



دانشگاه صنعتی شریف

دانشکده مهندسی کامپیوتر

یادگیری عمیق

نیم سال اول ۰۱-۰۲

استاد: حمید بیگی

گردآورندگان: مهدی کافی، امیر پورمند، فاطمه مهدوی

بررسی و بازبینی: محمد علی صدرایی جواهری

مهلت ارسال: ۱ آذر

شبکه های پیچشی

تمرین دوم

- مهلت ارسال پاسخ تا ساعت ۲۳:۵۹ روز مشخص شده است.
- در طول ترم امکان ارسال با تاخیر تمرین ها بدون کسر نمره تا سقف ۱۰ روز (تا سقف ۳ روز برای هر تمرین) وجود دارد. محل بارگزاری جواب تمرین ها بعد از ۵ روز بسته خواهد شد و پس از گذشت این مدت، پاسخ های ارسال شده پذیرفته نخواهند شد. همچنین، به ازای هر روز تأخیر غیر مجاز ۱۰ درصد از نمره تمرین به صورت ساعتی کسر خواهد شد.
- هم کاری و هم فکری شما در انجام تمرین مانعی ندارد اما پاسخ ارسالی هر کس حتما باید توسط خود او نوشته شده باشد.
- در صورت هم فکری و یا استفاده از هر منابع خارج درسی، نام هم فکران و آدرس منابع مورد استفاده برای حل سوال مورد نظر را ذکر کنید.
- لطفا تصویری واضح از پاسخ سوالات نظری بارگذاری کنید. در غیر این صورت پاسخ شما تصحیح نخواهد شد.
- پاسخ تمامی سوالات (تئوری و عملی) را در یک فایل فشرده به صورت `SPBio_Hw1_[firstName]_[lastName]_[StudentId]` نام گذاری کرده و ارسال کنید.

سوالات نظری (۷۰ نمره)

۱. طراحی شبکه (۵ نمره)

دسته بند شبکه عصبی کانولوشنی زیر را در نظر بگیرید.

$$Input = 128 \times 128 \times 3$$

$$Conv - 9 - 32$$

$$MaxPool - 2$$

$$Conv - 5 - 64$$

$$MaxPool - 2$$

$$Conv - 5 - 64$$

$$MaxPool - 2$$

$$FC - 3$$

Conv-K-N به معنای یک لایه کانولوشن با N فیلتر هر کدام با سایز $K \times K$ و پارامترهای stride و padding به ترتیب یک و صفر است. MaxPool-K به معنای لایه max pooling با سایز فیلتر $K \times K$ و stride و padding به ترتیب K و صفر می باشد. FC-N نیز به مفهوم لایه کاملاً متصل به هم با N نورون است.

(آ) تعداد وزن ها، بایاس ها و همچنین سایز feature maps را بدست آورید.

(ب) به چه دلیل قرار دادن non-linearity بین لایه های شبکه های عصبی مهم است؟

(ج) مزایا و معایب max pooling در شبکه های کانولوشنی چیست؟

(د) به دنبال آخرین لایه، ۳-FC چه نوع فعالسازی باید اعمال شود؟ با دادن بردار ورودی $[0.1, 0.5, 0.3]$ خروجی این لایه فعالساز چه خواهد بود؟

۲. تفسیرپذیری شبکه‌های کانولوشنی (۱۵ نمره)

به سوالات زیر پاسخ دهید.

- (آ) نحوه‌ی اشتراک گذاری پارامترها در لایه‌های کانولوشنی، باعث می‌شود این لایه‌ها، دارای ویژگی - equivariance نسبت به translation شوند. این ویژگی را شرح دهید و کاربرد آنرا را توضیح دهید.
- (ب) تفسیرپذیری شبکه‌های عصبی از اهمیت بالایی برخوردار است. هنگام دسته بندی تصاویر، علاقمندیم بدانیم کدام بخش‌های تصویر در دسته بندی، تأثیر بیشتری داشته‌اند. در مقاله ایده‌ی شبکه‌های - deconvolutional و در مقاله ایده‌ی شبکه‌های up-convolutional مطرح شده است. با بررسی این مقالات، توضیح دهید هر کدام از دو روش به چه صورت منجر به تفسیرپذیری شبکه کانولوشنی می‌شوند.

