(4) Redes Neurais Convolucionais - CNNs Redes Neurais e Aprendizado Profundo

Moacir Ponti ICMC, Universidade de São Paulo

www.icmc.usp.br/~moacir — moacir@icmc.usp.br

Agenda

Redes Neurais Convolucionais (CNNs)

Convolução

Camada convolucional para redes neurais

Exemplo

Número de parâmetros

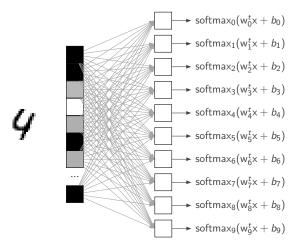
Pooling

Exemplo de problema: classificação de dígitos

▶ Imagens com $28 \times 28 = 784$ pixels,

Questões importantes sobre as redes MLP (densas)

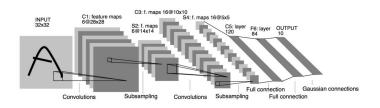
- 1. Valores de entrada (atributos) são considerados independentes
- 2. Não são aproveitadas relações locais entre os dados



Questões importantes sobre as redes MLP (densas)

- 1. Grande número de parâmetros: memória e processamento
 - **Exemplo:** entrada imagem de $28 \times 28 = 784$
 - Uma camada com 100 neurônios teria..
 - ▶ 78400 + 100 = 78500 parâmetros a serem aprendidos e mantidos na memória durante o treinamento

Redes Neurais Convolucionais (CNNs)



(Arquitetura LeNet)

Nova terminologia:

- Camada convolucional (convolutional layer)
- Subamostragem (pooling)
- Mapas de Ativação (activation/feature maps)
- Camada densa (dense/fully connected, tipo MLP)

- Operador que visa realizar uma combinação linear de valores locais da entrada
- ► Centrado em uma posição, e.g. (x, y), gera como saída um único valor de saída

		V	olume	de entr	ada 7	x 7								Volu	me de	saída		
	0	1	2	3	4	5	6		o W (3	,		0	1	2	3	4	5	6
0	2	2	2	2	3	3	3	-1	0.5	1	0							
1	1	0	1	1	1	1	0	-1	0	0	1							
2	1	1	3	3	0	0	0	0	0	0.5	2							
3	1	1	3	2	0	0	3				3							
4	1	1	3	2	0	0	3				4							
5	1	3	3	2	0	0	3				5							
6	3	3	3	2	0	0	3				6							

		V	olume	de entr	ada 7	¢ 7								Volu	me de	saída		
	0	1	2	3	4	5	6	Filtr	o W (3	x 3)		0	1	2	3	4	5	6
0	2	2	2	2	3	3	3	-1	0.5	1	0							
1	1	0	1	1	1	1	0	-1	0	0	1		1.5					
2	1	1	3	3	0	0	0	0	0	0.5	2							
3	1	1	3	2	0	0	3				3							
4	1	1	3	2	0	0	3				4							
5	1	3	3	2	0	0	3				5							
6	3	3	3	2	0	0	3				6							

		V	olume	de entr	ada 7	x 7								Volu	me de	saída		
	0	1	2	3	4	5	6	Filtr	o W (3	x 3)								
0	2	2	2	2	3	3	3	-1	0.5	1			0	1	2	3	4	
1	1	0	1	1	1	1	0	-1	0	0		0	1.5					
2	1	1	3	3	0	0	0	0	0	0.5		1						
3	1	1	3	2	0	0	3					2						
4	1	1	3	2	0	0	3					3						
5	1	3	3	2	0	0	3					4						
6	3	3	3	2	0	0	3											

		V	olume	de entr	ada 7	¢ 7								Volu	me de	saída		
	0	1	2	3	4	5	6	Filtr	o W (3	x 3)								
0	2	2	2	2	3	3	3	-1	0.5	1			0	1	2	3	4	
1	1	0	1	1	1	1	0	-1	0	0		0	1.5	2.5				
2	1	1	3	3	0	0	0	0	0	0.5		1						
3	1	1	3	2	0	0	3					2						
4	1	1	3	2	0	0	3					3						
5	1	3	3	2	0	0	3					4						
6	3	3	3	2	0	0	3											

		V	olume	de entr	ada 7	¢ 7								Volu	me de	saída		
	0	1	2 2 3 3		6	Filt	o W (3	x 3)										
0	2	2	2	2	3	3	3	-1	0.5	1			0	1	2	3	4	
1	1	0	1	1	1	1	0	-1	0	0		0	1.5	2.5	1			
2	1	1	3	3	0	0	0	0	0	0.5		1						
3	1	1	3	2	0	0	3					2						
4	1	1	3	2	0	0	3					3						
5	1	3	3	2	0	0	3					4						
6	3	3	3	2	0	0	3											

