

Inteligência Artificial na Visão Computacional

[PSI3471 aula 1 parte 1 - início]

I. Áreas envolvidas neste curso

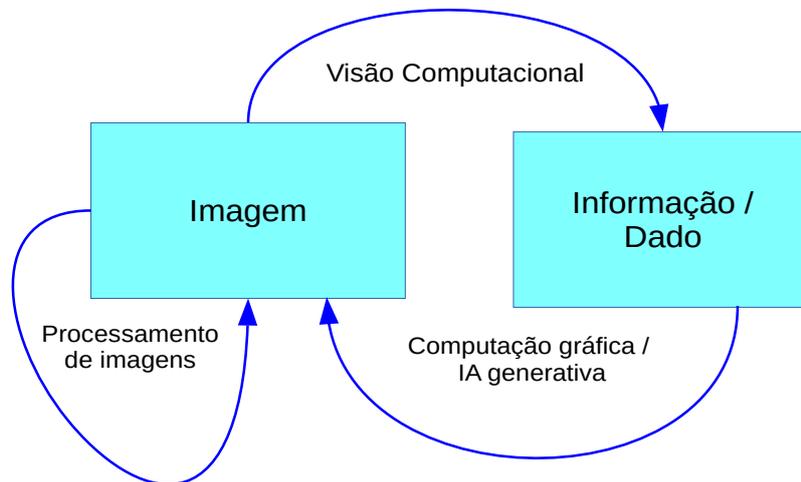


Figura 1: Processamento de imagens e visão computacional.

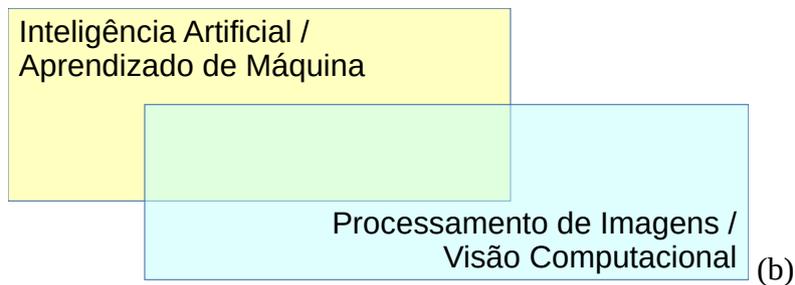
Processamento de imagens manipula imagens, de modo que a entrada e a saída do processo são imagens (figura 1). Por exemplo, reduzir ruído da imagem, ajustar brilho/contraste, aumentar resolução, rotacionar, etc.

Visão computacional busca extrair informações relevantes de imagens, de modo que a entrada do processo são imagens mas a saída são informações contidas na imagem (figura 1). Por exemplo, ler o número da placa de um automóvel, verificar se duas fotos se referem à mesma pessoa, analisar radiografia para verificar se o paciente tem Covid, etc.

Computação gráfica e IA generativa gera imagens a partir de alguma informação fornecida. Não veremos neste curso.



(a)



(b)

Figura 2: Relações entre diferentes áreas envolvidas neste curso.

II. Conteúdo do curso

Neste curso, vamos estudar:

1. Processamento de imagens e visão computacional “convencional” (sem usar aprendizado de máquina – figura 2b ciano sem intersecção).
2. Aprendizado de máquina “clássico”, profundo e rede neural convolucional (figura 2a).
3. Aprendizado de máquina usando rede neural convolucional, com aplicações em processamento de imagens e visão computacional (figura 2b, área em intersecção).

III. Inteligência artificial, aprendizado de máquina e aprendizado profundo

Inteligência artificial (IA) está realizando tarefas que pareciam impossíveis há apenas 10 anos atrás.

Estamos dando passos largos para chegar a um computador que:

- Entende o que vê;
- Lê e entende textos científicos, médicos e jurídicos;
- Compreende a fala humana e conversa em vários idiomas;
- Consegue gerar textos, imagens, músicas, vídeos e programas computacionais de acordo com a descrição fornecida;
- Aprende sozinho, a partir das experiências.

Praticamente todos os dias aparecem várias notícias sobre IA na grande mídia. É só colocar “inteligência artificial” num site de busca de notícias.

[PSI3471 aula 1 – pular a partir daqui]

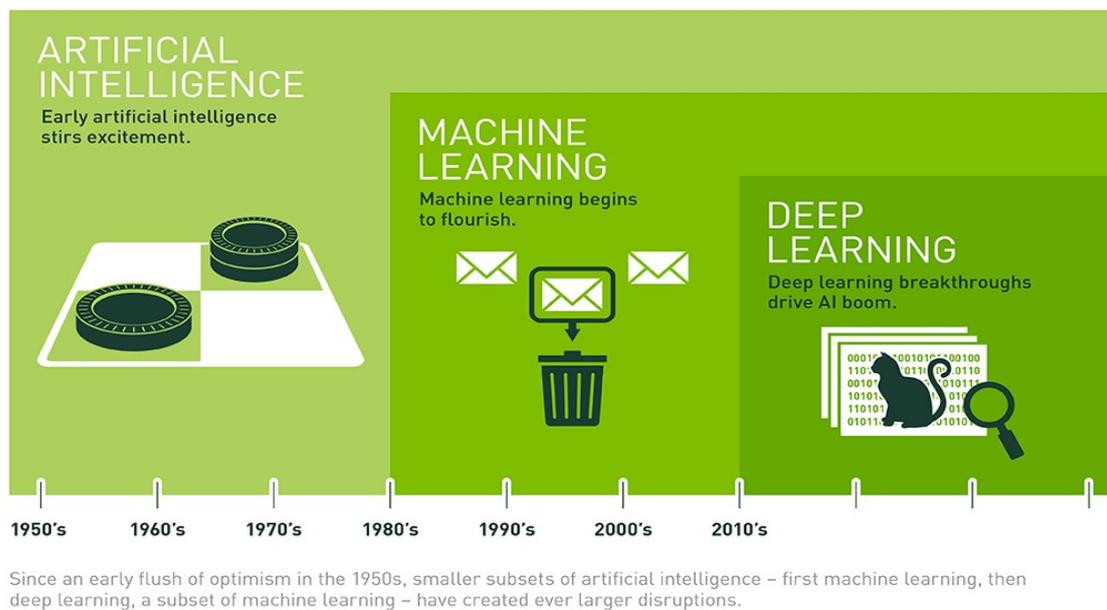


Figura 3: Épocas de pesquisas em inteligência artificial, aprendizado de máquina e aprendizado profundo.

