Aula 3
Redes Neurais Convolucionais #1

Eduardo L. L. Cabral

Objetivos

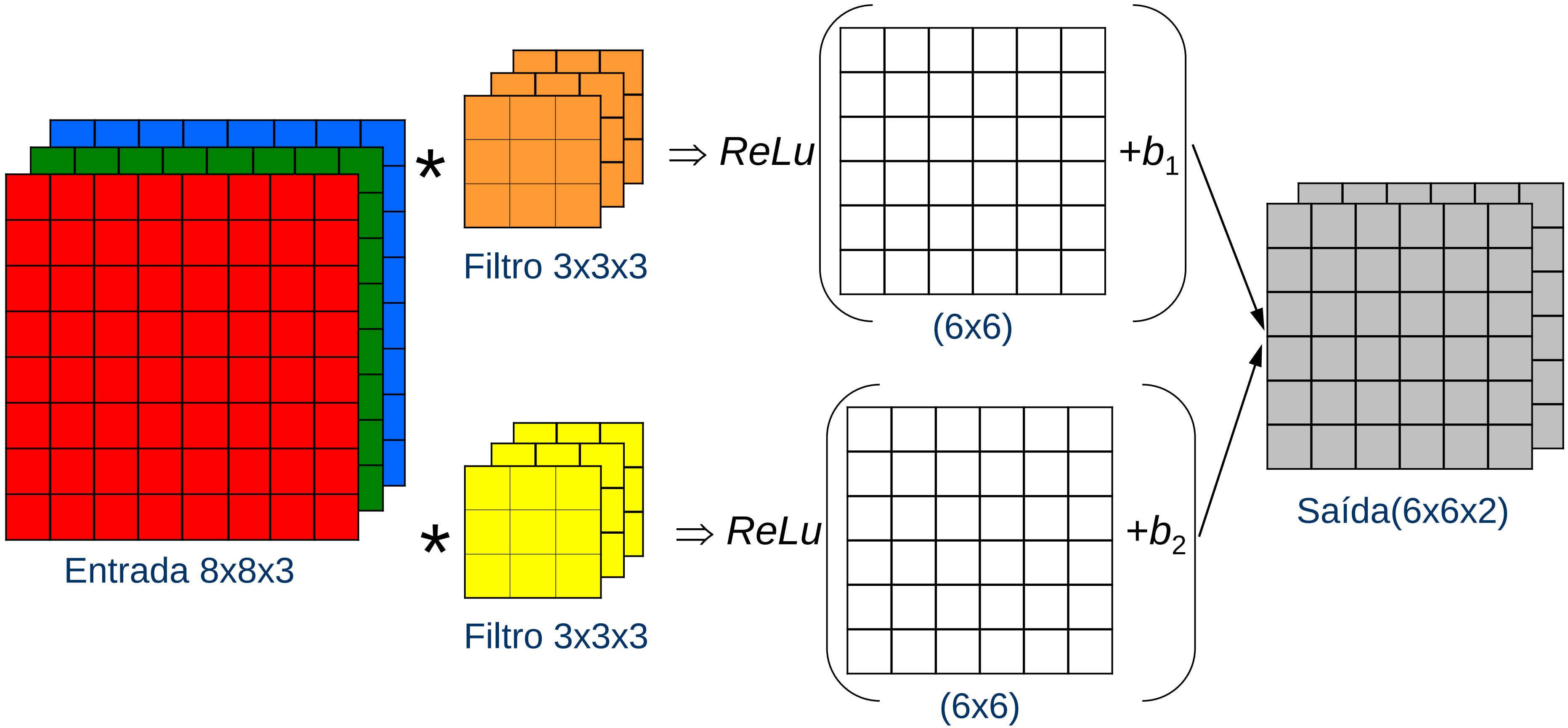
- Apresentar camada convolucional.
- Apresentar camada de “pooling”.
- Apresentar exemplo de uma RNA convolucional simples.
- Apresentar vantagens das RNAs convolucionais.

Camada convolucional

- Uma camada convolucional de uma RNA consiste de uma operação de convolução em um volume usando um conjunto de filtros de múltiplas dimensões.
- A diferença entre uma camada convolucional e um processo de convolução em volume é que na camada convolucional de uma RNA, é adicionado um viés e aplicada uma função de ativação no resultado do processo da convolução.

