

INF721

2023/2

UFV

Aprendizado em Redes Neurais Profundas

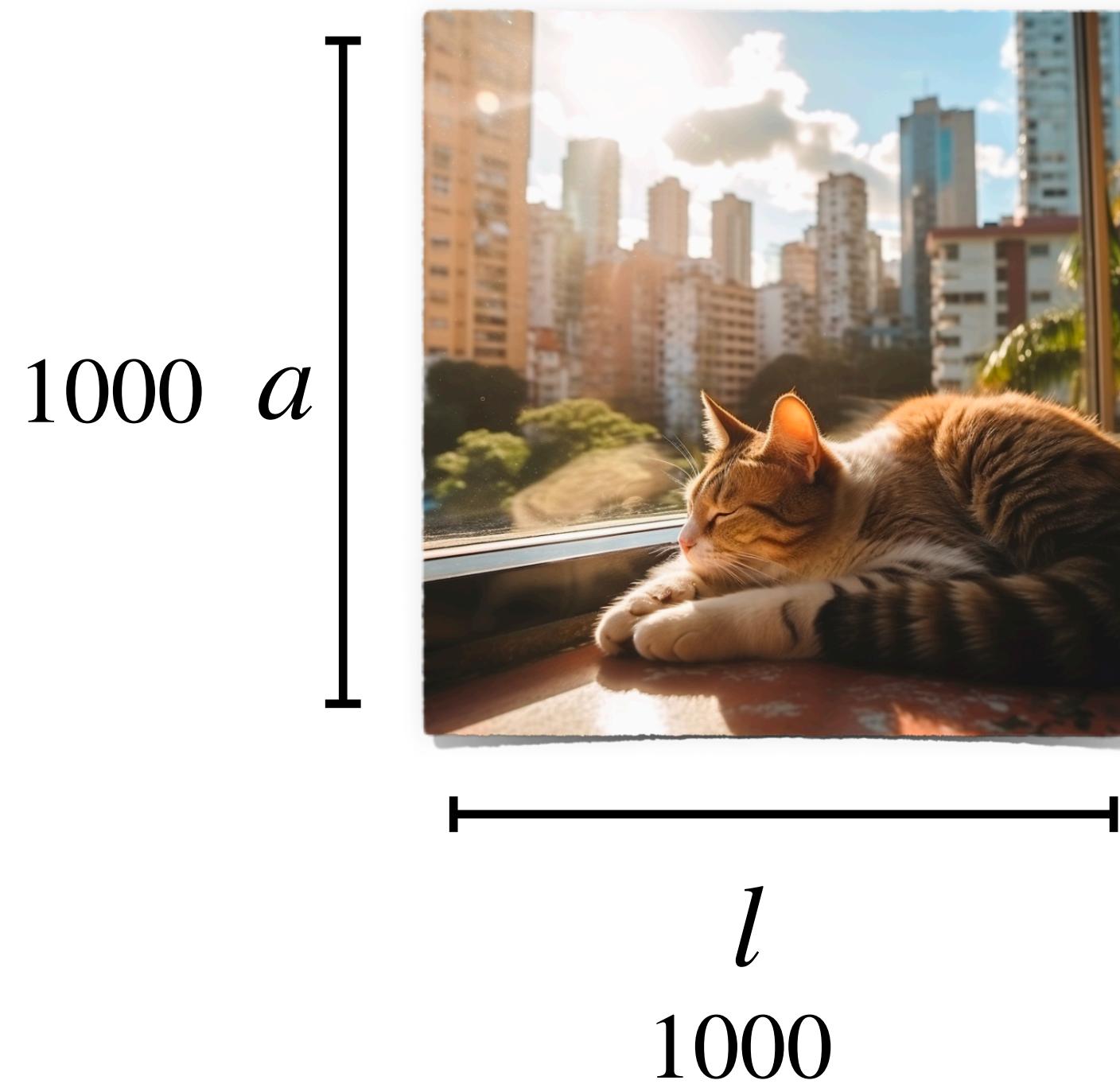
A13: Redes Neurais Convolucionais

Plano de Aula

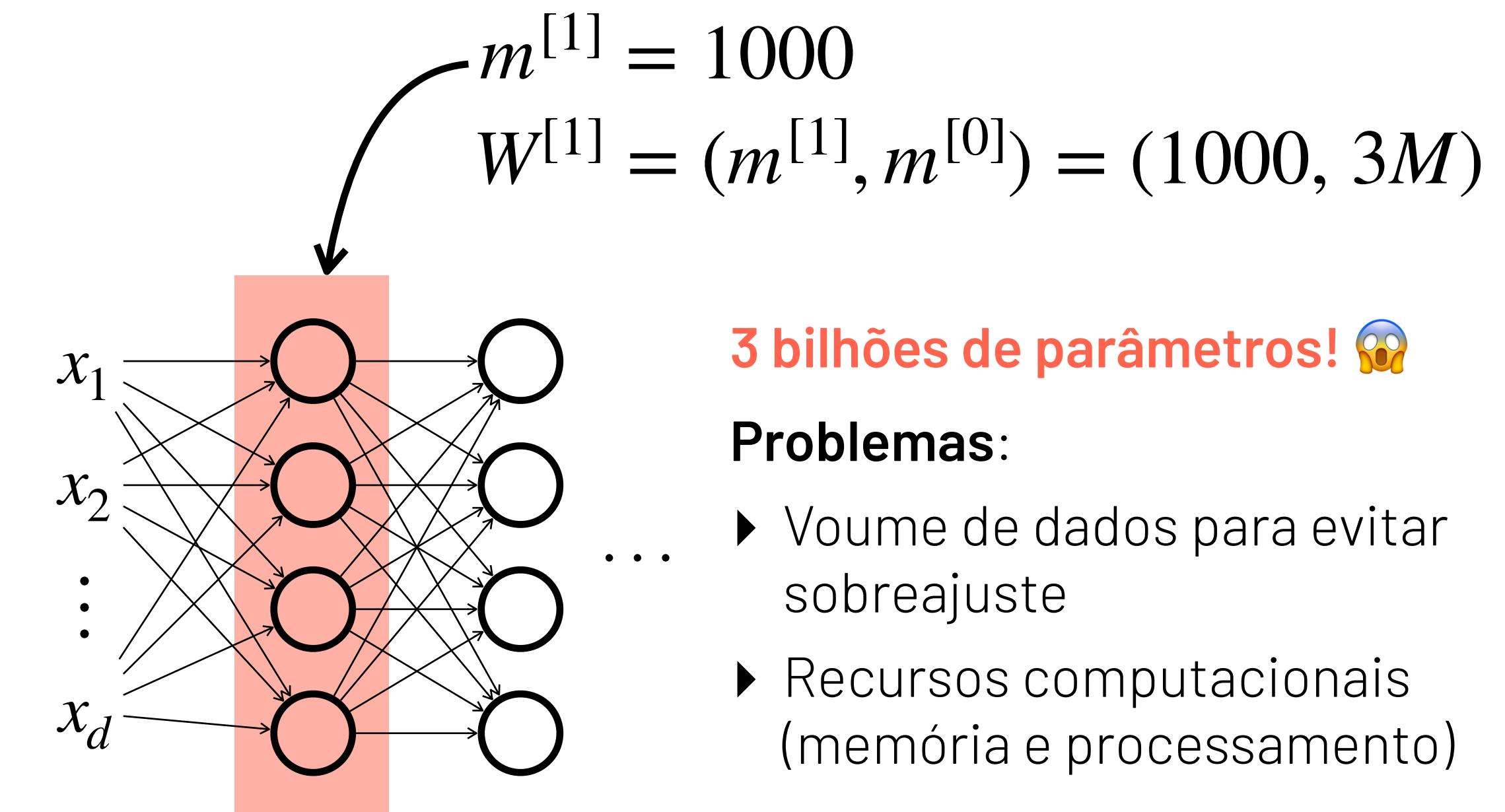
- ▶ Explosão de Parâmetros
- ▶ Filtros
- ▶ Convoluçãoes
 - ▶ Preenchimento (Padding)
 - ▶ Convoluçãoes Passadas (Strided Convolutions)
- ▶ Convoluçãoes em Volumes
- ▶ Redes Neurais Convolucionais (CNN)

Explosão de parâmetros

Para processar imagens com MLPs, temos que transformá-las em vetores de características.



$$x = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_d \end{bmatrix}$$



$$d = a \times l \times 3$$

$$d = 1000 \times 1000 \times 3 = 3M$$

3 bilhões de parâmetros! 😱

Problemas:

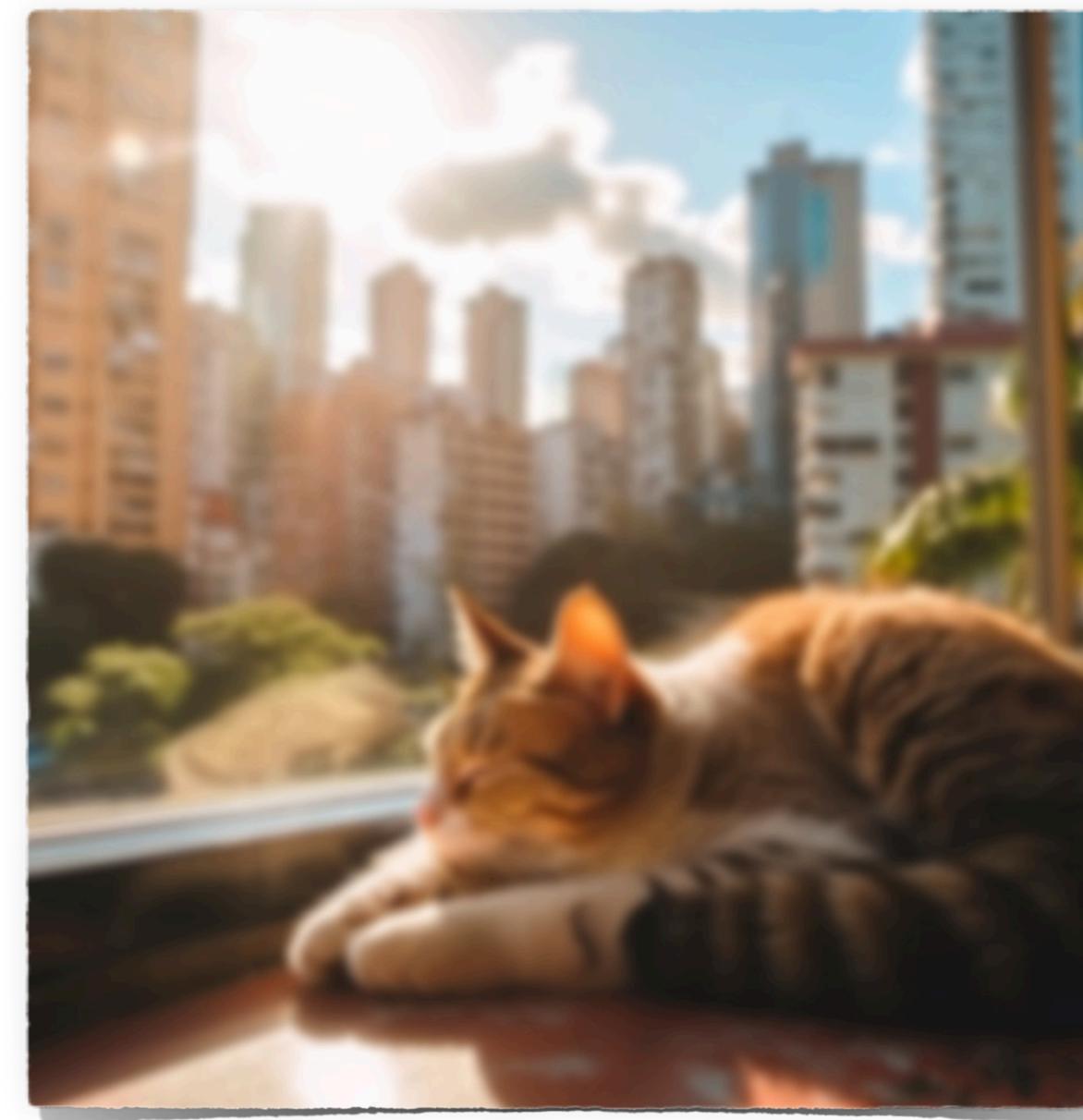
- ▶ Volume de dados para evitar sobreajuste
- ▶ Recursos computacionais (memória e processamento)

Convoluçãoes

Em processamento de imagens e visão computacional, convoluções são operações para aplicar **filtros** (i.e., transformações) em imagens.



