

# بسم الله الرحمن الرحيم

دانشگاه علم و صنعت ایران



زمستان ۱۳۹۹

تحويل: دوشنبه ۲۹ دی



تمرین سری پانزدهم

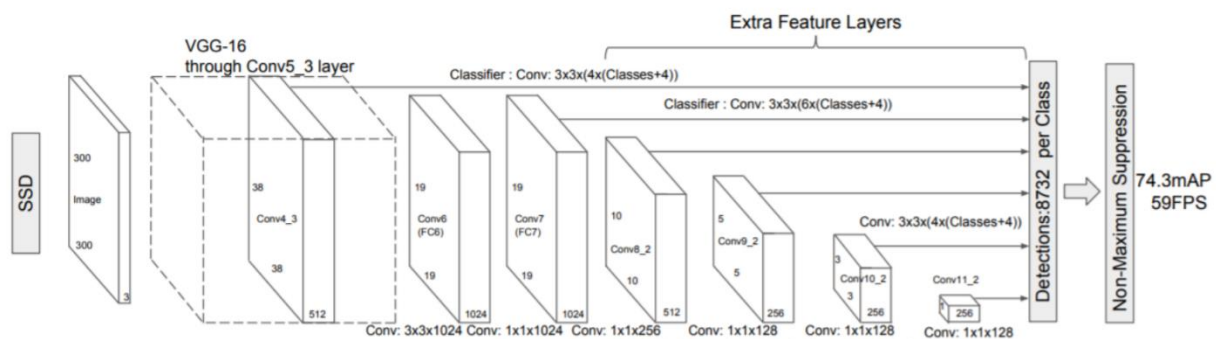
مبانی بینایی کامپیوتر

## توضیحات

- در ابتدای هر سوال یک نشانگر وجود دارد که مشخص می‌کند آن سوال تشریحی است یا برنامه‌نویسی؛ اگر نشانگر  یعنی سوال تشریحی است و اگر نشانگر  باشد یعنی سوال یک سوال برنامه‌نویسی است.
- برای تمامی سوال‌های پیاده‌سازی سعی شود توضیحی تشریحی در مورد نحوه پیاده‌سازی کد و بررسی و مقایسه نتایج در پاسخ‌های تشریحی آورده شود.
- خواهشمندیم اگر مطلبی را از اینترنت استفاده کردید حتماً منبع آن را مشخص کنید.
- لطفاً فایل‌های HW15Q4.ipynb و HW15Q5.ipynb نهایی خود را همراه با فایل pdf مربوط به پاسخ سوالات تشریحی فشرده کرده و ارسال کنید.
- **لطفاً در فایل گزارش خود به صورت دقیق، زمانی را که برای حل این تمرین صرف کردید، گزارش کنید.**
- **مهلت تحويل این تمرین قابل تمدید نیست و نمی‌توان در آن از زمان‌های تاخیر استفاده کرد. لطفاً این مورد را در نظر بگیرید.**

## سوالات

۱.  شبکه‌های خانواده RCNN را با شبکه‌های YOLO و SSD مقایسه کنید و نقاط ضعف و قوت هر کدام را بیان کنید (۱۵ نمره).
۲.  الف) در تصویر زیر نشان داده شده است که برای هر کلاس، ۸۷۳۲ ناحیه از تصویر ورودی بررسی می‌شود. نحوه محاسبه این عدد را به طور دقیق یادداشت کنید (۱۰ نمره).  
ب) در صورتیکه ابعاد تصویر ورودی این شبکه برابر با  $300 \times 600$  باشد، تعداد ناحیه‌های مورد بررسی چه عددی خواهد بود (۵ نمره)؟



۳. الف) یکی از شبکه‌های backbone که در بسیاری از شبکه‌های تشخیص شی (از جمله SSD) استفاده می‌شود، شبکه VGG16 است که جزئیات آن در جدول زیر نشان داده شده است. در صورتیکه لایه‌های کاملاً متصل این شبکه را به لایه‌های کانوولوشنی (بدون padding) تبدیل کنیم، و در ورودی شبکه یک تصویر  $512 \times 512 \times 3$  قرار دهیم، خروجی شبکه چه ابعادی خواهد داشت؟ (۲۰ نمره)

