



یادگیری ژرف

نیم سال دوم ۴۰۲-۱۴۰۱

مدرس: دکتر مهدیه سلیمانی

طرح تمرین: علی عبداللہی، سیدعلی یعقوب نژاد، زهرا رحیمی

تمرین سری دوم (۱۰۰ نمره + ۵ نمره امتیازی) نرمال سازی بچ و شبکه CNN زمان تحویل: ۲۶ اسفند (نظری)، ۲۸ اسفند (عملی)

لطفا نکات زیر را رعایت کنید:

- سوالات خود را از طریق پست مربوط به تمرین در Quera مطرح کنید.
- در هر کدام از سوالات، اگر از منابع خارجی استفاده کرده اید باید آن را ذکر کنید. در صورت همفکری با افراد دیگر هم باید نام ایشان را در سوال مورد نظر ذکر نمایید.
- پاسخ ارسالی واضح و خوانا باشد. در غیر این صورت ممکن است منجر به از دست دادن نمره شود.
- پاسخ ارسالی باید توسط خود شما نوشته شده باشد. به اسکرین شات از منابع یا پاسخ افراد دیگر نمره ای تعلق نمی گیرد.
- در صورتی که بخشی از سوال ها را جای دیگری آپلود کرده و لینک آن را قرار داده باشید، حتما باید تاریخ آپلود مشخص و قابل اعتنا باشد.
- تمام پاسخ های خود را در یک فایل با فرمت zip [Fullname]_[SID]_HW# روی کوثر قرار دهید.
- برای ارسال هر تمرین تا ساعت ۲۳:۵۹ روز ددلاین فرصت دارید. مهلت تاخیر (مجاز و غیر مجاز) برای این تمرین، ۱۰ روز است.

سوال ۱: (نظری) نرمال سازی بچ (۷ نمره + ۳ نمره امتیازی)

- (آ) استفاده از مقادیر بزرگ برای نرخ یادگیری در شبکه های ژرف سنتی می تواند منجر به انفجار یا محو شدن گرادیان شود. توضیح دهید چگونه نرمال سازی بچ به ما امکان استفاده از نرخ یادگیری بزرگتر از حد معمول را می دهد.
- (ب) زمانی که از نرمال سازی بچ استفاده می کنیم، در صورتی که وزن های شبکه در یک مقدار ثابت مانند α ضرب شوند، چه تاثیری بر گرادیان نسبت به وزن ها و گرادیان نسبت به سایر پارامترهای شبکه خواهد داشت؟
- (ج) یکی از اثرات جانبی نرمال سازی بچ منظم سازی^۱ است. با ذکر دلیل بیان کنید که بزرگ شدن ساینز بچ چه تاثیر بر ویژگی منظم سازی نرمال سازی بچ خواهد داشت.
- (د) (۳ نمره امتیازی) با مطالعه این مقاله نحوه انجام نرمال سازی بچ در شبکه های تماماً متصل و شبکه های پیچشی را با یکدیگر مقایسه نمایید.

سوال ۲: (نظری) کانولوشن گسترش یافته (۸ نمره)

در شبکه های پیچشی به صورت متداول از لایه های کانولوشن ساده استفاده می شود که با آن آشنا هستید. نوع دیگری از لایه ها که می توان از آنان در شبکه های پیچشی استفاده نمود، لایه های کانولوشن گسترش یافته^۲ است. در شکل ۱ یک تصویر شهودی از فیلتر کانولوشن گسترش یافته ارائه شده است؛ این فیلترها میان خانه هایی که فیلتر با استفاده از اطلاعات آن ها لایه بعد را محاسبه می کند، فاصله می اندازند یا به بیانی دیگر در زمان اعمال فیلتر و انجام عملیات ضرب کانولوشن، بر روی ورودی با طول گام^۳ بزرگتری حرکت می کنیم. توجه کنید که در اینجا طول گام مفهومی متفاوت نسبت به طول گام^۴ در لایه های شبکه کانولوشن دارد.

