آن نیز برابر با 224×224 می باشد.

Visualize

برای مشاهده خروجی لایه هایی که feature map را به ما می‌دهند، نیاز است تا یک مدل ساده تری به دست آوریم به طوریکه لایه های ورودی آن برابر اب لایه های ورودی همان مدل MobileNet باشد اما خروجی آن، برابر با لایه هایی باشد که feature map را در مدل MobileNet را به ما می‌دهد.

این کار باعث می‌شود تا تعداد لایه های مدل کاهش یابد اما وزن های هر لایه همان وزن های از پیش آموزش داده شبکه MobileNet می‌باشند که باعث می‌شود همان ویژگی هایی که در آن شبکه خروجی می‌دهند را نیز در این مدل ساده شده نیز خروجی دهند.

```
layers = ['conv_dw_1', 'conv_pw_1', 'conv_dw_2', 'conv_pw_2',  
          'conv_dw_3', 'conv_pw_3', 'conv_dw_4', 'conv_pw_4',  
          'conv_dw_5', 'conv_pw_5', 'conv_dw_6', 'conv_pw_6',  
          'conv_dw_7', 'conv_pw_7', 'conv_dw_8', 'conv_pw_8',  
          'conv_dw_9', 'conv_pw_9', 'conv_dw_10', 'conv_pw_10',  
          'conv_dw_11', 'conv_pw_11', 'conv_dw_12', 'conv_pw_12',  
          'conv_dw_13', 'conv_pw_13']  
outputs = [model.get_layer(i).output for i in layers]
```

شکل 3- لایه هایی با خروجی feature map در MobileNet

با استفاده از شکل 3، میتوانیم خروجی مدل ساده شده را لایه هایی قرار دهیم که به ما feature map ها را می‌دهد.

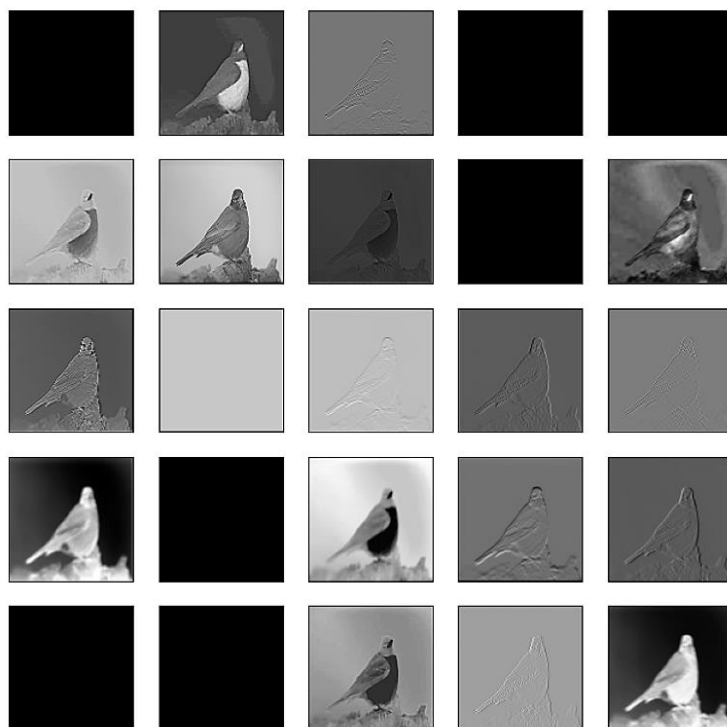
برای بررسی این لایه ها، از تصویر زیر استفاده می‌کنیم:



شکل 4- تصویر استفاده شده برای بررسی مدل

با بررسی لایه ها، خروجی به صورت زیر خواهد بود:

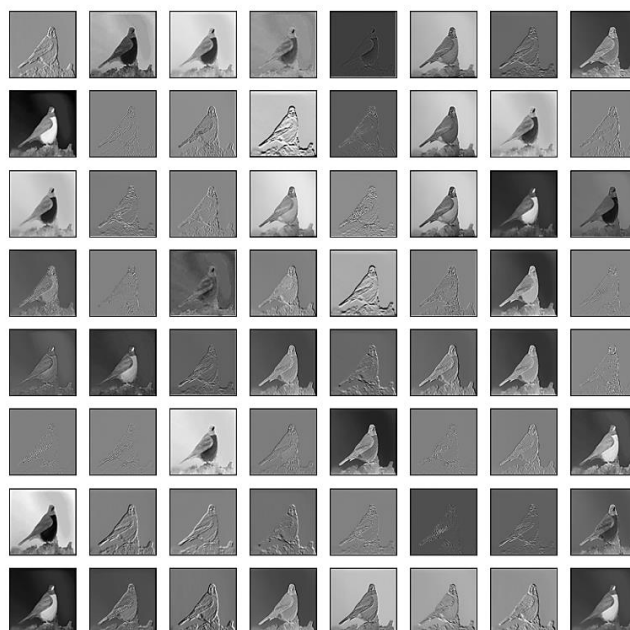
conv_dw_1 Visualization



شکل 5- خروجی لایه اول conv_dw

می‌توان دید که در این لایه، ویژگی‌های سطحی مانند لبه‌ها استخراج شده است و در برخی تنها قسمت‌های background مشخص شده است.

