به نام خدا



دانشگاه صنعتی امیرکبیر

(پلی‌تکنیک تهران)

درس شبکه‌های عصبی

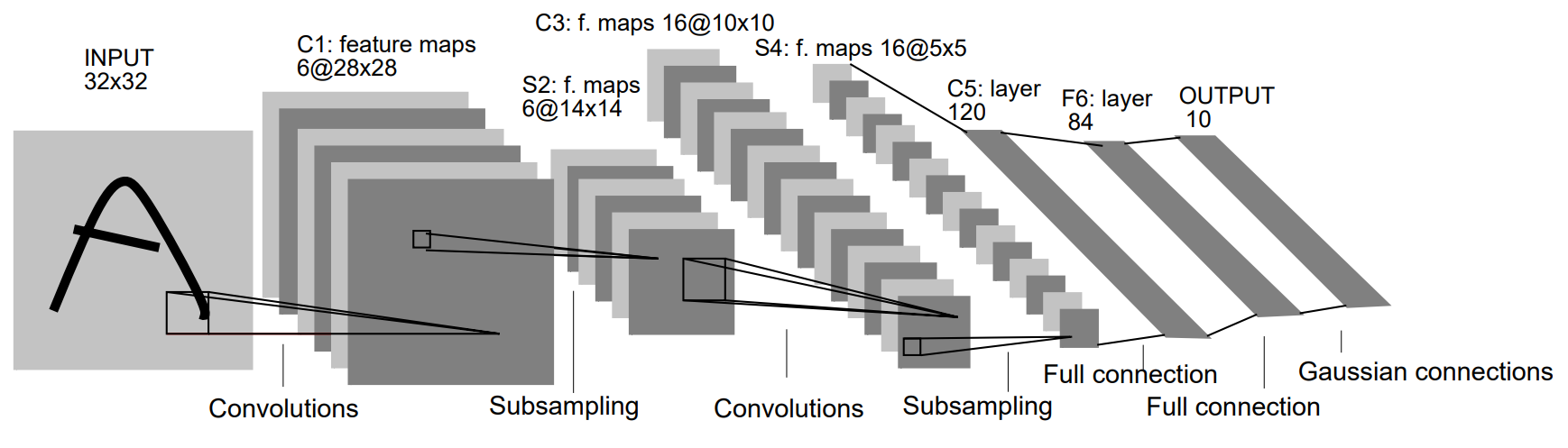
استاد صفابخش

تمرین چهارم

علیرضا مازوچی

۴۰۰۱۳۱۰۷۵

سوال ۱[[1]](#footnote-1)[[2]](#footnote-2)

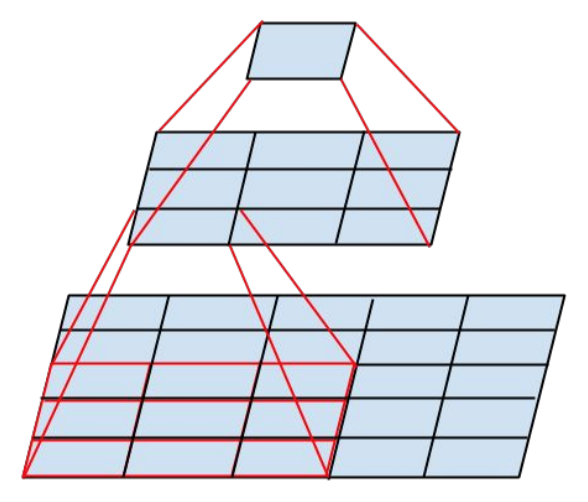
شبکه LeNET-5 دارای هفت لایه است: سه لایه کانوولوشنی، دو لایه نمونه‌برداری (Subsampling) و دو لایه از نوع تماما متصل (Fully Connected). این معماری در تصویر زیر قابل مشاهده است:

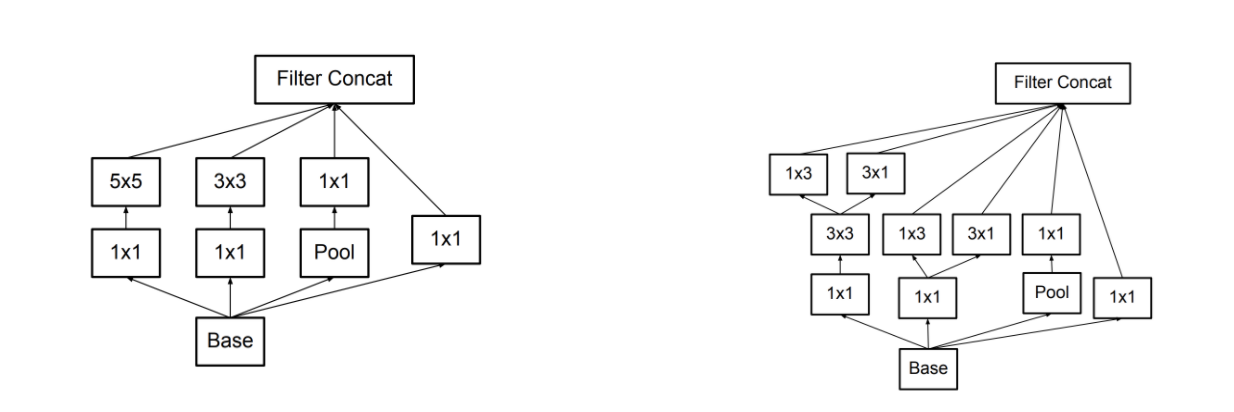
در معماری پیشنهادی آن‌ها ابعاد ورودی ۳۲×۳۲ است. کرنل استفاده‌شده در کلیه‌ی لایه‌های کانوولوشنی ۵×۵ به همراه تابع فعال‌سازی تانژانت هایپربولیک و در لایه‌های نمونه‌برداری ۲×۲ با یک تابع فعال‌سازی سیگموید است پس از اولین لایه کانوولوشنی تعداد ۶ نقشه ویژگی با ابعاد ۲۸×۲۸ تشکیل می‌شود. لایه نمونه‌برداری بعد از آن ابعاد هر نقشه ویژگی را به ۱۴×۱۴ کاهش می‌دهد. دومین لایه کانوولوشنی ۱۶ نقشه ویژگی با ابعاد ۱۰×۱۰ ایجاد می‌کند که لایه نمونه‌برداری بعد از آن ابعاد را به ۵×۵ کاهش می‌دهد. نهایتا آخرین لایه کانوولوشنی ۱۲۰ نقشه ویژگی با ابعاد ۱×۱ می‌سازد. در این مرحله عملا داده ورودی به یک داده یک بعدی با ۱۲۰ ویژگی تبدیل می‌شود. سپس با دو لایه تماما متصل ابعاد به ترتیب به ۸۴ و ۱۰ کاهش پیدا می‌کند.

شبکه‌ای که آن‌ها پیشنهاد داده‌اند برای تشخیص اعداد دست‌نویس انگلیسی است و لذا خروجی نهایی مشخص می‌کند که داده ورودی با چه احتمالی به کدام کلاس متعلق است. همچنین داده ورودی در اصل از نوع ۲۸×۲۸ است که با حاشیه‌گذاری (Padding) به ۳۲×۳۲ تبدیل می‌شود.

شبکه Inception-v3 تمرکز ویژه‌ای روی کاهش بار محساباتی داشته است. برای این امر از پنج تکنیک مهم استفاده کرده است که در ادامه آن را بررسی می‌کنم:

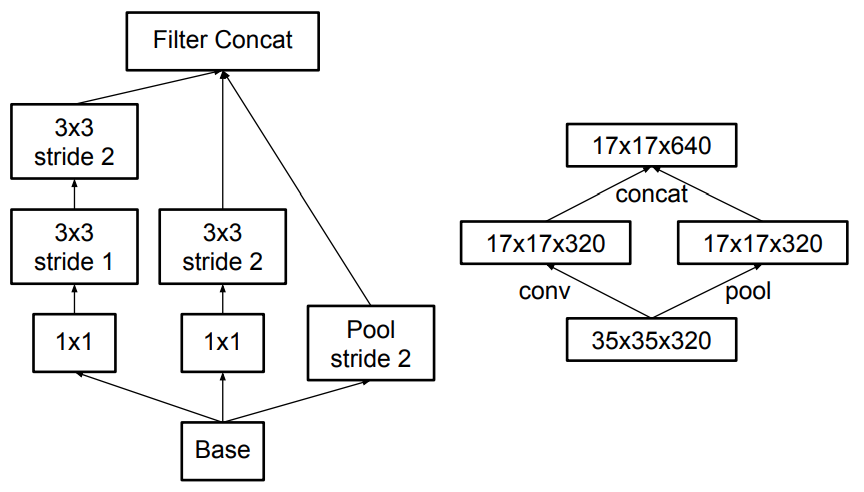
۱) کانوولوشن‌های تجزیه‌شده (Factorized Convolutions): با تجزیه کردن کانوولوشن‌ها پارامتر‌ها کاهش می‌یابد و کارایی حفظ می‌شود.