“Inteligência artificial” é estudada desde 1950 (figura 3). Em 1950, Alan Turing apresentou o conceito que hoje é conhecido como teste de Turing. Se uma máquina pudesse manter uma conversa (por meio de um teletipo) que fosse indistinguível de uma conversa com um ser humano, então seria razoável dizer que a máquina estava “pensando”. Em junho de 2014, uma chatbot desenvolvido na Rússia chamada Eugene passou no teste de Turing. Hoje, com certeza, os grandes modelos de linguagem como ChatGPT, LLaMA, Gemini, Gemma e outros passariam tranquilamente pelo teste de Turing.

Essencialmente, existem duas maneiras de fornecer conhecimento a um computador: design direto e aprendizado de máquina. No design direto, um projetista humano insere o conhecimento diretamente. É a forma convencional de programar o computador. Um programador pode usar instruções IF/ELSE para obter o comportamento desejado de um software. Esta é a forma como se tentava criar sistemas de IA antes do aprendizado de máquina (então chamado de sistema especialista).

Aprendizado de máquina começou a florescer a partir dos anos 1980 (figura 3). No aprendizado de máquina, o computador recebe apenas algumas regras sobre como extrair conhecimentos a partir dos dados. O computador então cria sozinho o modelo usando dados fornecidos pelo ser humano. Este processo normalmente usa algum método que procura diminuir a diferença entre o que a máquina faz e o que deveria fazer. O programa final terá muitas vezes um desempenho superior ao programa escrito pelos humanos através de design direto.

As redes neurais são uma das técnicas usadas no aprendizado de máquina (figura 2).

[PSI3471 aula 1 – pular até aqui]

Para exemplificar a diferença entre programação convencional e aprendizado de máquina, tente escrever um programa que classifica a imagem de um rosto humano em masculino ou feminino (figura 4). É praticamente impossível escrever esse programa! Tanto homem como mulher possui dois olhos, duas sobrancelhas, um nariz e uma boca. O que usaríamos como “condição” no “IF condição THEN homem ELSE mulher”?

Porém, usando aprendizado de máquina, este problema pode ser facilmente resolvido. Basta alimentar o algoritmo de aprendizado apropriado com centenas rostos de homens e outras centenas de mulheres que obteremos automaticamente um modelo capaz de distinguir rostos masculinos e femininos.

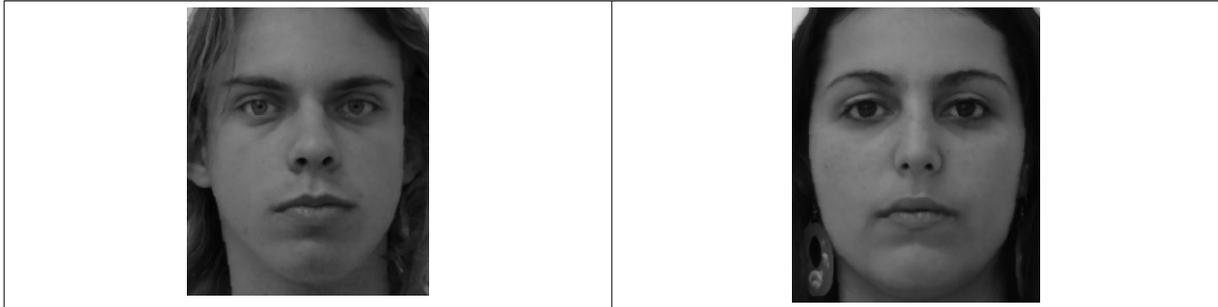


Figura 4: Um programa simples escrito pelos alunos (solução de um exercício-programa EP desta disciplina) comete 1% de erro ao classificar rostos em masculino/feminino.

[PSI3471 aula 1 – pular a partir daqui]

Entre os anos 1980 e 2010, o aprendizado de máquina (não profundo, isto é, clássico) só resolvia uma parte do problema de classificação de imagens. A extração de atributos das imagens, isto é, obter números que caracterizam as imagens, ainda era feito manualmente por seres humanos.

Por exemplo, considere classificar a imagem de uma lesão na pele em câncer maligna ou lesão benigna. Os seguintes atributos caracterizam câncer:

- Assimetria: uma metade da lesão é diferente da outra;
- Bordas irregulares: contorno mal definido;
- Cor variável: presença de várias cores numa mesma lesão;

Um programador humano escrevia uma função que calculava a assimetria da lesão (0=simétrico, 1=assimétrico), uma outra função que calculava irregularidade da borda (0=regular, 1=irregular), e um outro que calculava se havia várias cores na lesão (0=uma única cor, 1=várias cores). Aí, extraía os atributos de centenas de imagens de câncer e outras centenas de lesões benignas. Alimentava um algoritmo de aprendizado clássico com esses atributos para obter o sistema final. O aprendizado de máquina ajudava na classificação final mas não na extração dos atributos.