(آ) در مورد مزایا و معایب استفاده از فیلتر کانولوشن گسترش یافته تحقیق کنید و از هر کدام دو مورد ذکر نمایید.

(ب) در زمان استفاده از فیلتر کانولوشن گسترش یافته مفهومی به اسم محدوده دید^۵ معنای پررنگ تری پیدا می کند. فرض کنید ورودی $M \times N$ به شبکه ای شامل سه لایه کانولوشن گسترش یافته داده شده است؛ محدوده ای از ورودی که عنصر i, j نام خروجی مشاهده می کند را به صورت پارامتری مشخص کنید.

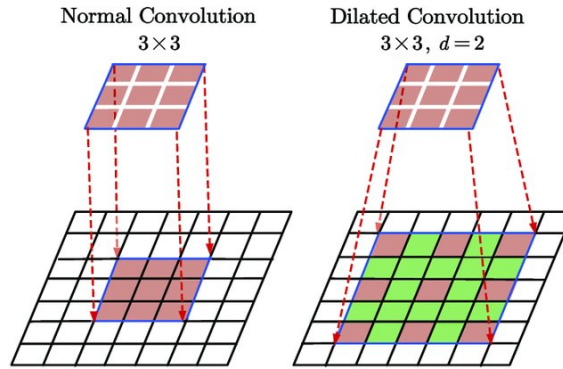
¹Regularization

²Dilated Convolution

³Step Size

⁴Stride

⁵Receptive Field



شکل ۱: فیلتر کانولوشن معمولی و کانولوشن گسترش یافته

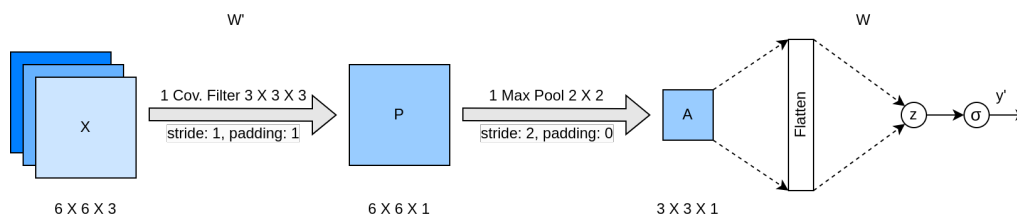
(ج) تصویری با ابعاد $I \times I \times 3$ به یک شبکه یک لایه با D فیلتر $K \times K \times 3$ با پارامتر گسترش N داده شده است. ابعاد خروجی این شبکه پس از اعمال بر ورودی را بر حسب I, D, K و N محاسبه کنید.

سوال ۳: (نظری) انتشار به عقب (۱۱ نمره)

یکی از اساسی‌ترین اجزا در یک شبکه ژرف، انتشار به عقب گرادیان^۶ در شبکه است. شبکه نشان داده شده در شکل ۲ را در نظر بگیرید. در این شکل وزن‌های لایه تماماً متصل تهایی و W' وزن‌های فیلتر کانولوشن لایه اول است. $P_{i,j}$ معرف عنصر سطر i ام و ستون j ام خروجی لایه کانولوشن و $W'_{i,j,k}$ بیانگر وزن در سطر i ام، ستون j ام و کانال k ام در فیلتر W' است. با توجه به شکل و توضیحات به سوالات زیر پاسخ دهید.

(آ) با استفاده از قاعده مشتق زنجیره‌ای عبارت $\frac{\partial loss}{\partial P_{i,j}}$ را بر حسب $\frac{\partial loss}{\partial z}$ محاسبه کنید.

(ب) حال با استفاده از روابط بدست آمده در قسمت اول و رابطه کانولوشن، عبارت $\frac{\partial loss}{\partial W'_{i,j,k}}$ را محاسبه کنید.



