



## یادگیری ژرف

نیم سال دوم ۴۰۲-۱۴۰۱

مدرس: دکتر مهدیه سلیمانی

طرح تمرین: علی عبداللہی، سیدعلی یعقوب‌نژاد، زهرا رحیمی

تمرین سری دوم (۱۰۰ نمره + ۳ نمره امتیازی) نرمال‌سازی 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 کرده‌اید 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 اسکرین‌شات از منابع یا پاسخ افراد دیگر نمره‌ای تعلق نمی‌گیرد.
- در صورتی که بخشی از سوال‌ها را جای دیگری آپلود کرده و لینک آن را قرار داده باشید، حتما باید تاریخ آپلود مشخص و قابل اعتنا باشد.
- تمام پاسخ‌های 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) نرمال‌سازی بچ (۷ نمره + ۳ نمره امتیازی)

- (آ) استفاده از مقادیر بزرگ برای نرخ یادگیری در شبکه‌های ژرف سنتی می‌تواند منجر به انفجار یا محو شدن گرادیان شود. توضیح دهید چگونه نرمال‌سازی بچ به ما امکان استفاده از نرخ یادگیری بزرگتر از حد معمول را می‌دهد.
- (ب) زمانی که از نرمال‌سازی بچ استفاده می‌کنیم، در صورتی که وزن‌های شبکه در یک مقدار ثابت مانند  $\alpha$  ضرب شوند، چه تاثیری بر گرادیان نسبت به وزن‌ها و گرادیان نسبت به سایر پارامترهای شبکه خواهد داشت؟
- (ج) یکی از اثرات جانبی نرمال‌سازی بچ منظم‌سازی<sup>۱</sup> است. با ذکر دلیل بیان کنید که بزرگ شدن ساینز بچ چه تاثیر بر ویژگی منظم‌سازی نرمال‌سازی بچ خواهد داشت.
- (د) (۳ نمره امتیازی) با مطالعه این مقاله نحوه انجام نرمال‌سازی بچ در شبکه‌های تماماً متصل و شبکه‌های پیچشی را با یکدیگر مقایسه نمایید.

### سوال ۲: (نظری) کانولوشن گسترش‌یافته (۷ نمره)

در شبکه‌های پیچشی به صورت متداول از لایه‌های کانولوشن ساده استفاده می‌شود که با آن آشنا هستید. نوع دیگری از لایه‌ها که می‌توان از آنان در شبکه‌های پیچشی استفاده نمود، لایه‌های کانولوشن گسترش‌یافته<sup>۲</sup> است. در شکل ۱ یک تصویر شهودی از فیلتر کانولوشن گسترش‌یافته ارائه شده است؛ این فیلترها میان خانه‌هایی که فیلتر با استفاده از اطلاعات آن‌ها لایه بعد را محاسبه می‌کند، فاصله می‌اندازند یا به بیانی دیگر در زمان اعمال فیلتر و انجام عملیات ضرب کانولوشن، بر روی ورودی با طول گام<sup>۳</sup> بزرگتری حرکت می‌کنیم. توجه کنید که در اینجا طول گام مفهومی متفاوت نسبت به طول گام<sup>۴</sup> در لایه‌های شبکه کانولوشن دارد.

(آ) درمورد مزایا و معایب استفاده از فیلتر کانولوشن گسترش‌یافته تحقیق کنید و از هر کدام دو مورد ذکر نمایید.

(ب) در زمان استفاده از فیلتر کانولوشن گسترش‌یافته مفهومی به اسم محدوده دید<sup>۵</sup> معنای پررنگ‌تری پیدا می‌کند. فرض کنید ورودی  $M \times N$  به شبکه‌ای شامل سه لایه کانولوشن گسترش‌یافته داده شده است؛ محدوده‌ای از ورودی که عنصر  $i, j$ ، نام خروجی مشاهده می‌کند را به صورت پارامتری مشخص کنید.