		V	olume	de entr	ada 7 ɔ	7								Volu	me de	saída		
	0	1	2	3	4	5	6	Filt	o W (3	x 3)								
0	2	2	2	2	3	3	3	-1	0.5	1			0	1	2	3	4	
1	1	0	1	1	1	1	0	-1	0	0		0	1.5	2.5	1	1.5		
2	1	1	3	3	0	0	0	0	0	0.5		1						
3	1	1	3	2	0	0	3					2						
4	1	1	3	2	0	0	3					3						
5	1	3	3	2	0	0	3					4						
6	3	3	3	2	0	0	3											

		V	olume	de entr	ada 7	¢ 7									Volu	me de	saída		
	0	1	2	3	4	5	6		Filtr	w (3	x 3)								
0	2	2	2	2	3	3	3		-1	0.5	1			0	1	2	3	4	
1	1	0	1	1	1	1	0		-1	0	0		0	1.5		1	1.5	0.5	
2	1	1	3	3	0	0	0		0	0	0.5		1						
3	1	1	3	2	0	0	3						2						
4	1	1	3	2	0	0	3						3						
5	1	3	3	2	0	0	3						4						
6	3	3	3	2	0	0	3												

		2 2 2 2 3 3 1 1 0 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1								Volu	me de	saída						
	0	_		3	4	_	6	Filtr	o W (3	x 3)								
0	2				_	3	3	-1	0.5	1			0	1	2	3	4	
1	1	0	1		1	1	0	-1	0	0		0	1.5	2.5	1	1.5	0.5	
2	1	1	3	3		0	0	0	0	0.5		1	0.5					
3	1	1	3	_		0	3					2						
4	1	_	-		0	0	3					3						
5	1	3	3	2		0	3					4						
6	3	3	3	2	0	0	3											

		V	olume	de entr	ada 7	x 7								Volu	me de	saída		
	0	1	2	3	4	5	6	Filtr	o W (3	x 3)								
0	2	2	2	2	3	3	3	-1	0.5	1			0	1	2	3	4	
1	1	0	1	1	1	1	0	-1	0	0		0	1.5	2.5	1	1.5	0.5	
2	1	1	3	3	0	0	0	0	0	0.5		1	0.5					
3	1	1	3	2	0	0	3					2						
4	1	1	3	2	0	0	3					3						
5	1	3	3	2	0	0	3					4						
6	3	3	3	2	0	0	3											

		V	olume	de entr	ada 7	x 7								Volu	me de	saída	
	0	1					_	Filtr	o W (3	x 3)							
0	2	2	2	2	3	3	3	-1	0.5	1			0	1	2	3	4
1	1	0	1	1	1	1	0	-1	0	0		0	1.5		1	1.5	0.5
2	1	1	3	3	0	0	0	0	0	0.5		1	0.5	1.5	-2.5		
3	1	1	3	2	0	0	3					2					
4	1	1	3	2	0	0	3					3					
5	1	3	3	2	0	0	3					4					
6	3	3	3	2	0	0	3										

		V	olume	de entr	ada 7	¢ 7								Volu	me de	saída		
	0	1	1 2 3 4 5 2 2 2 3 3					Filtr	o W (3	x 3)								
0	2		2	2	3		3	-1	0.5	1			0	1	2	3	4	
1	1	0	1	1	1	1	0	-1	0	0		0	1.5	2.5	1	1.5	0.5	
2	1	1	3	3	0	0	0	0	0	0.5		1	0.5	1.5	-2.5			
3	1	1	3	2	0	0	3					2						
4	1	1	3	2	0	0	3					3						
5	1	3	3	2	0	0	3					4						
6	3	3	3	2	0	0	3											

		2 2 2 3 3 3 3 0 1 1 1 1 1 0									Volu	me de	saída						
	0	_	_	3	4	_	6		Filtr	o W (3	x 3)								
0	2	2	2	2	3		3		-1	0.5	1			0	1	2	3	4	
1	1	0	1	_	1	1	0		-1	0	0		0	1.5	2.5	1	1.5	0.5	
2	1		3		0	0	0		0	0	0.5		1	0.5		-2.5		1	
3	1	1	3	2	0	0	3						2						
4	1	1	3	2	0	0	3						3						
5	1	3	3	2	0	0	3						4						
6	3	3	3	2	0	0	3												