شکل ۲: شبکه CNN

سوال ۴: (نظری) فشردن و تحریک (۶ نمره)

با مطالعه مقاله فشردن و تحریک^۷ به سوالات زیر پاسخ دهید:

(آ) آیا نقش یا اثر عملکرد بلوک SE در عمق‌های مختلف شبکه یکسان است؟ مختصر توضیح دهید.

(ب) در بخشی از محاسبات Excitation از فرمول زیر استفاده می‌شود:

$$s = F_{ex}(z, W) = \sigma(g(z, W)) = \sigma(W_2 \delta(W_1 z))$$

که در آن

$$W_1 \in \mathbb{R}^{\frac{C}{R} \times C}, W_2 \in \mathbb{R}^{C \times \frac{C}{R}}$$

علت استفاده از r در ابعاد ماتریس وزن چیست و چه اثری در عملکرد واحد Excitation دارد؟

سوال ۵: (نظری) پیدایش و کانولوشن عمقی (۱۳ نمره)

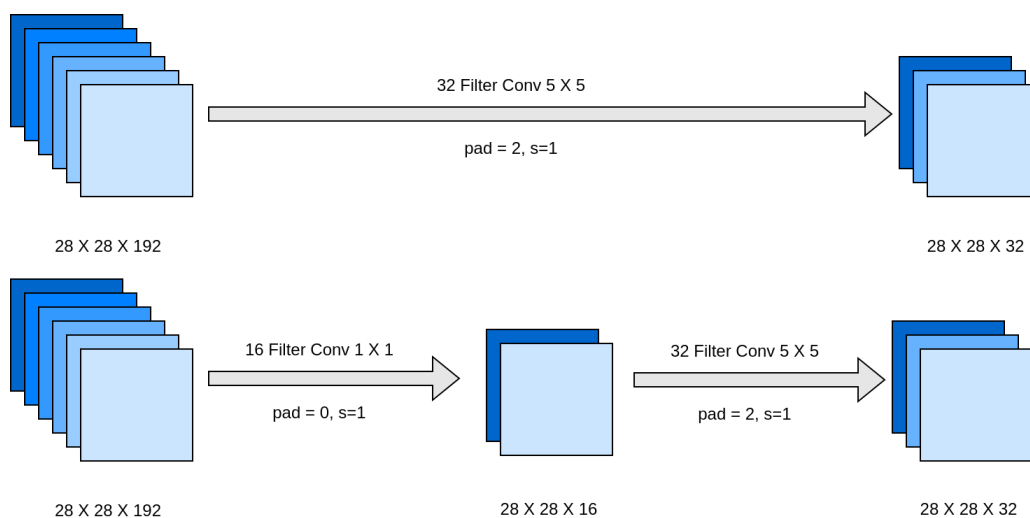
پیمانه پیدایش^۸ با هدف کاهش پیچیدگی محاسباتی بوجود آمد. در این سوال قصد داریم تا با پیمانه پیدایش و شبکه GoogleNet آشنا شویم.