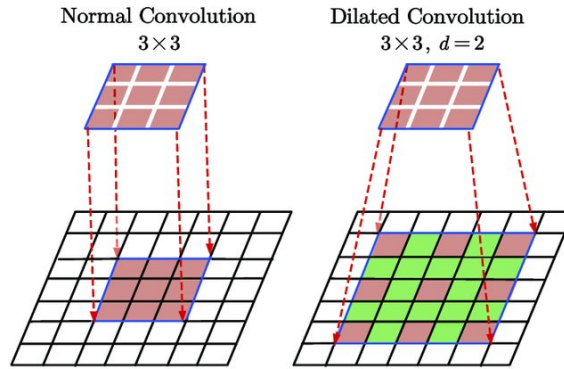
<sup>1</sup>Regularization

<sup>2</sup>Dilated Convolution

<sup>3</sup>Step Size

<sup>4</sup>Stride

<sup>5</sup>Receptive Field



شکل ۱: فیلتر کانولوشن معمولی و کانولوشن گسترش یافته

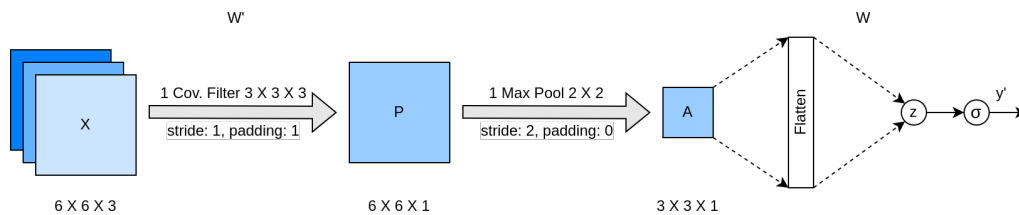
(ج) تصویری با ابعاد  $I \times I \times 3$  به یک شبکه یک لایه با  $D$  فیلتر  $K \times K \times 3$  با پارامتر گسترش  $N$  داده شده است. ابعاد خروجی این شبکه پس از اعمال بر ورودی را بر حسب  $I, D, K$  و  $N$  محاسبه کنید.

### سوال ۳: (نظری) انتشار به عقب (۱۰ نمره)

یکی از اساسی‌ترین اجزا در یک شبکه ژرف، انتشار به عقب گرادیان<sup>۶</sup> در شبکه است. شبکه نشان داده شده در شکل ۲ را در نظر بگیرید. در این شکل وزن‌های لایه تماماً متصل تهابی و  $W'$  وزن‌های فیلتر کانولوشن لایه اول است.  $P_{i,j}$  معرف عنصر سطر  $i$ ام و ستون  $j$ ام زام 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<sup>۷</sup> 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؟

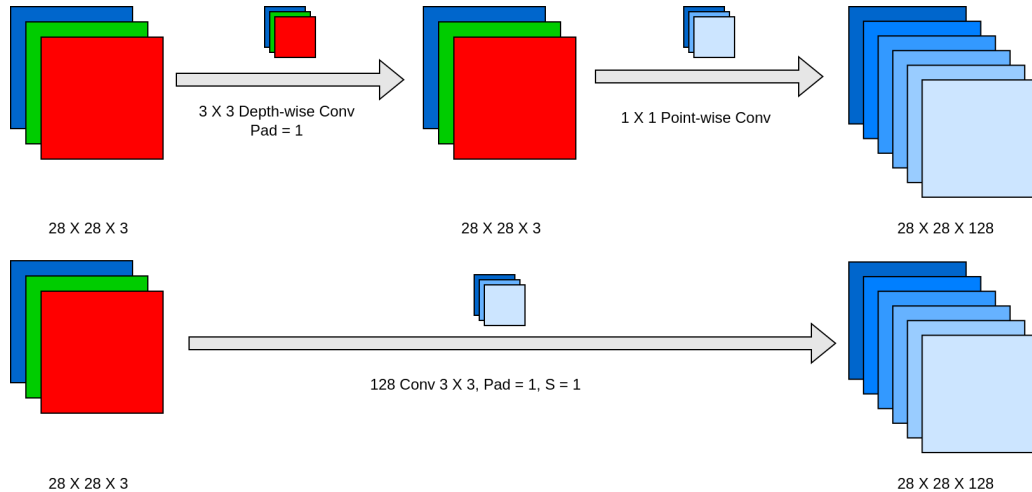
<sup>۶</sup>Gradient Back-propagation

<sup>۷</sup>Squeeze and Excitation

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

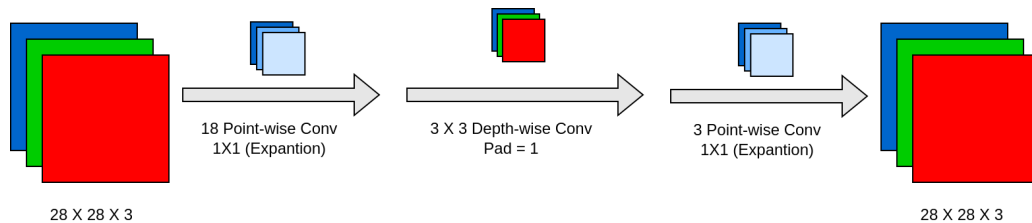
یکی دیگر از مشکلاتی که در ارتباط با شبکه‌های ژرف وجود دارد، تعداد بالای پارامترها و پیچیدگی بالای محاسباتی است. این مشکل استفاده از شبکه‌های CNN بر روی دستگاه‌های کوچک با پردازنده‌های محدود (مانند تلفن همراه) را با دشواری‌هایی همراه می‌سازد. برای حل این مشکل شبکه MobielNet-V1 استفاده می‌کند که تعداد پارامترها و پیچیدگی زمانی را کاهش دهد.