		V	olume	de entr	ada 7 :	x 7								Volu	me de	saída	
	0	1	2	3	4	5	6	Filtr	o W (3	x 3)							
0	2	2	2	2	3 3 3 1 1 0			-1	0.5	1			0	1	2	3	4
1	1	0	1	1	_	1 0		-1	0	0		0	1.5	2.5	1	1.5	0.5
2	1	1	3	3	0	0	0	0	0	0.5		1		1.5	-2.5		1
3	1	1	3	2	0	0	3					2	3				
4	1	1	3	2	0	0	3					3					
5	1	3	3	2	0	0	3					4					
6	3	3	3	2	0	0	3										
						***************************************	·										

		V	olume	de entr	ada 7	¢ 7								Volu	me de	saída		
	0	1 2 3 4 5 6 2 2 2 3 3 3						Filtr	o W (3	x 3)								
0	2		2	2	3			-1	0.5	1			0	1	2	3	4	
1	1	0	1	1	1	1	0	-1	0	0		0	1.5		1	1.5	0.5	
2	1	1	3	3	0	0	0	0	0	0.5		1		1.5	-2.5	-2.5	1	
3	1	1	3	2	0	0	3					2	3	3.5	-4.5	-5	1.5	
4	1	1	3	2	0	0	3					3	3	2.5	-5	-4	4.5	
5	1	3	3	2	0	0	3					4	3	0.5	-5	-4	4.5	
6	3	3	3	2	0	0	3											
																		Ī

- ➤ Zero-padding: para compensar a impossibilidade de computar todos os valores;
 - Amplia-se a entrada de forma que o volume de saída seja igual ao de entrada

		Vol	ume de	entra	da 7 x	7 + ze	ro pad	ding								V	olume	de saí	da		
	0	1	2	3	4	5	6	7	8	Filtr	o W (3	3 x 3)									
0	0	0	0	0	0	0	0	0	0						0	1	2	3	4	5	6
1	0	2	2	2	2	3	3	3	0	-1	0.5	1		0							
2	0	1	0	1	1	1	1	0	0	-1	0	0		1							
3	0	1	1	3	3	0	0	0	0	0	0	0.5		2							
4	0	1	1	3	2	0	0	3	0					3							
5	0	1	1	3	2	0	0	3	0					4							
6	0	1	3	3	2	0	0	3	0					5							
7	0	3	3	3	2	0	0	3	0					6							
8	0	0	0	0	0	0	0	0	0												

		v	olume	de en	trada 7	x 7 +	paddii	ng								V	olume	de saí	da 7 x	7		
	0	1	2	3	4	5	6	7	8		Filtr	o W (3	x 3)									
0	0	0	0	0	0	0	0	0	0							0	1	2	3	4	5	6
1	0	2	2	2	2	3	3	3	0		-1	0.5	1		0	0						
2	0	1	0	1	1	1	1	0	0		-1	0	0		1							
3	0	1	1	3	3	0	0	0	0		0	0	0.5		2							
4	0	1	1	3	2	0	0	3	0						3							
5	0	1	1	3	2	0	0	3	0						4							
6	0	1	3	3	2	0	0	3	0						5							
7	0	3	3	3	2	0	0	3	0						6							
8	0	0	0	0	0	0	0	0	0													

		v	olume	de ent	trada 7	x 7 +	paddii	ng							Volu	me de	saída			
	0	1	2	3	4	5	6	7	8	Filtr	o W (3	x 3)								
0	0	0	0	0	0	0	0	0	0					0	1	2	3	4	5	6
1	0	2	2	2	2	3	3	3	0	-1	0.5	1	0	0	-1.5					
2	0	1	0	1	1	1	1	0	0	-1	0	0	1							
3	0	1	1	3	3	0	0	0	0	0	0	0.5	2							
4	0	1	1	3	2	0	0	3	0				3							
5	0	1	1	3	2	0	0	3	0				4							
6	0	1	3	3	2	0	0	3	0				5							
7	0	3	3	3	2	0	0	3	0				6							
8	0	0	0	0	0	0	0	0	0											

		v	olume	de ent	trada 7	x 7 +	paddir	ng									Volu	me de	saída			
	0	1	2	3	4	5	6	7	8	F	iltro	W (3	x 3)									
0	0	0	0	0	0	0	0	0	0							0	1	2	3	4	5	6
1	0	2	2	2	2	3	3	3	0		-1	0.5	1		0	0	-1.5	-1.5				
2	0	1	0	1	1	1	1	0	0		-1	0	0		1							
3	0	1	1	3	3	0	0	0	0		0	0	0.5		2							
4	0	1	1	3	2	0	0	3	0						3							
5	0	1	1	3	2	0	0	3	0						4							
6	0	1	3	3	2	0	0	3	0						5							
7	0	3	3	3	2	0	0	3	0						6							
8	0	0	0	0	0	0	0	0	0													