Borrar
→
(blur)



Filtros

Um **filtro** (ou kernel) é uma pequena matriz (geralmente 3x3) de pesos que transforma um **pixel** com a soma ponderada dos pixels de sua vizinhança.

	206	205	247	
	144	161	137	
	192	154	75	

Pixel original (161)
e sua vizinhança

0,0625	0,125	0,0625
0,125	0,25	0,125
0,0625	0,0125	0,0625

*

Filtro (blur)

$$= \sum_{i=1}^3 \sum_{j=1}^3 = m_{i,j} * k_{i,j} =$$

$$\begin{aligned} & 206 * 0,0625 + 205 * 0,125 + 247 * 0,0625 + \\ & 144 * 0,125 + 161 * 0,25 + 137 * 0,125 + \\ & 192 * 0,0625 + 154 * 0,125 + 75 * 0,0625 = \\ & 178 \end{aligned}$$

	206	205	247	
	144	178	137	
	192	154	75	

Pixel transformado (178)
e sua vizinhança

Convoluçãoes

Uma **convolução** é uma operação entre uma imagem e um filtro que consiste em aplicar um filtro para cada pixel de uma imagem.

206	205	247	245	244
244	161	137	244	254
192	154	75	200	249
90	109	96	143	223
67	69	107	196	236

Imagen Original
 (5×5)

$$* \begin{matrix} 0,0625 & 0,125 & 0,0625 \\ 0,125 & 0,25 & 0,125 \\ 0,0625 & 0,0125 & 0,0625 \end{matrix}$$

Filtro (blur)

$$= \sum_{i=1}^3 \sum_{j=1}^3 = m_{i,j} * k_{i,j} =$$

Imagen Transformada
 (3×3)

Convoluçãoes

Uma **convolução** é uma operação entre uma imagem e um filtro que consiste em aplicar um filtro para cada pixel de uma imagem.

206	205	247	245	244
244	161	137	244	254
192	154	75	200	249
90	109	96	143	223
67	69	107	196	236

Imagen Original
(5 × 5)

$$* \quad \begin{matrix} 0,0625 & 0,125 & 0,0625 \\ 0,125 & 0,25 & 0,125 \\ 0,0625 & 0,0125 & 0,0625 \end{matrix}$$

Filtro (blur)

$$= \sum_{i=1}^3 \sum_{j=1}^3 = m_{i,j} * k_{i,j} =$$

178		

Imagen Transformada
(3 × 3)

Convoluçãoes

Uma **convolução** é uma operação entre uma imagem e um filtro que consiste em aplicar um filtro para cada pixel de uma imagem.

206	205	247	245	244
244	161	137	244	254
192	154	75	200	249
90	109	96	143	223
67	69	107	196	236

Imagen Original
 (5×5)

*

0,0625	0,125	0,0625
0,125	0,25	0,125
0,0625	0,0125	0,0625

Filtro (blur)

$$= \sum_{i=1}^3 \sum_{j=1}^3 = m_{i,j} * k_{i,j} =$$

178	175	

Imagen Transformada
 (3×3)

Convoluçãoes

Uma **convolução** é uma operação entre uma imagem e um filtro que consiste em aplicar um filtro para cada pixel de uma imagem.

206	205	247	245	244
244	161	137	244	254
192	154	75	200	249
90	109	96	143	223
67	69	107	196	236

Imagen Original
 (5×5)

*

0,0625	0,125	0,0625
0,125	0,25	0,125
0,0625	0,0125	0,0625

Filtro (blur)

$$= \sum_{i=1}^3 \sum_{j=1}^3 = m_{i,j} * k_{i,j} =$$

178	175	216

Imagen Transformada
 (3×3)

Convoluçãoes

Uma **convolução** é uma operação entre uma imagem e um filtro que consiste em aplicar um filtro para cada pixel de uma imagem.

206	205	247	245	244
244	161	137	244	254
192	154	75	200	249
90	109	96	143	223
67	69	107	196	236

Imagen Original
(5 × 5)

0,0625	0,125	0,0625
0,125	0,25	0,125
0,0625	0,0125	0,0625

*

Filtro (blur)

$$= \sum_{i=1}^3 \sum_{j=1}^3 = m_{i,j} * k_{i,j} =$$

178	175	216
141		

Imagen Transformada
(3 × 3)

Convoluçãoes

Uma **convolução** é uma operação entre uma imagem e um filtro que consiste em aplicar um filtro para cada pixel de uma imagem.

206	205	247	245	244
244	161	137	244	254
192	154	75	200	249
90	109	96	143	223
67	69	107	196	236

Imagen Original
 (5×5)

*

0,0625	0,125	0,0625
0,125	0,25	0,125
0,0625	0,0125	0,0625

Filtro (blur)

$$= \sum_{i=1}^3 \sum_{j=1}^3 = m_{i,j} * k_{i,j} =$$

178	175	216
141	133	183
106	117	167

Imagen Transformada
 (3×3)

Detecção de borda

Filtros podem ser utilizados para detecção de borda em imagens, o que é particularmente importante para extração de características.

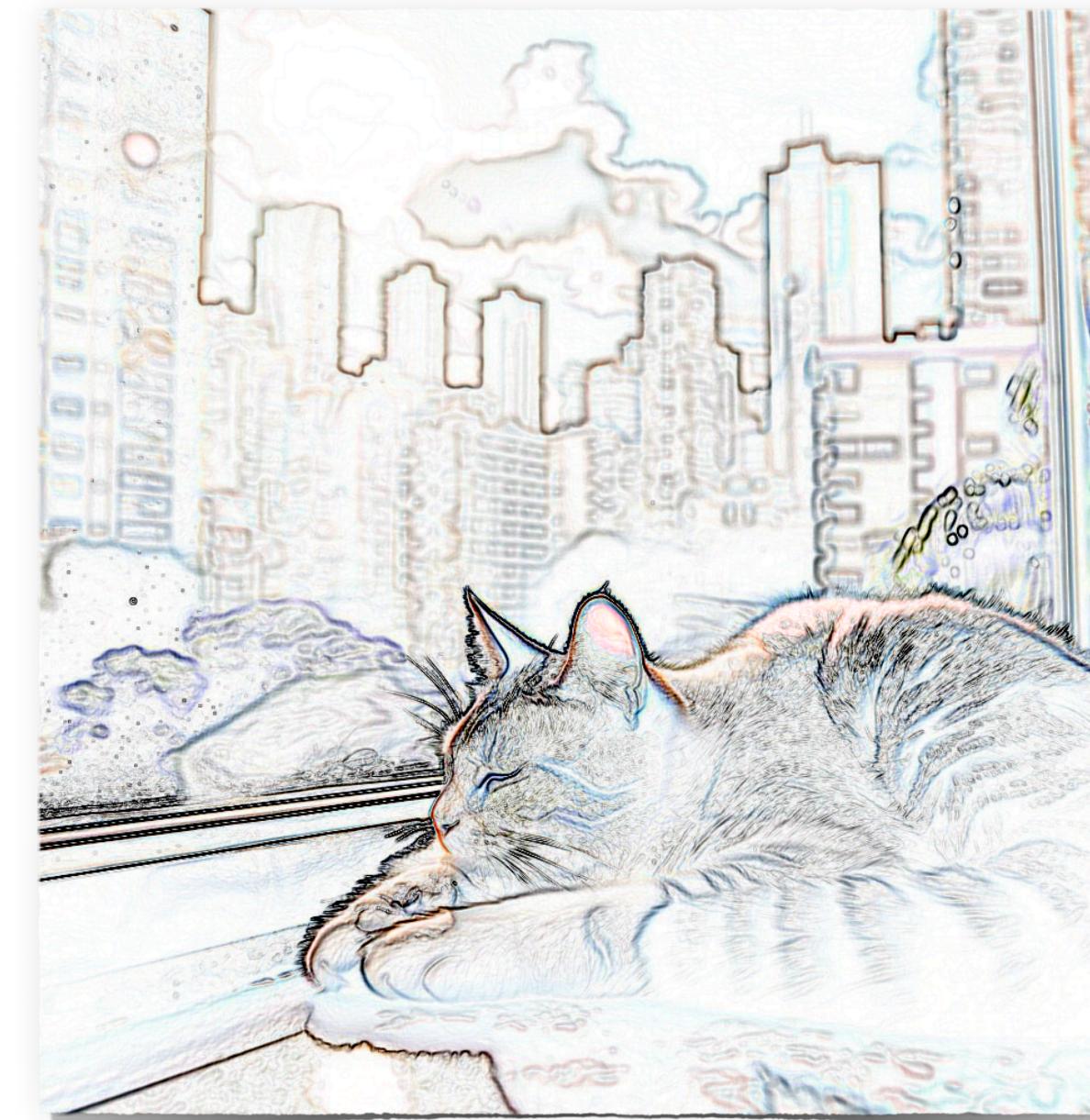


1	0	-1
1	0	-1
1	0	-1

Vertical

1	1	1
0	0	0
-1	-1	-1

Horizontal



Desenvolvendo Filtros

Diferentes filtros de detecção de borda foram desenvolvidos científicamente pela comunidade de processamento de imagens.

1	0	-1
1	0	-1
1	0	-1

1	0	-1
2	0	-2
1	0	-1

Sobel

3	0	-3
10	0	-10
3	0	-3

Scharr

Aprendendo Filtros

Redes Neurais Convolucionais (CNNs) **aprendem filtros** a partir de imagens e uma função de erro por meio do gradiente descendente.



