



Aula x* – Redes Neurais Artificiais

Análise de Imagens – 2021 2

*

Fora de qq ordem apenas incluída pra motivar a leitura de Preprocessamentos

Aura Conci
aconci@id.uff.br

Redes Neurais Artificiais (RNAs)

são um ramo da Inteligência Artificial (IA), fundada na década de 1940, quando McCulloch e Pitts desenvolveram o primeiro modelo neural artificial.



O problema básico resolvido pelas RNAs é:
a aquisição indutiva de conceitos a partir de exemplos.
Ou seja a capacidade de aprender e generalizar a partir de dados.

A RNA imita a capacidade humana de aprender com a experiência.

Redes Neurais Artificiais (RNAs)

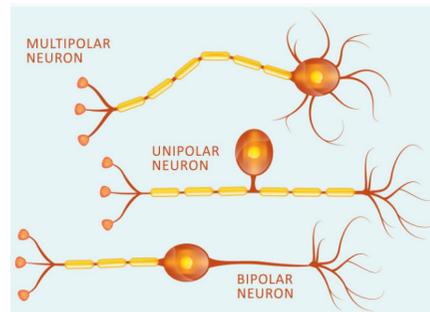
- As RNAs úteis na automatização do processo de aprendizado de regras de várias aplicações.



- Se baseiam em extrair conhecimento de dados com resposta conhecidamente corretas.

Características das RNAs

são esquema ou modelos de computação que representam parcialmente as redes neurais biológicas existentes nos cérebros, expressas por nós conectados (neurônios artificiais) organizados adequadamente em camadas.



- Todos os neurônios artificiais são conectados e capazes de transmitir sinais, geralmente números reais, inteiros ou booleanos, por meio de suas conexões (sinapses artificiais)



Neurônio artificial



é a unidade básica de uma RNA.

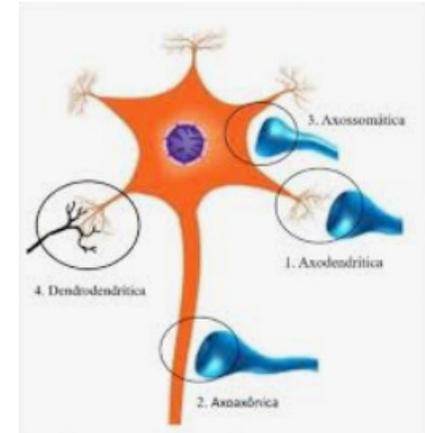
É fundamental para a construção de modelos mais poderosos.

- Em termos matemáticos, o neurônio artificial fornece uma saída para um determinado conjunto de entradas:

$$f(x) = \sigma \left(\sum_{i=1}^n x_i w_i + b \right)$$

Neurônio artificial cont.

$$f(x) = \sigma \left(\sum_{i=1}^n x_i w_i + b \right)$$



em que x_i é a **entrada** i , w_i é o **peso** sináptico associado a entrada i , b é o termo *bias* e σ é a função de ativação.

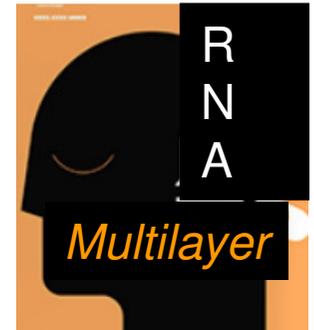
Portanto, cada neurônio combina suas entradas e, em seguida, passa por uma função de ativação, que pode ser um **filtro linear ou não-linear**.

No caso dos filtros não lineares, a função sigmóide e a função tangente hiperbólica são muito usados mas há outros exemplo.