		v	olume	de en	trada 7	x 7 +	paddii	ng								Volu	me de	saída			
	0	1	2	3	4	5	6	7	8	Filtr	o W (3	3 x 3)									
0	0	0	0	0	0	0	0	0	0						0	1	2	3	4	5	6
1	0	2	2	2	2	3	3	3	0	-1	0.5	1		0	0	-1.5	-1.5	-1.5			
2	0	1	0	1	1	1	1	0	0	-1	0	0		1							
3	0	1	1	3	3	0	0	0	0	0	0	0.5		2							
4	0	1	1	3	2	0	0	3	0					3							
5	0	1	1	3	2	0	0	3	0					4							
6	0	1	3	3	2	0	0	3	0					5							
7	0	3	3	3	2	0	0	3	0					6							
8	0	0	0	0	0	0	0	0	0												

		v	olume	de ent	trada 7	x 7 +	paddir	ng									Volu	me de	saída			
	0	1	2	3	4	5	6	7	8		Filtro	o W (3	x 3)									
0	0	0	0	0	0	0	0	0	0							0	1	2	3	4	5	6
1	0	2	2	2	2	3	3	3	0		-1	0.5	1		0	0	-1.5	-1.5	-1.5			
2	0	1	0	1	1	1	1	0	0		-1	0	0		1							
3	0	1	1	3	3	0	0	0	0		0	0	0.5		2							
4	0	1	1	3	2	0	0	3	0						3							
5	0	1	1	3	2	0	0	3	0						4							
6	0	1	3	3	2	0	0	3	0						5							
7	0	3	3	3	2	0	0	3	0						6							
8	0	0	0	0	0	0	0	0	0													

		ν	2 2 2 3 3				ng								Volu	me de	saída				
	0	1	2	3	4	5	6	7	8	Filt	o W (3 x 3)									
0	0	0	0	0	0	0	0	0	0						0	1	2	3	4	5	6
1	0	2	2	2	2	3	3	3	0	-1	0.5	1		0	0	-1.5	-1.5	-1.5	-1.5	-3	
2	0	1	0	1	1	1	1	0	0	-1	0	0		1							
3	0	1	1	3	3	0	0	0	0	0	0	0.5		2							
4	0	1	1	3	2	0	0	3	0					3							
5	0	1	1	3	2	0	0	3	0					4							
6	0	1	3	3	2	0	0	3	0					5							
7	0	3	3	3	2	0	0	3	0					6							
8	0	0	0	0	0	0	0	0	0												

- ► Convolução em profundidade: quando a entrada possui mais do que 1 canal
 - ▶ O filtro terá $k \times k \times p$, onde p é a quantidade de canais de entrada

Vol	lume d	e entra	ada 6 x	6 x 3	(RGB) + zei	o pado	ling	1	Filtro	W (3	х 3 х	3)					V	olume	e de sa	ída 6 x	6	
	0	1	2	3	4	5	6	7		-1	0.5	1											
0	0	0	0	0	0	0	0	0		-1	0	0						0	1	2	3	4	5
0	0	2	2	2	2	3	3	0		0	0	0.5					0						
1	0	1	0	1	1	1	1	0			1	0	1				1						
2	0	1	1	3	3	0	0	0			-1	1	0				2						
3	0	1	1	3	2	0	0	0			0	0	-0.5				3						
4	0	1	1	3	2	0	0	0				1	0	1			4						
5	0	1	3	3	2	0	0	0				1	0.5	-1			5						
7	0	0	0	0	0	0	0	0				0	1	0									
	0	1	2	3	4	5	6	7			0	1	2	3	4	5	6	7					
0	0	0	0	0	0	0	0	0		0	0	0	0	0	0	0	0	0					
1	0	3	3	1	1	1	1	0		1	0	0	3	2	2	3	2	0					
2	0	3	0	3	1	1	1	0		2	0	1	1	1	1	1	1	0					
3	0	3	3	3	3	0	0	0		3	0	1	2	3	3	0	0	0					
4	0	0	0	3	2	0	0	0		4	0	1	2	3	1	1	1	0					
5	0	1	2	2	2	0	0	0		5	0	1	3	3	2	1	1	0					
6	0	2	2	2	2	0	0	0		6	0	3	3	0	2	0	2	0					
7	0	0	0	0	0	0	0	0		7	0	0	0	0	0	0	0	0					