^۶Gradient Back-propagation

^۷Squeeze and Excitation

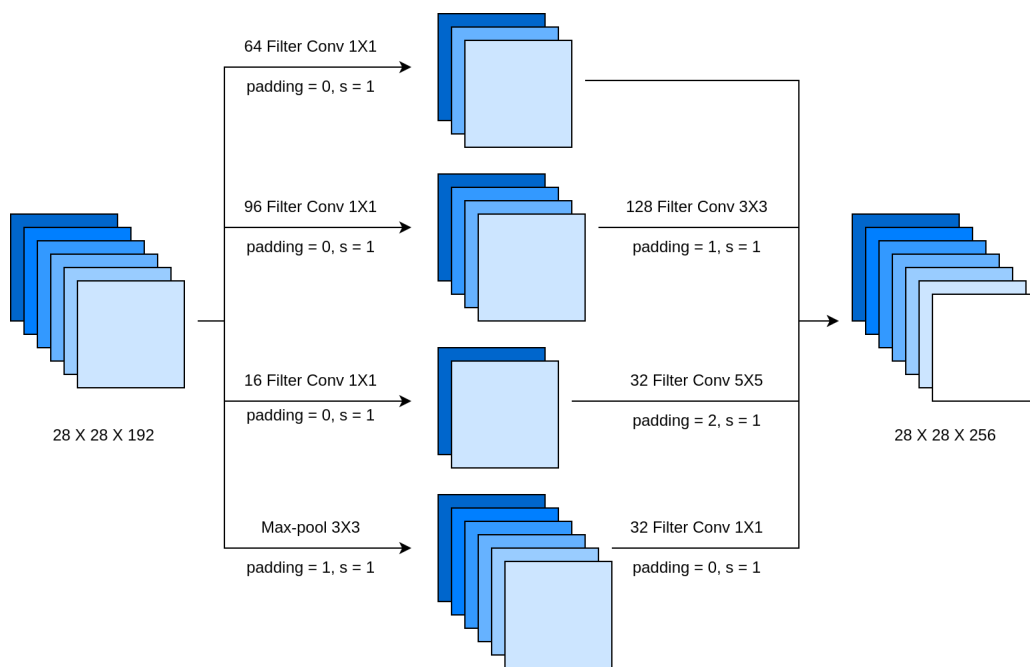
^۸Inception Module

(آ) در شکل ۳ یک لایه کانولوشن به دو صورت نشان داده شده است. ورودی هر دو حالت $28 \times 28 \times 192$ و خروجی هر دو $28 \times 28 \times 32$ است. با محاسبه تعداد کل عملیات‌های انجام شده در هر حالت، پیچیدگی محاسباتی دو حالت نشان داده شده را مقایسه کنید. مشخص کنید که افزودن فیلتر کانولوشن 1×1 ، پیچیدگی محاسباتی را چند درصد کاهش یا افزایش داده است؟



شکل ۳: لایه کانولوشن معمولی (بالا) و لایه کانولوشن با فیلتر 1×1 (پایین)

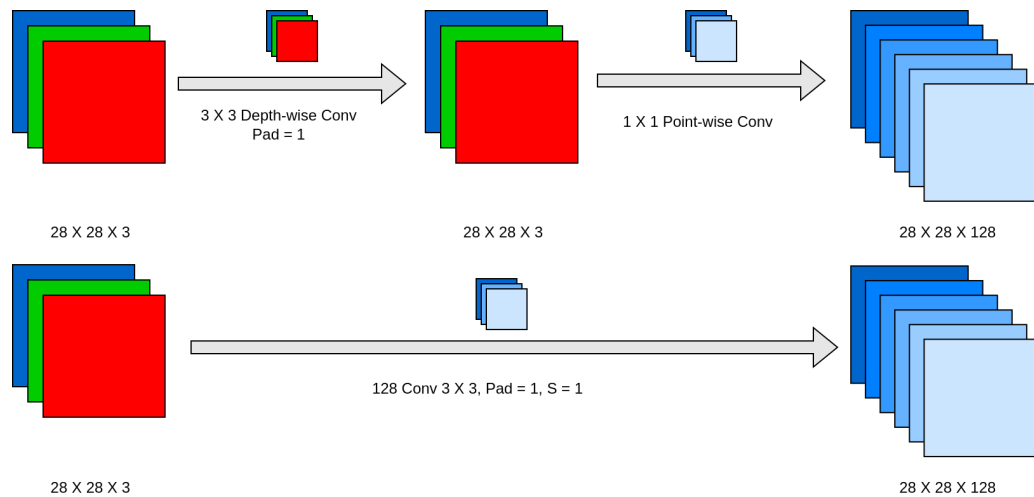
(ب) در شکل ۴ یک پیمانه پیدایش به صورت کامل نشان داده شده است. ترکیب متوالی این پیمانه با هم شبکه GoogleNet را تشکیل می‌دهد. توضیح دهید که دلیل استفاده از فیلترهایی با اندازه متفاوت (برای مثال 1×1 ، 3×3 و 5×5) و حتی فیلترهایی با انواع متفاوت (استفاده از Max-pooling در کنار کانولوشن) چیست؟ تعداد کل محاسبات در این شبکه را بدست آورید.