Camada convolucional

- Exemplo: entrada (8x8x3), com filtro (3x3x3x2), p = 0, s = 1:



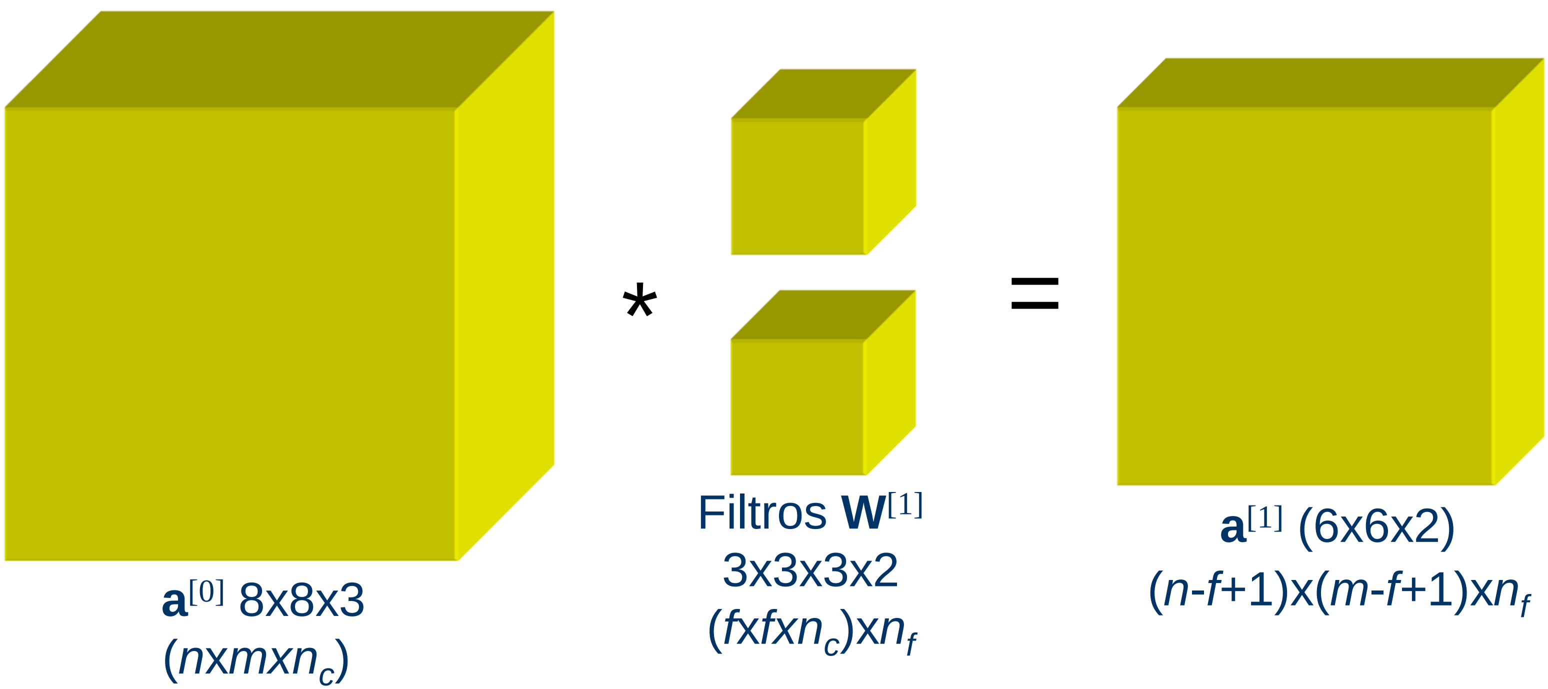
Camada convolucional

- Camada convolucional do exemplo anterior:
  - Entrada $\Rightarrow$ $\mathbf{a}^{[0]}$ = volume (8x8x3)
  - Filtros $\Rightarrow$ $\mathbf{W}^{[1]}$ = volume (3x3x3x2), p = 0, s = 1, $\mathbf{b}$ = vetor (2x1)
  - Saída $\Rightarrow$ $\mathbf{a}^{[1]}$ = volume (6x6x2)
- Convolução é uma multiplicação de uma matriz por um vetor (operação linear).
- Equação da camada:

$$\boxed{\begin{cases} \mathbf{z}^{[1]} = \mathbf{W}^{[1]}\mathbf{a}^{[0]} + \mathbf{b}^{[1]} \\ \mathbf{a}^{[1]} = g^{[1]}(\mathbf{z}^{[1]}) \end{cases}}$$