Vol	ume d	e entra	ıda 6 x	6 x 3	(RGB) + zer	o pado	ling			Filtro	W (3	х 3 х	3)					V	olume	de sa	ida 6 x	6	
	0	1	2	3	4	5	6	7			-1	0.5	1											
0	0	0	0_	_0_	-0-	- o-	- o-	_o_	_		-1	0	0						0	1	2	3	4	5
0	0	2	2	2	2	3	3	0			0	0	0.5					0	1					
1	0	1	0	1	1	1	1	0				1	0	1				1						
2	0	1	1	3	3_	_ Q_	_ 0_	- 0-			,7-	-1	1	0				2						
3	0	1	1	3	2	0	0	0	-			0	0	-0.5				3						
4	0	1	1	3	2	0	0	0					1	0	1			4						
5	0	1	3	3	2	0,	0	0				-	1	0.5	-1			5						
7	0	0	0	0	^مر	0	0	0				1	o	1	0									
				. "							"	1												
	0	1	2	3	4	5	6	7	,	1		ю	1	2	3	4	5	6	7					
0	٥	0	0	0	0	0	0	_0_			0	, 0	0	0	ο,'	0	0	0	0					
1	0	3	3	1	1	1	1	0			1	0	0	3	2	2	3	2	0					
2	0	3	0	3	1,	1	1	0			2	0	1	1	11	1	1	1	0					
3	0	3	3	3 1	3	0	0	0			3	0	1	2	3	3	0	0	0					
4	0	0	0	3	2	0	0	0			4	0	1	2	3	1	1	1	0					
5	0	1	2	2	2	0	0	0			5	0	1	3	3	2	1	1	0					
6	0	2	2	2	2	0	0	0			6	0	3	3	0	2	0	2	0					
7	0	0	0	0	0	0	0	0			7	0	0	0	0	0	0	0	0					

Vol	lume d	e entra	da 6 x	6 x 3	(RGB) + zei	o pado	ling	Filtro	W (3	х 3 х	3)					V	olume/	de sa	ída 6 x	6	
	0	1	2	3	4	5	6	7	-1	0.5	1											
0	0	0	0	0	0	0	0	0	-1	0	0						0	1	2	3	4	5
0	0	2	2	2	2	3	3	0	0	0	0.5					0	1	-2.5				
1	0	1	0	1	1	1	1	0		1	0	1				1						
2	0	1	1	3	3	0	0	0		-1	1	0				2						
3	0	1	1	3	2	0	0	0		0	0	-0.5				3						
4	0	1	1	3	2	0	0	0			1	0	1			4						
5	0	1	3	3	2	0	0	0			1	0.5	-1			5						
7	0	0	0	0	0	0	0	0			0	1	0									
	0	1	2	3	4	5	6	7		0	1	2	3	4	5	6	7					
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0					
1	0	3	3	1	1	1	1	0	1	0	0	3	2	2	3	2	0					
2	0	3	0	3	1	1	1	0	2	0	1	1	1	1	1	1	0					
3	0	3	3	3	3	0	0	0	3	0	1	2	3	3	0	0	0					
4	0	0	0	3	2	0	0	0	4	0	1	2	3	1	1	1	0					
5	0	1	2	2	2	0	0	0	5	0	1	3	3	2	1	1	0					
6	0	2	2	2	2	0	0	0	6	0	3	3	0	2	0	2	0					
7	0	0	0	0	0	0	0	0	7	0	0	0	0	0	0	0	0					

Volume de entrada 6 x 6 x 3 (RGB) + zero padding									Filtro W (3 x 3 x 3)								Volume de saída 6 x 6						
	0	1	2	3	4	5	6	7		-1	0.5	1											
0	0	0	0	0	0	0	0	0		-1	0	0						0	1	2	3	4	5
1	0	2	2	2	2	3	3	0		0	0	0.5					0	1	-2.5	-1	-1	0.5	2
2	0	1	0	1	1	1	1	0			1	0	1				1	11.5	5.5	16.5	9.5	8	3
3	0	1	1	3	3	0	0	0			-1	1	0				2	4.5	8	7	9	2.5	2.5
4	0	1	1	3	2	0	0	0			0	0	-0.5				3	5.5	14	22	5	0.5	2.5
5	0	1	1	3	2	0	0	-2				1	0	1			4	5.5	12.5	9	7	-0.5	4.5
6	0	1	3	3	2	0	0	-1				1	0.5	-1			5	4	11	6	7	-0.5	4.5
7	0	0	0	0	0	0	0	0				0	1	0									
	_		_	_		_	_	_									_						
	0	1	2	3	4	5	6	7			0	1	2	3	4	5	6	7					
0	0	0	0	0	0	0	0	0		0	0	0	0	0	0	0	0	0					
1	0	3	3	1	1	1	1	0		1	0	0	3	2	2	3	2	0					
2	0	3	0	3	1	1	1	0		2	0	1	1	1	1	1	1	0					
3	0	3	3	3	3	0	0	0		3	0	1	2	3	3	0	0	0					
4	0	0	0	3	2	0	0	0		4	0	1	2	3	1	1	1	0					
5	0	1	2	2	2	0	0	0		5	0	1	3	3	2	1	1	0					
6	0	2	2	2	2	0	0	0		6	0	3	3	0	2	0	2	0					
7	0	0	0	0	0	0	0	0		7	0	0	0	0	0	0	0	0					