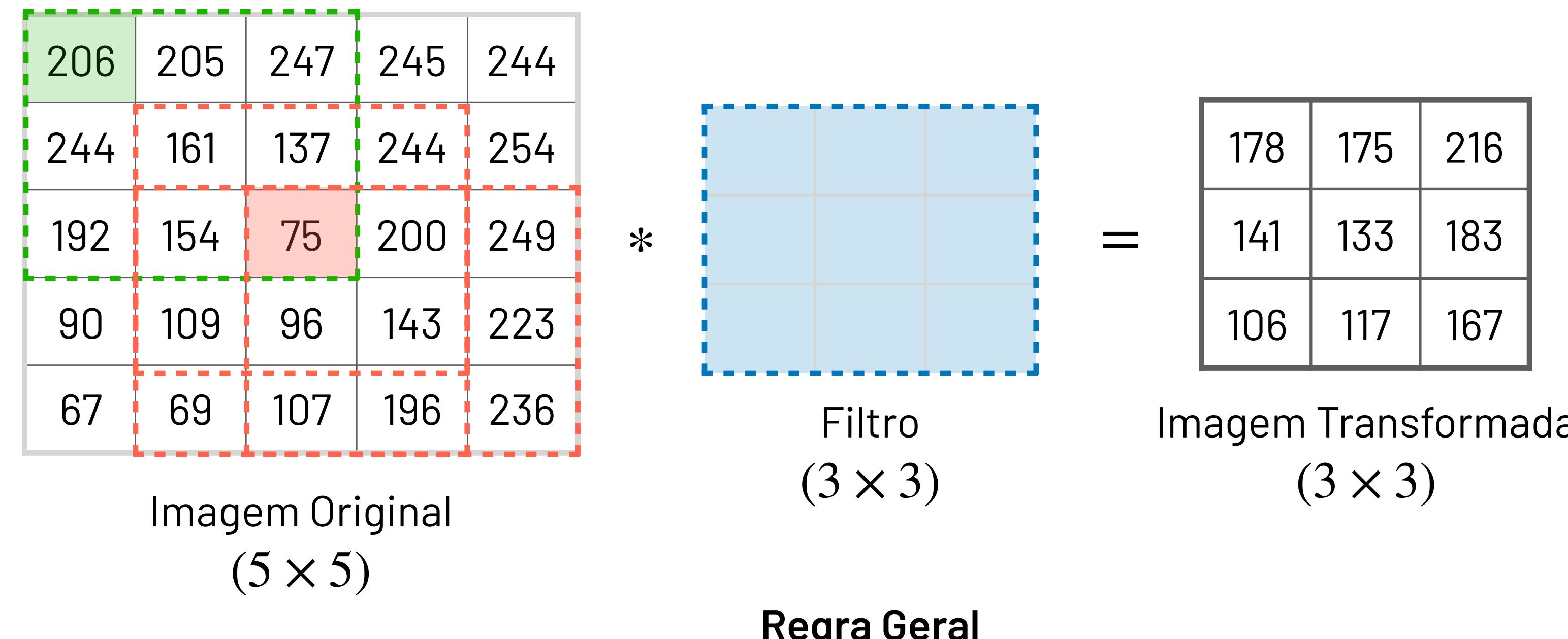
*

W1	W2	W3
W4	W5	W6
W7	W8	W9

Os pesos da uma CNN são organizados em filtros de convolução

Convoluçãoes reduzem o tamanho da imagem

- ▶ Aplicações consecutivas de convoluções podem tornar a imagem muito pequena (e.g., 1x1)
- ▶ Pixels dos cantos são menos compartilhados que pixels do meio



Preenchimento (Padding)

Adicionar uma **borda** com p pixels na imagem original.

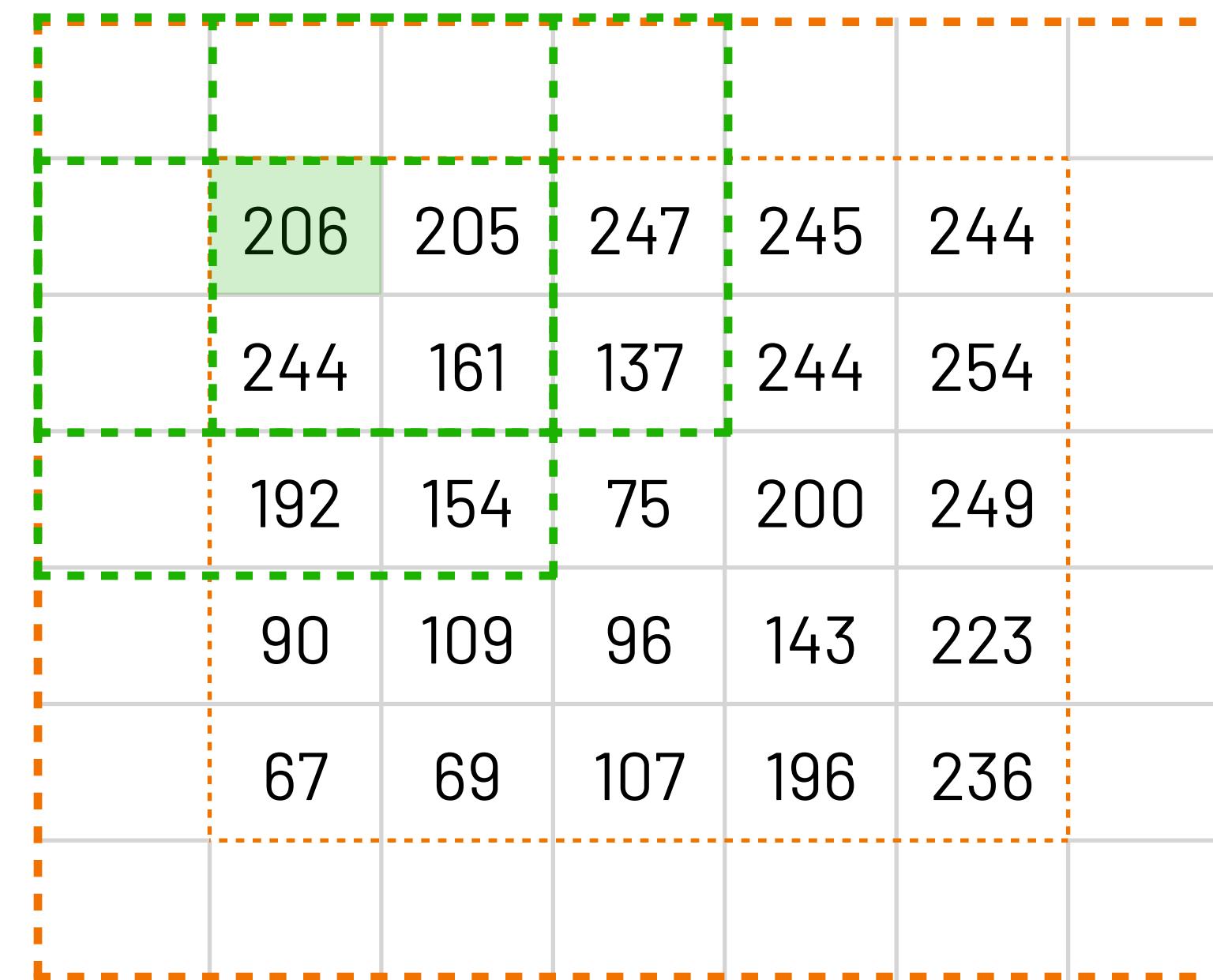
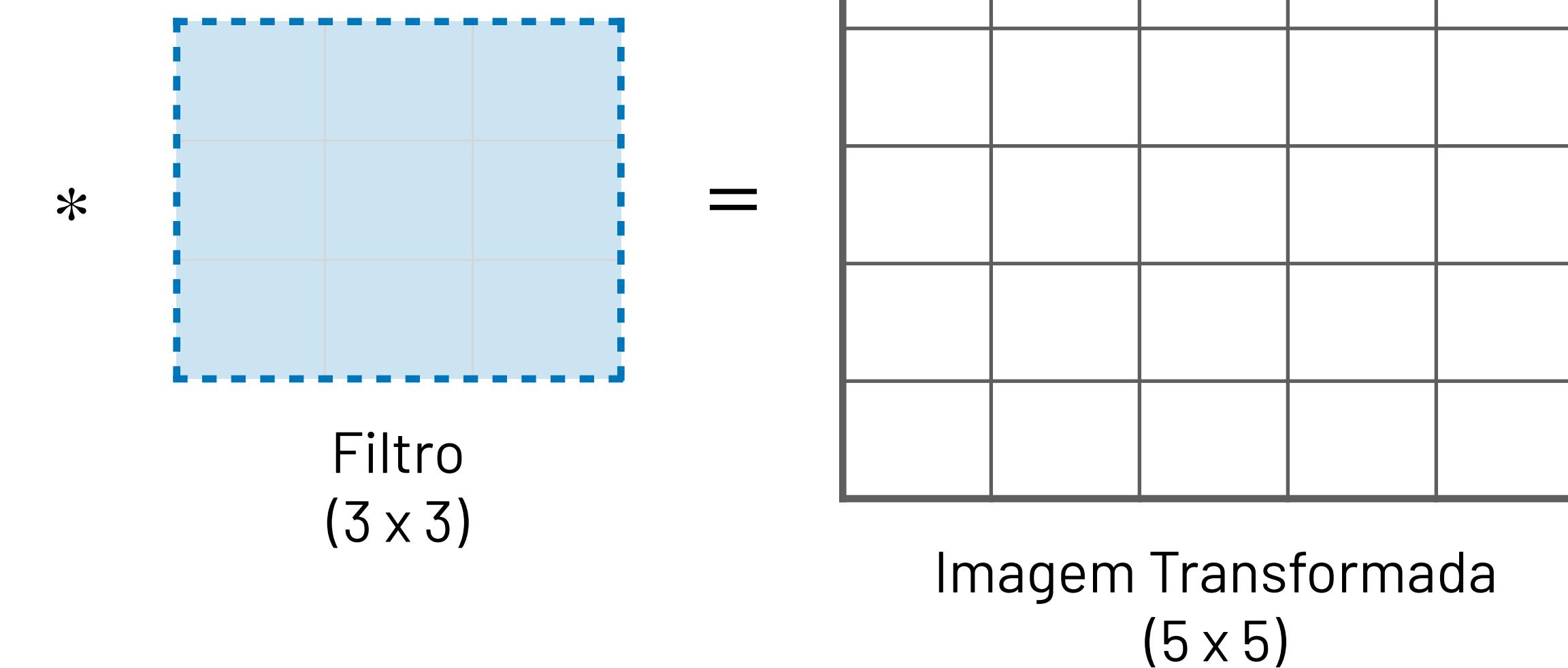


Imagen Original
(5 x 5)



Regra Geral

$$(n \times n) * (f \times f) = (n + 2p - f + 1 \times n + 2p - f + 1)$$

Preenchimento (Padding)

Para encontrar o valor de p que mantém o tamanho de uma imagem $n \times n$ após a convolução com um filtro de tamanho f (ímpar), basta resolver a seguinte equação:

$$n + 2p - f + 1 = n$$

$$2p - f + 1 = 0$$

$$2p = f - 1$$

$$p = \frac{f - 1}{2}$$

Convoluçãoes Passadas (*Strided Convolutions*)

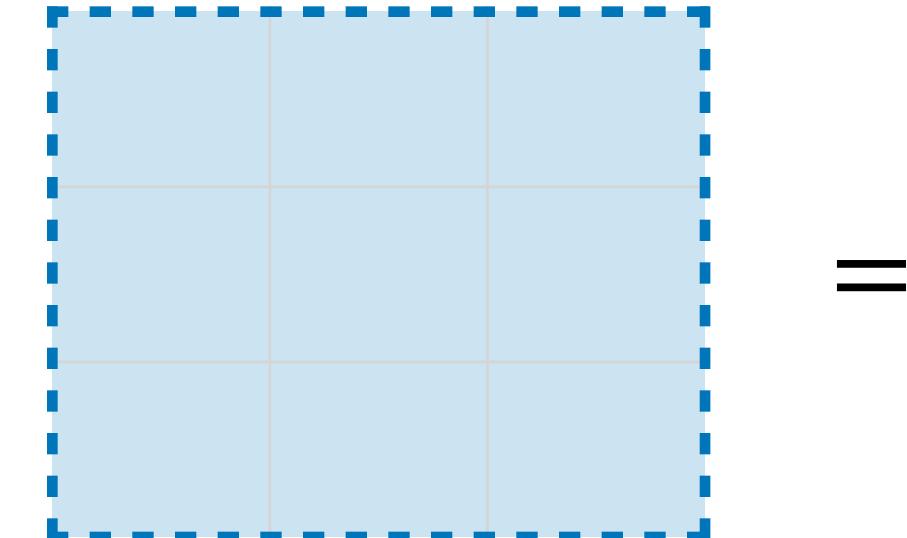
Convoluçãoes podem ser executadas com passos (strides) maiores do que 1.

