به نام خدا

مبانی بینایی کامپیوتر

دکتر محمدی

تمرین شش

علی عطاریان – ۹۹۵۲۱۴۵۱

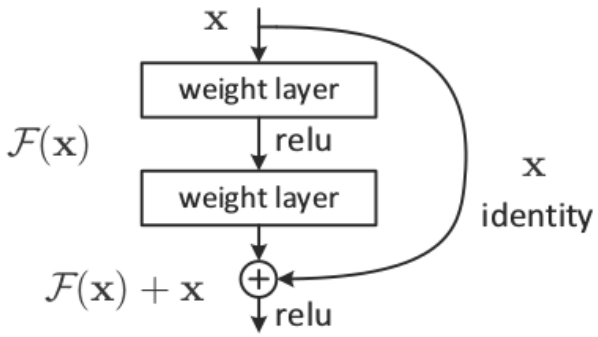
سوال ۱) الف)

هنگام استفاده از روش gradient learning با backpropagtion، برای بهبود وزن‌ها به ترتیب از لایه آخر تا لایه اولیه، با قاعده زنجیره‌ای، مشتق گرفته می‌شود. در n لایه n مشتق در یکدیگر ضرب می‌شوند، اگر این مشتق‌ها بزرگ باشد به exploding gradient و اگر مشتق ها کوچک باشند به vanishing gradient منجر می‌شود. در حالت انفجار گرادیان مدل بسیار ناپایدار می‌شود و یادگیری انجام نمی‌گیرد. همچنین در موارد حاد، به مشکل overflow برمی‌خوریم که وزن ها مقدار NaN می‌گیرند. در حالت ناپدیدشدن گرادیان چون مقدار مشتق‌ها بسیار کوچک است تغییر ملموسی در وزن‌ها بوجود نمی‌آید و باز هم مدل یاد نمی‌گیرد. در بدترین حالت مشتق صفر منجر به گرادیانت صفر می‌شود و وزن‌ها صفر می‌شوند.

چند راه حل برای این مشکل مطرح است: با کاهش تعداد لایه ها تعداد ضرب‌ها کم می‌شود و احتمال هر دو مشکل کاهش می‌یابد اما پیچیدگی شبکه کم می‌شود. همچنین با gradient clipping حد بالا و پایین برای گرادیان تعریف می‌کنیم. علاوه بر این دو روش با مقداردهی اولیه صحیح وزنها (استفاده از Xavier یا He) می‌توان از این مشکل جلوگیری کرد.

سوال ۱) ب)

شبکه‌های قبل ResNet با مشکل vanishing gradient مواجه بودند. ResNet با معرفی Residual blocks و روش skip connections این مشکل را حل کرد. البته باید توجه داشت استفاده از residual network تنها برای حل مشکل VGP نیست و یادگیری مدل را نیز بهبود می‌بخشد. یک residual block در واقع تشکیل شده از چندین لایه کانولوشنی و یک shortcut connection است. خروجی هر بلاک مجموع خروجی لایه‌های کانولوشنی و ورودی است.



در این حالت شبکه بر روی خروجی هر بلاک train می‌شود و این یعنی بسیاری از لایه ها skip می‌شوند و تعداد ضربها کم می‌شود. در واقع اضافه کردن ورودی به خروجی نهایی اجازه می‌دهد هنگام backpropagation گرادیان‌ها مستقیما از طریق shortcut connection (ididtity path) پروپگیت شوند. با این معماری می‌توانیم شبکه‌هایی با صدها لایه بسازیم.

سوال ۲) الف)

پارامترها: در هر کدام از لایه‌های Conv 1*1، برای هر کدام از ۳ کانال ۳۲ ضریب داریم که با احتساب bias می‌شود 128=32*(3+1). در مورد Conv 3*3 می‌توان گفت که ورودی همه آنها ۳۲ کاناله است و هر کانولوشن ۹ ضریب دارد که با احتساب bias می‌شود:
9248=32*(32*9+1). لایه آخر نیز یک ورودی ۹۶ کاناله دارد (با فرض کانکت کردن ۳تا خروجی ۳۲ کاناله) که با در نظر گرفتن بایاس می‌شود 24832=256*(96*1+1) پس جمعا داریم: 128+128+128+9248+9248+9248+24832=52960

RF: در هر ۳ لایه Conv 1*1 هر پیکسل معادل یک پیکسل است پس میدان تاثیر ۱ است. در دوتا لایه ابتدایی Conv 3*3 پنجره ۳*۳ داریم و هر پیکسل ورودی معادل ۱ پیکسل است پس میدان تاثیر 3*3 است. در لایه Conv 3*3 بعدی پنجره ۳*۳ است و هر پیکسل ورودی نماینده ۹ پیکسل تصویر اصلی است پس میدان تاثیر 5*5 است. در لایه آخر نیز پنجره ۱*۱ است و از بین ورودی‌های این لایه نیز بیشترین میدان تاثیر 5*5 است پس میدان تاثیر این لایه همان 5*5 است. برای کل شبکه نیز بیشترین RF لایه‌ها را در نظر می‌گیریم که 5*5 است.

سوال ۲) ب) A)

پارامتر: مشابه توضیحات بخش اول برای لایه اول داریم: 448=16*(9*3+1) و برای لایه دوم داریم 4640=32*(9*16+1) پس 448+4640=5088

RF: در لایه اول میدان تاثیر ۳*۳ و در لایه دوم ۵*۵ است و در کل ۵*۵ است.

سوال ۲) ب) B)

پارامتر: برای لایه اول داریم 16*(n-2)*(n-2)*(3*9+1)

و برای لایه دوم داریم که 32*(n-4)*(n-4)*(16*9+1)

RF: مانند مورد قبلی لایه اول ۳*۳ و لایه دوم ۵*۵ و کل شبکه ۵*۵ است.