Camada convolucional

- Camada convolucional ⇒ entrada $\mathbf{a}^{[0]}$ (8x8x3), filtros $\mathbf{W}^{[1]}$ (3x3x3x2), com p = 0, s = 1, função de ativação ⇒ saída $\mathbf{a}^{[1]}$ (6x6x2)



- Se tivesse 10 filtros ⇒ $\mathbf{a}^{[1]}$ seria um volume (6x6x10)

Camada convolucional

- Exemplo de número de parâmetros de uma camada convolucional $\Rightarrow$ camada com 10 filtros de dimensão 3x3x3:
  - Cada filtro tem 3x3x3 = 27 parâmetros + 1 viés = total de 28 parâmetros;
  - Os 10 filtros tem 10x28 = 280 parâmetros.
- Fórmula geral para número de parâmetros de uma camada convolucional:

$$\boxed{N = f^{[l]} \times f^{[l]} \times n_c^{[l-1]} \times n_f^{[l]} + n_f^{[l]}}$$

- Não importa a dimensão da entrada da camada, o número de parâmetros da camada só depende da dimensão e do número de filtros.
- O mesmo conjunto de filtros pode ser usado para detectar características em imagens de qualquer dimensão.

Camada convolucional

- Notação e parâmetros da *l*-ésima camada convolucional:
  - Parâmetros da camada:

    $f^{[l]}$ = dimensão dos filtros;

    $p^{[l]}$ = tamanho do “padding”;

    $s^{[l]}$ = “stride”;

    $n_c^{[l]}$ = número de filtros (antigo n_f);

  - Filtros da camada:

    Dimensão de cada filtro $\Rightarrow$ ($f^{[l]} \text{x} f^{[l]} \text{x} n_c^{[l-1]}$);

    $\mathbf{W}^{[l]}$ = parâmetros dos filtros, dimensão ($f^{[l]} \text{x} f^{[l]} \text{x} n_c^{[l-1]} \text{x} n_c^{[l]}$);

    $\mathbf{b}^{[l]}$ = viéses dos filtros, dimensão ($n_c^{[l]} \text{x} 1$) $\Rightarrow$ tensor de dimensão (1,1,1, $n_c^{[l]}$).

Camada convolucional

- Notação e parâmetros da *l*-ésima camada convolucional:
  - Entrada da camada:

    $\mathbf{a}^{[l-1]} \Rightarrow$ tensor de dimensão $(n_H^{[l-1]} \mathrm{x} n_W^{[l-1]} \mathrm{x} n_c^{[l-1]})$

    Para m exemplos $\Rightarrow$ $\mathbf{A}^{[l-1]}$ tensor de dimensão $(m \mathrm{x} n_H^{[l-1]} \mathrm{x} n_W^{[l-1]} \mathrm{x} n_c^{[l-1]})$.
  - Saída da camada (ativações):

    $\mathbf{a}^{[l]} \Rightarrow$ tensor de dimensão $(n_H^{[l]} \mathrm{x} n_W^{[l]} \mathrm{x} n_c^{[l]})$;

    Para m exemplos $\Rightarrow$ $\mathbf{A}^{[l]}$ tensor de dimensão $(m \mathrm{x} n_H^{[l]} \mathrm{x} n_W^{[l]} \mathrm{x} n_c^{[l]})$.

$$\boxed{n_{H,W}^{[l]} = \left\lfloor \frac{n_{H,W}^{[l-1]} + 2p^{[l]} - f^{[l]}}{s^{[l]}} + 1 \right\rfloor}$$

n_H = altura do tensor, n_W = largura do tensor.

RNA convolucional simples

- Uma RNA convolucional pode ser formada por várias camadas convolucionais.
- Exemplo de uma RNA convolucional para tarefa de classificação:
  - Imagens (64x64x3) $\Rightarrow$ Entrada da RNA $\Rightarrow$ $n_H^{[0]} = n_W^{[0]} = 64,\ n_c^{[0]} = 3$
  - 1ª camada $\Rightarrow$ $f^{[1]} = 3, s^{[1]} = 1, p^{[1]} = 0, n_c^{[1]} = 10$

    Saída da 1ª camada $\Rightarrow$ dimensão (62x62x10)

$$n_{H,W}^{[1]} = \left\lfloor \frac{n_{H,W}^{[0]} + 2p^{[1]} - f^{[1]}}{s^{[1]}} + 1 \right\rfloor = \left\lfloor \frac{64+0-3}{1} + 1 \right\rfloor = 62 \quad (n_H^{[1]} = n_W^{[1]} = 62,\ n_c^{[1]} = 10)$$

