تمرین سری هشتم درس تصویربرداری رقمی

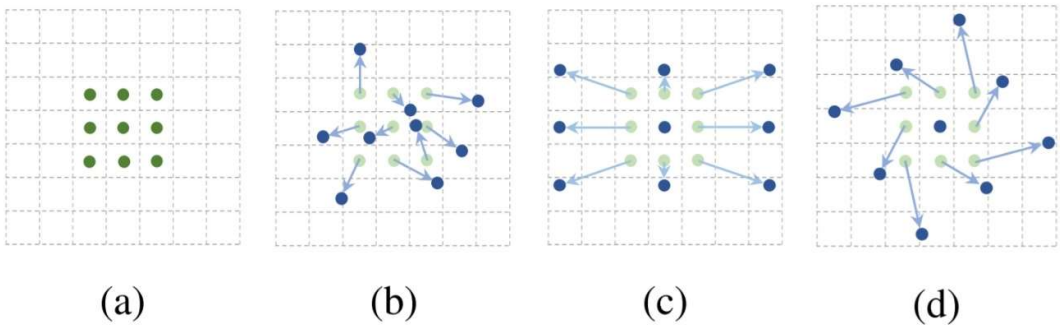
پوریا محمدی نسب

(۴۰۰۷۲۲۱۳۸)

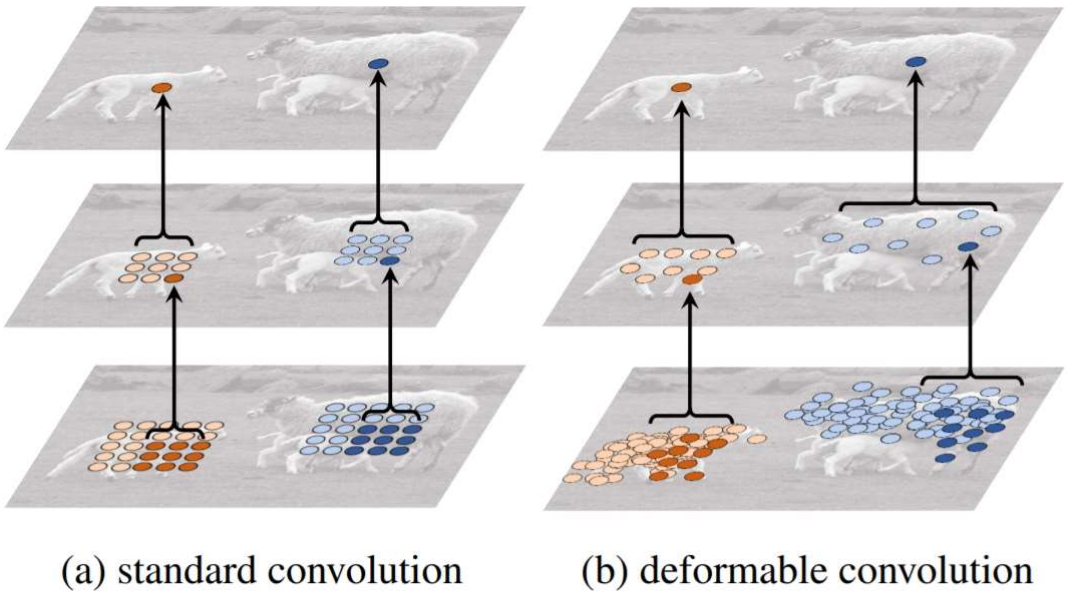
۱- مقاله زیر را باد دقت مطالعه کنید و به صورت خلاصه Deformable Convolution را توضیح دهید(برای توضیح بهتر از شکل های مقاله هم استفاده کنید).

Dai, Jifeng, et al. "Deformable convolutional networks." Proceedings of the IEEE international conference on computer vision. 2017.

یکی از چالش های مهم در مبحث visual recognition مدلسازی تبدیلات و تغییرات هندسی است برای مثال تغییر ابعاد شی، موقعیت و زاویه نگاه کردن به شی میتواند دشواری هایی برای ترین کردن یک CNN به وجود بیاورد. دو راه برای ساختن دیتاستی با میزان تغییرات بیشتر داریم. راه اول استفاده از تکنیک augmentation است. راه دوم این است که از ویژگی ها و الگوریتم های تغییرناپذیر استفاده کنیم. مشکل این راه حل ها این است که اولا نیاز به دانش قبلی از مسئله دارد و این باعث میشود قدرت تعمیم برای مسائل جدید از الگوریتم گرفته شود و ثانیا در حالتی که transformation پیچیده ای داشته باشیم داده افزایی برای فیچرها و الگوریتم ها میتواند سخت باشد. برای غلبه به این مشکلات در مقاله ذکر شده دو ماژول جدید معرفی شد که تغییرات اساسی در CNN ها و ظرفیت حل مسئله ی آنها ایجاد کرد. اولین مورد Deformable convolution است که offset های دو بعدی به مکان های نمونه گیری اضافه میکند که باعث میشود این شبکه و لایه ی کانولوشن فارق از تغییرات هندسی شود. شکل زیر به صورت واضح تری مفهوم deformable convolution را منتقل میکند.