stride = 2

206	205	247	245	244
244	161	137	244	254
192	154	75	200	249
90	109	96	143	223
67	69	107	196	236

Imagen Original
 (5×5)

*



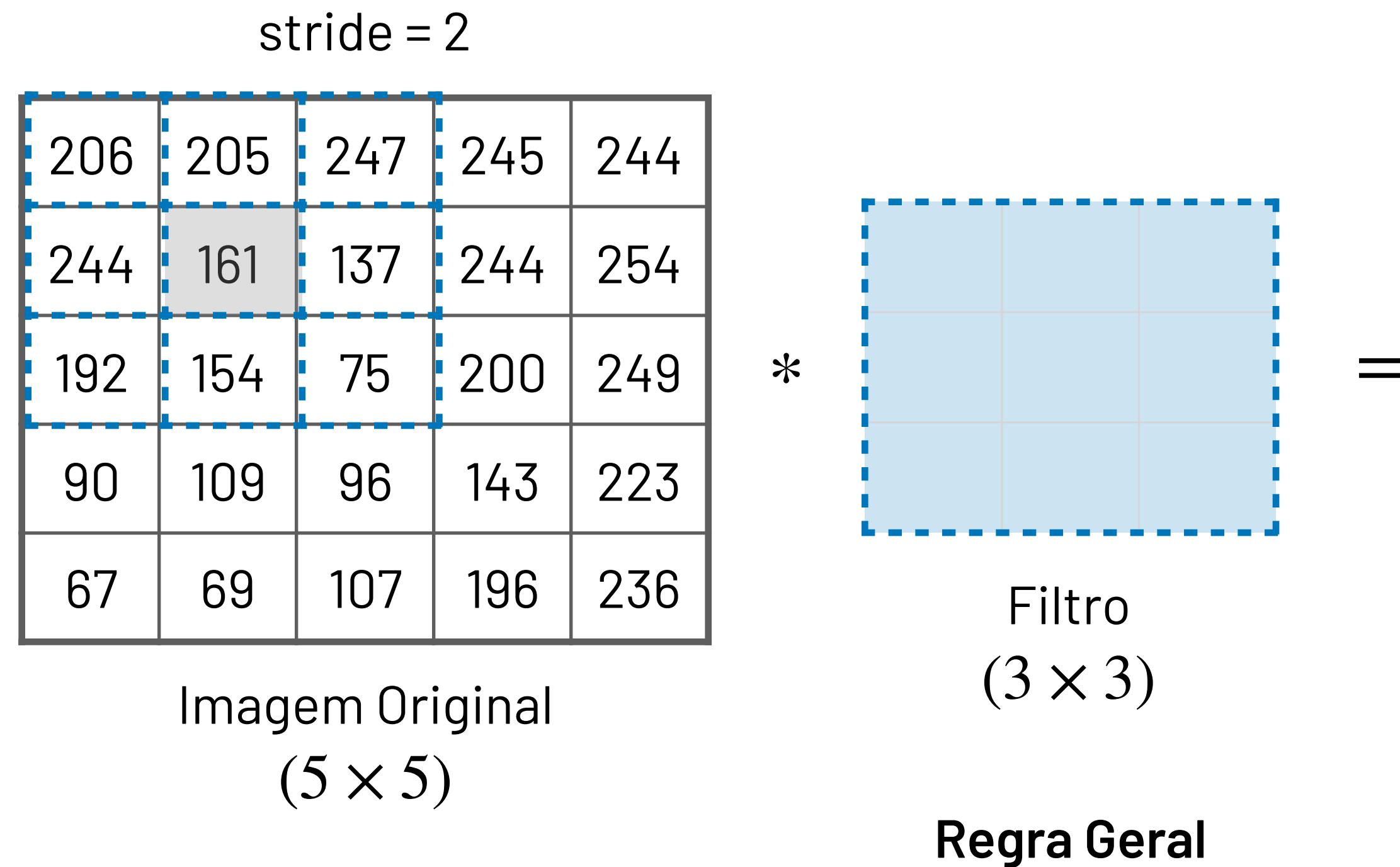
Filtro
 (3×3)

Regra Geral

$$(n \times n) * (f \times f) = \left(\frac{n + 2p - f}{s} + 1 \times \frac{n + 2p - f}{s} + 1 \right)$$

Convoluçãoes Passadas (*Strided Convolutions*)

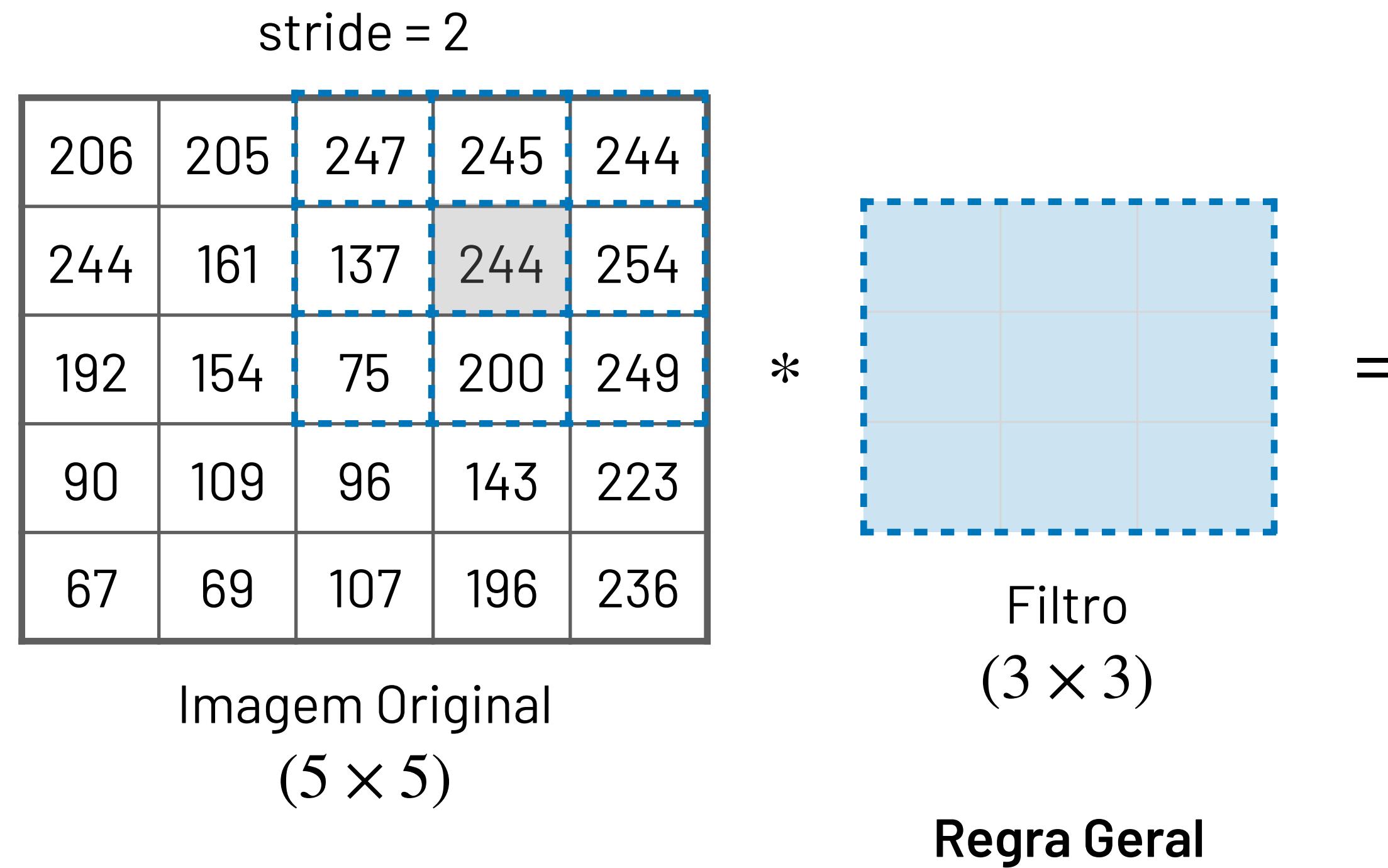
Convoluçãoes podem ser executadas com passos (strides) maiores do que 1.



$$(n \times n) * (f \times f) = \left(\frac{n + 2p - f}{s} + 1 \times \frac{n + 2p - f}{s} + 1 \right)$$

Convoluçãoes Passadas (*Strided Convolutions*)

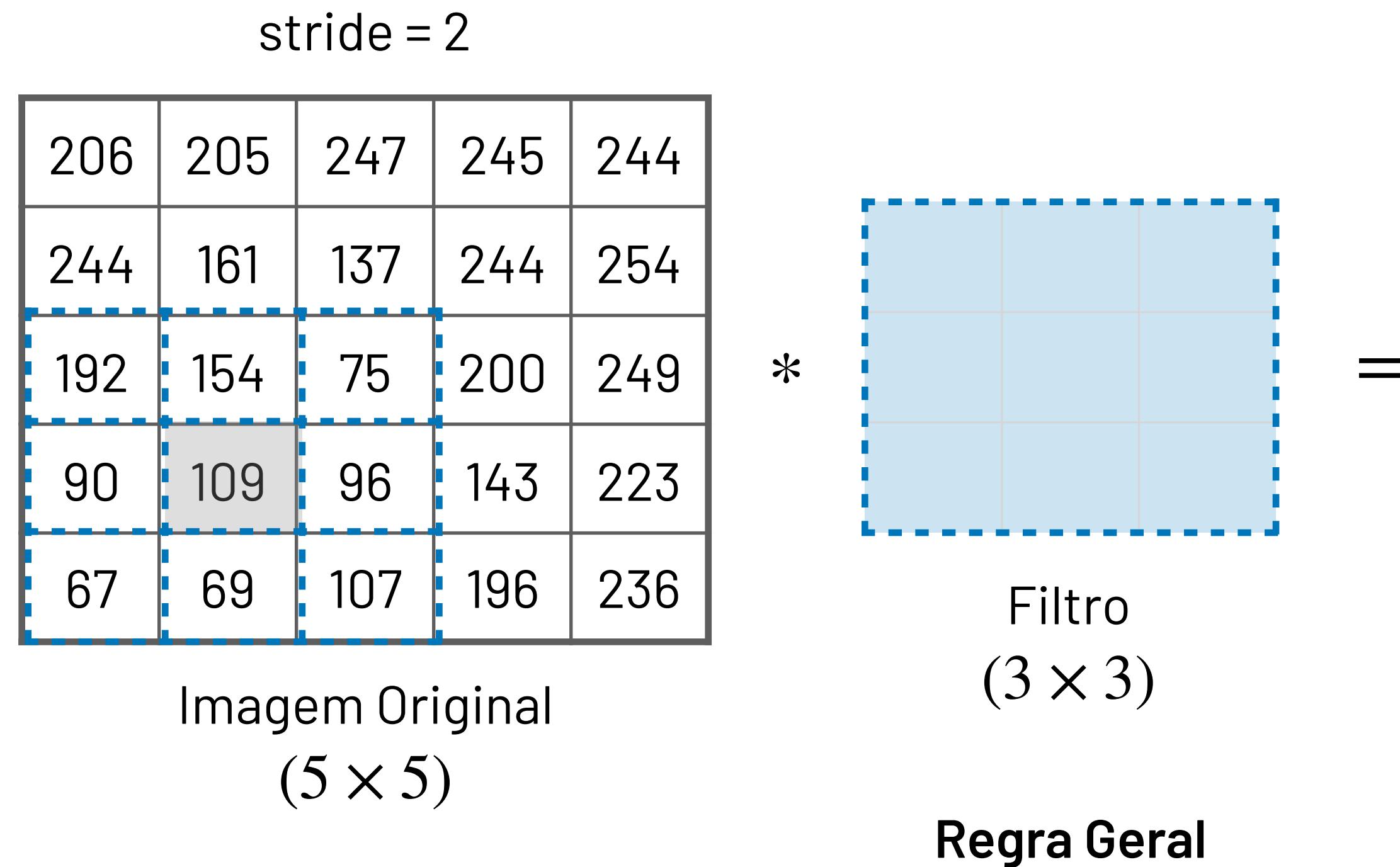
Convoluçãoes podem ser executadas com passos (strides) maiores do que 1.