  - 2ª camada $\Rightarrow f^{[2]} = 5, s^{[2]} = 2, p^{[2]} = 0, n_c^{[2]} = 20$

    Saída da 2ª camada $\Rightarrow$ dimensão (29x29x20)

$$n_{H,W}^{[2]} = \left\lfloor \frac{n_{H,W}^{[1]} + 2p^{[2]} - f^{[2]}}{s^{[2]}} + 1 \right\rfloor = \left\lfloor \frac{62+0-5}{2} + 1 \right\rfloor = 29 \quad (n_H^{[2]} = n_W^{[2]} = 29,\ n_c^{[2]} = 20)$$

RNA convolucional simples

- Exemplo de uma RNA convolucional para uma tarefa de classificação:
  - 3ª camada $\Rightarrow f^{[3]} = 5, s^{[3]} = 2, p^{[3]} = 0, n_c^{[3]} = 40$

    Saída da 3ª camada $\Rightarrow$ dimensão (13x13x40)

$$n_{H,W}^{[3]} = \left\lfloor \frac{n_{H,W}^{[2]} + 2p^{[3]} - f^{[3]}}{s^{[3]}} + 1 \right\rfloor = \left\lfloor \frac{29+0-5}{2} + 1 \right\rfloor = 13 \ (n_H^{[3]} = n_W^{[3]} = 13,\ n_c^{[3]} = 40)$$

  - 4ª camada $\Rightarrow$ camada de redimensionamento (“flattening”) da saída da 3ª camada para transformar o volume de dimensão (13x13x40) em um vetor de dimensão (6760x1) $\Rightarrow$ objetivo é preparar a saída da 3ª camada para ser processada por uma camada densa.

    Saída da 4ª camada $\Rightarrow$ dimensão (6760x1)

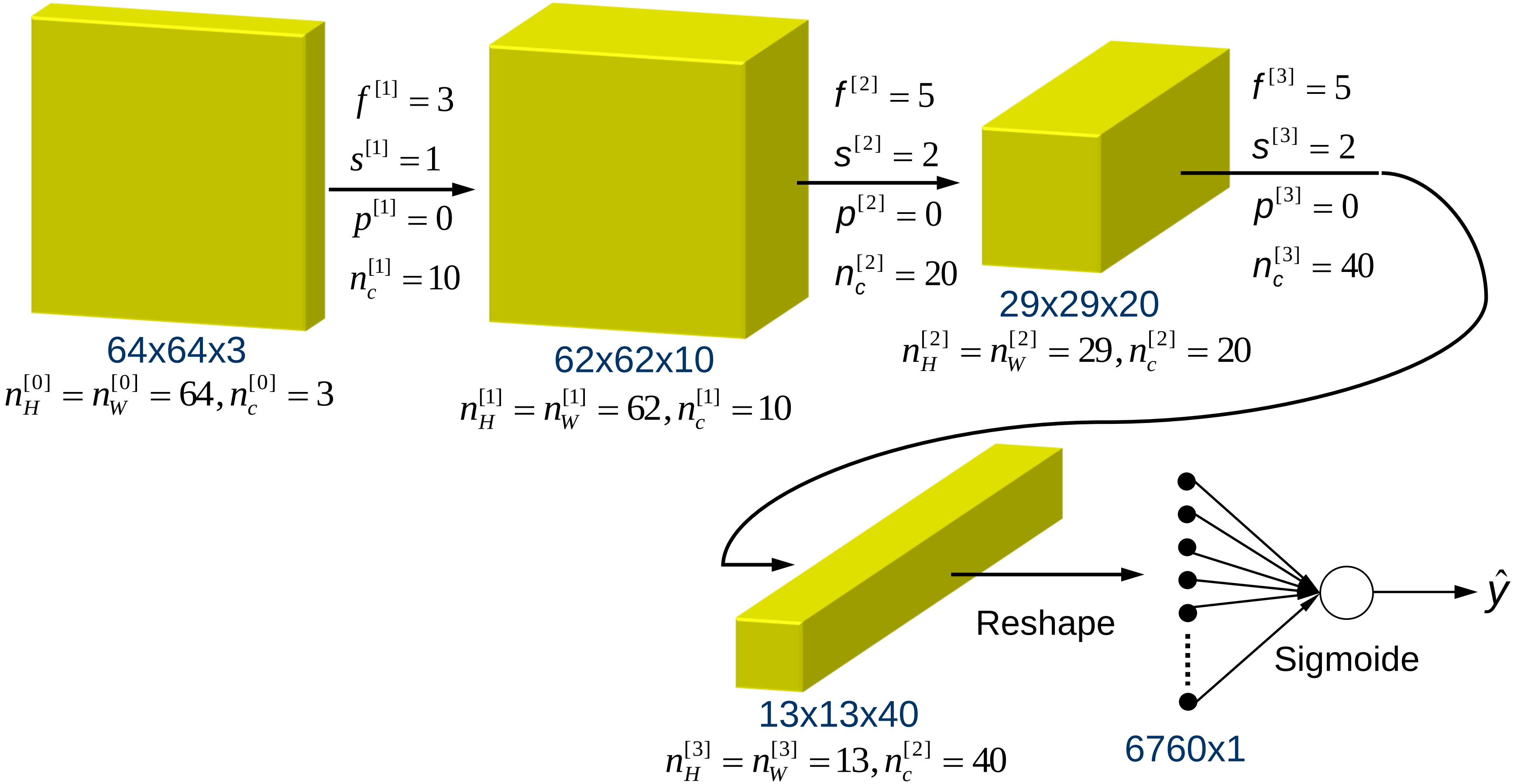
  - 5ª camada (camada de saída) $\Rightarrow$ camada densa de classificação:

    Classificação binária $\Rightarrow$ 1 neurônio com função de ativação sigmóide;

    Classificação multiclasse $\Rightarrow$ camada softmax.

RNA convolucional simples

- Exemplo de uma RNA convolucional para uma tarefa de classificação:



RNA convolucional simples

- Na configuração de uma RNA convolucional deve-se escolher os parâmetros da RNA $\Rightarrow$ número de camadas (L), tamanhos dos filtros (f), número de filtros (n_c), “paddings” (p), “strides” (s), função de ativação, como fazer?
  - ➢ Em geral dimensões das ativações diminuem na medida em que avançamos nas camadas da RNA.
- Tipos de camadas mais comuns em uma RNA convolucional:
  - Convolucional;
  - “Pooling”;
  - Totalmente conectada (densa).
  - Podem existir RNAs convolucionais sem camadas densas.

Camada de “pooling”

- Camadas de “pooling” são essenciais nas RNAs convolucionais.
- Uma camada de “pooling” é usada após uma operação de convolução para detectar onde na imagem uma determinada característica aparece com mais freqüência e de forma mais ressaltada.
- Uma camada de “pooling” reduz o número de ativações.
- Existem dois tipos de camada de “pooling”:
  - “Max pooling” (seleção pelo máximo);
  - “Average pooling” (seleção pela média).
- Uma camada de “pooling” não tem parâmetros a serem treinados.

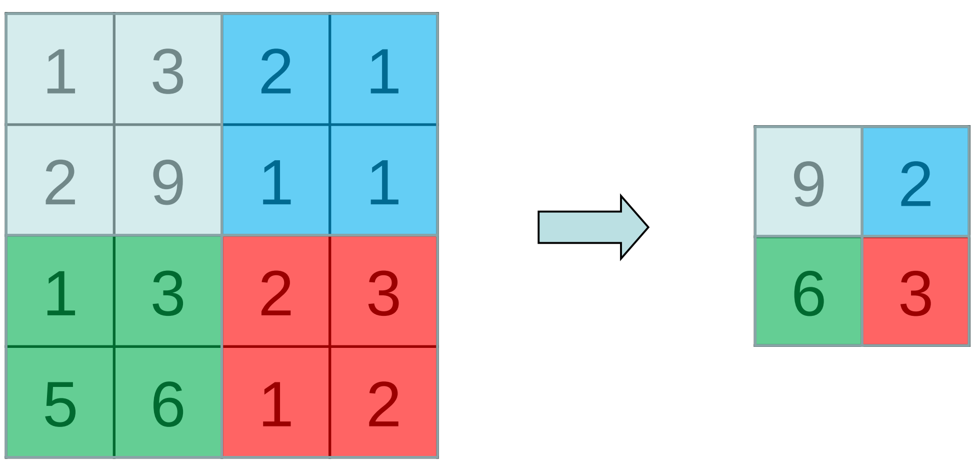
Camada de “max pooling”