۳. تشخیص شی (۵ نمره)

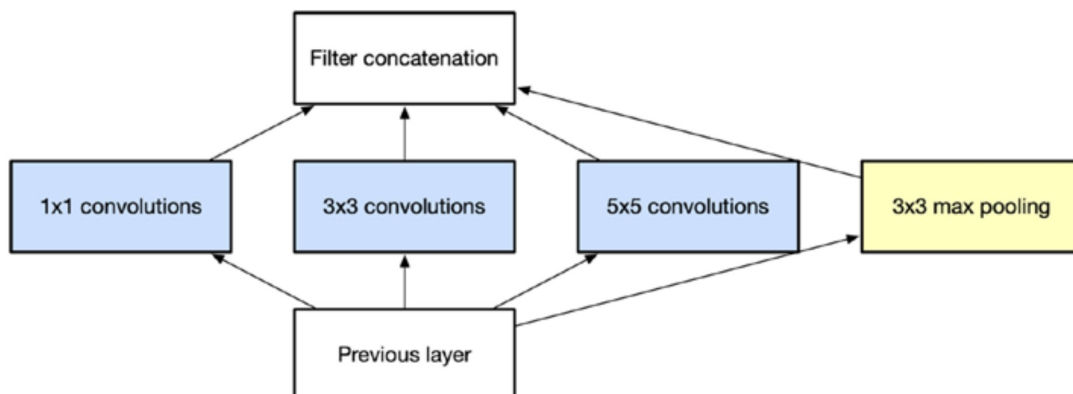
یکی از کاربردهای شبکه‌های کانولوشنی، آشکارسازی اشیا است. مقالات مرتبط را مطالعه کرده و به سوالات زیر پاسخ دهید.

- (آ) تفاوت بین آشکارسازهای تک مرحله ای و دو مرحله ای را توضیح دهید.
- (ب) نحوه آشکارسازی توسط Yolo را به طور خلاصه بیان کنید. با روش RCNN مقایسه کنید.
- (ج) RetinaNet چه نوع آشکارسازی است و ویژگی pyramid network چگونه عملکرد آن را بهبود می بخشد؟

۴. شبکه‌های کانولوشنی شناخته شده (۱۰ نمره)

به سوالات زیر پاسخ دهید.

- (آ) یکی از شبکه‌های کانولوشنی شناخته شده، شبکه Inception است که در مقاله معرفی شد. با بررسی این مقاله، هدف ایجاد این شبکه را بنویسید. سپس دو شبکه‌ی زیر را در نظر بگیرید.
- شبکه‌ای کانولوشنی که در هر لایه دارای ۳۲ کرنل با فیلترهای به ترتیب یک در یک، سه در سه و پنج در پنج است.
 - شبکه‌ای کانولوشنی مانند شکل زیر که در هر لایه دارای ۳۲ کرنل است.

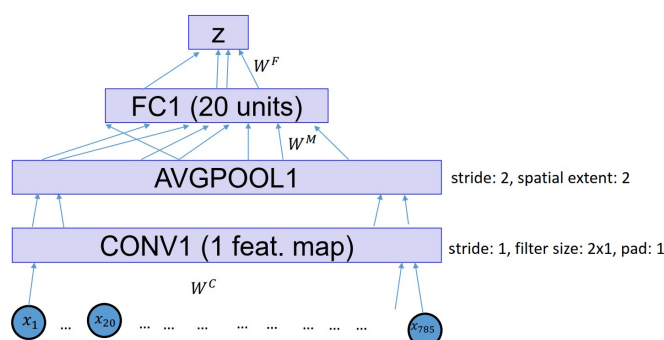


با فرض اینکه ورودی این شبکه‌ها دارای یک کانال باشد، تعداد پارامترهای هر لایه را با نوشتن محاسبات، محاسبه کنید.

(ب) در ابتدا مشکل محوشدگی گرادین در شبکه‌های عمیق را توضیح دهید. سپس شرح دهید چگونه شبکه‌ی ResNet سعی در پاسخ دادن به این مشکل دارد.

۵. بررسی شبکه کانولوشنی (۲۰ نمره)

شبکه کانولوشنی زیر را در نظر بگیرید.