A dificuldade principal para criar sistemas deste tipo está justamente na extração dos atributos. Que atributos permitem distinguir homem/mulher, gato/cachorro, chimpanzé/orangotango, etc.? Consequentemente, esses problemas de classificação eram considerados praticamente insolúveis.

O aprendizado profundo usando rede neural convolucional (figuras 2 e 3) conseguiu automatizar também a extração dos atributos. Usando aprendizado profundo, não é mais necessário escrever manualmente funções que extraem bons atributos (assimetria, borda irregular, cor variável, etc). O próprio algoritmo de aprendizado profundo irá descobrir quais são os melhores atributos para fazer a classificação, a partir dos exemplos de treinamento.

Isto foi uma revolução! Os problemas de visão computacional antes considerados insolúveis passaram a ser resolvidos. A introdução do aprendizado profundo foi por volta de 2010, mas popularizou-se somente a partir de 2015. Portanto, a atual onda de IA tem menos de 10 anos de vida! O progresso foi super rápido.

[PSI3471 aula 1 – pular até aqui]

VI. Algumas aplicações que veremos no curso

Abaixo, apresento algumas aplicações de IA em visão computacional que veremos nos cursos PSI3471/PSI3472 ou PSI5790.

1) Classificação de imagens.

Como desenvolver um sistema que reconhece que na foto tem chimpanzé, orangotango, etc. (figura 5)? Antes do aprendizado profundo, este impossível resolver este problema. Vamos estudar no curso como isto pode ser feito.



Figura 5: Classificar imagens pelo conteúdo.

Vários problemas de classificação de imagens foram objeto de exercícios programas (EPs) deste curso. Num deles, pedi para classificar imagens de rostos em masculino ou feminino (figura 4). As boas soluções cometiam somente 1% de erro.

Num outro EP, pedi para classificar a variedade de grãos de arroz (figura 6).

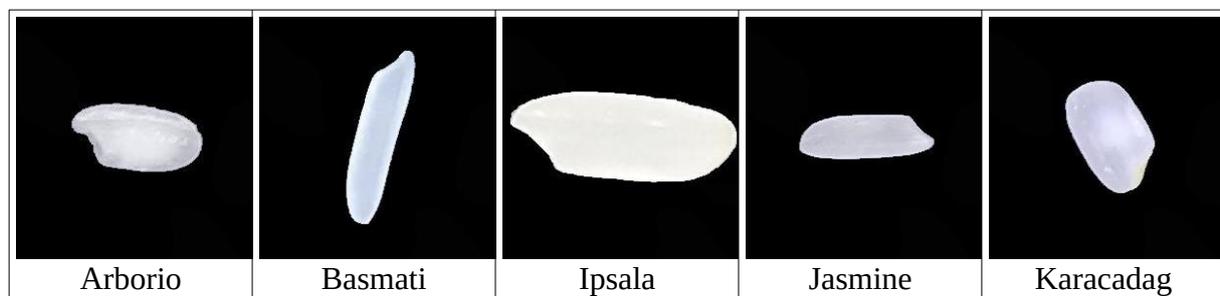


Figura 6: Os 5 tipos de grãos de arroz do conjunto de dados.

Um outro EP consistiu em classificar radiografias de pulmão em “normal” (sadio), com “co-vid-19” e “non-covid” (está com alguma infecção viral ou bacteriana mas não é covid) – figura 7. A taxa de acerto foi aproximadamente 95%.