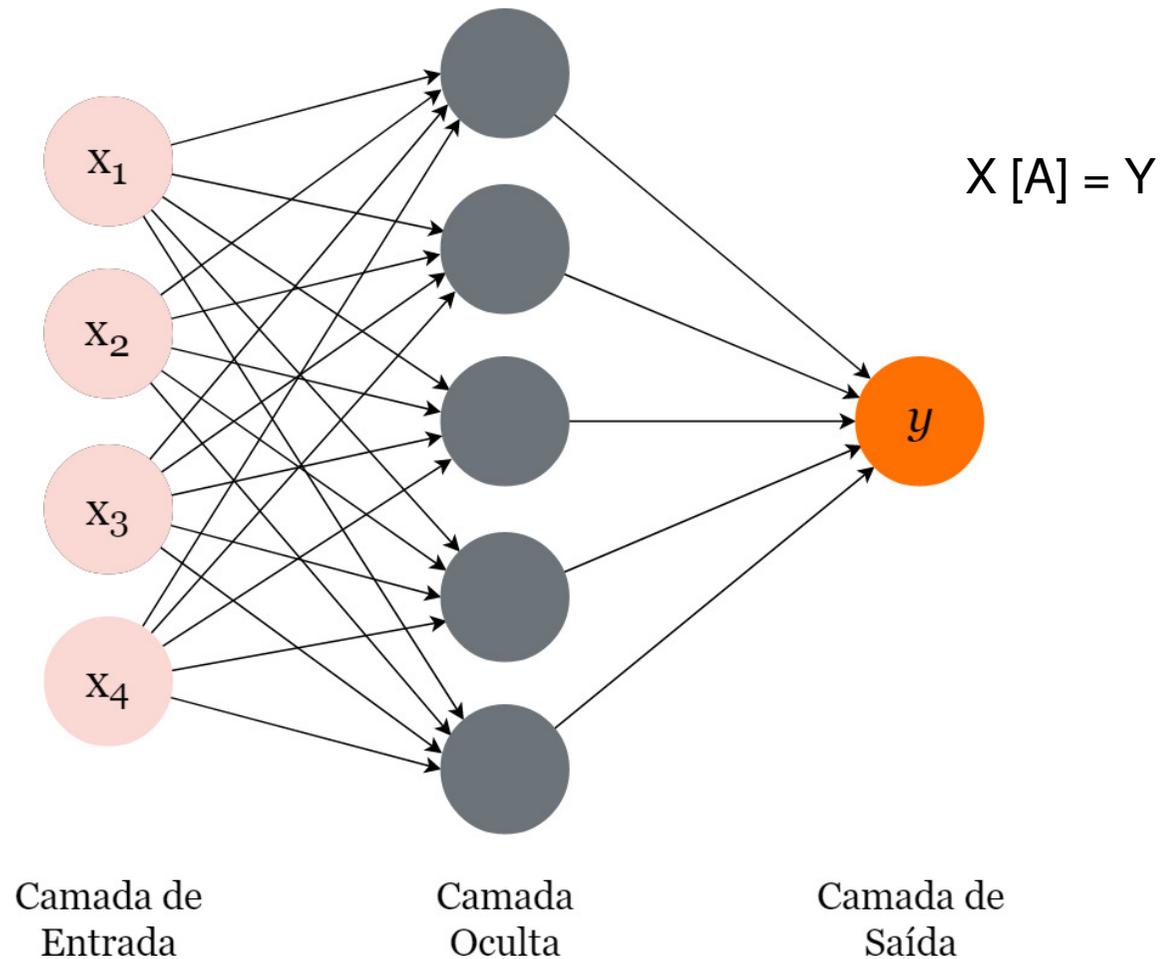
Recentemente, novas funções não lineares têm sido propostas, como as *Rectified Linear Units* (ReLU) e *Leaky Rectified Linear Unit* (Leaky ReLU), cuja velocidade de convergência em relação é até 6 vezes mais rápida.

Redes *Multilayer Perceptron*

- Embora existam inúmeras arquiteturas RNA, a arquitetura *Multilayer Perceptron* (MLP) é a mais frequente.
- Essas são compostas por multi camadas completamente conectadas às camadas subsequentes.
- MLP é composta por uma **camada inicial**, que recebe os dados de entrada do modelo, seguida por uma ou mais **camadas intermediárias** (ocultas) e uma **camada de saída**



Arquitetura básica de MLP com apenas uma camada oculta



Definição dos pesos: feed forward

- Os neurônios de cada camada são totalmente conectados com sua camada adjacente e produz um vetor de saída de acordo com o vetor da camada anterior.
- A saída de cada camada é calculada aplicando a função de ativação de cada neurônio a todos os neurônios da camada:

$$y^l = \sigma(W^l y^{l-1} + b^l)$$

- em que y^l é o vetor de saída, W^l é a matriz dos pesos aplicado a cada par de neurônio da camada l e $l - 1$, e b^l é o vetor do termo *bias* de cada neurônio da camada l .
- Para treinar efetivamente as MLPs, o algoritmo de *backpropagation* é aplicado.

Backpropagation

- Durante o treinamento de uma MLP, uma função de erro é definida para calcular o erro das previsões do modelo com base na saída esperada para os dados de entrada fornecidos. O objetivo do treinamento é minimizar o erro total da rede (soma das funções de erro aplicadas a todos os exemplos de uma base de imagens).
- Para isso, as MLPs usam um algoritmo denominado *backpropagation*.

Backpropagation

(ou retro substituição dos pesos)

- O algoritmo *backpropagation* é caracterizado por dois passos consecutivos:
 - primeiro, os dados são enviados para a camada de entrada da rede, e trafegam pela rede camada por camada, sendo transformados por meio de cálculos realizados até que, na camada de saída, uma resposta seja produzida;
 - segundo, a saída obtida é comparada com a **saída desejada** da entrada correspondente e um **erro** é calculado.



