



تمرین سری هفتم  
درس مبانی بینایی کامپیوتر

نام مدرس: دکتر محمدی  
دستیاران آموزشی مرتبط: نادری فرد، غیوری،  
میرآخوخلو

مهلت تحویل: ۱۴۰۳/۳/۲۷

۱- با مطالعه لینک زیر به سوالات پاسخ دهید. (۲۰)

<https://medium.com/jun94-devpblog/dl-13-convolution-and-pooling-variants-dilated-convolution-spp-aspp-a954a282ff5c>

الف) یک لایه کانولوشنی با کرنل سایز  $k$  و  $dilation\ rate = d$  در نظر بگیرید و ابعاد  $dilated\ kernel$

را بر حسب  $k, d$  بنویسید.  

$$k + (k-1) \times (d-1)$$

$$k + d \times (k-1)$$

ب) اگر  $dilation\ rate$  لایه کانولوشنی را ۳ برابر کنیم تعداد پارامترهای قابل آموزش لایه چند برابر می شود؟

ج) در جدول زیر مقادیر مشخص شده را بیابید.

Layer	1	2	3	4	5	6	7	8
Convolution	3*3	3*3	3*3	3*3	3*3	5*5	5*5	$E * E = 7 \times 7$
Dilation rate	1	1	4	$B = 11$	8	3	2	6
Receptive field	3*3	5*5	$A$	35*35	$C$	$D$	71*71	107*107
			13x13		51x51	63x63		

د) شبکه ای با دو لایه کانولوشنی با کرنل سایز ۵ و سه لایه ادغام بیشینه را در نظر بگیرید. در صورتی که بخواهیم

به حداقل میدان تاثیر  $107 \times 107$  برسیم (مشابه با لایه ۸ بخش ج) حداقل گام مورد نیاز برای لایه های ادغام بیشینه را بدست آورید.

(لایه های ادغام بیشینه مشابه هم هستند).

۲- الف) با مطالعه لینک زیر تعداد پارامترها و عملیات ضرب لایه کانولوشنی معمولی و  $depthwise\ separable$

را بدست آورید و باهم مقایسه کنید. (۱۰)

<https://towardsdatascience.com/a-basic-introduction-to-separable-convolutions-b99ec3102728>

input : (۱۲۸, ۱۲۸, ۳)      kernel size = ۵      filters = ۶۴

ب) اگر ابعاد ورودی لایه کانولوشنی معمولی (۱۲۸ و ۱۲۸)، کرنل سایز  $3 \times 3$  و تعداد فیلترهای خروجی آن ۳۲ باشد، تعداد پارامترهای این لایه در صورتیکه از کانولوشن  $depthwise$  استفاده کنیم چند برابر می شود؟

(از بایاس صرف نظر کنید)



تمرین سری هفتم  
درس مبانی بینایی کامپیوتر

نام مدرس: دکتر محمدی  
دستیاران آموزشی مرتبط: نادری فرد، غیوری،  
میرآخوخلو

مهلت تحویل: ۱۴۰۳/۳/۲۷

۳- الف) تصویر ورودی و کلیشه زیر را در نظر بگیرید و با ذکر دلیل متد تطبیق کلیشه مناسب را مشخص کنید. (۲۰)

تصویر

۵۳	۱۸۱	۲۰۰	۱۵۱	۲۵۵	۲۵۳	۲۵۵
۲۰۰	۷	۲۹	۳	۲۰۰	۱۹۹	۲۱۸
۹۱	۲۱	۱۸	۵	۲۲۰	۲۵۵	۲۵۵
۹۶	۲	۱۴	۱۱	۲۱۹	۲۵۴	۲۵۳
۱۱۲	۲۱۷	۲۱۱	۱۹۹	۲۱۸	۲۵۱	۲۵۳

کلیشه

۶	۲۷	۳
۲۱	۱۸	۵
۲	۱۵	۱۱

$$R(x, y) = \sum_{x', y'} (T(x', y') \cdot I(x + x', y + y'))$$

$$R(x, y) = \frac{\sum_{x', y'} (T'(x', y') \cdot I'(x + x', y + y'))}{\sqrt{\sum_{x', y'} T'(x', y')^2 \cdot \sum_{x', y'} I'(x + x', y + y')^2}}$$

ب) با استفاده از تطبیق کلیشه تمامی سکه های موجود در تصویر زیر را بیابید.