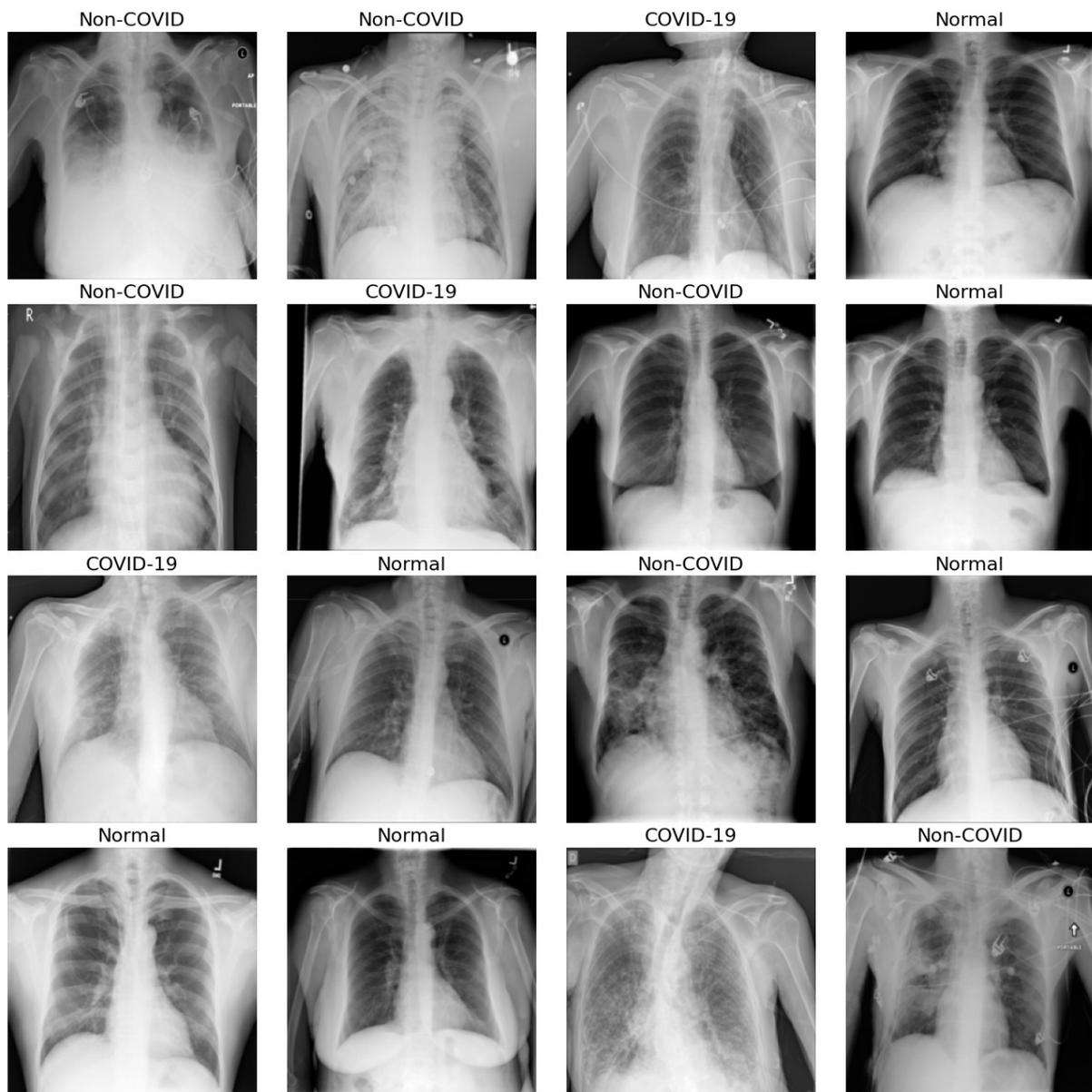


Figura 7: Exemplos de imagens do conjunto de imagens COVID-QU-Ex.

2) Detecção de objetos:

Também vamos estudar a detecção de objetos (figura 8): Localizar na imagem onde está cada objeto.

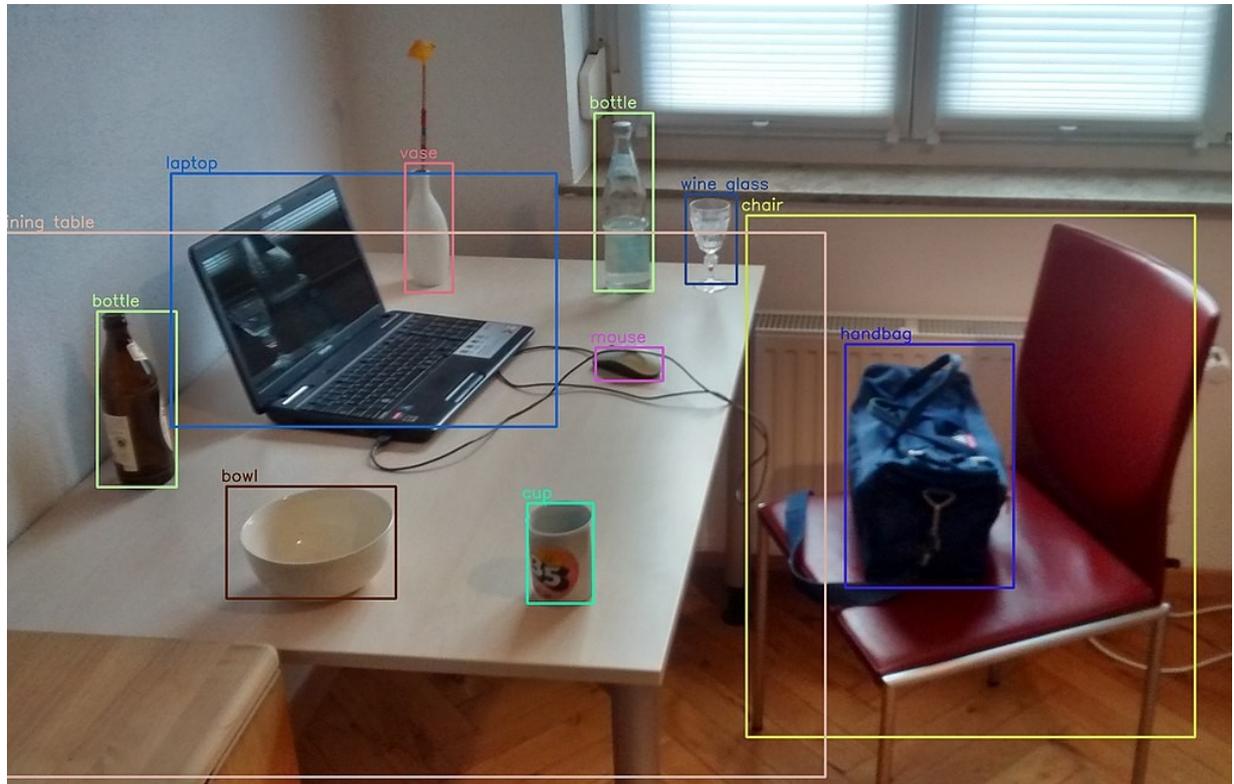


Figura 8: Detecção de objetos.

Um EP pedia para detectar bananas, inseridas em posições, escalas e rotações aleatórias dentro das imagens.