در این تصویر a نشان دهنده یک لایه کانولوشنی استاندارد است و تصاویر c،b و d نشان دهنده ی قدرت تعمیم این لایه بر روی transformation های مختلف روی عکس میباشد.



تصویر بالا نیز به طور واضحی تفاوت بین کانولوشن استاندارد و حالت deformable را نشان میدهد که از دو activation map استفاده شده است. لایه ای که در بالای همه قرار دارد فعال ساز روی feature map است که دو شی با ابعاد و شکل متفاوت را تشخیص میدهد. لایه ی میانی لایه ی نمونه برداری است که در واقع از ۹ نقطه در لایه ی پایینی یکی را انتخاب میکند.

برای مقایسه و مشاهده ی عملکرد deformable convolution از آن در لایه های مختلف کانولوشنی در معماری های معروفی که برای تسک VOC 2007 test طراحی شدند استفاده کردند و نتایج در یک جدول قابل مشاهده است.

usage of deformable convolution (# layers)	DeepLab		class-aware RPN		Faster R-CNN		R-FCN	
	mIoU@V (%)	mIoU@C (%)	mAP@0.5 (%)	mAP@0.7 (%)	mAP@0.5 (%)	mAP@0.7 (%)	mAP@0.5 (%)	mAP@0.7 (%)
none (0, baseline)	69.7	70.4	68.0	44.9	78.1	62.1	80.0	61.8
res5c (1)	73.9	73.5	73.5	54.4	78.6	63.8	80.6	63.0
res5b,c (2)	74.8	74.4	74.3	56.3	78.5	63.3	81.0	63.8
res5a,b,c (3, default)	**75.2**	**75.2**	74.5	57.2	78.6	63.3	81.4	64.7
res5 & res4b22,b21,b20 (6)	74.8	75.1	**74.6**	**57.7**	**78.7**	**64.0**	**81.5**	**65.4**

در این جدول ۴ معماری معروف Deep lab، RPN، R-CNN و R-FCN مورد بررسی قرار گرفتند. در ردیف های این جدول مشخص شده است که deformable convolution روی کدام لایه های کانولوشنی اعمال شده. همانطور که واضح است اعمال این لایه در معماری بشدت میتواند عملکرد را بهبود ببخشد.

۲- الف) شبکه VGG19 یکی از شبکه های پرکاربرد برای دسته بندی تصویر است. جزئیات لایه های کانولوشنی این شبکه در شکل زیر نشان داده شده است.

Layer name	input shape			padding	stride	kernel size	Filters
Input	256	256	3				
block1_conv1	256	256	3	same	1	(3×3)	64
block1_conv2	256	256	64	same	1	(3×3)	64
block1_pool	256	256	64		2	(2×2)	
block2_conv1	128	128	64	same	1	(3×3)	128
block2_conv2	128	128	128	same	1	(3×3)	128
block2_pool	128	128	128		2	(2×2)	
block3_conv1	64	64	128	same	1	(3×3)	256
block3_conv2	64	64	256	same	1	(3×3)	256
block3_conv3	64	64	256	same	1	(3×3)	256
block3_conv4	64	64	256	same	1	(3×3)	256
block3_pool	64	64	256		2	(2×2)	
block4_conv1	32	32	256	same	1	(3×3)	512
block4_conv2	32	32	512	same	1	(3×3)	512
block4_conv3	32	32	512	same	1	(3×3)	512
block4_conv4	32	32	512	same	1	(3×3)	512
block4_pool	32	32	512		2	(2×2)	
block5_conv1	16	16	512	same	1	(3×3)	512
block5_conv2	16	16	512	same	1	(3×3)	512
block5_conv3	16	16	512	same	1	(3×3)	512
block5_conv4	16	16	512	same	1	(3×3)	512
block5_pool	16	16	512		2	(2×2)	

بعد از این لایه ها ، برای تبدیل تنسور خروجی به بردار میتوانیم از حالتهای زیر استفاده کنیم و بعد از آن هم با استفاده از یک لایه کاملا متصل دسته بندی را انجام دهیم. اگر مسئله دسته بندی ۳۰ کلاسه باشد، برای هر کدام از حالتهای زیر تعداد پارامترهای لایه مربوطه و همچنین لایه کاملا متصل بعد از آن را محاسبه کنید.

