



## Assignment NO.8 Solutions

Digital Image Processing | Fall 1400 | Dr.Mohammadi

Teacher Assistant : Fatemeh Anvari

---

Student name : **Amin Fathi**

Student id : **400722102**

## Problem 1

شبکه‌های عصبی کانولوشنال (CNN) به دلیل ساختارهای هندسی ثابت در ماژول‌های ساختمان خود، ذاتاً به مدل‌سازی تبدیل‌های هندسی محدود می‌شوند. دو ماژول جدید برای افزایش قابلیت مدل‌سازی تبدیل CNN معرفی می‌شود، یعنی کانولوشن تغییر شکل‌پذیر (Deformable Convolutional Networks) و RoI pooling.

هر دو بر اساس ایده افزایش مکان‌های نمونه‌برداری فضایی در ماژول‌ها با افست‌های اضافی و یادگیری افست‌ها از وظایف هدف هستند. ماژول‌های جدید می‌توانند به آسانی جایگزین هم‌تایان ساده خود در CNN‌های موجود شوند و شبکه‌های کانولوشن قابل تغییر شکل را ایجاد کنند.

ایده اول offset‌های دو بعدی به مکان‌های نمونه‌گیری اضافه می‌کند که باعث می‌شود این شبکه و لایه کانولوشن فارغ از تغییرات هندسی شود. شکل زیر به صورت واضح‌تری مفهوم deformable convolution را منتقل می‌کند:

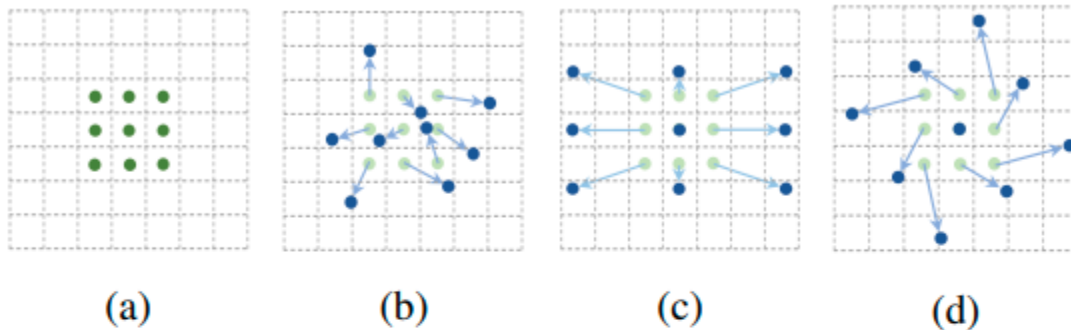
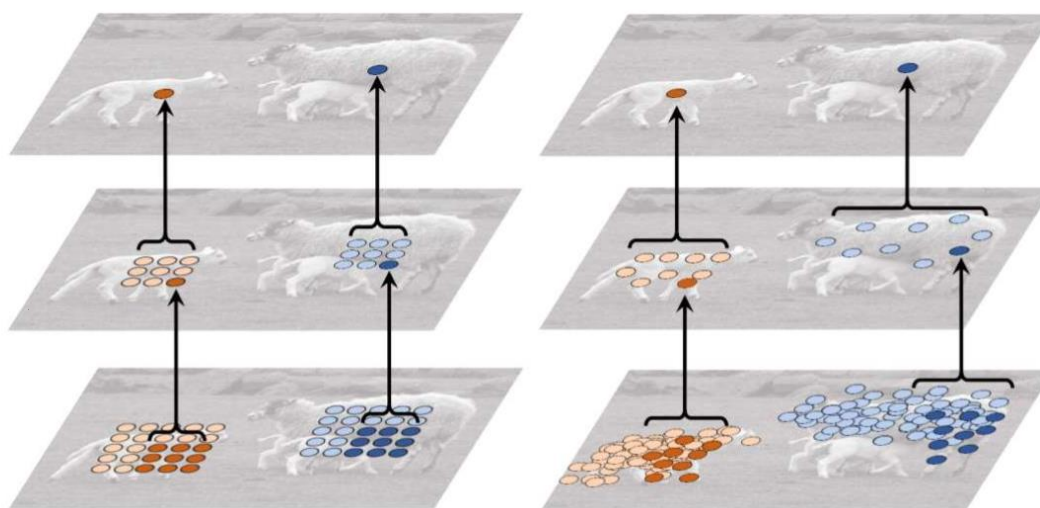


Figure 1: Illustration of the sampling locations in  $3 \times 3$  standard and deformable convolutions. (a) regular sampling grid (green points) of standard convolution. (b) deformed sampling locations (dark blue points) with augmented offsets (light blue arrows) in deformable convolution. (c)(d) are special cases of (b), showing that the deformable convolution generalizes various transformations for scale, (anisotropic) aspect ratio and rotation.

deformable convolutional تعداد کمی پارامتر و محاسبات را برای یادگیری افست اضافه می‌کنند و به راحتی می‌توان به صورت end-to-end با backpropagation استاندارد آن‌ها را آموزش داد.