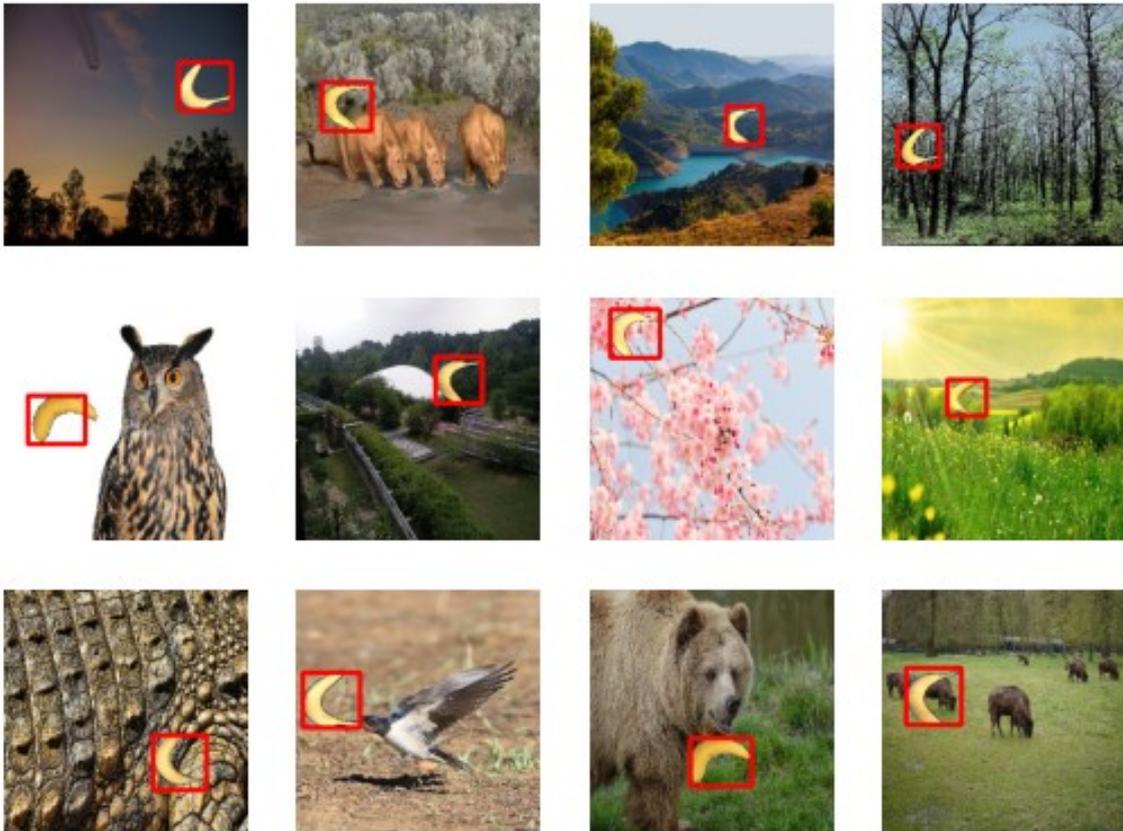


Figura F: Saídas geradas pelo programa que detecta bananas. Erro médio entre posição verdadeira e detectada foi 2,79 pixels.

Na figura 9, além de classificar mamografia (raio-x da mama) em câncer/não-câncer, é possível criar “heatmap” indicando onde provavelmente está a lesão.

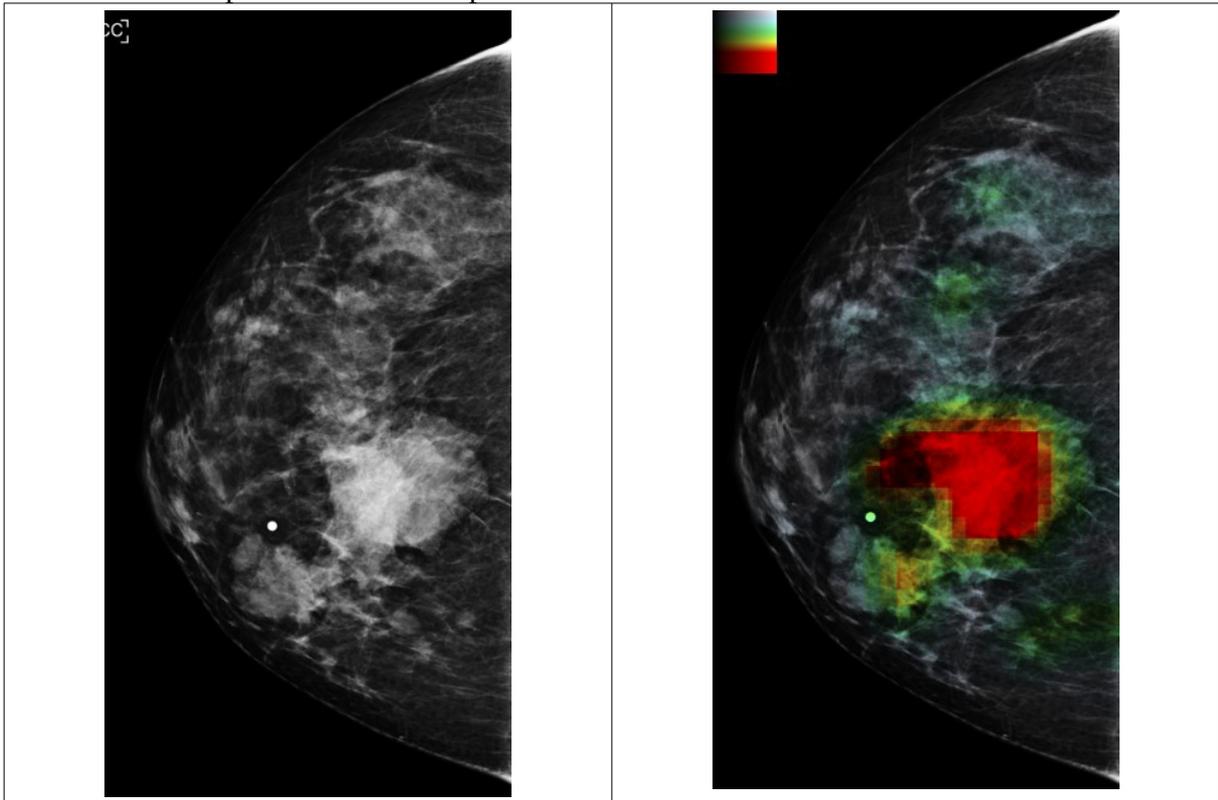


Figura 9: Computador consegue classificar mamografia em lesão maligna, lesão benigna ou sem lesão. Também pode gerar “heatmap” indicando o local provável de câncer.