- Em seguida, os erros obtidos são relacionados aos parâmetros da rede e esses são propagados novamente, mas agora da **camada de saída** até a camada de **entrada**, desta forma, os pesos das conexões nas camadas ocultas são modificados à medida que o erro é retro propagado.

$$\begin{aligned}X [A] [A]^{-1} &= Y [A]^{-1} \\X [I]^{-1} &= Y [A]^{-1} \\X &= Y [A]^{-1}\end{aligned}$$

- O processo continua várias vezes até que uma melhoria desejada na previsão do modelo (tolerância) seja alcançada.

Os passos do algoritmo

1. Iniciar os valores dos pesos (sinápses) de cada neurônio aleatoriamente;
2. Apresentar as entradas da rede em um vetor x_1, x_2, \dots, x_n de características e especificar um vetor d_1, d_2, \dots, d_n de saídas desejadas;

3. Calcular as saídas da rede y_1, y_2, \dots, y_n conforme:

$$f(x) = \sigma \left(\sum_{i=1}^n x_i w_i + b \right)$$

4. Reajustar os pesos começando pelos neurônio da camada de saída, retropropagando até a camada de entrada, de acordo com

$$w_{ij}(t + 1) = w_{ij}(t) + \eta \delta_j x_i$$

Os passos do algoritmo

cont.

$$w_{ij}(t + 1) = w_{ij}(t) + \eta \delta_j x_i$$

Onde:

w_{ij} é o peso do neurônio j em uma iteração t , x_i corresponde a um neurônio de saída ou de entrada, η é a taxa de aprendizagem e δ_j é um erro de gradiente para o neurônio j .

Se j for um neurônio de saída, então δ_j é definido por

$$\delta_j = y_j(1 - y_j)(d_j - y_j)$$

onde d_j denota a saída desejada e y_j é a saída desejada da rede.

Se o neurônio j for um neurônio oculto, então δ_j é definido por

$$\delta_j = x_j(1 - x_j) \sum_k \delta_k w_{jk}$$

onde k denota todos os neurônios acima do neurônio j ;

Os passos do algoritmo cont .

5. Retornar ao passo 2 até que uma determinada condição seja satisfeita.

É importante ressaltar que a taxa de aprendizado (η) influencia a magnitude das mudanças dos pesos, desempenhando papel fundamental no aprendizado do modelo.

Taxas de aprendizados muito pequenas implicam em pequenas variações, o que torna o modelo lento e aumenta as chances de parar em mínimos locais.

No entanto, altas taxas de aprendizado tendem a produzir grandes oscilações, o que compromete o processo de aprendizado do modelo.

A taxa de aprendizado é introduzida na rede para permitir uma convergência mais rápida ao erro desejado e ao mesmo tempo evita que a rede oscile, diminuindo a taxa de aprendizado quando o erro tende a aumentar.

Aprendizado Profundo

- Durante o processo de aprendizagem, nós humanos somos inicialmente conduzidos a interpretar e compreender conceitos mais simples, a fim de aprender conceitos mais complexos a partir de conceitos previamente observados ao longo da vida.
- O processo de aprendizagem, denominado aprendizado profundo, sugere uma divisão em camadas hierárquicas do cérebro com diferentes responsabilidades.
- O aprendizado profundo é uma subárea do aprendizado de máquina que usa arquitetura hierárquicas para aprender abstrações de alto nível em um conjunto de imagens.
- É uma abordagem em evolução e tem sido amplamente usada em domínios tradicionais de inteligência artificial, como análise semântica, transferência de aprendizado, processamento de linguagem natural e visão computacional.
- Diferente das RNAs tradicionais, essas técnicas permitem a extração automática de características do conjunto de treinamento, sem a necessidade de técnicas de processamento de imagens e reconhecimento de padrões.
- Existem três fatores importantes para o crescimento do uso aprendizado profundo em Imagens e computação Visual:
 - o aumento da capacidade de processamentos dos *chips* de processamento gráfico,
 - os avanços consideráveis nos algoritmos de aprendizado de máquina e
 - o custo significativamente reduzido do *hardware* de computação (e computação distribuída) .