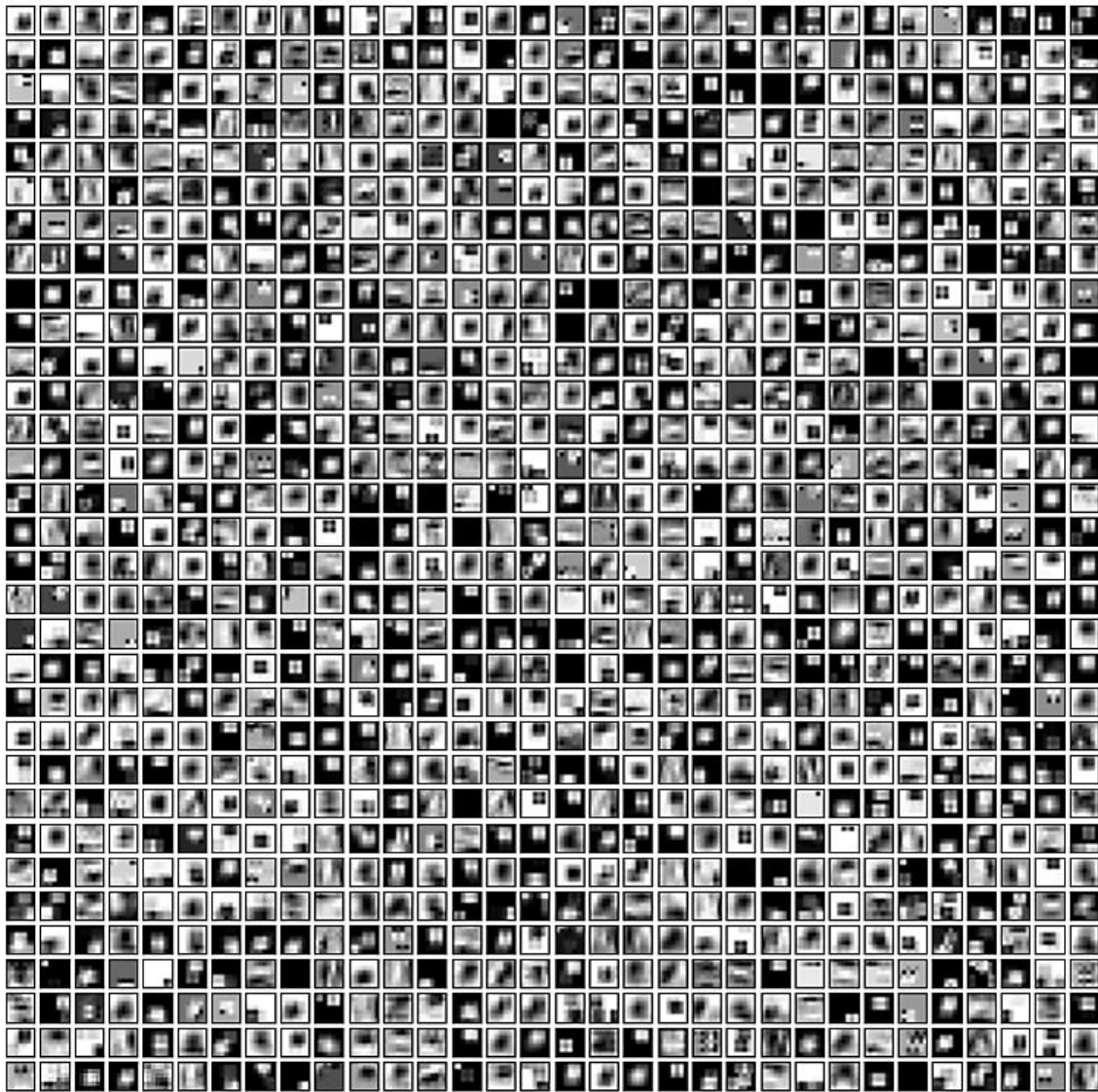
conv_pw_1 Visualization



شکل 6- خروجی لایه اول conv_pw

می‌توان دید که در این لایه، ویژگی‌ها همانند لایه conv_dw می‌باشد اما پیچیده‌تر شده‌اند.
با جلو رفتن هر لایه، ویژگی‌های استخراج شده از تصاویر بیشتر می‌شود به طوریکه برای لایه آخر داریم:

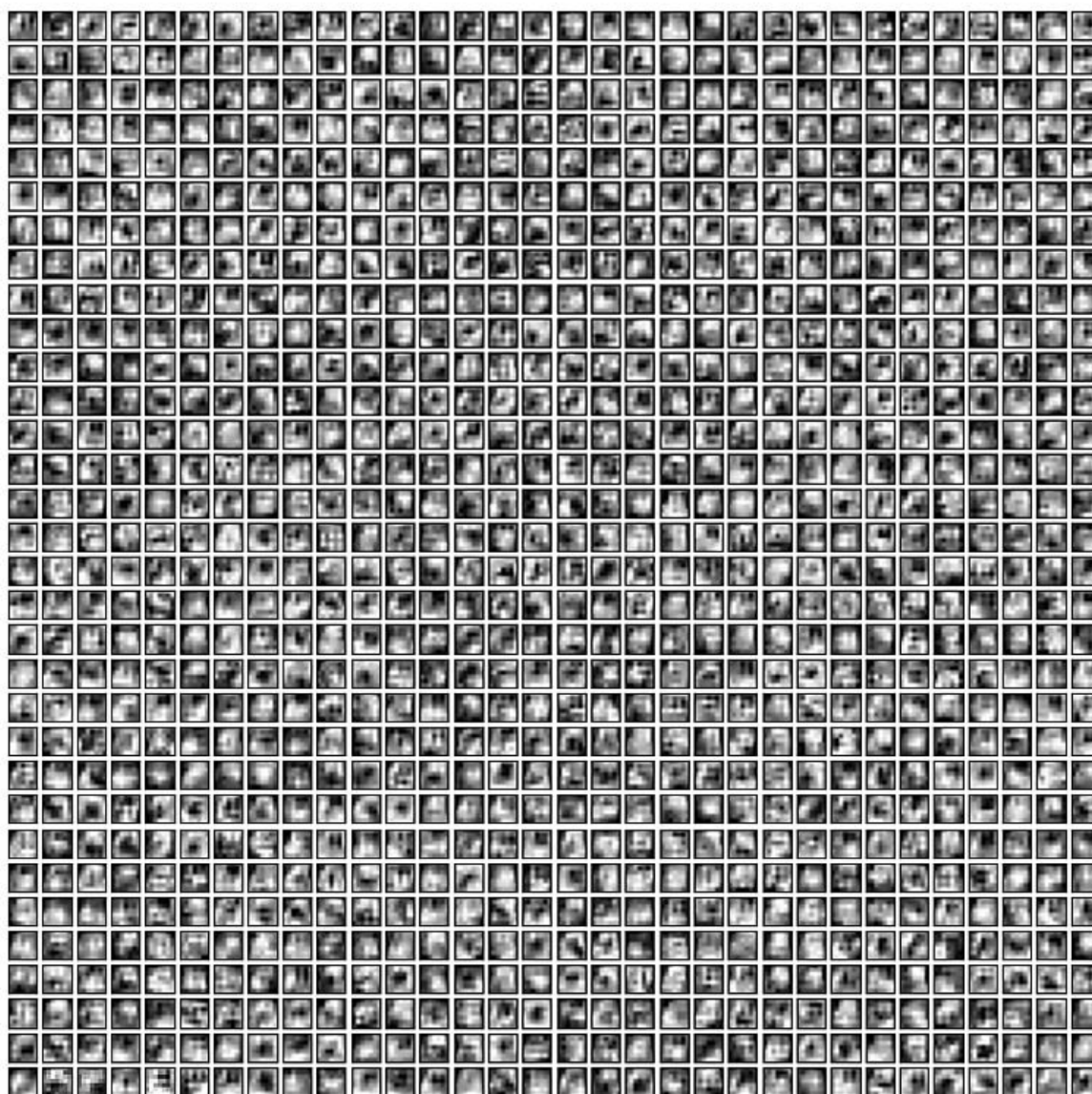
conv_dw_13 Visualization



شکل 7- خروجی لایه 13م conv_dw

می‌توان دید که در این لایه، ویژگی‌ها بسیار دقیق‌تر شده‌اند به طوریکه ویژگی‌هایی همانند لبه‌های
بال، منقار و موقعیت هر کدام از اجزای پرند نیز استخراج شده است.

conv_pw_13 Visualization



شکل 8-خروجی لایه 13 ام conv_pw

در خروجی conv_pw می‌توان دید که تمامی ویژگی‌های ریز تصویر به دست آمده است.

Conclusion

در این پروژه دریافتیم که خروجی لایه‌هایی که feature map ها را به دست می‌آورند، در هر لایه ویژگی‌های خاصی از تصویر را استخراج می‌کنند که این ویژگی‌ها در ابتدا ویژگی‌های سطحی می‌باشد و رفته رفته این ویژگی‌ها پیچیده تر می‌شود که باعث می‌شود تا شبکه بتواند با بررسی این ویژگی‌ها، نوع تصویر ورودی را پیش بینی کند.

در این پروژه از مدل MobileNet استفاده شد، بدیهی است که اگر از مدل های پیچیده تری استفاده شود (عمق بیشتر)، ویژگی های استخراج شده از تصویر بیشتر شده و ویژگی ها دقیق تر می باشد که منجر به آن می شود تا شبکه پیش بینی دقیق تری نسبت به تصویر ورودی داشته باشد.

مراجع

<https://machinelearningmastery.com/how-to-visualize-filters-and-feature-maps-in-convolutional-neural-networks/>

<https://www.kaggle.com/code/arpitjain007/guide-to-visualize-filters-and-feature-maps-in-cnn>

<https://keras.io/api/applications/>