(آ) در شکل ۳ 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.



شکل ۳: لایه کانولوشن معمولی (پایین) و لایه از شبکه MobielNet-V1 (پایین)

(ب) در شبکه MobielNet-V2 از دو فیلتر کانولوشن  $1 \times 1$  استفاده شده است. شمای کلی معماری استفاده شده از این شبکه در شکل ۴ نشان داده شده است. با محاسبه تعداد پارامترها برای این شبکه، این شبکه را با شبکه MobielNet-V1 مقایسه نمایید.



شکل ۴: لایه از شبکه MobielNet-V2

سوال ۶: (نظری) شبکه باقی مانده (۷ نمره)

یکی از مشکلاتی که آموزش شبکه‌های ژرف را با مشکل مواجه می‌کند مساله محوشدن گرادیان<sup>۹</sup> 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<sup>۱۰</sup> است. در این شبکه یک یال جانبی<sup>۱۱</sup> به مدل اضافه شده است.

(آ) برای درک بهتر مکانیزم بهبود مشکل محوشدگی گرادیان تصویر ۵ 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.)

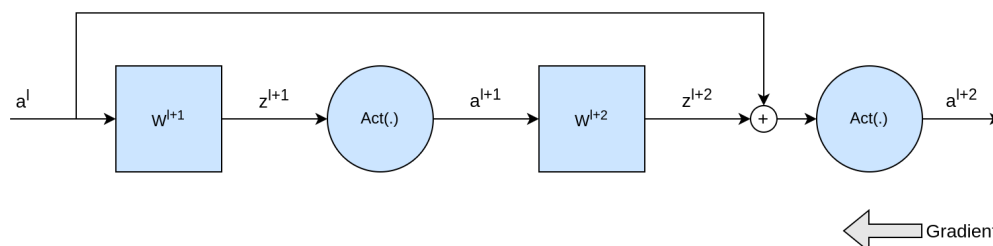
<sup>۸</sup>Depth-wise Convolution

<sup>۹</sup>Gradient Vanishing

<sup>۱۰</sup>ResNet

<sup>۱۱</sup>Bypass

(ب) بر اساس رابطه محاسبه شده در بخش قبل، تحلیل کنید که چگونه اضافه کردن یال جانبی در پیمانه باقی مانده باعث بهبود مشکل محو شدن 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 ۵: 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<sup>۱۲</sup> را برای وظیفه دسته بندی<sup>۱۳</sup> تصاویر با دادگان Fashion-MNIST پیاده سازی می‌کنید. نوتبکی که در اختیار شما قرار گرفته شامل بخش‌های:

- بارگذاری دادگان و تقسیم بندی آن به سه بخش آموزش<sup>۱۴</sup>، اعتبارسنجی<sup>۱۵</sup> و آزمون<sup>۱۶</sup>
- پیاده سازی ساختار شبکه پیچشی با استفاده از PyTorch
- انتخاب و تعریف نرخ یادگیری<sup>۱۷</sup>، تابع زیان<sup>۱۸</sup> و بهینه ساز<sup>۱۹</sup> مناسب
- آموزش مدل و رسم نمودار زیان
- پیش بینی خروجی مدل روی دادگان آزمون

است. با مراجعه به نوتبک مورد نظر بخش های مشخص شده از کد را تکمیل نمایید.

### سوال ۸: (عملی) قطعه بندی معنایی تصاویر با استفاده از شبکه UNet (۳۰ نمره)

در این تمرین قصد داریم تا ضمن پیاده سازی شبکه UNet با نمونه ای از قطعه بندی معنایی<sup>۲۰</sup> تصاویر آشنا گردیم. در شکل ۶ نمونه ای از قطعه بندی 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 می انجامد. در این تمرین با موارد زیر آشنا می گردید:

- پیاده سازی شبکه UNet
  - تفاوت شبکه UNet و CNN معمولی
  - پیاده سازی قطعه بندی معنایی بر CARLA self-driving car dataset
- برای انجام این تمرین با مراجعه به فایل مورد نظر بخش های مشخص شده از کد را تکمیل نمایید.

<sup>12</sup>Convolutional Neural Network

<sup>13</sup>Classification

<sup>14</sup>Train

<sup>15</sup>Validation

<sup>16</sup>Test

<sup>17</sup>Learning-rate

<sup>18</sup>Loss Function

<sup>19</sup>Optimizer

<sup>20</sup>Semantic Segmentation



شکل ۶: قطعه‌بندی معنایی تصاویر