Camada convolucional

Entrada
$$(m \times n \times p)$$



e.g.
$$32 \times 32 \times 3$$

Filtro (kernel ou neurônio convolutional) w com tamanho $k \times k \times p$, e.g. $5 \times 5 \times 3$

 Cada neurônio realiza a convolução da entrada e gera um volume (matriz/tensor) de saída

Centrado em um pixel específico, temos, matematicamente

$$w^t x + b$$

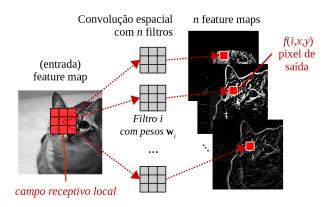
- sim, há a soma de bias para além dos pesos da convolução.

Camada convolucional

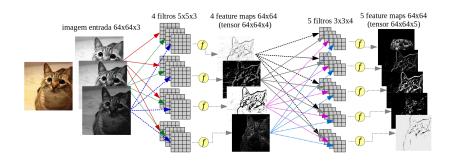
- Mapas de ativação (ou características) são obtidos após convolução e função de ativação (e.g. ReLU);
- Empilhados formam um tensor que será a entrada da próxima camada.



Camada convolucional: campo receptivo local



Camada convolucional: feature maps



Camada convolucional: entrada, filtro, passo

A camada convolucional tem que levar em conta:

- tamanho da entrada (largura, altura, profundidade)
- tamanho do filtro
 - a profundidade deve ser igual à da entrada
 - ▶ altura e largura afetam o campo receptivo local
- stride (passo)
 - ▶ 1 : todos os pixels são filtrados pelo neurônio
 - > 1 : salta um número de pixels em determinada direção, a cada convolução.
 - nesse caso o volume de saída tem tamanho reduzido, ex. com passo 2

Entrada 6 x 6 x 3 (RGB) + zero padding + stride (2,2)										Filtro	W (3	х 3 х	3)					V	olume	de sa	ída 3	х 3	
	0	1	2	3	4	5	6	7		-1	0.5	1											
0	0	0	0	0	0	0	0	0		-1	0	0							0	1	2		
1	0	2	2	2	2	3	3	0		0	0	0.5						0					
2	0	1	0	1	1	1	1	0			1	0	1					1					
3	0	1	1	3	3	0	0	0			-1	1	0					2					
4	0	1	1	3	2	0	0	0			0	0	-0.5										
5	0	1	1	3	2	0	0	0				1	0	1									
6	0	1	3	3	2	0	0	0				1	0.5	-1									
7	0	0	0	0	0	0	0	0				0	1	0									
	0	1	2	3	4	5	6	7			0	1	2	3	4	5	6	7					
0	0	0	0	0	0	0	0	0		0	0	0	0	0	0	0	0	0					
1	0	3	3	1	1	1	1	0		1	0	0	3	2	2	3	2	0					
2	0	3	0	3	1	1	1	0		2	0	1	1	1	1	1	1	0					
3	0	3	3	3	3	0	0	0		3	0	1	2	3	3	0	0	0					
4	0	1	0	3	2	0	0	0		4	0	1	2	0	2	0	0	0					
5	0	1	2	2	2	0	0	0		5	0	1	3	3	2	1	1	0					
6	0	2	2	2	2	0	0	0		6	0	3	3	0	2	0	2	0					
7	0	0	0	0	0	0	0	0		7	0	0	0	0	0	0	0	0					

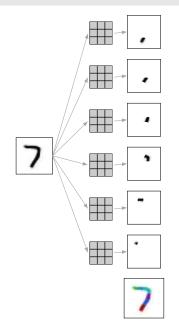
Entrada 6 x 6 x 3 (RGB) + zero padding + stride (2,2)										Filtro	W (3	х 3 х	3)		Bi	as	Volume de saída 3 x 3						
	0	1	2	3	4	5	6	7		-1	0.5	1			0.	.5							
0	0	0	0	0	0	0	0	0		-1	0	0											
0	0	2	2	2	2	3	3	0		0	0	0.5							0	1	2		
1	0	1	0	1	1	1	1	0			1	0	1					0	1.5				
2	0	1	1	3	3	0	0	0			-1	1	0					1					
3	0	1	1	3	2	0	0	0			0	0	-0.5					2					
4	0	1	1	3	2	0	0	0				1	0	1									
5	0	1	3	3	2	0	0	0				1	0.5	-1									
7	0	0	0	0	0	0	0	0				0	1	0									
	0	1	2	3	4	5	6	7			0	1	2	3	4	5	6	7					
0	0	0	0	0	0	0	0	0		0	0	0	0	0	0	0	0	0					
1	0	3	3	1	1	1	1	0		1	0	0	3	2	2	3	2	0					
2	0	3	0	3	1	1	1	0		2	0	1	1	1	1	1	1	0					
3	0	3	3	3	3	0	0	0		3	0	1	2	3	3	0	0	0					
4	0	1	0	3	2	0	0	0		4	0	1	2	0	2	0	0	0					
6	0	1	2	2	2	0	0	0		6	0	1	3	3	2	1	1	0					
7	0	2	2	2	2	0	0	0		7	0	3	3	0	2	0	2	0					
8	0	0	0	0	0	0	0	0		8	0	0	0	0	0	0	0	0					