3) Segmentação semântica:

Segmentação semântica consiste em classificar cada pixel da imagem em classes. Figura 10 mostra a segmentação da imagem em objeto de interesse (chimpanzé) e o fundo da imagem. Figura 11 segmenta as membranas das células, pintando-as de preto.

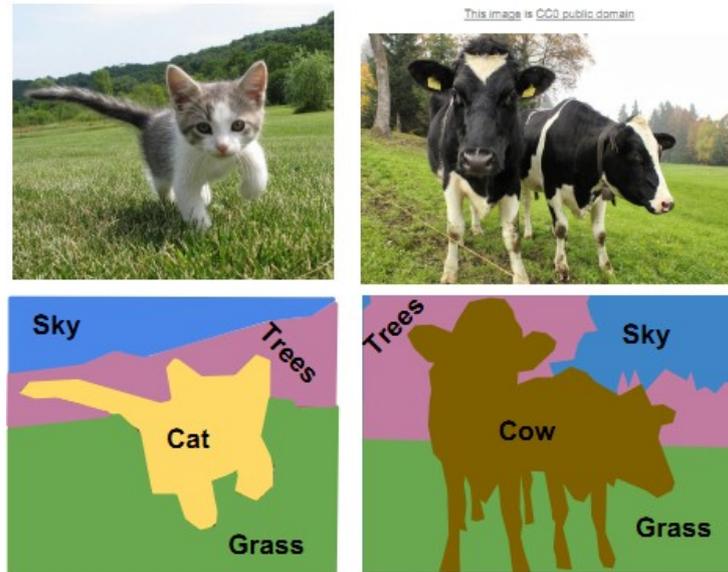


Figura: Segmentação semântica



Figura 10: Segmentação da imagem em objeto de interesse (chimpanzé) e fundo da imagem.

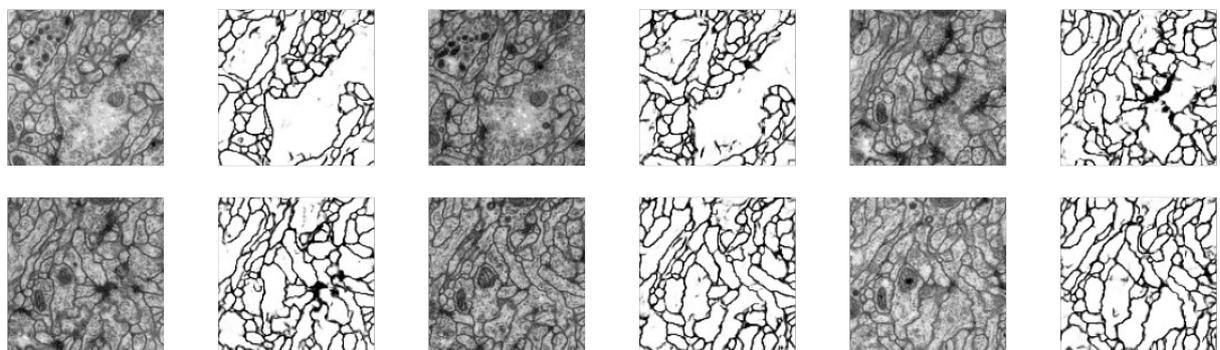


Figura 11: Algumas células (imagens da esquerda) e as membranas das células segmentadas.

4) Colorir fotos preto-e-branco.

É possível usar aprendizado profundo para colorir automaticamente fotos preto-e-branco (figuras 12 e 13). Na internet, há vários documentários coloridos sobre a segunda guerra mundial, criados a partir de filmes originalmente gravados em preto-e-branco.



Figura 12: Exemplos de imagens coloridos artificialmente.

Como colorir uma foto preto-e-branco? Mostrar ao sistema IA muitas imagens coloridas. A partir de uma imagem colorida, o sistema gera imagem em níveis de cinzas correspondente. Tendo muitos pares de imagens de treino (níveis de cinza, coloridas), o sistema aprende a converter foto em tons de cinza para colorida.

