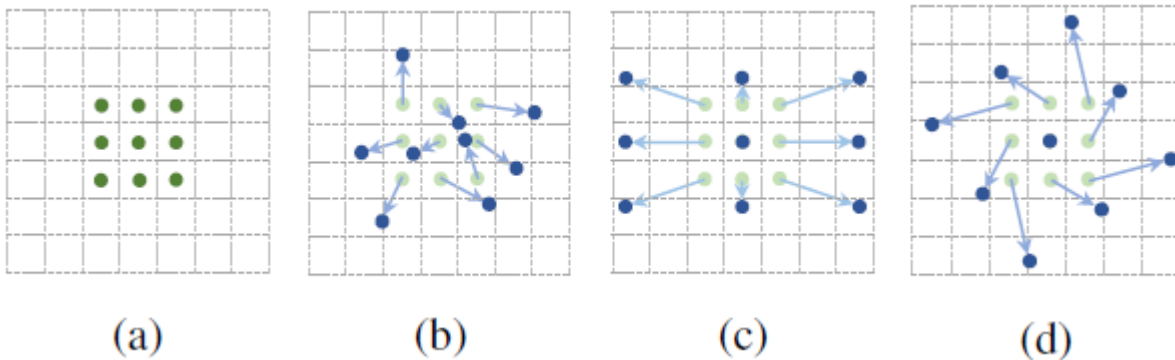


در deformable convolution به منظور داشتن receptive field های متفاوت افست های دو بعدی به مکان های نمونه گیری منظم grid در کانولوشن استاندارد افزوده می شود.



در شکل فوق در قسمت a یک کانولوشن استاندارد را می بینیم که به صورت معمولی مطابق با نقاط کانولوشن استاندارد نمونه برداری می شود یعنی وزن های کانولوشن را در نقاط منظم (سبز) ضرب و نتایج را جمع می کند. اما در deformable convolution به جای نقاط سبز آفست هایی دو بعدی به مختصات نقاط کانولوشن استاندارد (نقاط سبز) افزوده می شود که این آفست ها قابل یادگیری هستند و وزن های کانولوشن در نقاط جدید ضرب و نتایج جمع می شود (b, c, d)

$$\mathcal{R} = \{(-1, -1), (-1, 0), \dots, (0, 1), (1, 1)\}$$

با توجه به کرنل R فوق کانولوشن استاندارد به صورت زیر محاسبه می گردد:

$$y(\mathbf{p}_0) = \sum_{\mathbf{p}_n \in \mathcal{R}} w(\mathbf{p}_n) \cdot x(\mathbf{p}_0 + \mathbf{p}_n)$$

اما در deformable convolution مقدار  $\Delta \mathbf{p}_n$  که همان آفست هست به  $\mathbf{p}_0 + \mathbf{p}_n$  افزوده می شود و حاصل نهایی به صورت زیر محاسبه می شود.

$$y(\mathbf{p}_0) = \sum_{\mathbf{p}_n \in \mathcal{R}} w(\mathbf{p}_n) \cdot x(\mathbf{p}_0 + \mathbf{p}_n + \Delta \mathbf{p}_n).$$

برای اجرای معادله فوق از درونیابی دوخطی (bilinear interpolation) استفاده می کنیم زیرا با اضافه کردن آفست به موقعیت های نمونه موجود، نقاط اعشاری را به دست می آوریم که مکان های مشخصی در شبکه

نیستند و برای تخمین مقادیر پیکسل آن‌ها از درونیایی دوخطی استفاده می‌کنیم که از شبکه  $2 \times 2$  پیکسل مجاور برای تخمین مقدار موقعیت پیکسل تغییر شکل یافته جدید استفاده می‌کند.

$$\mathbf{x}(\mathbf{p}) = \sum_{\mathbf{q}} G(\mathbf{q}, \mathbf{p}) \cdot \mathbf{x}(\mathbf{q}),$$

که  $\mathbf{p} = \mathbf{p}_0 + \mathbf{p}_n + \Delta \mathbf{p}_n$  و  $\mathbf{q}$  پیکسل‌های مجاور هستند و  $G$  کرنل درونیایی دو خطی است.  $G$  دو بعدی است و می‌توانیم به صورت زیر آن را به دو کرنل یک بعدی تقسیم می‌شود

$$G(\mathbf{q}, \mathbf{p}) = g(q_x, p_x) \cdot g(q_y, p_y)$$

(۲)

(الف)

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)	

ابتدا باید shape خروجی block5\_pool را محاسبه کرد:

$\text{block5\_pool.shape} = 8, 8, 512$

Flatten:

$\text{flatten\_output.shape} = 8 * 8 * 512 = 32768$

$\text{flatten\_number\_of\_parameters} = 0$

$\text{dense\_number\_of\_parameters} = 30 * (32768 + 1) = 983070$

GAP:

$\text{gap\_output.shape} = 512$

$\text{gap\_number\_of\_parameters} = 0$

$\text{dense\_number\_of\_parameters} = 30 * (512 + 1) = 1530$

GWAP: با وزن یکسان برای تمام کانالها

$\text{GWAP\_e\_output.shape} = 512$

$\text{GWAP\_e\_number\_of\_parameters} = 8 * 8 = 64$

$\text{dense\_number\_of\_parameters} = 30 * (512 + 1) = 1530$

GWAP: با وزن متفاوت برای هر کانال

$\text{GWAP\_d\_output.shape} = 512$

$\text{GWAP\_d\_number\_of\_parameters} = 8 * 8 * 512 = 32768$

$\text{dense\_number\_of\_parameters} = 30 * (512 + 1) = 1530$

$\text{bin} = 4$ : هیستوگرام قابل آموزش با

Histogram\_layer\_parameters:  $512 * 4 * 2 = 4096$

Histogram\_layer\_output.shape =  $512 * 4 = 2048$

$\text{dense\_number\_of\_parameters} = 30 * (2048 + 1) = 61470$

$\text{bin} = 8$ : هیستوگرام قابل آموزش با

$\text{Histogram\_layer\_parameters}: 512 * 8 * 2 = 8192$

$\text{Histogram\_layer\_output.shape} = 512 * 8 = 4096$

$\text{dense\_number\_of\_parameters} = 30 * (4096 + 1) = 122910$

(ب)

**Flatten:**

مزیت: اطلاعات مکانی را دور نمی‌ریزد و می‌تواند با استفاده از آن‌ها عملکرد بهتری داشته باشد و مثلاً تشخیص می‌دهد یک چیزی در بالای تصویر است یا پایین تصویر

عیب: تعداد پارامترهای لایه کاملاً متصل بعد از آن بسیار زیاد است.

**GAP:**

مزیت: تعداد پارامترهای لایه کاملاً متصل بعد از آن را به شدت کاهش می‌دهد ولی با این وجود عملکرد خوبی دارد.

عیب: اطلاعات مکانی را دور می‌ریزد. ضمن اینکه ممکن است با توجه به داده و وظیفه مورد نظر میانگین گرفتن مناسب نباشد و انتخاب  $\max$  مناسبتر باشد.

**GWAP** با وزن یکسان برای تمام کانالها:

مزیت: تعداد پارامترهای لایه کاملاً متصل بعد از آن را به شدت کاهش می‌دهد ولی با این وجود عملکرد خوبی دارد. نسبت به **GAP** اطلاعات مکانی بیشتری را ذخیره می‌کند با اینکه پارامترهای افزوده شده زیاد نیست.

عیب: اطلاعات مکانی را دور می‌ریزد. روی اطلاعات تمام کانال‌ها میانگین گرفته می‌شود و اهمیت هر کانال به یک اندازه در نظر گرفته می‌شود. ضمن اینکه ممکن است با توجه به داده و وظیفه مورد نظر میانگین گرفتن مناسب نباشد و انتخاب  $\max$  مناسبتر باشد.

**GWAP** با وزن متفاوت برای هر کانال:

مزیت: تعداد پارامترهای لایه کاملاً متصل بعد از آن را به کاهش می‌دهد. نسبت به GWAP با وزن یکسان برای تمام کانالها اطلاعات مکانی بیشتری را ذخیره می‌کند اطلاعات هر کانال مستقل از کانالهای دیگر است. عملکرد آن به شدت نسبت به GAP و GWAP با وزن یکسان برای تمام کانالها بهتر است.

معایب: ممکن است با توجه به داده و وظیفه مورد نظر میانگین گرفتن مناسب نباشد و انتخاب max مناسبتر باشد. کمی از اطلاعات مکانی دور ریخته می‌شود.

هیستوگرام قابل آموزش با  $\text{bin}=4$ :

مزیت: به جای میانگین گرفتن هیستوگرام را ذخیره می‌کند بنابراین اطلاعات بهتری را نگه می‌دارد. تعداد پارامترها برای لایه کاملاً متصل بعد نسبت به flatten کاهش میابد اما اطلاعات بهتری نسبت به روشهای GAP و حالت‌های متفاوت GWAP ذخیره می‌کند.

معایب: پارامترها برای لایه کاملاً متصل بعد از آن از روش GAP و حالت‌های متفاوت GWAP بسیار بیشتر است. اطلاعات مکانی کمتری نسبت به flatten دارد.

هیستوگرام قابل آموزش با  $\text{bin}=8$ :

مزیت: نسبت به هیستوگرام قابل آموزش با  $\text{bin}=4$  با بیشتر شدن bin عملکرد بهتری دارد. تعداد پارامترها برای لایه کاملاً متصل بعد نسبت به flatten کاهش میابد اما اطلاعات بهتری نسبت به روشهای GAP و حالت‌های متفاوت GWAP ذخیره می‌کند.

معایب: نسبت به هیستوگرام قابل آموزش با  $\text{bin}=4$  با بیشتر شدن bin پارامترهای بیشتری دارد. اطلاعات مکانی کمتری نسبت به flatten دارد.

(۳)

شبکه unet عملکرد خوبی نسبت به dlinknet ندارد اما سریعتر است. شبکه dlinknet به دلیل استفاده از resnet50 حدوداً ۲ برابر پارامتر دارد و آموزش آن به زمان بیشتری نیاز دارد و عملکرد موفق‌تری دارد.