

In the name of god

هر کس به زبانی تو را حمد گوید

به نام خدا

امیر شکری، فرشاد اصغر زاده همپا، علیرضا غلام نیا

استاد راهنما: دکتر فدائی اسلام

تمرین شماره ۴ – بینایی کامپیوتر

از معماری های

- AlexNet
- VGGNet16
- GoogLeNet
- ResNet18
- DenseNet121

در مجموع پنج سوال به همراه جواب طرح نمایید.

به سوالات مفهومی که تکرار کمتری (در میان سوالات طرح شده) داشته باشند نمره بیشتری تعلق خواهد گرفت.

هدف: یادگیری دقیق این معماری ها است.

حل به صورت گروهی

سوال ۱ :

نکات مهمی که معماری alexNet را نسبت به باقی معماری های cnn برجسته تر کرده است چیست ؟

پاسخ سوال ۱ :

- در معماری alexNet به جای توابع سیگموئیدی یا هایپربولیک از توابع فعالسازی ReLU استفاده می شود که سرعت ReLU در برابر دقت مورد نظر ۵ برابر بیشتر و بهتر می باشد.
- برای جلوگیری و مقابله از overfitting به جای regularization از Dropout استفاده می شود اما نکته اینجاست که زمان آموزش با نسبت ۰/۵ دوبرابر بیشتر می شود.
- برای داده های بزرگ از ۷ لایه با 650k units و ۶۰ میلیون پارامتر استفاده می شود.

سوال ۲ :

وقتی ResNet لایه های عمیق تری نسبت به VGG دارد پس چرا پیچیدگی زمانی VGG از ResNet بالاتر است ؟

پاسخ سوال ۲ :

VGG16 تقریباً ۱۳۸ میلیون پارامتر دارد و ResNet فقط ۲۵/۵ میلیون پارامتر دارد که این نمی تواند حاکی از سرعت بیشتر آن باشد و این مورد را باید دقت کنید که تعداد پارامترها فضای مورد نیاز برای ذخیره سازی در شبکه را کاهش می دهد اما باز هم دلیل سرعت بیشتر ResNet نسبت به VGG نیست.

این تفاوت سرعت محاسبات به شدت به نوع اجرا بستگی دارد؛ می توان از شمارش FLOPS ها در عملیات های فعالسازی و لایه ی pooling صرف نظر کرد زیرا هزینه ی آنها نسبتاً کم است.

اول از همه سرعت convolution به اندازه ی ورودی بستگی دارد.

در آخرین مرحله هر بلاگ طول و عرض با دو فاکتور کاهش پیدا می کند به طوری که در دو لایه ی اول ResNet موفق با کاهش height و width می شود و عرض تصویر به ضریب ۴ تبدیل می شود.

در CGG دو لایه ی اول از از همبستگی کامل استفاده می کند که عمل بسیار گرانی به حساب می آید؛ برای جلوگیری از مشکل محاسباتی در ResNet در لایه ی اول به این مسائل می پردازیم و در آن width و height را با یک ضریب کاهش می دهیم و حداکثر عمل pooling بعدی را با فاکتور همان ضریب کاهش می دهیم.

فیلترهای کانولوشن در ResNet به آرامی جمع می شوند و در ResNet با توجه به مسائل و شواهد kernel کمتری نسبت به VGG استفاده می شود اما تعداد بیشتری convolution و active function غیر خطی وجود دارد و دلیل آن هم این است که از شبکه های کوچک تری با عمق زیاد استفاده می کنند.

لایه های بعدی ResNet هم مثل قبل از همان استراتژی پیروی می کنند و سعی می کنند کوچک تر شوند و عمق را بیشتر کنند؛ در حالی که از شبکه های با تعداد فیلتر بیشتر در لایه های convolution استفاده شده ولی این امر به منظور کاهش بیشتر عملیات انجام شده است.

در مجموع ایده ی استفاده از لایه های convolution با ابعاد ۱ در ۱ این اجازه را می دهد که عمق کانال را قبل از استفاده از convolution با ابعاد ۳ در ۳ کاهش دهید و بعد دوباره آنرا به مقادیر بعدی تبدیل کنید.

سوال ۳ :

چرا می توانیم لایه های fully connected را در VGGNet حذف کنیم؟

پاسخ سوال ۳ :

یک نکته ی مهم این است که در VGGNet نمی توان هر دو لایه ی fully connected را حذف کرد و فقط اولین لایه ی fully connected می توان حذف کرد و دلیل آن این است که می توانید یک مدل با ظرفیت کمتر ایجاد کنید.

این موارد با توجه به شواهد تجربی مبتنی بر معماری ها می باشد و حتی معماری هایی هستند که بدون دو لایه ی fully connected به خوبی کار می کنند.

سوال ۴ :

چرا softmax چندانکه وجود دارد؟

پاسخ سوال ۴ :

معماری های عمیقی مثل GoogleNet ۲۲ لایه دارند که در معرض خطر از بیت رفات شیب ها در خین فرآیند training هستند. با توجه به این مورد لایه های میانی این مسئله را حل می کنند به طوری که ضرر کلی تشکیل شده از ضرر میانی و ضرر نهایی می باشد به همین دلیل است که در مجموع به جای یک لایه برای آخرین لایه ی شبکه در کل سه لایه ضرر وجود دارد.

سوال ۵ :

در GoogleNet در گره ورودی در هنگام عبور روبه عقب یا backwards چه اتفاقی رخ می دهد؟

پاسخ سوال ۵ :

جمع اتفاقی می افتد؛ رفتار محاسبه ی backward pass ها به دلیل chain rule در هر گره مستقل است بنابراین حتی مهم نیست که بعدا یک اتصال وجود داشته باشد و هر زمان گره ها را به یک روش تقسیم می کنیم و gradients اضافه می شود.

منابع

1. <https://medium.com/@RaghavPrabhu/cnn-architectures-lenet-alexnet-vgg-googlenet-and-resnet-7c81c017b848>
2. <https://stats.stackexchange.com/questions/280179/why-is-resnet-faster-than-vgg>
3. <https://stats.stackexchange.com/questions/403863/why-can-we-remove-fc-layers-in-vggnet>
4. <https://stats.stackexchange.com/questions/274286/google-inception-modelwhy-there-is-multiple-softmax/274623#274623>
5. <https://stats.stackexchange.com/questions/473256/what-happens-at-the-input-node-in-an-inception-module-during-the-backwards-pass>