$$(n \times n) * (f \times f) = \left(\frac{n + 2p - f}{s} + 1 \times \frac{n + 2p - f}{s} + 1 \right)$$

Convoluçãoes Passadas (*Strided Convolutions*)

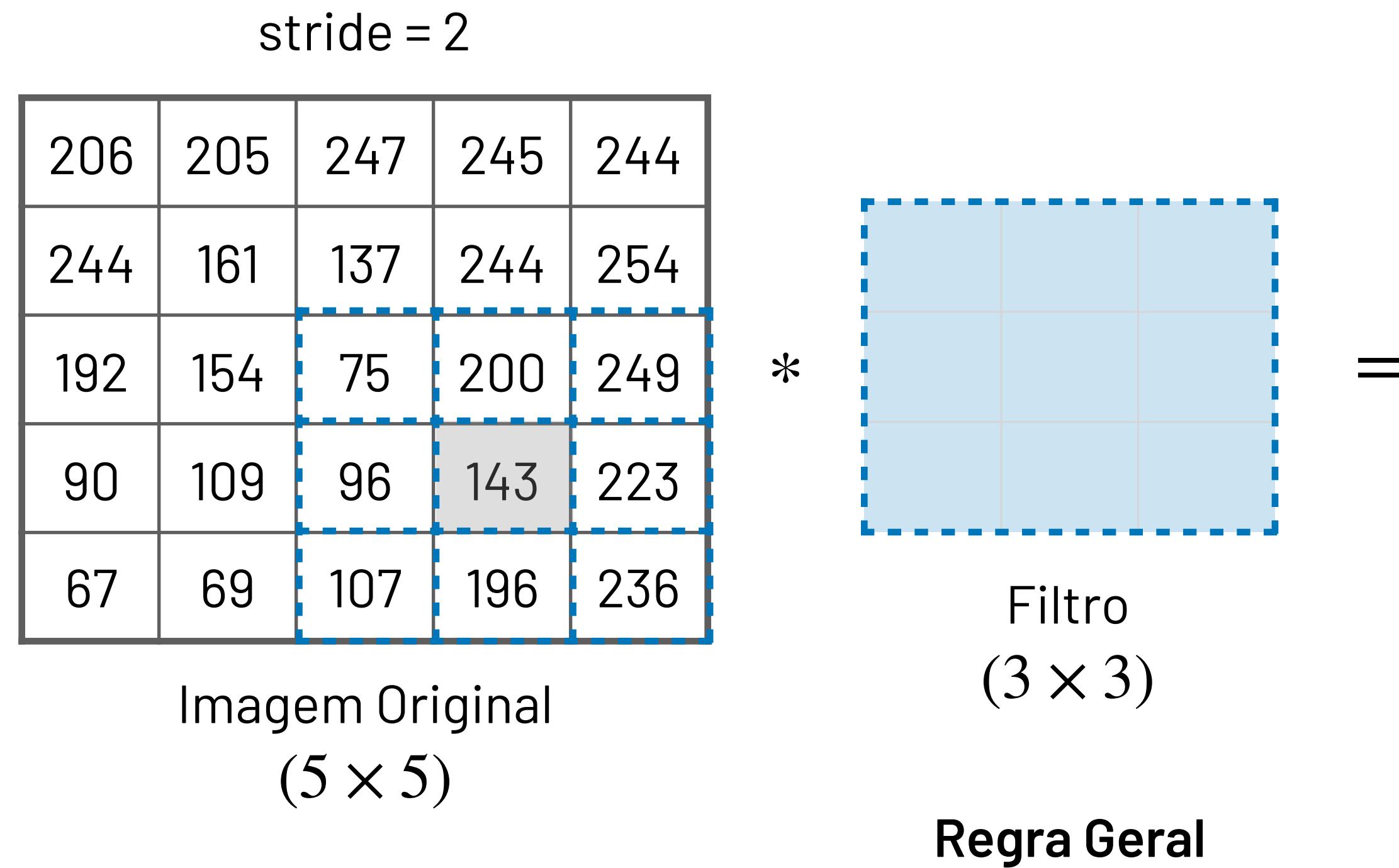
Convoluçãoes podem ser executadas com passos (strides) maiores do que 1.



$$(n \times n) * (f \times f) = \left(\frac{n + 2p - f}{s} + 1 \times \frac{n + 2p - f}{s} + 1 \right)$$

Convoluçãoes Passadas (*Strided Convolutions*)

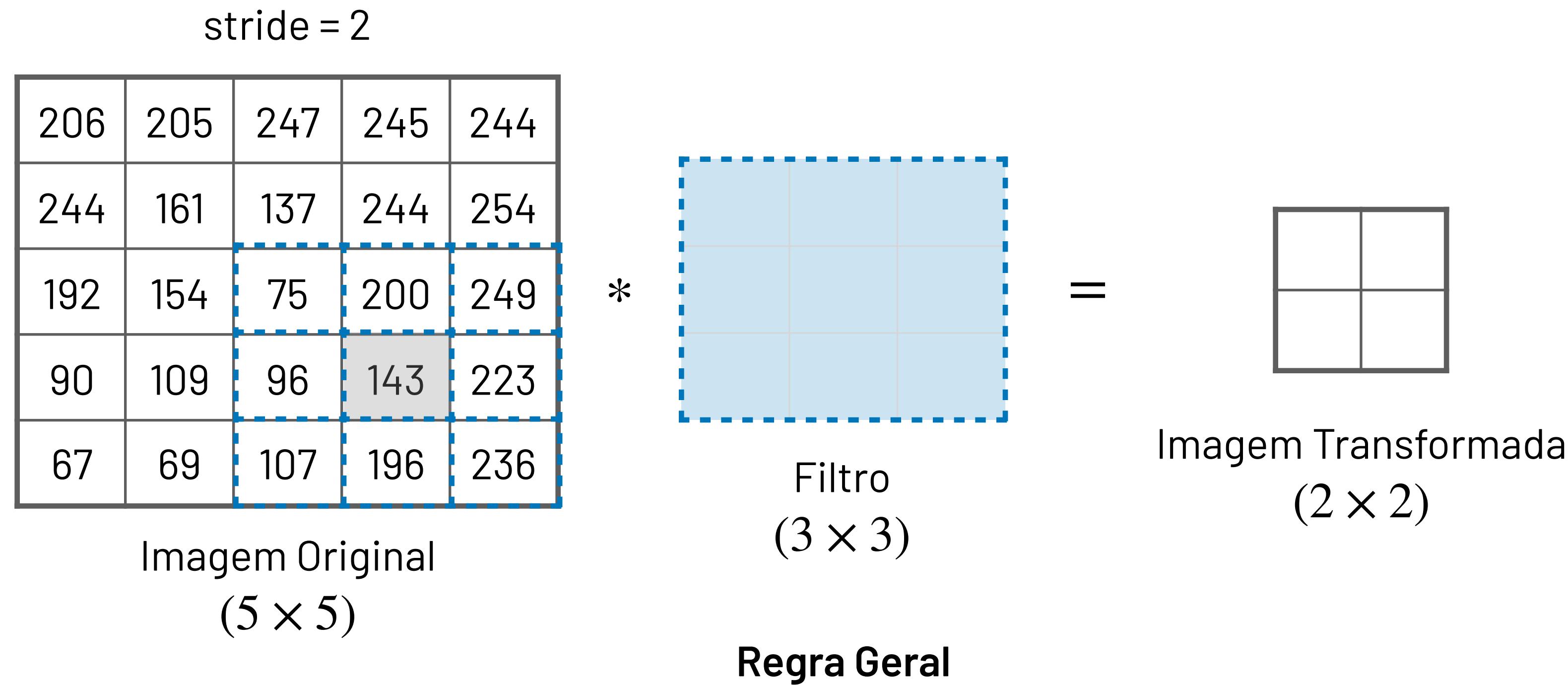
Convoluçãoes podem ser executadas com passos (strides) maiores do que 1.



$$(n \times n) * (f \times f) = \left(\frac{n + 2p - f}{s} + 1 \times \frac{n + 2p - f}{s} + 1 \right)$$

Convoluçãoes Passadas (*Strided Convolutions*)

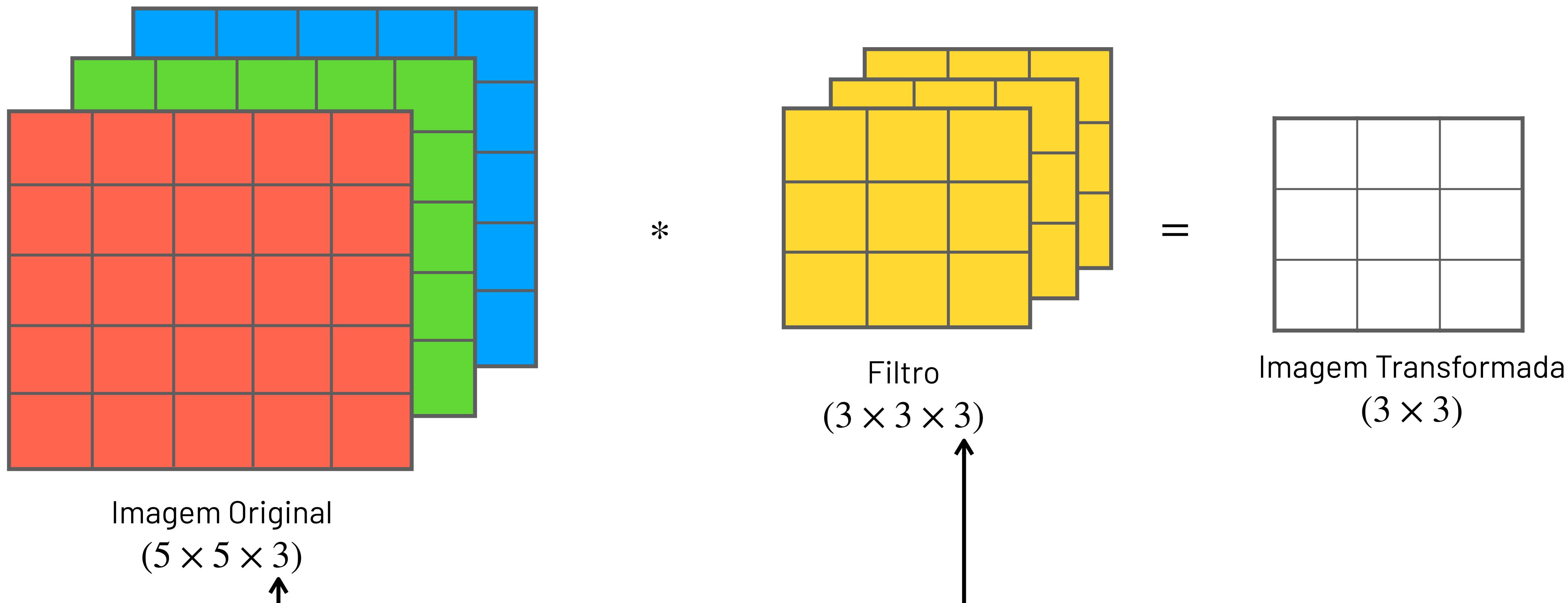
Convoluçãoes podem ser executadas com passos (strides) maiores do que 1.