Entrada 6 x 6 x 3 (RGB) + zero padding + stride (2,2)										Filtro		Bi	as	Volume de saída 3 x 3								
	0	1	2	3	4	5	6	7		-1	0.5	1			0.	.5						
0	0	0	0	0	0	0	0	0		-1	0	0										
0	0	2	2	2	2	3	3	0		0	0	0.5							0	1	2	
1	0	1	0	1	1	1	1	0			1	0	1					0	1.5	-0.5		
2	0	1	1	3	3	0	0	0			-1	1	0					1				
3	0	1	1	3	2	0	0	0			0	0	-0.5					2				
4	0	1	1	3	2	0	0	0				1	0	1								
5	0	1	3	3	2	0	0	0				1	0.5	-1								
7	0	0	0	0	0	0	0	0				0	1	0								
	0	1	2	3	4	5	6	7			0	1	2	3	4	5	6	7				
0	0	0	0	0	0	0	0	0		0	0	0	0	0	0	0	0	0				
1	0	3	3	1	1	1	1	0		1	0	0	3	2	2	3	2	0				
2	0	3	0	3	1	1	1	0		2	0	1	1	1	1	1	1	0				
3	0	3	3	3	3	0	0	0		3	0	1	2	3	3	0	0	0				
4	0	1	0	3	2	0	0	0		4	0	1	2	0	2	0	0	0				
5	0	1	2	2	2	0	0	0		5	0	1	3	3	2	1	1	0				
6	0	2	2	2	2	0	0	0		6	0	3	3	0	2	0	2	0				
7	0	0	0	0	0	0	0	0		7	0	0	0	0	0	0	0	0				

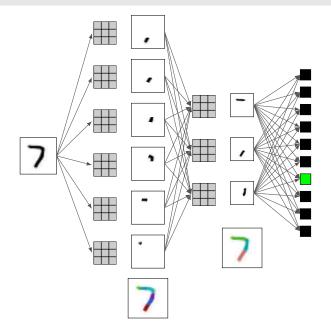
Entrada 6 x 6 x 3 (RGB) + zero padding + stride (2,2)										Filtro	W (3	х 3 х	3)		Bi	as	Volume de saída 3 x 3						
	0	1	2	3	4	5	6	7		-1	0.5	1			0.	.5							
0	0	0	0	0	0	0	0	0		-1	0	0											
1	0	2	2	2	2	3	3	0		0	0	0.5							0	1	2		
2	0	1	0	1	1	1	1	0			1	0	1					0	1.5	-0.5	1		
3	0	1	1	3	3	0	0	0			-1	1	0					1					
4	0	1	1	3	2	0	0	0			0	0	-0.5					2					
5	0	1	1	3	2	0	0	0				1	0	1									
6	0	1	3	3	2	0	0	0				1	0.5	-1									
7	0	0	0	0	0	0	0	0				0	1	0									
	0	1	2	3	4	5	6	7			0	1	2	3	4	5	6	7					
0	0	0	0	0	0	0	0	0		0	0	0	0	0	0	0	0	0					
1	0	3	3	1	1	1	1	0		1	0	0	3	2	2	3	2	0					
2	0	3	0	3	1	1	1	0		2	0	1	1	1	1	1	1	0					
3	0	3	3	3	3	0	0	0		3	0	1	2	3	3	0	0	0					
4	0	1	0	3	2	0	0	0		4	0	1	2	0	2	0	0	0					
5	0	1	2	2	2	0	0	0		5	0	1	3	3	2	1	1	0					
6	0	2	2	2	2	0	0	0		6	0	3	3	0	2	0	2	0					
7	0	0	0	0	0	0	0	0		7	0	0	0	0	0	0	0	0					

