

Assignment NO.8 Solutions

Digital Image Processing | Fall 1400 | Dr.Mohammadi

Teacher Assistant : Fatemeh Anvari

Student name : **Amin Fathi**

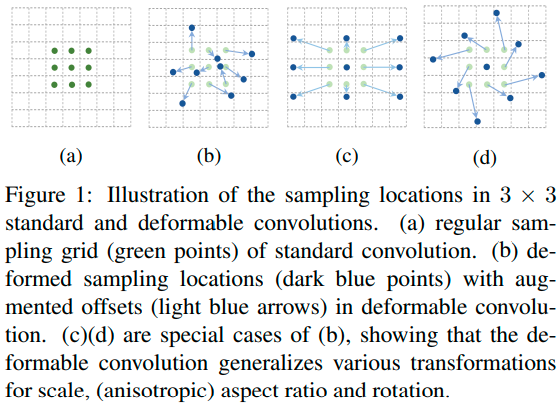
Student id : **400722102**

**Problem 1**

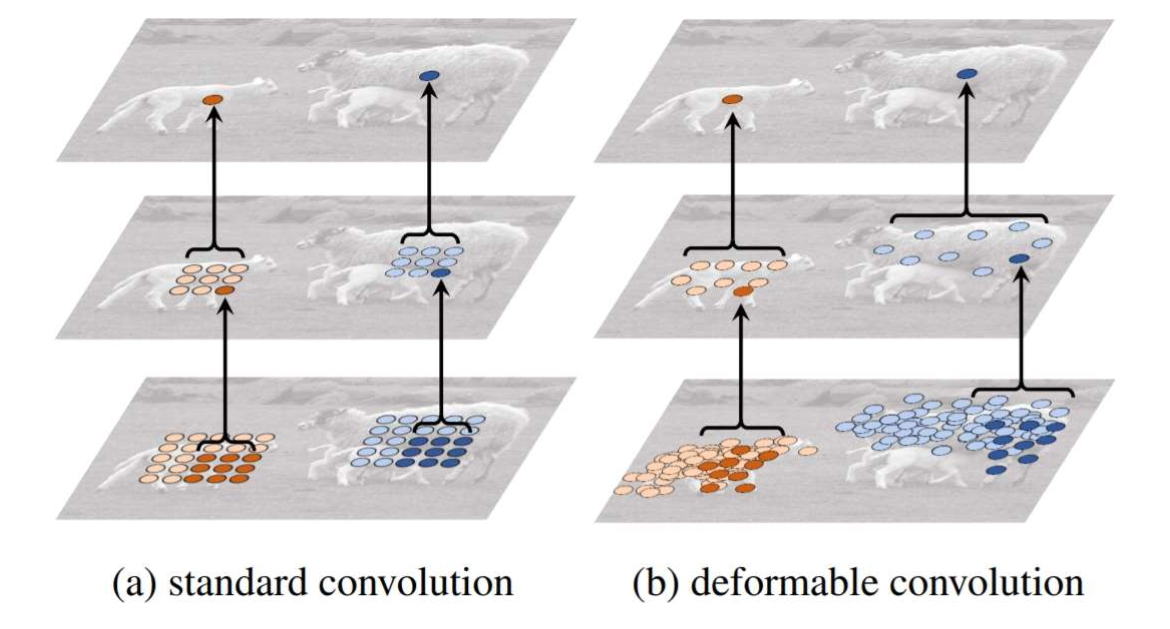
شبکه‌های عصبی کانولوشنال (CNN) به دلیل ساختارهای هندسی ثابت در ماژول‌های ساختمان خود، ذاتاً به مدل‌سازی تبدیل‌های هندسی محدود می‌شوند. دو ماژول جدید برای افزایش قابلیت مدل‌سازی تبدیل CNN معرفی می‌شود، یعنی کانولوشن تغییر شکل‌پذیر (Deformable Convolutional Networks) وRol pooling .

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 offsetهاي دو بعدي به مكان هاي نمونه گيري اضافه ميكند كه باعث ميشود اين شبكه و لايه ي كانولوشن فارغ از  
تغييرات هندسي شود. شكل زير به صورت واضح تري مفهوم deformable convolutionرا منتقل ميكند:



deformable convolutional تعداد کمی پارامتر و محاسبات را برای یادگیری افست اضافه می کنند و به راحتی می توان به صورت end-to-end با backpropagation استاندارد آن ها را آموزش داد.



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 برداراست ، 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 30 \* (32768 +1) ( +1 برای بایاس برای هر کلاس خروجی )

**GAP**

GAP در واقع میانگین همه درایه های ماتریس ویژگی های ورودی اش را میگیرد وبه 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 تعدا پارامتر های آن برابر است با :

30 \* (512 + 1) = 15390

**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. پرامتر های لایه اخر و شکل خروجی همانند مورد قبلی است .

**GAP با وزن ها متفاوت برای هرکانال**

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 512 \* 4 . لایه FC هم دارای 30 نورون خروجی است و 2048 = 512 \* 4 نورون ورودی

تعداد پارامتر های این لایه برابر است با :

512 \* 4 \* 2 = 4096

تعداد پارامتر های FC هم برابر است با :  
(2048 + 1 ) \* 30 = 61470

**هیستوگرام قابل آموزش با 8bin**

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

تعداد پارامتر های این لایه برابر است با :

512 \* 8 \* 2 = 8192

تعداد پارامتر های FC هم برابر است با :  
(4096 + 1 ) \* 30 = 122910

مقایسه

Flatten مزیتی که دارد حفظ تمام ویژگلی های لایه قبلی است اما این کار به بهای افزایش ابعاد لایه بعدی و هزینه محاسباتی آن می باشد .GAP اطلاعات زیادی دور می ریزد که باعث می شود کاهش ابعاد و پارامتر ها را داشته باشیم و هزینه محاسباتی کمتر . GWAP با وزن های مساوی همانند GAP است ( هر چند تعداد پارامتر های بیشتری دارد ) GWAP با وزن ها متفاوت اطلاعات و پارامتر های بیشتری را منتقل کرده و تولید می کند و بار محاسباتی دارد اما میتواند عیب دو مورد قبلی را ( حذف اطلاعات کلیدی ) را با تنظیم کردن وزن ها رفع کرد . معایب هیستوگرام قابل اموزشی ، سربار محاسباتی آن است که میتوان با bin های اولیه آن را بهبود بخشید ، در کل هم توانایی بالاتری برای مسایل computer vision دارد .