$$(n \times n) * (f \times f) = \left(\frac{n + 2p - f}{s} + 1 \right) \times \left(\frac{n + 2p - f}{s} + 1 \right)$$

Convoluçãoes em Volumes

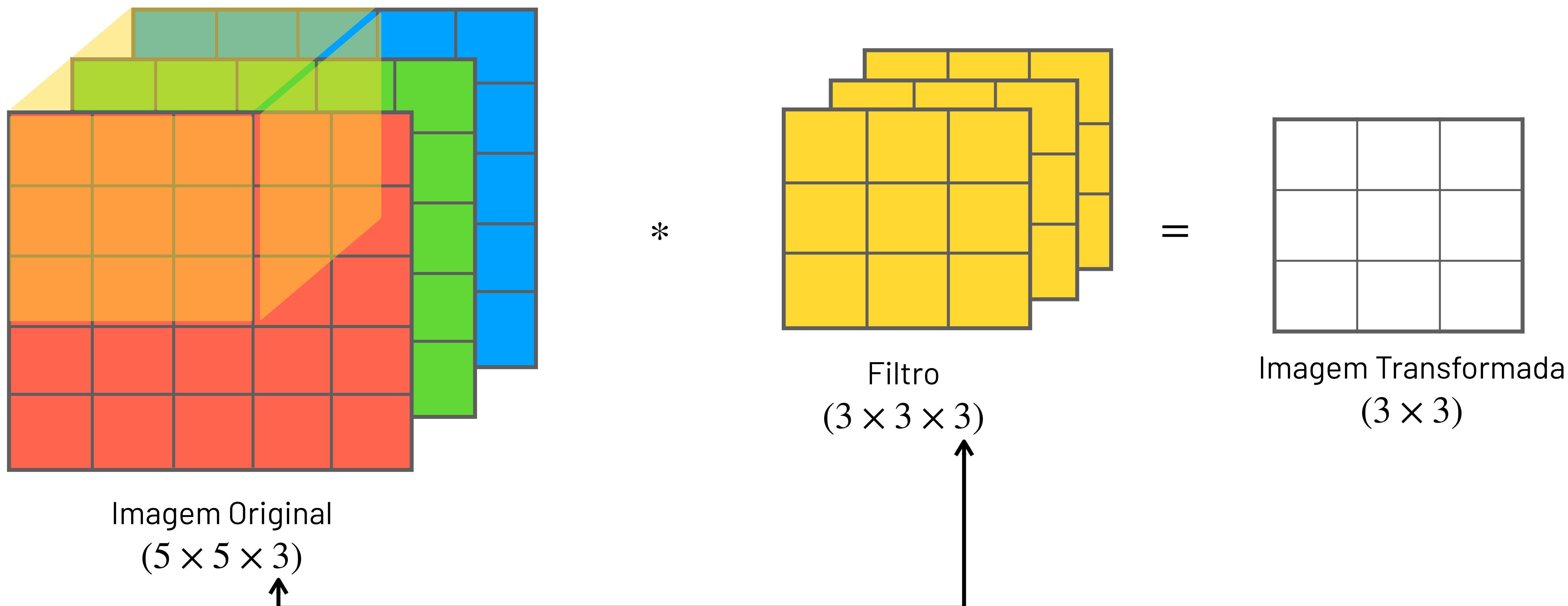
Convoluçãoes em imagens coloridas (R,G,B) necessitam filtros com 3 canais



O número de canais deve ser o mesmo na imagem e no filtro!

Convoluçãoes em Volumes

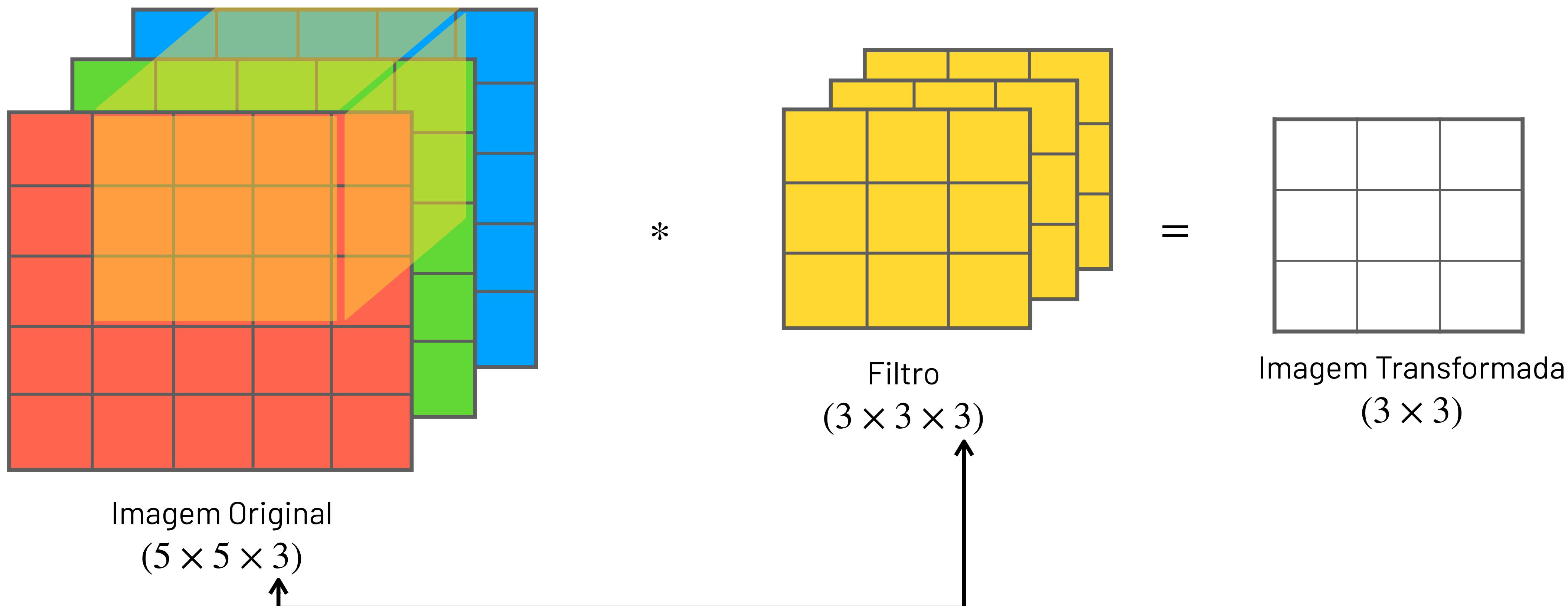
Convoluçãoes em imagens coloridas (R,G,B) necessitam filtros com 3 canais



O número de canais deve ser o mesmo na imagem e no filtro!