شکل ۴: شمای کلی پیمانه پیدایش

(ج) یکی دیگر از مشکلاتی که در ارتباط با شبکه‌های ژرف وجود دارد، تعداد بالای پارامترها و پیچیدگی بالای محاسباتی است. این مشکل استفاده از شبکه‌های CNN بر روی دستگاه‌های کوچک با پردازنده‌های محدود (مانند تلفن همراه) را با دشواری‌هایی همراه می‌سازد. برای حل این مشکل شبکه MobielNet-MobielNet-V1 است. شبکه MobielNet-V1 از کانولوشن عمقی^۹ استفاده می‌کند که تعداد پارامترها و پیچیدگی زمانی را کاهش دهد. در شکل ۵ یک پیمانه از شبکه MobielNet-V1 با یک لایه از شبکه CNN معمولی نمایش داده شده است. در هر روش تعداد پارامترها را محاسبه کنید.

^۹Depth-wise Convolution



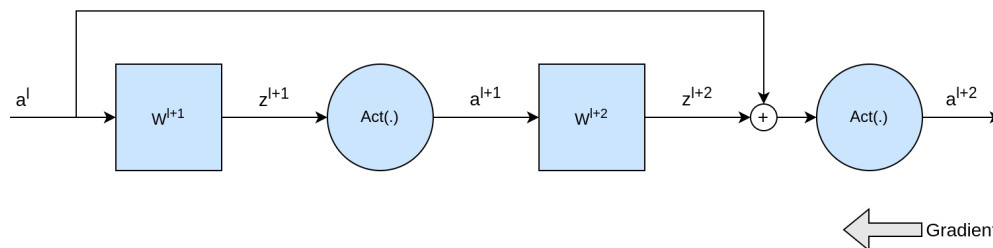
شکل ۵: لایه کانولوشن معمولی (پایین) و لایه از شبکه MobielNet-V1 (پایین)

سوال ۶: (نظری) شبکه باقی مانده (۵ نمره + ۲ نمره امتیازی)

یکی از مشکلاتی که آموزش شبکه‌های ژرف را با مشکل مواجه می‌کند مساله محوشدن گرادیان^{۱۰} است. در زمان انتشار به عقب، گرادیان با عبور از لایه‌های متوالی و ضرب شدن در بردارهای گرادیان و وزن‌ها (ضرایب فیلترهای کانولوشن در شبکه‌های CNN) به تدریج کاهش می‌یابد و به صفر میل می‌کند. این مساله باعث توقف آموزش در لایه‌های ابتدایی شبکه می‌گردد. یکی از محدودیت‌هایی که مساله محوشدن گرادیان به مدل تحمیل می‌کند، محدودیت بر عمق شبکه ژرف است؛ چراکه با عمیق شدن شبکه مشکل محوشدن گرادیان پررنگ‌تر می‌گردد. یکی از روش‌های حل مشکل محوشدن گرادیان استفاده از معماری شبکه باقی مانده^{۱۱} است. در این شبکه یک یال جانبی^{۱۲} به مدل اضافه شده است.