- Hiperparâmetros da operação de “pooling”:
  - f = dimensão da janela;
  - s = deslocamento (“stride”) da janela.
- Exemplo de operação de “max pooling” (f = 2, s = 2)

$$\begin{bmatrix} 1 & 3 & 2 & 1 \\ 2 & 9 & 1 & 1 \\ 1 & 3 & 2 & 3 \\ 5 & 6 & 1 & 2 \end{bmatrix} \Rightarrow \begin{bmatrix} 9 & 2 \\ 6 & 3 \end{bmatrix}$$

Camada de “max pooling”

- O que a operação de “max-pooling” realiza:
  - Detecta a característica mais ressaltada dentro da janela;
  - Um valor grande significa que uma dada característica está mais presente;
  - Um valor pequeno significa que uma dada característica está menos presente;
  - A operação de “max-pooling” seleciona os locais da imagem onde uma dada característica está mais ressaltada.



Camada de “max pooling”

- Exemplo de “max-pooling” com f = 3 e s = 1:

1	3	2	1	3
2	9	1	1	5
1	3	2	3	2
8	3	5	1	0
5	6	1	2	9

5x5xn_c

9		

3x3xn_c

Camada de "max pooling"

- Exemplo de "max-pooling" com f = 3 e s = 1:

1	3	2	1	3
2	9	1	1	5
1	3	2	3	2
8	3	5	1	0
5	6	1	2	9

5x5xn_c

9	9	

3x3xn_c

Camada de "max pooling"

- Exemplo de "max-pooling" com $f = 3$ e $s = 1$:

1	3	2	1	3
2	9	1	1	5
1	3	2	3	2
8	3	5	1	0
5	6	1	2	9

$5 \text{x} 5 \text{x} n_c$

9	9	5

$3 \text{x} 3 \text{x} n_c$

Camada de “max pooling”

- Exemplo de “max-pooling” com f = 3 e s = 1:

1	3	2	1	3
2	9	1	1	5
1	3	2	3	2
8	3	5	1	0
5	6	1	2	9

5x5xn_c

9	9	5
9		

3x3xn_c

Camada de “max pooling”

- Exemplo de “max-pooling” com f = 3 e s = 1:

1	3	2	1	3
2	9	1	1	5
1	3	2	3	2
8	3	5	1	0
5	6	1	2	9

5x5xn_c

9	9	5
9	9	

3x3xn_c

Camada de “max pooling”

- Exemplo de “max-pooling” com $f = 3$ e $s = 1$:

1	3	2	1	3
2	9	1	1	5
1	3	2	3	2
8	3	5	1	0
5	6	1	2	9

$5 \times 5 \times n_c$

9	9	5
9	9	5

$3 \times 3 \times n_c$

Camada de “max pooling”

- Exemplo de “max-pooling” com f = 3 e s = 1:

1	3	2	1	3
2	9	1	1	5
1	3	2	3	2
8	3	5	1	0
5	6	1	2	9

5x5xn_c

9	9	5
9	9	5
8		

3x3xn_c

Camada de “max pooling”

- Exemplo de “max-pooling” com f = 3 e s = 1:

1	3	2	1	3
2	9	1	1	5
1	3	2	3	2
8	3	5	1	0
5	6	1	2	9

5x5xn_c

9	9	5
9	9	5
8	6	

3x3xn_c

Camada de "max pooling"

- Exemplo de "max-pooling" com f = 3 e s = 1:

1	3	2	1	3
2	9	1	1	5
1	3	2	3	2
8	3	5	1	0
5	6	1	2	9

5x5xn_c

9	9	5
9	9	5
8	6	9

3x3xn_c

Camada de “max pooling”

- Operação de “pooling” em um volume:
  - Em cada canal do volume de entrada a operação de “pooling” é realizada de forma independente;
  - Preserva o número de canais do volume de entrada.
- Normalmente na camada de “pooling” usa-se “padding”, p = 0.
- Dimensão da saída de uma camada de “pooling $\Rightarrow$ fórmula para cálculo da dimensão da saída é a mesma que para uma camada convolucional (p = 0).

$$\boxed{n_{H,W}^{[l]} = \left\lfloor \frac{n_{H,W}^{[l-1]} - f^{[l]}}{s^{[l]}} + 1 \right\rfloor}$$

Camada de “average pooling”

- A operação de “average pooling” funciona quase da mesma forma que “max pooling”, mas na saída ao invés de se calcular o valor máximo na janela, calcula-se a média dos valores na janela.
- Exemplo de operação de “average pooling” (f = 2, s = 2)