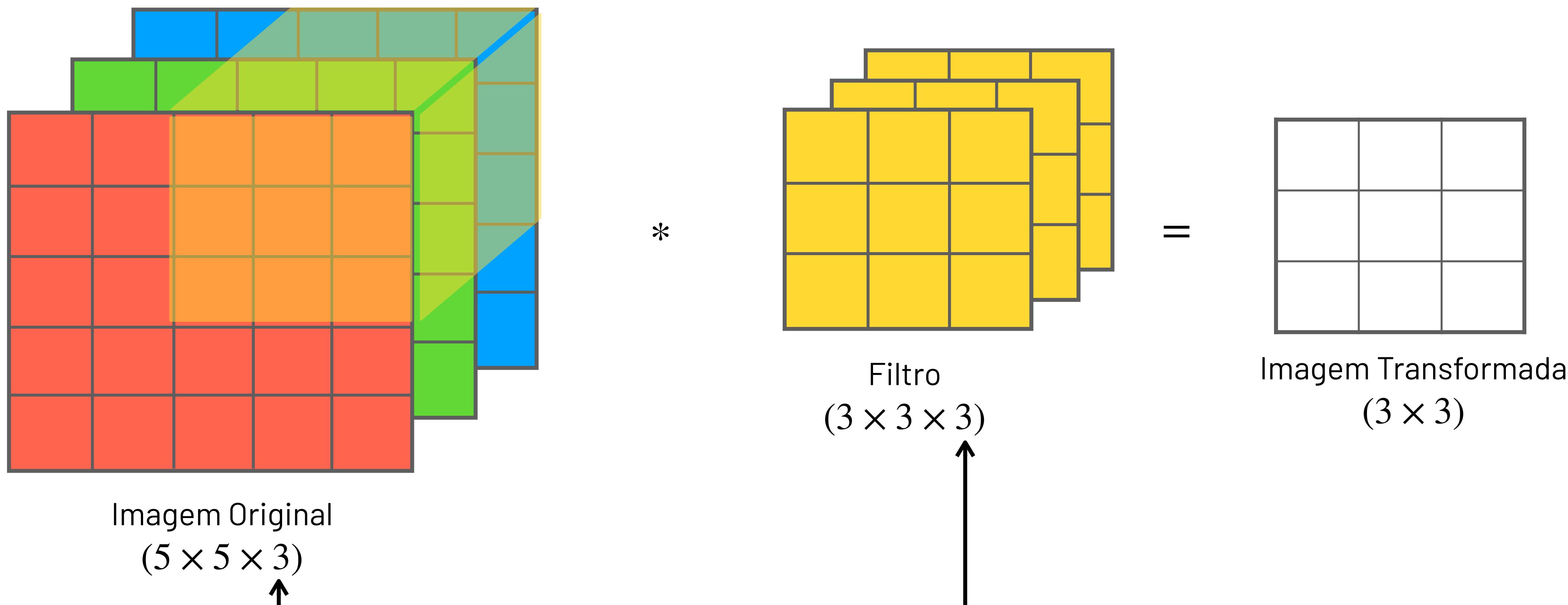
Convoluçãoes em Volumes

Convoluçãoes em imagens coloridas (R,G,B) necessitam filtros com 3 canais



Convoluçãoes em Volumes

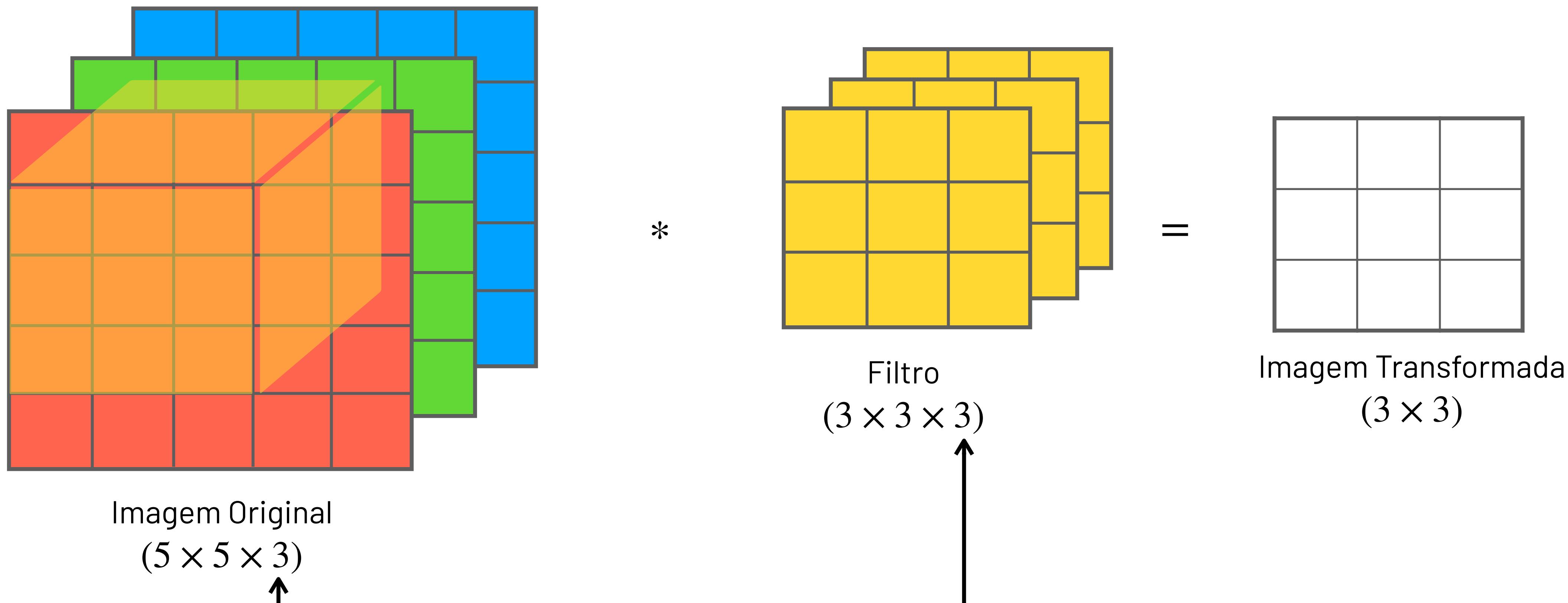
Convoluçãoes em imagens coloridas (R,G,B) necessitam filtros com 3 canais



O número de canais deve ser o mesmo na imagem e no filtro!

Convoluçãoes em Volumes

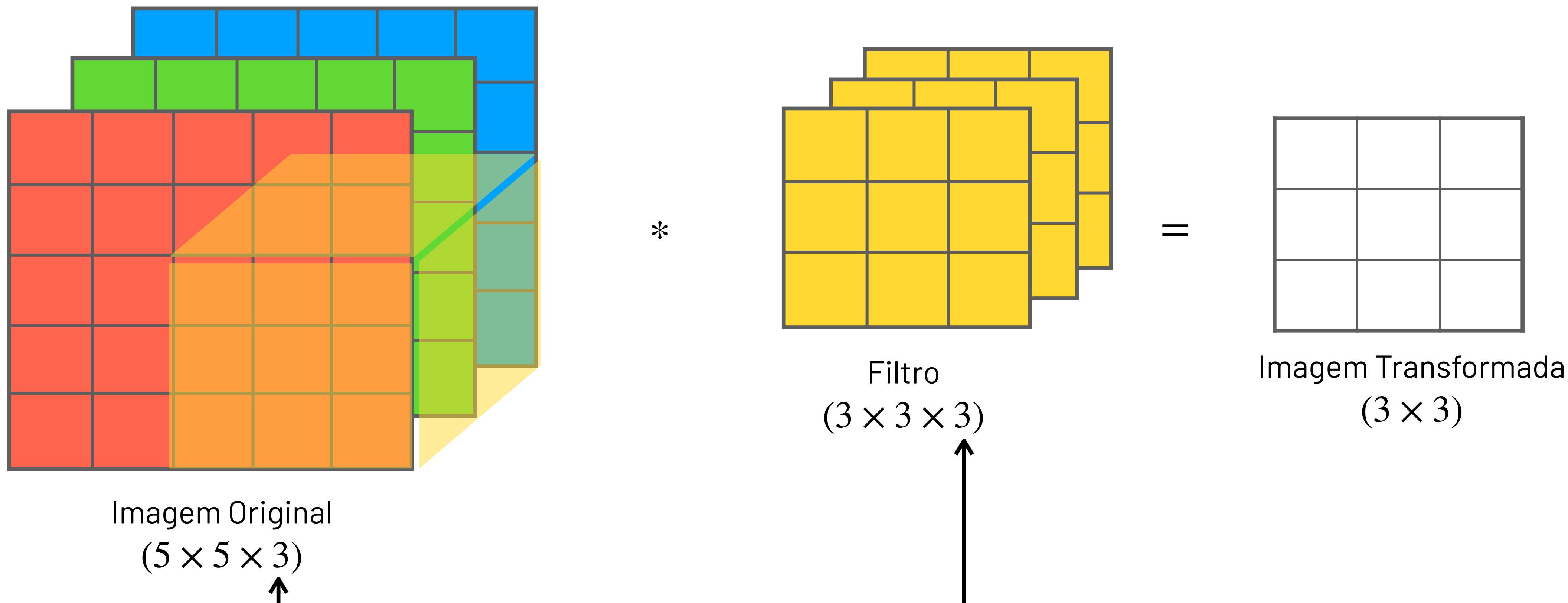
Convoluçãoes em imagens coloridas (R,G,B) necessitam filtros com 3 canais



O número de canais deve ser o mesmo na imagem e no filtro!

Convoluçãoes em Volumes

Convoluçãoes em imagens coloridas (R,G,B) necessitam filtros com 3 canais



O número de canais deve ser o mesmo na imagem e no filtro!

Múltiplos Filtros

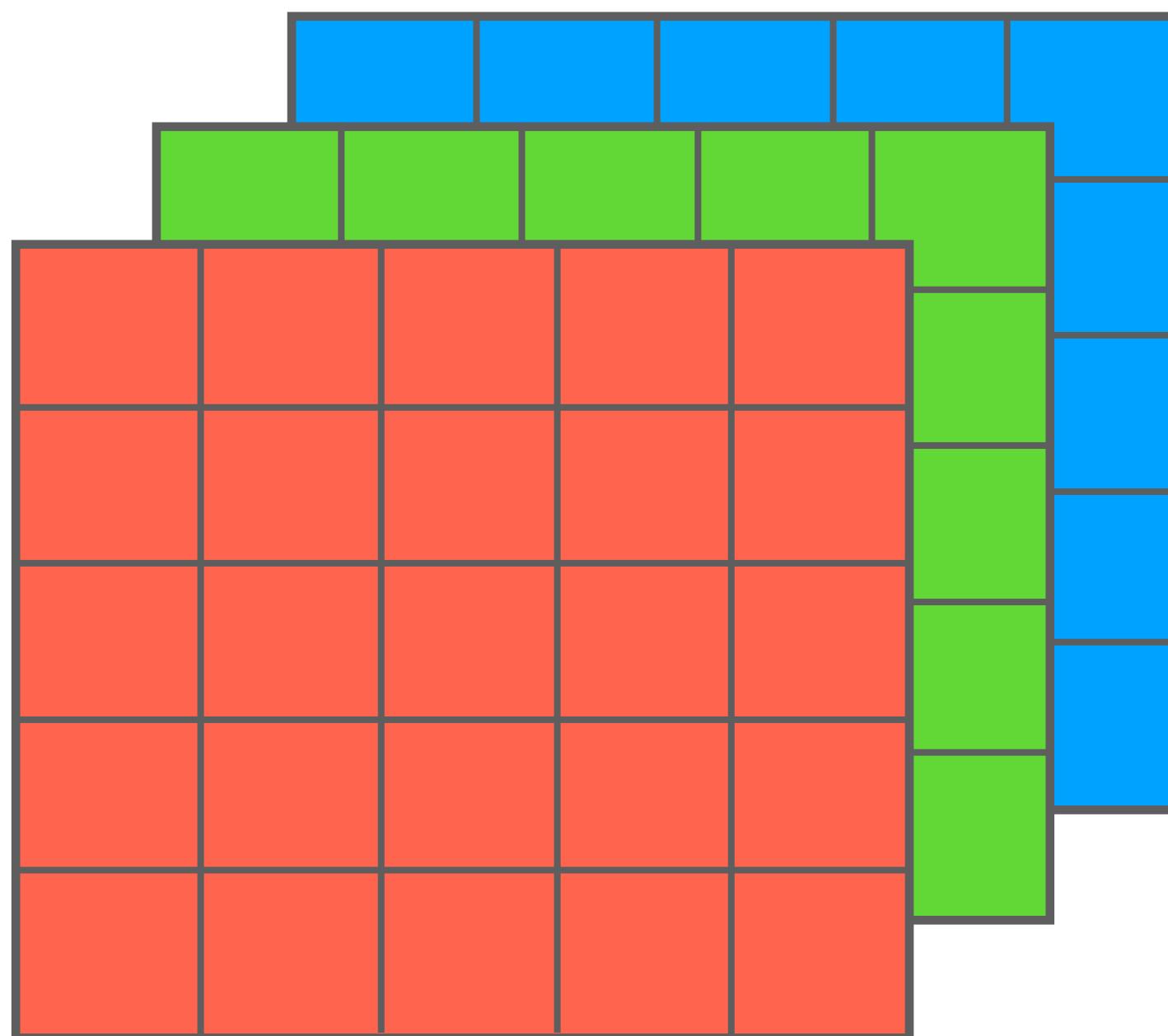
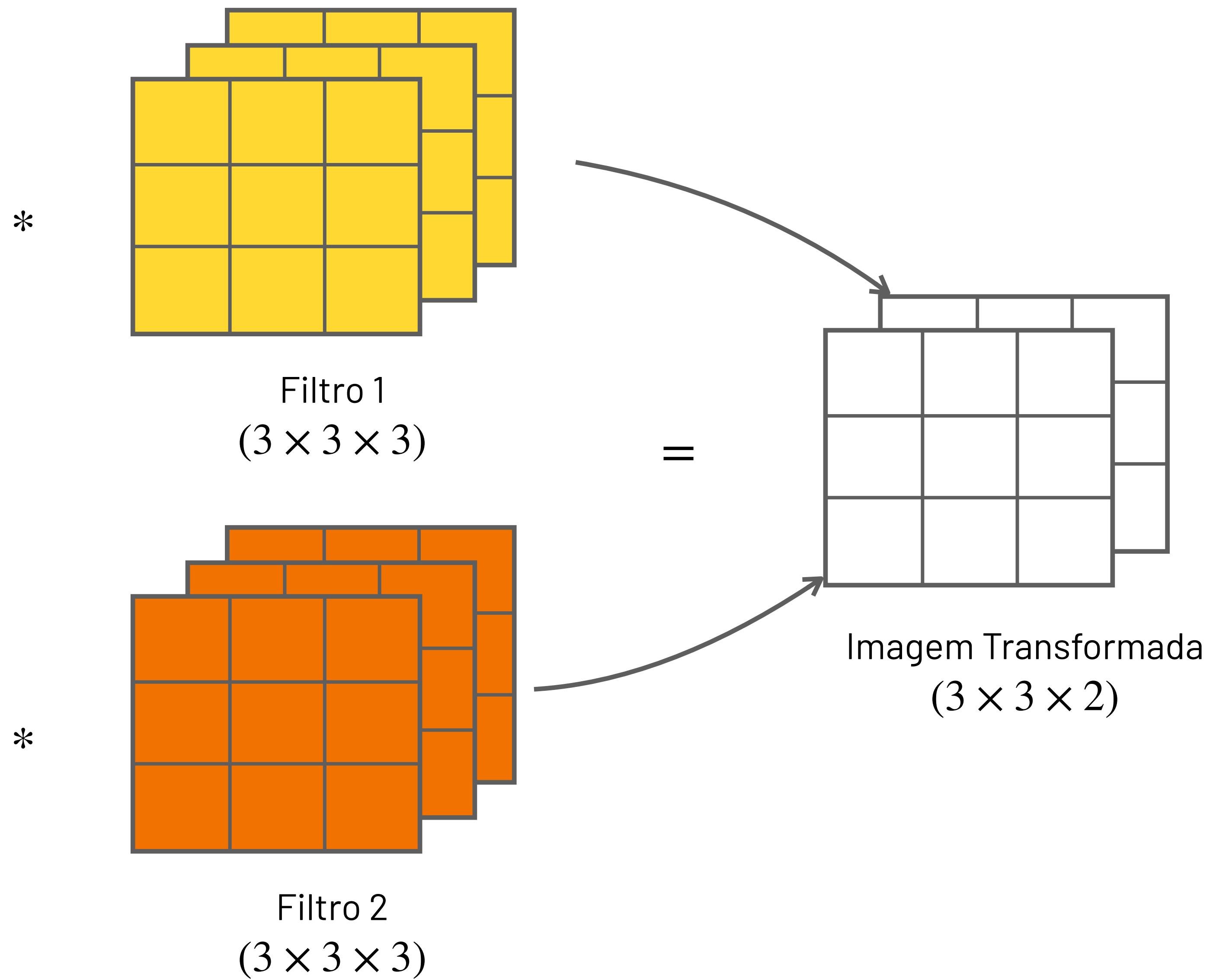
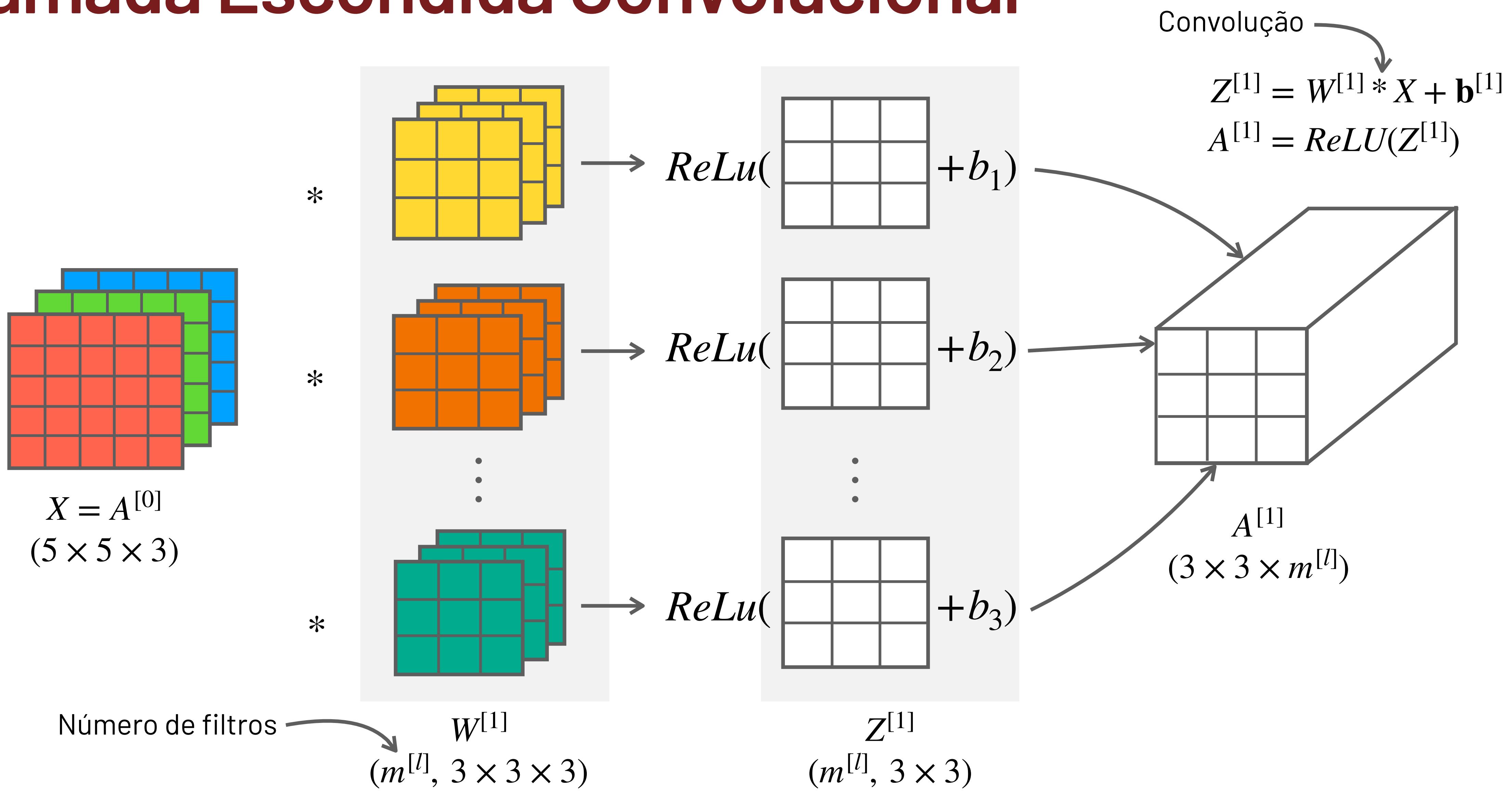


