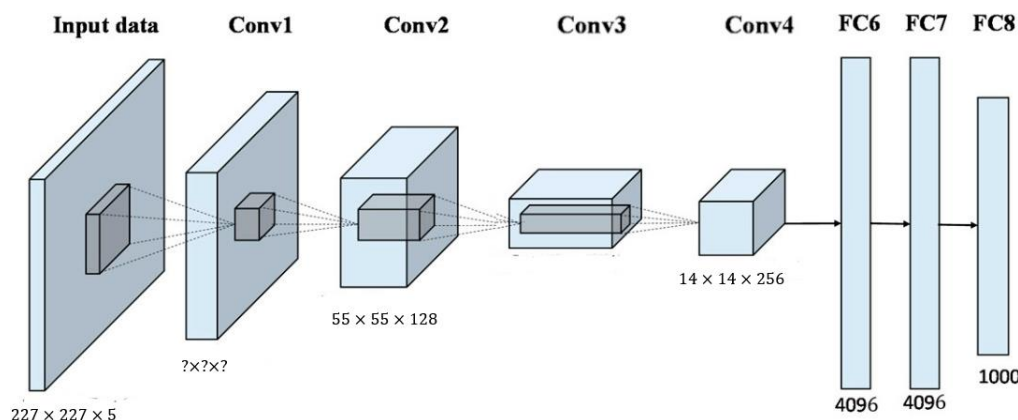




### تمرین اول: (۴۵ نمره)

شبکه ی زیر را که ساختاری مشابه شبکه Alex-Net دارد اما نوع و تعداد فیلتر های آن متفاوت است در نظر بگیرید:



الف) اگر طبق شکل ابعاد ورودی شبکه پس از اعمال padding ،  $227 \times 227 \times 5$  باشد و یک مجموعه 64 تایی فیلتر  $7 \times 7$  با  $\text{stride}=2$  به آن اعمال شود، ابعاد خروجی این بخش را محاسبه کنید. (ابعاد feature map لایه conv1) (۵ نمره)

ب) تعداد ضرب های لازم برای ایجاد لایه خروجی بعدی (conv1) را به دست آورید. (ضرب در مقادیر صفر را نیز در نظر بگیرید). (۵ نمره)

ج) اگر در گذر از لایه conv1 به conv2 ابتدا از یک لایه  $3 \times 3$  maxpooling با  $\text{stride} = 2$  بگذریم و سپس با pad-zero کردن به اندازه 2 از هر طرف و اعمال 128 فیلتر، خروجی ما  $55 \times 55 \times 128$  باشد، ابعاد فیلتر ها و میزان stride مورد استفاده این بخش را بیابید. (۵ نمره)

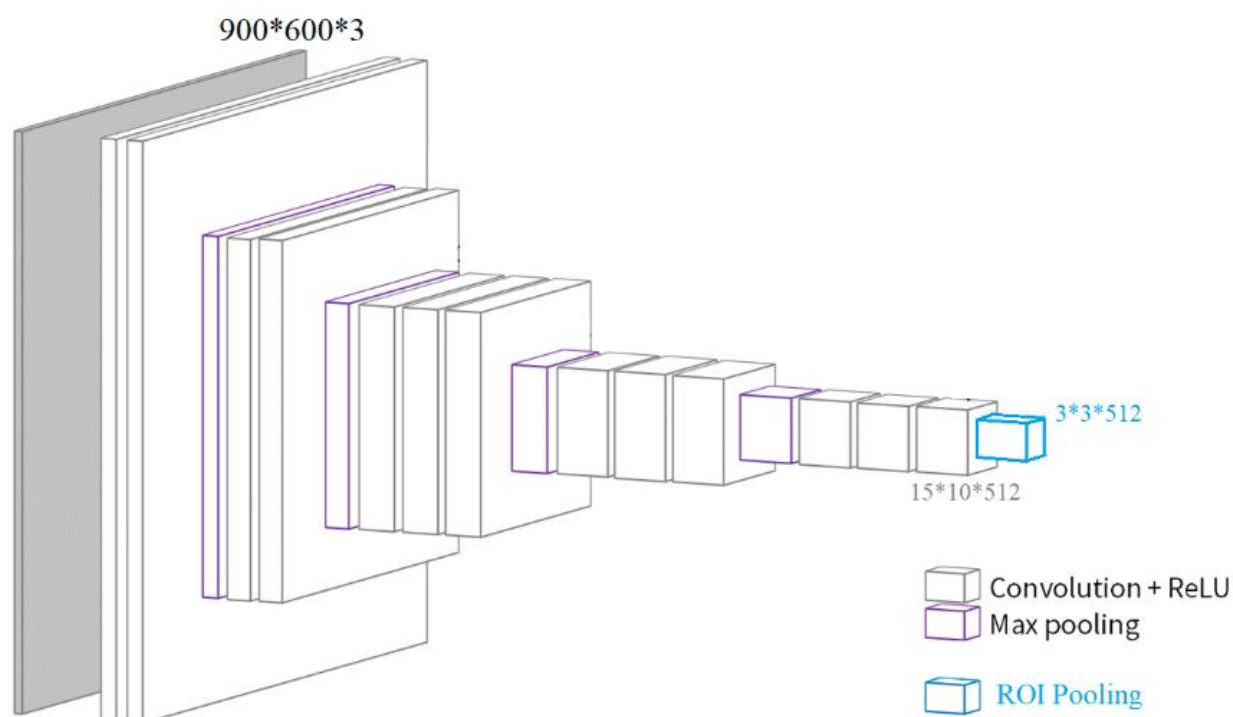
د) اگر ابعاد feature map لایه conv4 پس از گذر از فیلترهایی به سائز  $3 \times 3$  و  $\text{stride} = 1$  (با اعمال zero-pad به اندازه 1 از هر طرف قبل اعمال فیلتر ها)  $14 \times 14 \times 256$  شده باشد، فیلتر هایی که از لایه conv2 به conv3 استفاده شده اند را مشخص کنید. (برای این بخش الزاما جواب یکتا وجود ندارد و ترکیبی از چند عملیات بین دو لایه میتواند مورد استفاده قرار بگیرد، صرفا میدانیم عمق لایه conv3 384 است). (۷ نمره)