تمرین سری هفتم  
درس مبانی بینایی کامپیوتر

نام مدرس: دکتر محمدی  
دستیاران آموزشی مرتبط: نادری فرد، غیوری،  
میرآخوخلو

مهلت تحویل: ۱۴۰۳/۳/۲۷

۴- مدل SAM (segment anything) مدلی برای بخش بندی تصاویر است که ماسک هایی با کیفیت بالا از

اشیاء درون تصویر تولید میکند و برای کاربردهای مختلفی از قطعه بندی مورد استفاده قرار میگیرد. در نوتبوک SAM میتوان نمونه ای از ماسک های تولید شده برای تصویر را مشاهده کرد.

بخش های مشخص شده را تکمیل کنید و پس از آن با اجرای کدها به خروجی نهایی برسید. (۱۰)

۵- به نوتبوک semantic segmentation مراجعه کنید و بخش های خواسته شده را کامل کنید. (۲۵)

همچنین callback های مناسب جهت ذخیره بهترین مدل را به کد اضافه کنید. در این سوال ابتدا یک

شبکه Unet را از صفر آموزش می دهید. سپس در بخش دوم یک شبکه Unet با انکودر پیش آموخته

را آموزش می دهید. در مرحله اول انکودر کاملاً فریز است و تنها دیکودر آموزش می بیند و در مرحله انتهایی

کل شبکه تنظیم دقیق می شود. در گزارش خود علاوه بر توضیح کد خودتان، توضیحات مختصری در

مورد قسمت های آماده کد ارائه دهید.

۶- در این تمرین قصد داریم تشخیص اشیا با الگوریتم Fast R-CNN را بررسی کنیم. برای این کار طبق روند

زیر عمل کنید و نوتبوک Q6 را تکمیل و به همراه گزارش کامل از توضیحات بخش های مختلف کد و تفسیر

نتایج ارسال کنید. (۳۵)

• برای این تمرین از محیط tensorflow استفاده کنید.

• مجموعه ی Pascal VOC را از این لینک دانلود کرده و در پروژه خود بارگزاری (load) کنید.

<https://paperswithcode.com/dataset/pascal-voc>

• با تبدیل حاشیه نویسی ها به فرمت مورد نظر مانند لیبل و... داده را برای آموزش آماده کنید.



تمرین سری هفتم  
درس مبانی بینایی کامپیوتر

نام مدرس: دکتر محمدی  
دستیاران آموزشی مرتبط: نادری فرد، غیوری،  
میرآخوخلو

مهلت تحویل: ۱۴۰۳/۳/۲۷

- تابعی بنویسید که ۱۰ تا از این تصاویر و حاشیه مویشی های موجودشان که در حال حاضر به عنوان لیبل و جعبه های محدود کننده (bounding boxes) در دسترس قرار می گیرند را نمایش دهد.

- مدل zoo یک مدل از پیش آموزش دیده Fast R-CNN را به صورت زیر فراخوانی کنید.

```
model = tf.keras.applications.ResNet50(include_top=False,  
input_shape=(None, None, 3))
```

- تابعی بنویسید که روی هر عکس با استفاده از مدل Fast R-CNN استنتاج انجام دهد.
- کارایی مدل را با استفاده از validation set ارزیابی کنید. در این بخش معیار های زیر را محاسبه و گزارش کنید.

1. میانگین دقت متوسط (mAP)

2. دقت (precision)

3. Recall

4. F1 score

5. Confusion matrix

- نمودار Precision-recall را رسم کنید.

لطفا سند قوانین انجام و تحویل تمرین های درس را مطالعه و موارد خواسته شده را رعایت فرمایید.

موفق و سلامت باشید