(آ) برای درک بهتر مکانیزم بهبود مشکل محوشدگی گرادیان تصویر ۶ را مورد بررسی قرار می‌دهیم. برای ساده‌سازی فرض کنید پیمانه باقی مانده، به جای اعمال بر یک شبکه کانولوشن، بر یک شبکه MLP اعمال شده است. در ابتدا $\frac{\partial a^{l+2}}{\partial a^l}$ را محاسبه کنید سپس با فرض داشتن $\frac{\partial loss}{\partial a^{l+2}}$ و بر اساس قاعده مشتق زنجیره‌ای، $\frac{\partial loss}{\partial a^l}$ را محاسبه کنید. (در این سوال فرض کنید که تابع فعال‌ساز هر لایه با $Act(\cdot)$ مشخص می‌گردد).

(ب) (۲ نمره امتیازی) بر اساس رابطه محاسبه شده در بخش قبل، تحلیل کنید که چگونه اضافه کردن یال جانبی در پیمانه باقی مانده باعث بهبود مشکل محوشدن گرادیان می‌گردد. (راهنمایی: برای ارائه بهتر تحلیل می‌توانید مشابه روند محاسبات قسمت قبل را برای دو لایه متوالی از شبکه MLP انجام دهید و روابط بدست آمده را مقایسه کنید. همچنین می‌توانید بررسی کنید که در صورت اضافه کردن یک پیمانه باقی مانده دیگر به انتهای یک شبکه، در زمان انتشار به عقب، مسیر عبور گرادیان چه تغییری می‌کند.)



شکل ۶: شمای کلی پیمانه باقی مانده

سوال ۷: (عملی) پیاده سازی شبکه عصبی پیچشی - دسته بندی تصاویر (۲۰ نمره)

در این سوال شما یک شبکه عصبی پیچشی^{۱۳} را برای وظیفه دسته بندی^{۱۴} تصاویر با دادگان Fashion-MNIST پیاده سازی می‌کنید. نوتبوکی که در اختیار شما قرار گرفته شامل بخش‌های:

¹⁰Gradient Vanishing

¹¹ResNet

¹²Bypass

¹³Convolutional Neural Network

¹⁴Classification

- بارگذاری دادگان و تقسیم‌بندی آن به سه بخش آموزش^{۱۵}، اعتبارسنجی^{۱۶} و آزمون^{۱۷}
 - پیاده‌سازی ساختار شبکه پیچشی با استفاده از PyTorch
 - انتخاب و تعریف نرخ یادگیری^{۱۸}، تابع زیان^{۱۹} و بهینه‌ساز^{۲۰} مناسب
 - آموزش مدل و رسم نمودار زیان
 - پیش‌بینی خروجی مدل روی دادگان آزمون
- است. با مراجعه به نوت‌بوک مورد نظر بخش‌های مشخص شده از کد را تکمیل نمایید.

سوال ۸: (عملی) قطعه‌بندی معنایی تصاویر با استفاده از شبکه UNet (۳۰ نمره)

در این تمرین قصد داریم تا ضمن پیاده‌سازی شبکه UNet با نمونه‌ای از قطعه‌بندی معنایی^{۲۱} تصاویر آشنا گردیم. در شکل ۷ نمونه‌ای از قطعه‌بندی معنایی نشان داده شده است. توجه داشته باشید که پس از تکمیل فایل مربوطه، آموزش شبکه برای ۳۰ epoch بر بستر Google Colab با استفاده از GPU در حدود ۳۰ دقیقه به طول می‌انجامد.



شکل ۷: قطعه‌بندی معنایی تصاویر

در این تمرین با موارد زیر آشنا می‌گردید:

- پیاده‌سازی شبکه UNet
 - تفاوت شبکه UNet و CNN معمولی
 - پیاده‌سازی قطعه‌بندی معنایی بر CARLA self-driving car dataset
- برای انجام این تمرین با مراجعه به فایل مورد نظر بخش‌های مشخص شده از کد را تکمیل نمایید.

¹⁵Train

¹⁶Validation

¹⁷Test

¹⁸Learning-rate

¹⁹Loss Function

²⁰Optimizer

²¹Semantic Segmentation