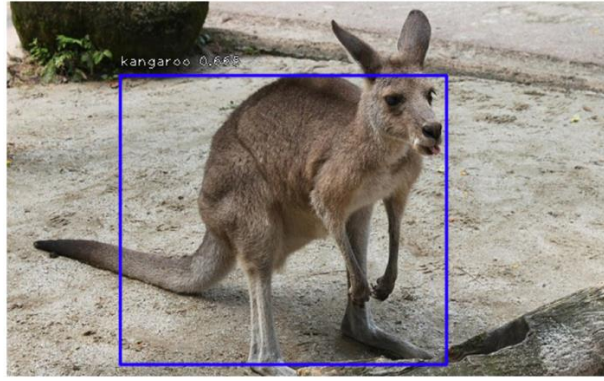
ب) لایه ابتدایی این شبکه شامل  $64 \times 3 \times 3$  فیلتر است. در حالتی که ورودی شبکه  $512 \times 512 \times 3$  باشد، این لایه چه تعداد ضرب و جمع نیاز دارد (به طور دقیق محاسبه کنید و مراحل محاسبه را بنویسید)؟ (۱۰ نمره)

پ) اگر بجای پیاده‌سازی کانوولوشنی، شبکه VGG16 را بر روی همان مکان‌هایی که در قسمت قبل اعمال شده است، به صورت عادی اعمال کنیم، تعداد ضرب و جمع لازم چه مقدار خواهد بود (در این قسمت تعداد ضرب و جمع در حالت عادی با ورودی  $224 \times 224 \times 3$  را محاسبه کنید و این عدد را در تعداد دفعاتی که لازم است شبکه اجرا شود ضرب کنید؟ نتیجه این قسمت را با قسمت قبل مقایسه کنید) (۱۰ نمره).

VGG16 - Structural Details													
#	Input Image			output			Layer	Stride	Kernel		in	out	Param
1	224	224	3	224	224	64	conv3-64	1	3	3	3	64	1792
2	224	224	64	224	224	64	conv3064	1	3	3	64	64	36928
	224	224	64	112	112	64	maxpool	2	2	2	64	64	0
3	112	112	64	112	112	128	conv3-128	1	3	3	64	128	73856
4	112	112	128	112	112	128	conv3-128	1	3	3	128	128	147584
	112	112	128	56	56	128	maxpool	2	2	2	128	128	65664
5	56	56	128	56	56	256	conv3-256	1	3	3	128	256	295168
6	56	56	256	56	56	256	conv3-256	1	3	3	256	256	590080
7	56	56	256	56	56	256	conv3-256	1	3	3	256	256	590080
	56	56	256	28	28	256	maxpool	2	2	2	256	256	0
8	28	28	256	28	28	512	conv3-512	1	3	3	256	512	1180160
9	28	28	512	28	28	512	conv3-512	1	3	3	512	512	2359808
10	28	28	512	28	28	512	conv3-512	1	3	3	512	512	2359808
	28	28	512	14	14	512	maxpool	2	2	2	512	512	0
11	14	14	512	14	14	512	conv3-512	1	3	3	512	512	2359808
12	14	14	512	14	14	512	conv3-512	1	3	3	512	512	2359808
13	14	14	512	14	14	512	conv3-512	1	3	3	512	512	2359808
	14	14	512	7	7	512	maxpool	2	2	2	512	512	0
14	1	1	25088	1	1	4096	fc		1	1	25088	4096	102764544
15	1	1	4096	1	1	4096	fc		1	1	4096	4096	16781312
16	1	1	4096	1	1	1000	fc		1	1	4096	1000	4097000
Total													138,423,208

۴. با استفاده از [تطبیق کلیشه](#) برنامه‌ای بنویسید که شماره یک پلاک را بخواند. کلیشه مربوط به اعداد در پوشه Templates موجود است و تصاویر مربوط به پلاک نیز در پوشه Plates ذخیره شده است (برای این سوال لازم است فایل HW15Q4.ipynb را تکمیل کنید. این فایل را لطفاً بر روی دستگاه خودتان تکمیل کنید نه بر روی کولب) (۳۰ نمره).

۵. یکی از شبکه‌های معروف برای تشخیص شی در تصویر، شبکه [RetinaNet](#) است (این شبکه تا حد زیادی شبیه به SSD است). در این تمرین از شما می‌خواهیم با مطالعه بخش Training در این [صفحه](#) و همچنین استفاده از این [لینک](#) آموزشی، یک مدل تشخیص کانگورو در تصویر آموزش بدهید. برای آموزش مدل از مجموعه داده [kangaroo](#) استفاده کنید (۸۰ درصد از تصاویر را برای آموزش و ۲۰ درصد را برای ارزیابی شبکه استفاده کنید). پس از آموزش مدل، آن را بر روی تصاویر ارزیابی اعمال کنید و خروجی‌های بدست آمده را گزارش کنید (نمونه خروجی یک داده به صورت زیر است و برای تمامی داده‌های ارزیابی خروجی را در گزارش ثبت کنید). همچنین، تمام مراحل انجام شده و فایل‌های تولید شده برای آموزش و ارزیابی مدل را ارسال کنید (برای این سوال لازم است فایل HW15Q5.ipynb را بر روی کولب تکمیل کنید). (۲۵ نمره).



• موفق باشید.