Vantagens do deep learning

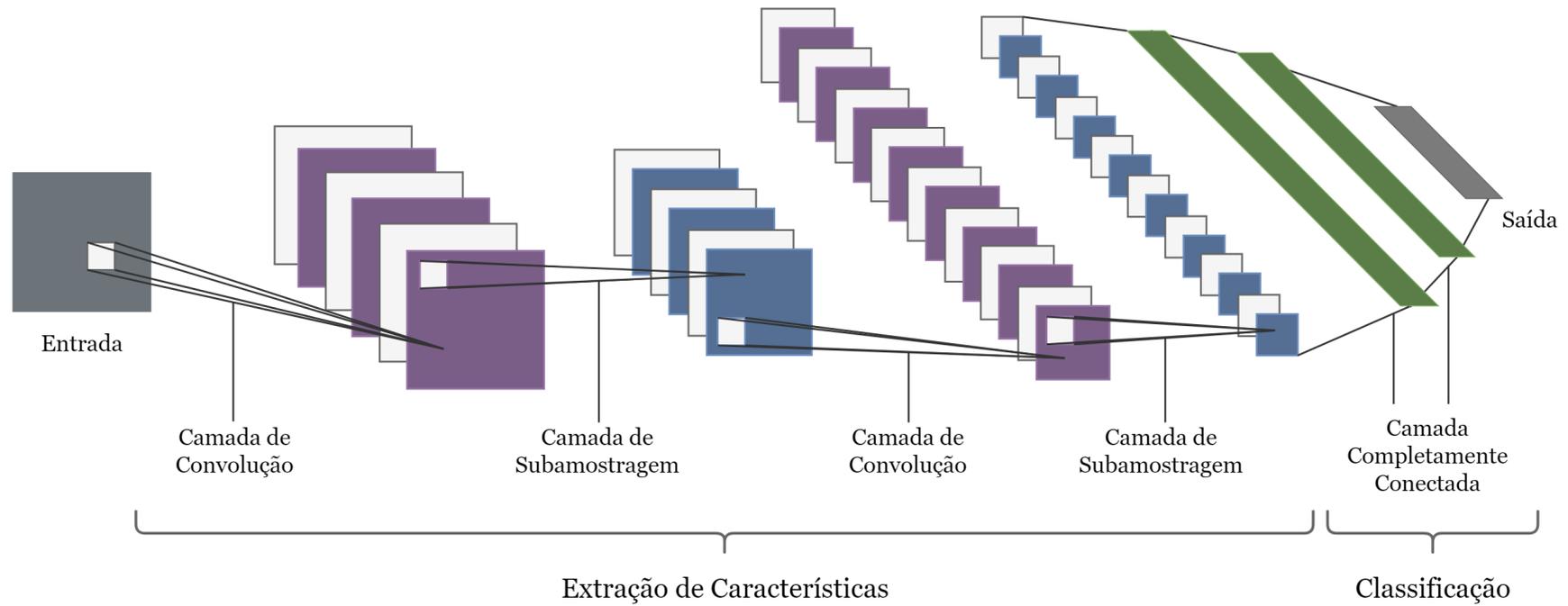
- As etapas de extração, seleção e classificação de características são abstraídas no próprio modelo, com pouca intervenção humana.
(ótimo para serem usadas na forma de uma caixa preta, ou receita de execução)
- Atualmente, existem muitos modelos e técnicas de aprendizagem profundas na literatura, como as redes neurais convolucionais, redes neurais recorrentes, redes de crença profunda, redes de memória de longo prazo, os auto-codificadores esparsos empilhados, entre outras.
- As diferentes arquiteturas têm diferentes vantagens com base na aplicação e nas características dos dados envolvidos.
- Em na **visão computacional**, são mais usadas: as redes neurais convolucionais (CNN) e (em seqüência) as redes neurais recorrentes (RNN) .

Redes Neurais Convolucionais

Convolutional Neural Networks - CNN são capazes de construir um aprendizado hierárquico de características.

- É bastante popular em tarefas de visão computacional (detecção, segmentação e classificação de imagens e vídeos).
- Geralmente tem três tipos de camadas:
 - convolução,
 - subamostragem e
 - conexão completa.