Entra	Entrada 6 x 6 x 3 (RGB) + zero padding + stride (2,2)										W (3	х 3 х	3)		Bi	as	Volume de saída 3 x 3						
	0	1	2	3	4	5	6	7		-1	0.5	1			0.	.5							
0	0	0	0	0	0	0	0	0		-1	0	0											
1	0	2	2	2	2	3	3	0		0	0	0.5							0	1	2		
2	0	1	0	1	1	1	1	0			1	0	1					0	1.5	-0.5	1		
3	0	1	1	3	3	0	0	0			-1	1	0					1	5				
4	0	1	1	3	2	0	0	0			0	0	-0.5					2					
5	0	1	1	3	2	0	0	0				1	0	1									
6	0	1	3	3	2	0	0	0				1	0.5	-1									
7	0	0	0	0	0	0	0	0				0	1	0									
	0	1	2	3	4	5	6	7			0	1	2	3	4	5	6	7					
0	0	0	0	0	0	0	0	0		0	0	0	0	0	0	0	0	0					
1	0	3	3	1	1	1	1	0		1	0	0	3	2	2	3	2	0					
2	0	3	0	3	1	1	1	0		2	0	1	1	1	1	1	1	0					
3	0	3	3	3	3	0	0	0		3	0	1	2	3	3	0	0	0					
4	0	1	0	3	2	0	0	0	1	4	0	1	2	0	2	0	0	0					
5	0	1	2	2	2	0	0	0		5	0	1	3	3	2	1	1	0					
6	0	2	2	2	2	0	0	0		6	0	3	3	0	2	0	2	0					
7	0	0	0	0	0	0	0	0		7	0	0	0	0	0	0	0	0					

Classificação de dígitos com conv.layers



Classificação de dígitos com conv.layers



Número de parâmetros em CNNs

$$[(k \times k \times p) + 1] \times d$$
:

- ightharpoonup pesos dos filtros: k imes k imes p , p é dado pela profundidade da entrada
- número de filtros/neurônios: d (cada um gera um mapa de ativação)
- ▶ +1 é o termo bias de cada filtro

Ex: entrada $32 \times 32 \times 3$ e 3 camadas:

- ► Conv.Layer 1: k = 5, d = 8
- ► Conv.Layer 2: k = 3, d = 16
- ► Conv.Layer 3: k = 1, d = 32
- \blacktriangleright # parametros Conv.layer 1: $[(5 \times 5 \times 3) + 1] \times 8 = 608$
- ▶ # parameters Conv.layer 2: $[(3 \times 3 \times 8) + 1] \times 16 = 1168$
- \blacktriangleright # parameters Conv.layer 3: $[(1 \times 1 \times 16) + 1] \times 32 = 544$

Subamostragem: Pooling layer

Opera sobre cada mapa de ativação, reduzindo a dimensão lateral

- max pooling: aplica a operação de máximo local
- average pooling: aplica operação de média local

Ex.: max pooling com tamanho de pool 2 e passo 2.



Usar camadas convolucionais com passo/stride > 1 pode substituir pooling

Pooling layer

Reduzir o tamanho da entrada permite que o filtro opere em regiões maiores da imagem.

Empilhamento de camadas convolucionais aumenta o campo receptivo local não necessitando manter a resolução de entrada



128 x 128



64 x 64 32x32 16x16





(uso de filtro de mesmo tamanho em imagens progressivamente menores)

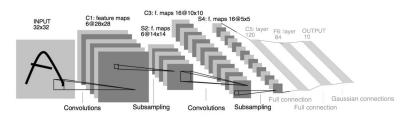
Global pooling

Obtém um valor por canal, como se o tamanho de pool fosse igual às dimensões laterais

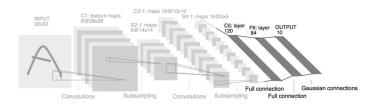
Ex. numa entrada com $40 \times 40 \times 100$, a saída será 100 dimensões.

Voltando à arquitetura: camadas convolucionais

- Extraem características espaciais



Camadas densas e saída



Dense/fully connected (FC) layer:

- ▶ similar à de uma MLP
- pode ser vista como uma projeção dos dados em uma dimensionalidade arbitrária

Saída: comumente densa (ex: classificação e regressão)

- ▶ pode ser vista como um vetor de distribuição de probabilidades
- não é densa em redes completamente convolucionais (Fully Convolutional Networks)

Bibliography I

Moacir A. Ponti, Gabriel Paranhos da Costa. Como funciona o Deep Learning

SBC, 2017. Book chapter.

https://arxiv.org/abs/1806.07908

Moacir A. Ponti, Leo Ribeiro, Tiago Nazaré, Tu Bui, John Collomosse. Everything You Wanted to Know About Deep Learning for Computer Vision but were Afraid to Ask. SIBGRAPI-T, 2017. Tutorial.