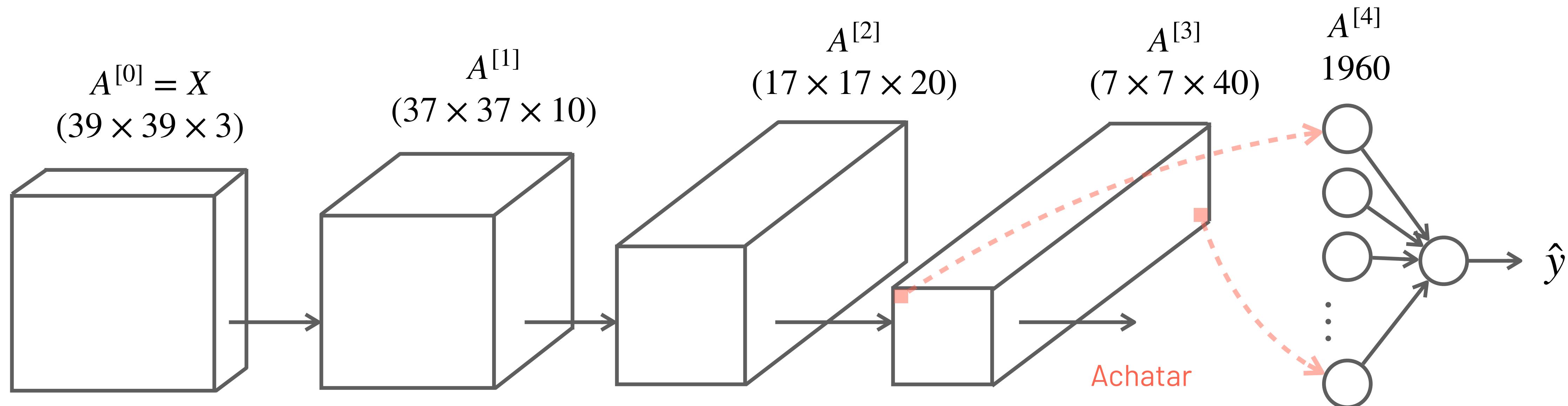
Imagen Original
 $(5 \times 5 \times 3)$



Camada Escondida Convolucional



Rede Neural Convolucional (CNN)



$$\begin{array}{ll} n^{[0]} = 39 & f^{[1]} = 3 \\ s^{[1]} = 1 & \\ p^{[1]} = 0 & \\ m^{[1]} = 10 & \end{array}$$

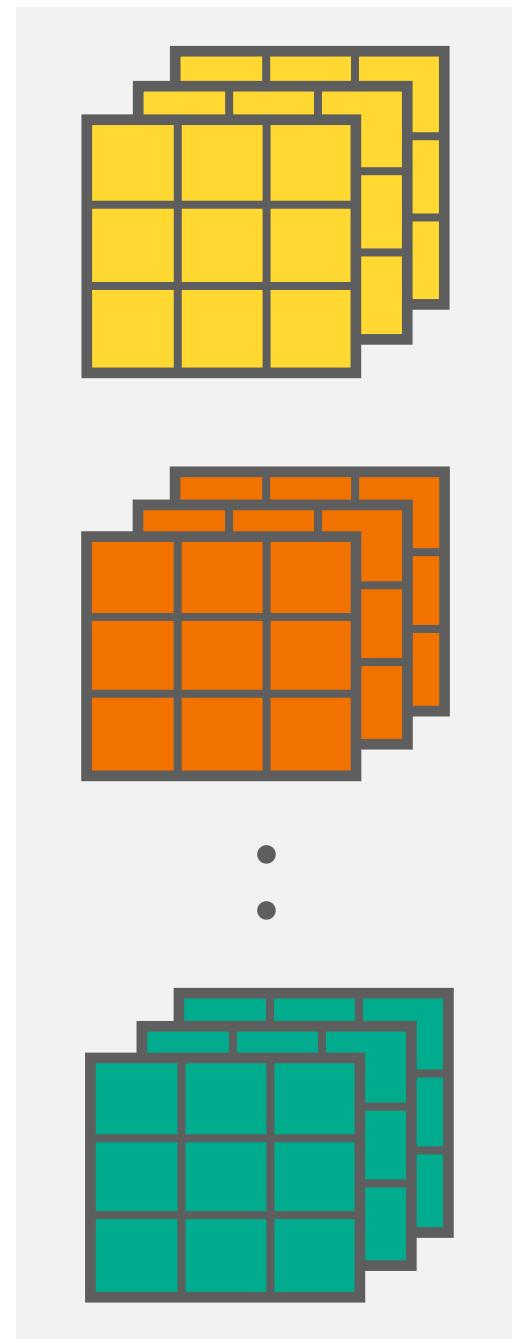
$$\begin{array}{ll} f^{[2]} = 5 & f^{[3]} = 5 \\ s^{[2]} = 2 & s^{[3]} = 2 \\ p^{[2]} = 0 & p^{[3]} = 0 \\ m^{[2]} = 20 & m^{[3]} = 40 \end{array}$$

Notação:

- ▶ $f^{[l]}$ tamanho dos filtros da camada l
- ▶ $s^{[l]}$ tamanho do stride da camada l
- ▶ $p^{[l]}$ tamanho do padding da camada l
- ▶ $m^{[l]}$ número de filtros na camada l

Número de parâmetros

Quantos parâmetros uma camada com 10 filtros ($3 \times 3 \times 3$) tem?



$W^{[1]}$
 $(10, 3 \times 3 \times 3)$

$$\begin{aligned} 3 \times 3 \times 3 &= 27 \\ &\quad + 1 \\ &= 28 \\ &\quad \times 10 \\ &= \underline{\underline{280}} \text{ parâmetros} \end{aligned}$$

Próxima aula

A12: Estudo de Casos de CNNs

Resnet, Inception Network, MobileNet e Efficient Net.