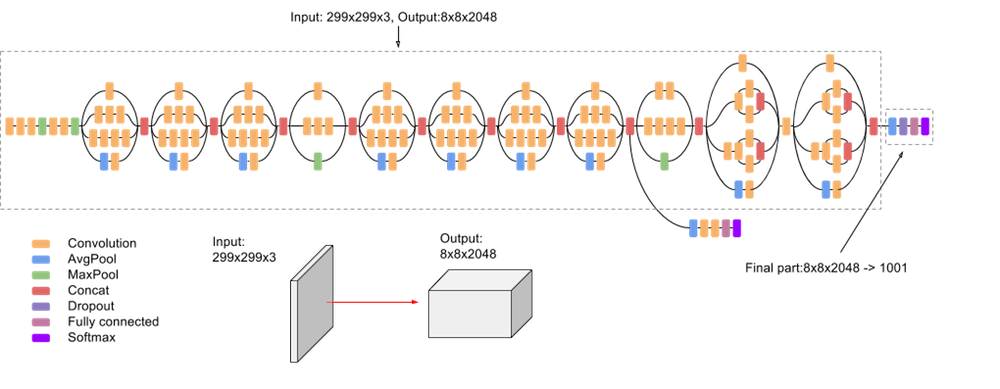
۲) کانوولوشن‌های کوچک‌تر: استفاده از کانوولوشن‌های کوچکتر تعداد پارامتر کمتری دارد و محاسبات کمتری را رقم خواهد زد. در LeNET-5 از کانوولوشن‌های ۵×۵ استفاده شده است که ۲۵ پارامتر دارد. در Inception-v3 با جایگزین کردن دو کانولوشن ۳×۳ با حفظ کارایی تعداد پرامتر‌ها را به ۱۸ رسانده‌اند. در تصویر روبرو می‌توان دید که چگونه یک کانوولوشن ۵×۵ با دو مرحله کانوولوشن ۳×۳ جایگزین شده است.

۳) کانوولوشن‌های نامتقارن (Asymmetric Convolutions): در Inception-v3 از کانوولوشن‌های نامتقارن کمک گرفته می‌شود که زمان آموزش را کاهش می‌دهد. یک کانوولوشن ۳×۳ را اگر بخواهیم با تکنیک کانوولوشن‌های کوچک‌تر بشکنیم نیاز به دو کانوولوشن ۲×۲ خواهیم داشت که چندان کاهش پارامتر ندارد ولی اگر از یک کانوولوشن ۱×۳ و یک کانوولوشن ۳×۱ استفاده کنیم مفید است. معماری تصویر راست جایگزین‌شده معماری تصویر چپ با بهره‌گیری از این تکنیک و تکنیک قبلی است.

۴) دسته‌بند کمکی: دسته‌بند کمکی یک شبکه کانوولوشنی کوچک است که در میان لایه‌های شبکه اصلی در حین آموزش قرار می‌گیرد. تابع هزینه در زمان آموزش شامل تابع هزینه این زیرشبکه‌های کمکی هم می‌شود. این دسته‌بند کمکی در نقش یک منظم‌ساز برای شبکه عمل می‌کند. بدیهی است که چنین چیزی در LeNET-5 وجود نداشته است.

۵) کاهش ابعاد (Grid Size Reduction) بهینه: در Inception-v3 برای کاهش ابعاد ورودی از نحوه دیگری از ترکیب لایه‌ها استفاده کرده است که در مجموع تعداد محاسبات کمتری را خواهد داشت. در تصویر زیر معماری مربوط به این تکنیک آورده شده است.



در تصویر زیر معماری نهایی مربوط به Inception-v3 آورده شده است:

سوال ۲

1. <https://towardsdatascience.com/understanding-and-implementing-lenet-5-cnn-architecture-deep-learning-a2d531ebc342> [↑](#footnote-ref-1)
2. <https://blog.paperspace.com/popular-deep-learning-architectures-resnet-inceptionv3-squeezenet/> [↑](#footnote-ref-2)