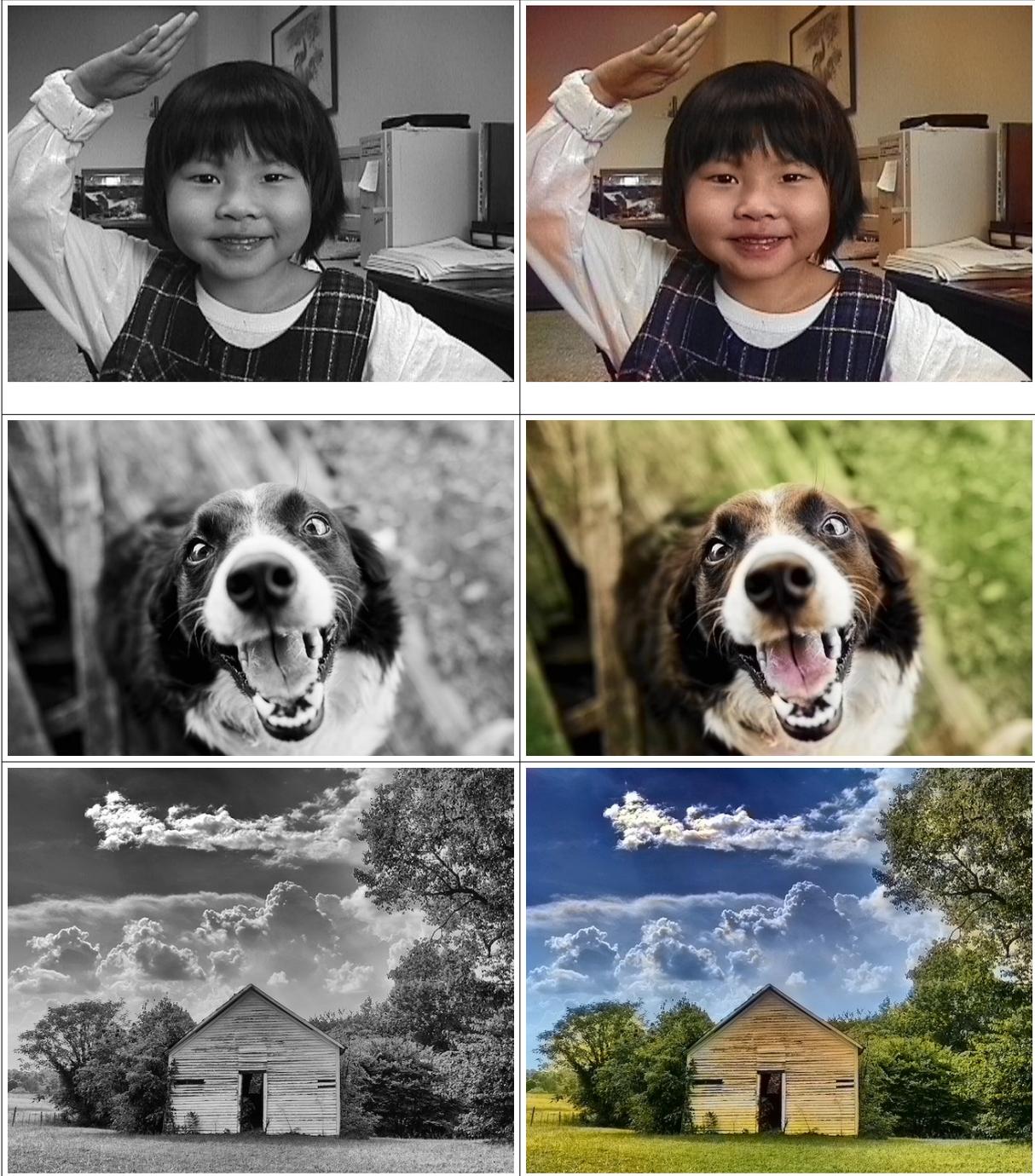


Figura 13: Outros exemplos de colorir automaticamente imagens preto-e-branco.

Figura 14 mostra um EP desta disciplina para colorir imagens. Há casos em que é impossível descobrir a cor original de uma certa região da imagem, por exemplo, a cor do cachecol da moça da última linha. A original era verde mas o sistema pintou de marrom, pois não é possível descobrir a cor original a partir da imagem em níveis de cinzas.



Figura 14: Colorir imagens automaticamente.

5) Super-resolução:

Uma outra aplicação é aumentar a resolução das imagens, para obter imagens mais nítidas. As figuras 15 e 16 mostram o aumento da resolução de imagens 8×8 para 32×32 . A figura 16 corresponde a um EP desta disciplina.

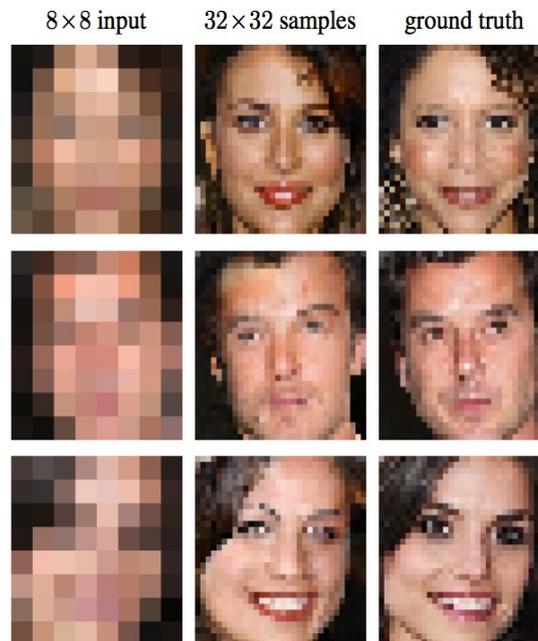


Figura 15: Super-resolução de imagens. Esquerda: Imagem em baixa resolução. Centro: Imagem gerada pela rede neural. Direita: Imagem em alta resolução ideal.



Figura 16: Imagens de super-resolução de um EP desta disciplina. Da esquerda para direita: Imagem original 8×8 ; interpolação bilinear para 32×32 ; interpolação bicúbica para 32×32 ; super-resolução com aprendizado profundo; saída ideal 32×32 .

6) Reconhecimento facial:

Um problema comum é o reconhecimento facial, isto é, dada uma imagem facial, achar a pessoa correspondente num banco de dados (BD) de faces ou dizer que a pessoa não está no BD. Aqui, não podemos usar técnicas tradicionais de classificação, pois não é possível construir um BD com centenas de fotos de cada pessoa.

Num EP da disciplina, dada uma imagem facial, pedia-se que classifique a pessoa num dos 5 “vingadores”. Foi possível atingir taxa de erro de 1,2%.



Figura: Exemplos de imagens faciais dos 5 atores de vingadores.

[PSI3471 aula 1 – pular a partir daqui]

V. Algumas outras aplicações que não veremos no curso

Abaixo, apresento algumas outras aplicações de IA em visão computacional que não veremos neste curso.