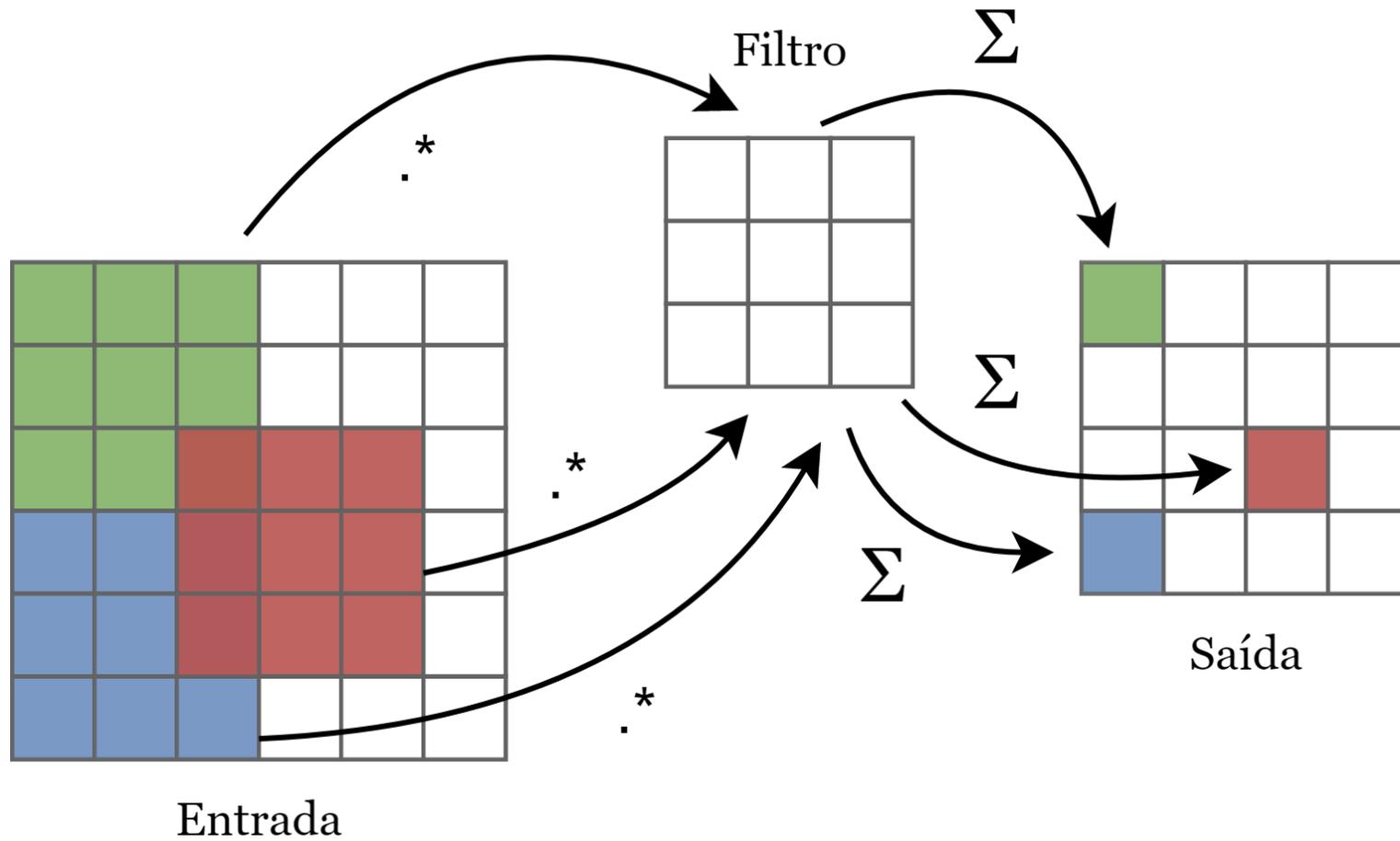
Exemplo de CNN adaptado de (Liu et al., 2017).



Camadas de convolução

- extraem características de baixo nível importantes da imagem de entrada, como textura simples e bordas.
- Usam a idéia de como se aplicam os **filtros por convolução no domínio** do espaço nas bandas das imagens
- À medida que mais camadas de convolução são adicionadas, características de nível mais alto são extraídas.