(a) standard convolution

(b) deformable convolution

همچنین در شکل بالا به طور کامل میتوان تفاوت میان deformable و استاندارد کانولوشن را دید .

## Problem 2

### Flatten

با توجه به اینکه لایه آخر maxpooling است و ورودی  $16*16*512$  پس خروجی maxpooling به صورت  $8*8*512$  خواهد بود که این خروجی ورودی flatten می باشد که خروجی flatten هم برابر با ضرب ابعاد ورودی در هم به صورت بردار است ، یعنی برابر با  $8 * 8 * 512 = 32768$  خواهد بود ، همچنین flatten پارامتر ندارد . لایه خروجی هم با توجه به این که سی کلاس است در واقع یک FC با سی نورون خروجی و 32768 نورون ورودی خواهد بود و تعداد پارامتر های آن برابر است با:

$$983070 = (32768 + 1) * 30 \quad (+1 \text{ برای بایاس برای هر کلاس خروجی})$$

### GAP

GAP در واقع میانگین همه درایه های ماتریس ویژگی های ورودی اش را میگیرد و به نوعی فقط عمق ورودی را در خروجی میدهد ، اینجا با توجه به اینکه ورودی اش  $8*8*512$  است ، خروجی اش 512 خواهد بود و در واقع FC در نهایت 30 نورون خروجی و 512 نورون ورودی دارد و در مجموع تعداد پارامتر های آن برابر است با :

$$15390 = (512 + 1) * 30$$

### GAP با وزن ها یکسان برای تمام کانال ها

همانند مورد قبل عمل می کند ، پارامتر های این لایه برابر است با  $8 * 8 = 64$  و شکل خروجی 512 است. پارامتر های لایه آخر و شکل خروجی همانند مورد قبلی است .

### GAP با وزن ها متفاوت برای هر کانال

شکل ورودی و خروجی FC و تعداد پارامتر هایش همانند دو مورد قبلی است و تنها تفاوتش این است که بعضی کانال ها تاثیر بیشتری در نتیجه دارند و بعضی کمتر ( بسته به وزنشان ) تعداد پارامتر های لایه GWAP هم برابر است با :  $8*8*512 = 32768$

### هیستوگرام قابل آموزش با 4bin

ابعاد خروجی این لایه برابر است با  $4 * 512$  . لایه FC هم دارای 30 نورون خروجی است و  $4 * 512 = 2048$  نورون ورودی  
تعداد پارامتر های این لایه برابر است با :

$$4096 = 4 * 512$$

تعداد پارامتر های FC هم برابر است با :

$$61470 = (2048 + 1) * 30$$

## هیستوگرام قابل آموزش با 8bin

ابعاد خروجی این لایه برابر است با  $8 * 512$ . لایه FC هم دارای 30 نورون خروجی است و  $8 * 512 = 4096$  نورون ورودی

تعداد پارامترهای این لایه برابر است با :

$$512 * 8 * 2 = 8192$$

تعداد پارامترهای FC هم برابر است با :

$$(4096 + 1) * 30 = 122910$$

## مقایسه

Flatten مزیتی که دارد حفظ تمام ویژگی‌های لایه قبلی است اما این کار به بهای افزایش ابعاد لایه بعدی و هزینه محاسباتی آن می‌باشد. GAP اطلاعات زیادی دور می‌ریزد که باعث می‌شود کاهش ابعاد و پارامترها را داشته باشیم و هزینه محاسباتی کمتر. GWAP با وزن‌های مساوی همانند GAP است (هر چند تعداد پارامترهای بیشتری دارد) GWAP با وزن‌ها متفاوت اطلاعات و پارامترهای بیشتری را منتقل کرده و تولید می‌کند و بار محاسباتی دارد اما میتواند عیب دو مورد قبلی را (حذف اطلاعات کلیدی) را با تنظیم کردن وزن‌ها رفع کرد. معایب هیستوگرام قابل آموزشی، سربار محاسباتی آن است که میتوان با bin های اولیه آن را بهبود بخشید، در کل هم توانایی بالاتری برای مسایل computer vision دارد.