ابعاد ورودی شبکه 785×1 و خروجی شبکه 1×1 است. شبکه از لایه‌های زیر تشکیل شده است.

- لایه ورودی X با zero-padding با طول ۱
- لایه کانولوشن یک بعدی CONV1 با یک کرنل 2×1 و تابع فعالسازی ReLU
- لایه AVGPOOL1 average-pooling
- لایه تمام متصل FC1 با تابع فعالسازی ReLU
- لایه خروجی Z که به لایه FC1 تماماً متصل است و از تابع فعالسازی sigmoid استفاده می‌کند

وزنی را که واحد i ام از لایه FC1 را به Z متصل می‌کند با W_i^F و بایاس Z را با b^F نشان می‌دهیم. وزنی را که واحد j ام از لایه AVGPOOL1 را به واحد i ام از FC1 متصل می‌کند با W_{ji}^A و بایاس i امین واحد از FC1 را با b_i^M نشان می‌دهیم. بردار W^C برابر با $[W_1^C, W_2^C]$ و بایاس لایه کانولوشنی، b^C است. داده‌های مجموعه‌ی آموزش به صورت $X = [X^{(1)}, X^{(2)}, \dots, X^{(N)}]$ و خروجی‌های مورد انتظار به صورت $Y = [Y^{(1)}, Y^{(2)}, \dots, Y^{(N)}]$ است. خروجی‌های لایه‌های شبکه را به ترتیب $f(X^{(i)})$, $a(X^{(i)})$, $c(X^{(i)})$ و $z(X^{(i)})$ می‌نامیم. تابع هزینه به صورت زیر تعریف می‌شود.

$$\text{cost}(X, Y) = \sum_n \text{cost}(X^{(n)}, Y^{(n)}) = \sum_n (-Y^{(n)} \log(z(X^{(n)})) - (1 - Y^{(n)}) \log(1 - z(X^{(n)}))).$$

با توجه به اطلاعات بالا به پرسش‌های زیر پاسخ دهید.

- (آ) تعداد پارامترهای شبکه بالا را با ذکر جزئیات محاسبه کنید.
- (ب) برای فقط یک داده‌ی آموزشی مقدار $\frac{\partial \text{Cost}}{\partial W_{ji}^A}$ را با جزئیات محاسبات، محاسبه کنید.
- (ج) برای فقط یک داده‌ی آموزشی مقدار $\frac{\partial \text{Cost}}{\partial W_1^C}$ را با جزئیات محاسبات، محاسبه کنید.

۶. کانولوشن متسع (۱۵ نمره)

کانولوشن متسع (Dilated Convolution) یک روش کم هزینه برای افزایش محدوده دید (Receptive Field) (تعداد پیکسل های ورودی که یک نورون را تحت تاثیر قرار می دهند) شبکه های کانولوشنی است که به صورت زیر تعریف می شود:

$$(K \star_D I)(i, j) = \sum_{m=-\infty}^{\infty} \sum_{n=-\infty}^{\infty} K(m, n) I(i + Dm, j + Dn)$$

(خروجی تنها برای اندیس های که کرنل و تصویر همپوشانی کامل دارند محاسبه می شود)

(آ) (۴ نمره) در یک شبکه کانولوشنی، با 1 لایه کانولوشن $K \times K$ با طول گام یک، عرض محدوده دید بر حسب 1 به دست آورید.

(ب) (۴ نمره) برای ورودی $I \in \mathbb{R}^{M \times N}$ و کرنل $K \in \mathbb{R}^{F \times F}$ نشان دهید خروجی عملگر کانولوشن متسع دارای ابعاد $(M - DF + D) \times (N - DF + D)$ است.

(ج) (۲ نمره) نشان دهید کانولوشن متسع معادل کانولوشن با کرنل متسع شده $K' = K \otimes A$ است. ماتریس A را مشخص کنید. (\otimes عملگر kronecker product است)

(د) (۵ نمره) در مورد Masked Convolution، کاربرد و محدودیت آن تحقیق کنید. چگونه می توان با استفاده از کانولوشن متسع محدودیت Masked Convolution را بهبود داد؟

سوالات عملی (۳۰ نمره)

۱. پیاده سازی شبکه پیچشی

برای حل این سوال به نوتبوک ضمیمه شده مراجعه کنید!