و) تعداد پارامتر های قابل یادگیری لایه ی FC6,FC8 را محاسبه کنید. (۸ نمره)

ه) با توجه به ساختار کامل شده، تعداد کل پارامترهای قابل یادگیری شبکه را محاسبه کنید. (۱۵ نمره)

## تمرین دوم: (۵۵ نمره)

بخشی از یک شبکه عصبی را در شکل زیر مشاهده میکنید. ورودی شبکه تصویر  $(w = 900, h = 600)$  و  $900 \times 600 \times 3$  و ابعاد feature map در ورودی لایه ROI Pooling مطابق شکل  $15 \times 10 \times 512$  میباشد که در شکل دوم مقادیر نرمال شده گوشه بالا چپ یکی از این feature map 512 به شما داده شده است.



شکل ۱ - بخشی از یک شبکه عصبی

0.2	0.4	0.9	0.15	0.5	0.2	0.3	0.35
0.75	0.1	0.2	0.7	0.55	0.25	0.8	0.6
0.9	0.35	0.6	0.45	0.3	0.15	0.4	0.75
0.3	0.8	0.5	0.6	0.2	0.9	0.1	0.85
0.1	0.7	0.35	0.8	0.4	0.55	0.5	0.7
0.4	0.2	0.9	0.7	0.25	0.3	0.6	0.8
0.3	0.55	0.7	0.4	0.1	0.7	0.75	0.5
0.6	0.4	0.35	0.8	0.35	0.2	0.9	0.55

شکل ۲ - مقادیر گوشه بالا چپ feature map در ورودی لایه ROI Pooling

با استفاده از روش selective search توانسته ایم proposal های متفاوتی از تصویر استخراج کنیم. یکی از این proposal ها در موقعیت  $(x_{top-left}, y_{top-left}, w, h) = (160, 20, 300, 260)$  قرار دارد. (x محور افقی و y محور عمودی تصویر است).

الف) میخواهیم خروجی لایه ROI Pooling برای این proposal را حساب کنیم. به همین منظور: (۲۵ نمره)

۱. مختصات ۴ گوشه باکس مورد نظر را در feature map بدست آورید. اگر اعداد بدست آمده نیاز به رند کردن داشتند از این قاعده استفاده کنید:

مرزهای بالا و چپ را به بالا (ceiling) و مرزهای پایین و راست به پایین (floor) رند کنید.

مختصات جدید ۴ گوشه باکس را که حالا به صورت دقیق بر روی گرید feature map قرار گرفته اند گزارش کنید.

۲. با توجه به اندازه خروجی لایه ROI Pooling گرید مناسبی بر روی ROI بدست آمده در قسمت قبل فیت کنید. از آنجا که احتمال دارد ابعاد bin های گرید نیاز به رند کردن داشته باشند از این قاعده استفاده کنید:

برای هر bin، مرزهای بالا و چپ را به پایین (floor) و مرزهای پایین و راست را به بالا (ceiling) رند کنید.

محدوده bin ها را بر روی قسمتی از feature map که مربوط به proposal است نشان دهید.

۳. ماکزیمم مقادیر هر bin را بدست آورده و خروجی لایه ROI Pooling را مشخص کنید.

ب) همانطور که میدانیم اگر به جای لایه ROI Pooling از لایه ROI Align استفاده کنیم، دیگر نیازی به رند کردن اعداد در دو مرحله اول نیست و مقادیر با دقت بین پیکسلی محاسبه میشوند. میخواهیم روند ایجاد خروجی را برای bin (0,0) طی کنیم. (۳۰ نمره)

۱. گرید مناسب را بر روی مختصات اولیه بدست آورده در الف-۱ فیت کنید. و مقادیر مرزهای bin ها را گزارش کنید.

۲. برای bin (0,0) ۴ نقطه با فواصل یکسان  $\frac{h}{3}, \frac{w}{3}$  مشخص کنید و مختصات این نقاط را بنویسید.

۳. با استفاده از bilinear interpolation مقادیر feature map در این نقاط را بدست آورید.

۴. خروجی این bin از ماکزیمم گیری بین ۴ مقدار محاسبه شده بدست می آید، آن را گزارش کنید.