operação de convolução



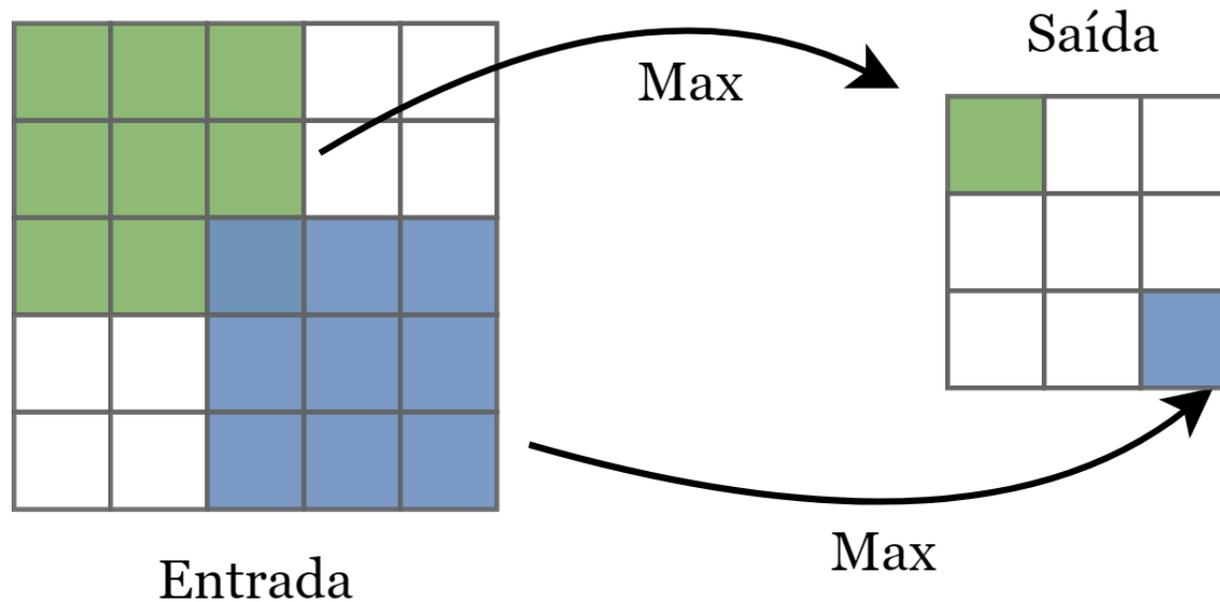
As camadas de convolução

- são composta por filtros treináveis aplicados individualmente na imagem de entrada da rede, gerando diversos mapas de características.
- Os filtros são PREPROCESSAMENTOS LOCAIS definidos por uma pequena área (3×3 , 5×5 , 7×7 *pixels*) e cada neurônio é ligado apenas aos neurônios nas proximidades da camada anterior.
- Para cada filtro aplicado a uma imagem de entrada, pode existir um neurônio conectado com a saída de um subconjunto de neurônios da camada anterior.
- Os pesos são compartilhados entre os neurônios, fazendo com que os filtros aprendam os padrões que ocorrem em qualquer parte da imagem.
- Após gerar mapas de características (FEATURES MAPS) , uma camada de ativação não linear é aplicada.
- Ao fim do treinamento da rede, cada KERNEL de filtro será responsável por detectar uma característica específica da imagem, ou de parte dela.

Camada de Subamostragem

- As camadas de sub amostragem são usadas para reduzir as dimensões dos mapas de características produzidos pela camadas anteriores e selecionar as características desejadas (geralmente devem ser invariantes a rotações, deslocamentos e distorções).
- A subamostragem consiste em aplicar filtros nos mapas de características e usar operações como média (*average pooling*) ou valor máximo (*max-pooling*) para extrair valores dessas sub-regiões, diminuindo sua area.
- Eles passam a fornecer 1 valor que representa toda a area local do kernal usada par ponderar a convolução
- Ou seja, em uma máscara 3 x 3 com 9 valores, apenas o valor do *pixel central* entre eles permanecerá.
- Teremos uma redução dependente da área e de como a janela será considerado na imagem com deslizamento de 1 por 1 pixel ou na forma de áreas não interceptadas. (overleap) .
- Também há operações de *upsampling* que aumentam a resolução ao invés de diminuir.

Exemplo de uma operação com um kernel 3x3 onde apenas o valor central é mantido na janela que escorrega na imagem de pixel a pixel