1	3	2	1
2	9	1	1
1	7	2	3
4	6	1	2

→

3,75	1,25
4,5	2

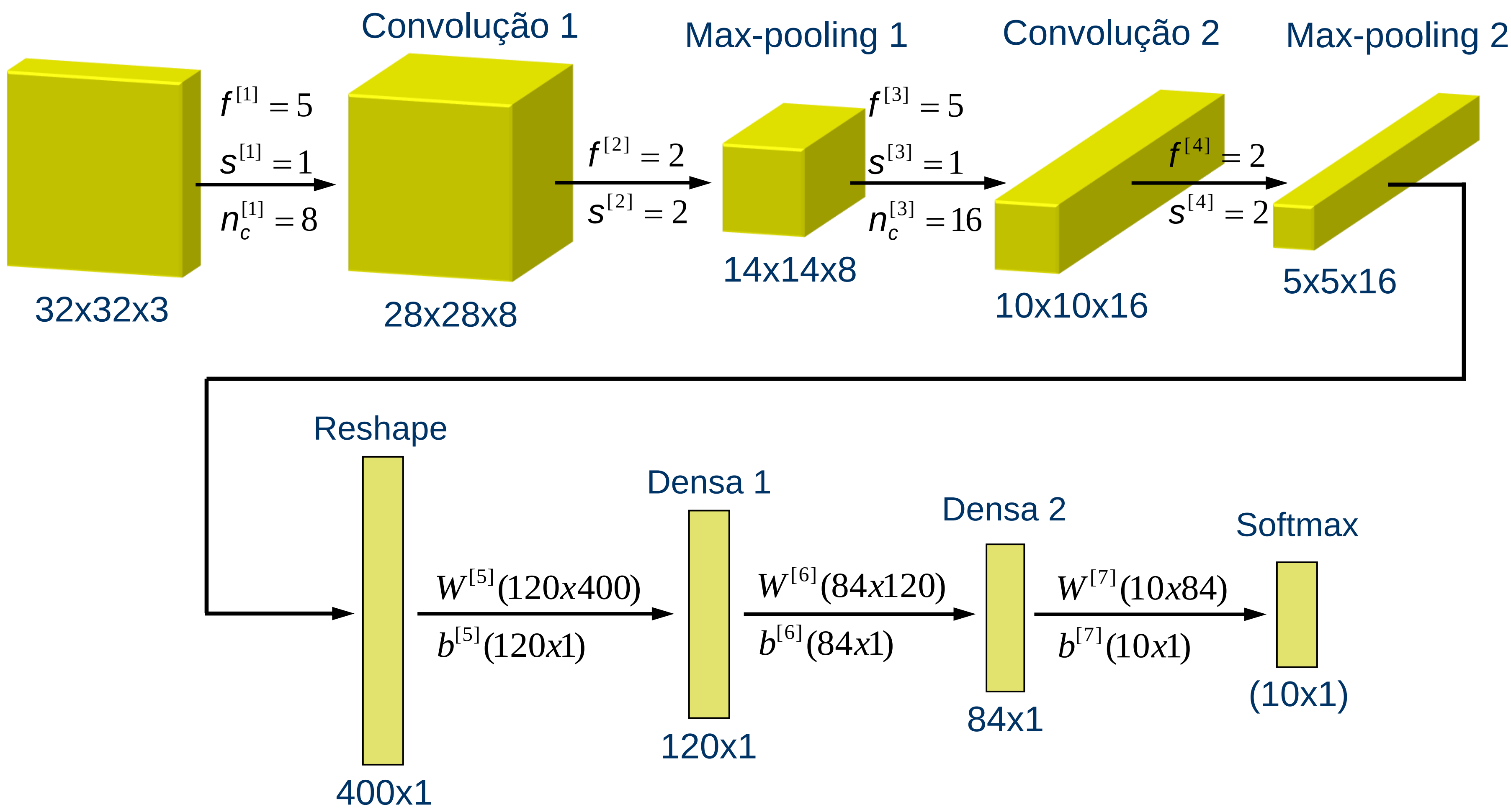
Camada de “average pooling”

- A operação de “average pooling” não é muito usada nas RNAs.
- Operação de “max-pooling” é muito mais usada nas RNAs convolucionias, com exceção das redes muito profundas.
- A operação de “padding” é pouco usada nas camadas de “pooling”, mas pode ser utilizada.
- Importante:
  - Operação de “pooling” não altera número de canais;
  - Camada de “pooling” não tem parâmetros para aprender.