- Flatten

در مورد عملگر Flatten به آخرین لایه ی VGG19 توجه میکنیم که یک MaxPooling است و خروجی با توجه به وردی نوشته شده در جدول (8,8,512) است. با ورود این تنسور به Flatten خروجی یک وکتور با طول (512*8*8) یعنی 32768 است. پس در قسمت آخر معماری یک FC با 32768 ورودی و 30 نرون خروجی خواهیم داشت.

- GAP

GAP مخفف شده ی اسم لایه ی Global Average Pooling است. واضح است که این لایه میانگین تمام پیکسل ها را گرفته و در وکتوری قرار میدهد. پس بنابراین اگه شکل ورودی را (8,8,512) در نظر بگیریم خروجی این لایه یک وکتور به اندازه عمق تنسور ورودی است یعنی 512. پس لایه ی FC در این حالت 512 ورودی و 30 خروجی خواهد داشت.

- GWAP با وزن یکسان برای تمام کانال ها

اگر در Global Weighted Average Pooling وزن تمام کانال ها یکسان باشد عملکرد آن دقیقا مشابه GAP است. بنابراین خروجی این نوع لایه (GWAP) با لایه ی GAP هیچ تفاوتی ندارد.

- GWAP با وزن متفاوت برای هر کانال

در این حالت که وزن کانال ها با هم برابر نیس تنها تفاوت در محاسبات این لایه است که برای مثال یک لایه وزن بیشتری دارد و تاثیر بیشتری در خروجی میانگین گرفته شده دارد و یک کانال وزن کمتری دارد. اما این بار هم شکل خروجی با دو حالت قبلی هیچ تفاوتی ندارد.

- هیستوگرام قابل آموزش با ۴ Bin

هیستوگرام قابل آموزش به تعداد bin هایش نقشه فعال سازی دارد. برای مثال ما ورودی (8,8,512) است و چون 4 عدد bin داریم پس خروجی ما به صورت (512*4) یعنی یک وکتور 2048 تایی است. پس ساختار قسمت FC شبکه به صورت 2048 ورودی در 30 خروجی است.

- هیستوگرام قابل آموزش با ۸ Bin

به طور مشابه با حالت قبلی خروجی در این حالت یک وکتور (512*8) یعنی 4096 تایی است پس FC ما 4096 ورودی و 30 خروجی دارد.

۲ - ب) مزایا و معایب هر کدام از حالتهای بالا را بیان کنید.

Flatten که یکی از رایج ترین حالت ها برای تبدیل تنسور به وکتور است یک مزیت نسبت به سایر روش ها دارد که همه ی اطلاعات تنسور قبلی را در خود نگه داری میکند و داده ای از دست نمیرود اما همین نگهداری تمام اعداد از تنسور قبلی حجم حافظه و پارامترهای FC را افزایش میدهد. GAP دقیقا نقطه مقابل Flatten است به این معنی که حجم حافظه و تعداد پارامتر ها را کاهش میدهد اما در مقابل ممکن است مقدار زیادی از دیتا را از دست بدهد. همانطور که گفته شد GWAP با اوزان مساوی تفاوتی با GAP ندارد. اما GWAP با اوزان متفاوت میتواند مشکلاتی که در GAP کلاسیک وجود داشتند را تا حد خوبی حل کند(برای مثال GAP در تشخیص دقیق ناحیه ها و اشیاء ایراداتی دارد.) و در انتها هیستوگرام قابل آموزش با انگیزه افزایش کارای در دو تسک semantic segmentation و object detection پیشنهاد شدند و نشان دادند در این تسک ها عملکرد بسیار خوبی دارند.

References

1) https://arxiv.org/pdf/1809.08264.pdf (GWAP)

2) https://keras.io/api/layers/pooling_layers/global_average_pooling2d/

3) https://machinelearningmastery.com/use-pre-trained-vgg-model-classify-objects-photographs/

4) https://keras.io/api/applications/vgg/

5) https://arxiv.org/pdf/1804.09398.pdf

6) mathalope.co.uk/2015/03/20/descriptive-statistics-on-histogram-what-are-the-pros-and-cons-of-large-vs-small-binwidth/

7) quora.com/Why-was-global-average-pooling-used-instead-of-a-fully-connected-layer-in-GoogLeNet-and-how-was-it-different

8) https://paperswithcode.com/method/global-average-pooling