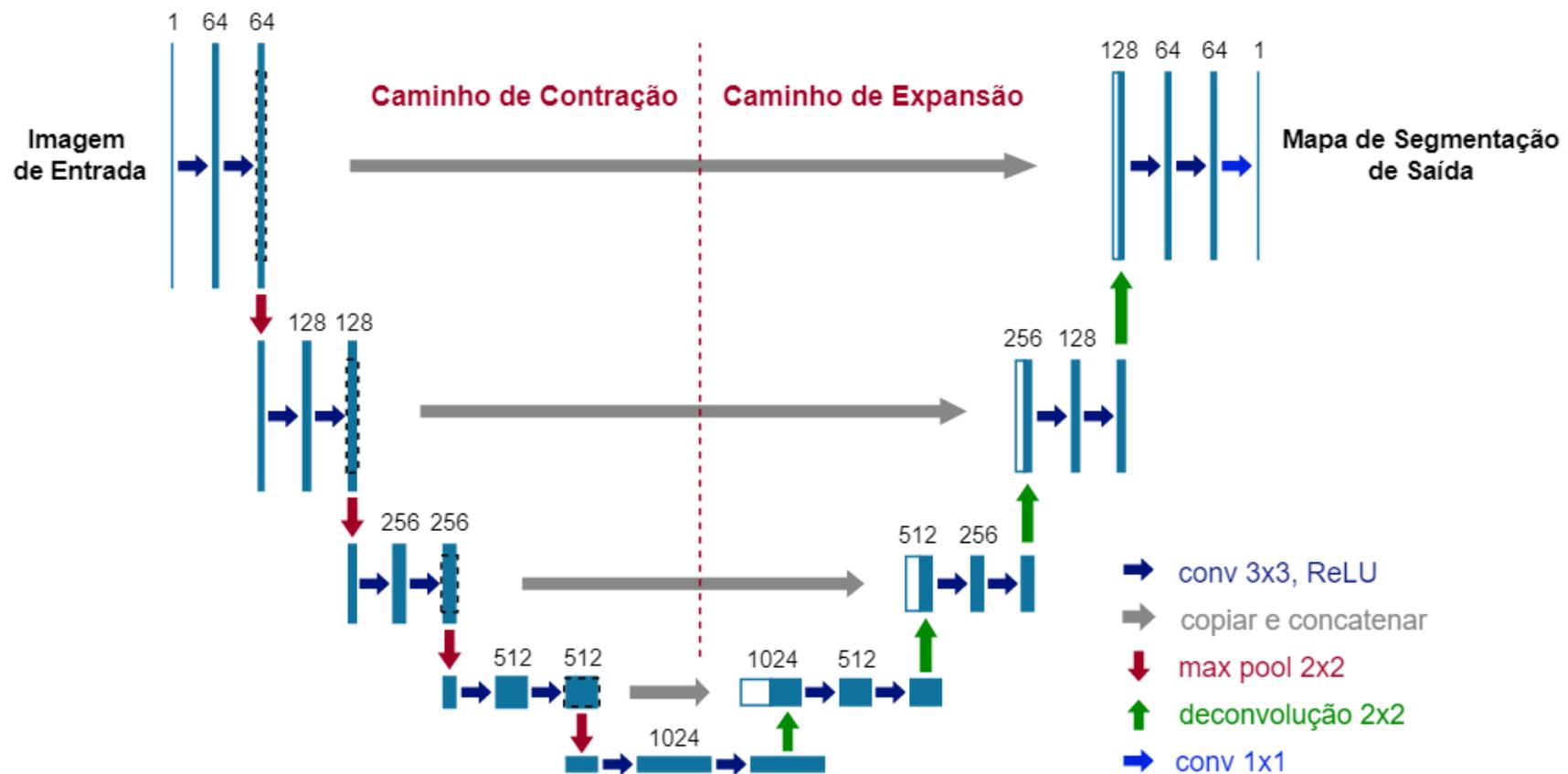
conexão completa

- Os neurônios em uma camada totalmente conectada têm conexões com todas as ativações na camada anterior.
- Geralmente a camada totalmente conectada conecta a rede à camada discriminante (camada de saída), que fornece a saída desejada.
- Se a rede estiver treinada em um nível satisfatório, a rede pode ser usada como uma ferramenta para classificar novos dados.

Exemplificando: U-Net

- é uma RNA completamente convolucional proposta em 2015 para **segmentação de imagens biomédicas**.
- Tem um caminho inicial de **contração** e um caminho posterior de **expansão** simétrico.
- A principal estratégia da U-Net é a combinação entre os mapas de características (features) no estágio de contração e seus **correspondentes simétricos** de expansão.
- O caminho de contração (início) captura o contexto e o caminho de expansão (final) simétrico permite uma localização na imagem, segmentando essa.

Adaptado de (Ronneberger et al. 2015).



Para melhorar o tempo de busca

Aos melhores pesos em cada etapa e local
outras características de configuração das
redes (*hiper parameters*) são geralmente
utilizadas técnicas como:

- ***Transfer learning*** (+ ***fine tuning***) que usa uma rede já treinada para um objetivo em outro similar ; e
- ***Data Augmentation*** que é o fazer artificialmente mais dados de entrada e saída.

Data Augmentation

- técnica comumente usada na literatura para aumentar o número (cardinalidade do conjunto) e a diversidade dos conjuntos de treinamento (pares entrada-saída) .
- Os métodos de *data augmentation*, podem “imaginar” alterações nas imagens, para que tenham um melhor entendimento delas.
- Técnicas de processamento de imagens são usadas para aumentar a diversidade da base de imagens aplicando transformações aleatórias (mas realistas), como:
 - transformações geométricas (por exemplo, rotação, translação e espelhamento) e
 - filtros (passa baixa, alta, mudança de resolução espacial e tonal, etc.) para realçar as imagense modifica-las na entrada sem alterar a saída, por exemplo.

Detalhes desta U-Net exemplificada

- O caminho de contração tem várias aplicações repetidas de:
 - duas convoluções 3x3, cada uma
 - acompanhada por uma função de ativação ReLU e
 - uma operação de sub-amostragem com operação de *max-pooling* 2x2, reduzindo a largura e altura da imagem, com o passo (*stride*) 2 para sub-amostragem.
- Em cada etapa de sub-amostragem, o número de canais de características dobra.
- No caminho de contração as camadas de 1 a 5 são compostas por **convoluções com kernel quadrados** de 64, 128, 256, 512 e 1024, pixels respectivamente.

E no final o que se faz

- Os pesos usados e outras opções da rede são comparados com o resultado de saída conhecido.
- E as respostas corretas e erradas comparada.
- Há quatro tipos de possíveis respostas a essas comparações. que são identificadas em uma tabela chamada de....

Matriz de confusão

A onde se leva em consideração 4 variáveis comparados os resultados obtidos pela rede com os conhecidos e considerados como corretos (groundtruth ou benchmarks) chamados de “padrão ouro” :

- (1) Verdadeiro Positivo (VP) indica os casos realmente positivos detectados;
- (2) Falso Positivo (FP) denota os casos negativos erroneamente detectados como positivos;
- (3) Verdadeiro Negativo (VN) denota os casos negativos verdadeiramente detectados; e
- (4) Falso Negativo (FN) denota os casos positivos detectados erroneamente como negativos.