Exemplo de RNA convolucional

- Rede LeNet-5 usada para classificação de dígitos em imagens coloridas.
- Arquitetura:
  - Dimensão das imagens de entrada $\Rightarrow$ 32x32x3;
  - Duas camadas convolucionais, seguidas de camadas de “max-pooling”;
  - Duas camadas densas e uma camada softmax;
  - Todos os “paddings” são iguais a zero.

RNA convolucional LeNet-5



RNA convolucional LeNet-5

Camada	Dimensão	Número de ativações	Número de parâmetros
Entrada	(32,32,3)	3,072	0
Convolução 1 (f=5, s=1, p = 0, n_c=8)	(28,28,8)	6,272	608
"Pooling" 1 (f=2, s=2)	(14,14,8)	1,568	0
Convolução 2 (f=5, s=1, p = 0, n_c=16)	(10,10,16)	1,600	3.216
"Pooling" 1 (f=2, s=2)	(5,5,16)	400	0
Redimensionamento	(400x1)	400	0
Densa 1	(120,1)	120	48.120
Densa 2	(84,1)	84	10.164
Softmax	(10,1)	10	850

RNAs convolucionais

- Arquitetura comum nas RNAs convolucionais:

  conv + *pool* $\rightarrow$ *conv* + *pool* $\rightarrow$ *conv* + *pool* $\rightarrow$ *camadas densas* $\rightarrow$ *camada de classificação*
  - Número de ativações das camadas diminui muito com a profundidade.
  - Camadas densas tem muitos parâmetros.
  - Camadas convolucionais tem poucos parâmetros.
  - Camadas “pooling” não tem parâmetros.
- Uma RNA convolucional tem muitos hiperparâmetros $\Rightarrow$ dificuldade de criar uma RNA convolucional é grande.
- É difícil combinar todos esses tipos de camadas e definir seus hiperpâmetros de forma que a rede consiga realizar a sua tarefa com o desempenho desejado.

RNAs convolucionais

- No lugar de se criar uma nova RNA convolucional o ideal é analisar RNAs convolucionais existentes, que apresentam bom desempenho e a partir dessas modificá-las.
- **Cuidado** $\Rightarrow$ confusão de nomenclatura;
  - Convenção #1 $\Rightarrow$ denominar uma camada convolucional seguida de uma camada de “pooling” como sendo somente uma única camada.
  - Convenção #2 $\Rightarrow$ a camada convolucional seguida de uma camada de “pooling” são consideradas duas camadas.
  - Essas duas convenções são usadas normalmente.

Vantagens das RNAs convolucionais

- RNAs convolucionais são usadas para:
  - Processamento de imagens;
  - Processamento de séries temporais.
- A grande vantagem das RNAs convolucionais é o compartilhamento de parâmetros:
  - Uma camada densa para processar uma imagem de dimensão 32x32x3 (3.072 entradas) e produzir uma saída de 6.272 ativações possui 3.072x6.272 + 6.272 = 19.273.856 parâmetros $\Rightarrow$ uma quantidade muito grande para uma imagem pequena.
  - Uma camada convolucional com 8 filtros de dimensão 5x5 usada para processar a mesma imagem e produzir uma saída com mesmo número de ativações 6.272 (dimensão 28x28x8) possui 5x5x3x8 + 8 = 608 parâmetros.

Vantagens das RNAs convolucionais

- Razões para uma camada convolucional possuir poucos parâmetros:
  - Compartilhamento de parâmetros $\Rightarrow$ um filtro usa os mesmos parâmetros para detectar uma determinada característica em toda a imagem;
  - Os mesmos tipos de características estão presentes em todas as imagens e em princípio em qualquer posição nas imagens;
  - Em uma imagem cada resultado de um filtro depende somente dos valores locais da imagem onde está posicionado o filtro.
- Uma RNA convolucional é invariante a translações $\Rightarrow$ operação de convolução translada um filtro em toda a imagem detectando a presença de uma determinada característica.

Treinamento das RNAs convolucionais

- O treinamento de uma RNA convolucional é realizado da mesma forma que nas RNAs com somente camadas densas.
  - Define-se uma função de custo;
  - Escolha da função de custo depende da tarefa que se quer realizar;
  - Utiliza-se um algoritmo de otimização para minimizar a função de custo;
  - O resultado do treinamento é o cálculo de todos os parâmetros da RNA, que no caso das camadas convolucionais são os valores dos parâmetros dos seus filtros.