Com esses resultados se monta uma matriz onde o desejável é só haver valores na diagonal principal, ou seja ser algo como:

Matriz de confusão

		Valor Verdadeiro (confirmado por análise)	
		Positivos	Negativos
Valor Previsto (predito pelo teste)	Positivos	VP Verdadeiro Positivo	FP Falso Positivo
	Negativos	FN Falso Negativo	VN Verdadeiro Negativo

Com esses resultados também se calcula os avaliadores de adequação da rede como as denominadas:

Acurácia (Acc)

- é definida como a razão entre o número de casos classificados corretamente e o número total de casos na amostra do estudo:

$$Acc = \frac{VP + VN}{VP + VN + FP + FN}$$

sensibilidade (Sen)

é a capacidade do sistema em prever corretamente a condição para casos que realmente a tem, ou seja, indica a proporção de verdadeiros positivos.

E definida como

$$Sen = \frac{VP}{VP + FN}$$

A especificidade (Esp)

é a capacidade do sistema em prever corretamente a ausência da condição para casos que realmente não a tem, ou seja, a proporção de verdadeiros negativos

Definida como

$$Spe = \frac{VN}{VN + FP}$$

O índice Dice, também conhecido como F-score e F-measure, É uma estatística usada para comparar a similaridade de duas amostras.

Definida como

$$Dice = \frac{2VP}{2VP + FP + FN}$$

O índice de Jaccard (Jacc),

também conhecido como *Intersection over Union* (IoU).

Ele busca apresentar de forma objetiva o nível de similaridade entre duas amostras.

Pode ser calculado pela razão entre a intersecção de dois conjuntos (marcação obtida pelo método e marcação do especialista) e sua união.

Definido como

$$Jacc = \frac{VP}{VP + FP + FN}$$

São também importantes coisas como o teste de significância

Que é um procedimento estatístico que permite tomar uma decisão entre duas ou mais hipóteses (aceitar ou rejeitar), usando os dados observados de um determinado experimento.

Para isso, é necessário encontrar o *p-value*, que é uma espécie de teste de significância.

Inicialmente, devem ser especificados os elementos para a formulação de um plano de análise, são eles: o método de teste e o nível de significância.

Um método de teste usado pode ser o teste z de duas proporções, pois determina a diferença hipotética entre duas abordagens.

O teste z é apropriado quando o tamanho da amostra é relativamente grande e possui uma distribuição normal.

E o nível de significância adotado pode ser 0,05, onde $p > 0,05$ aceita a hipótese nula, o que indica que não há diferença significativa entre as abordagens.

Mas por hora

- Essa introdução informal já deve motivar a entender pelo menos um pouco o que é filtragem no domínio do espaço por convolução.....
- Que é muito falada na seção seguinte do nosso curso e uma das técnicas de pré-processamento muito usada....

Referencias

- Ronneberger, O.; Fischer, P.; Brox, T. U-net: Convolutional networks for biomedical image segmentation. In: SPRINGER. **International Conference on Medical image computing and computer-assisted intervention**. [S.l.], 2015. v. 9351, p. 234–241. ISBN 978-3-319-24573-7.
- Liu, W.; Wang, Z.; Liu, X.; Zeng, N.; Liu, Y.; Alsaadi, F. E. A survey of deep neural network architectures and their applications. **Neurocomputing**, v. 234, p. 11 – 26, 2017. ISSN 0925-2312.
- M. Moran , M. Faria, G. Giraldi, L. Bastos, L. Oliveira, A. Conci , **Classification of Approximal Caries in Bitewing Radiographs Using Convolutional Neural Networks**, <https://www.mdpi.com/1424-8220/21/15/5192>, Sensors 2021, 21(15), 5192; <https://doi.org/10.3390/s21155192>
- Bruno Barros, Paulo Lacerda, Celio Albuquerque, Aura Conci, "Pulmonary COVID-19: Learning Spatiotemporal Features Combining CNN and LSTM Networks for Lung Ultrasound Video Classification" , Sensors, 2021, 21(16), 5486; Special Issue Image Sensing and Processing with Convolutional Neural Networks, available online: doi: 10.3390/s21165486
- Paulo Lacerda, Bruno Barros, Celio Albuquerque, Aura Conci, Hyperparameter Optimization for COVID-19 Pneumonia Diagnosis Based on Chest CT, Biosensors, Sensors 2021, 21(6), 2174; <https://www.mdpi.com/journal/sensors/sections/biosensors> Biomedical Signal Processing for Disease Diagnosis https://www.mdpi.com/journal/sensors/special_issues/BSPDD , **Sensors MDPI** , ISSN : 1424-82