1) Computador soluciona melhor algumas “captchas” do que o ser humano.



Figura 17: Exemplos de antigas “captchas”.

“Captchas” são letras e números distorcidos projetados para serem lidos facilmente por seres humanos mas difíceis de serem lidos por um computador (figura 17). Usando classificação de imagens por aprendizado profundo, o computador consegue ler algumas captchas melhor que seres humanos, de forma que perderam a sua utilidade.

2) Gerar faces falsas:

Existe uma ampla área de pesquisa em aprendizado profundo que consiste gerar imagens (figura 18).

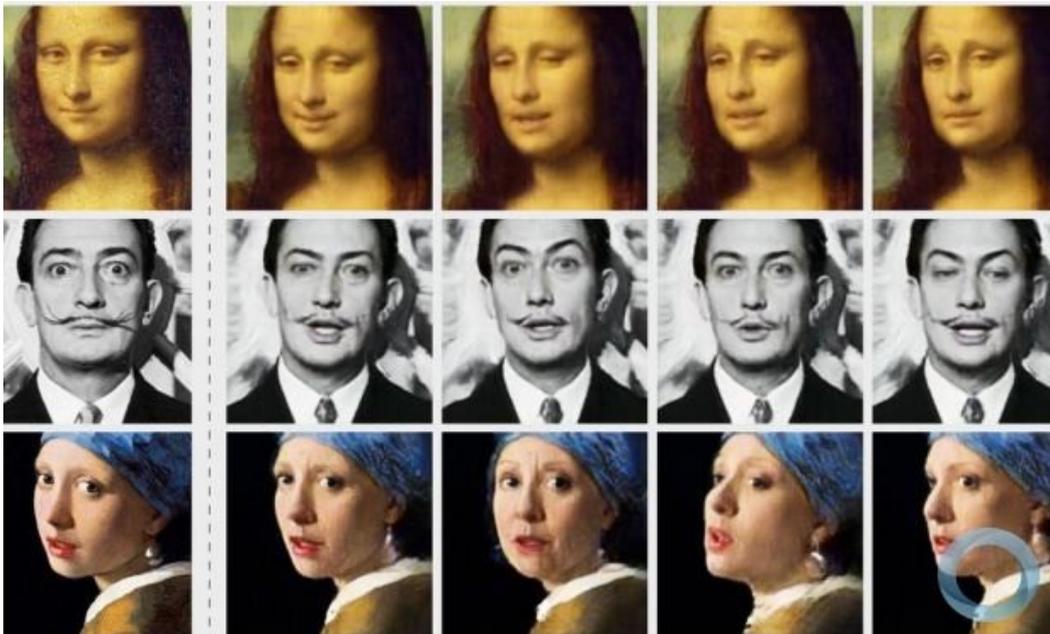


Figura 18: Aprendizado profundo consegue gerar faces falsas.

3) Transferir estilo de pintura

Nesta aplicação, uma foto é redesenhada de acordo com o estilo de pintura de diferentes artistas (figura 19).

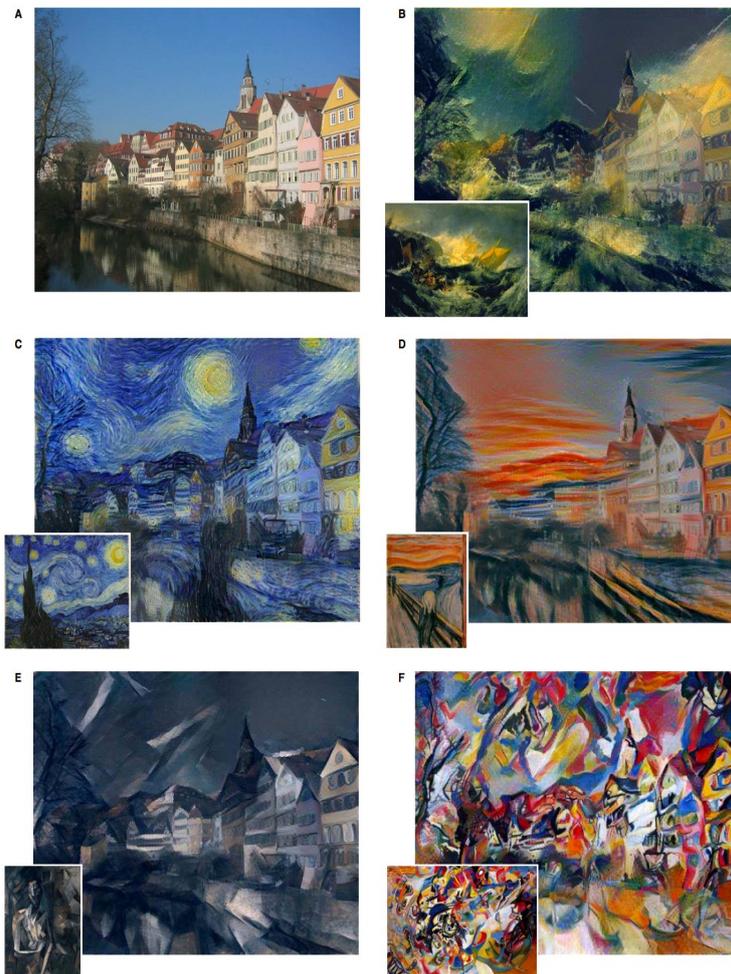


Figura 19: Transferência de estilo.

4) Gerar imagens de objetos, animais ou rostos de pessoas inexistentes:

As pessoas da figura 20 não existem. Essas fotos foram geradas pelo aprendizado profundo. Na verdade, IA juntou as características de milhares rostos reais para criar um modelo que gera novos rostos a partir de números aleatórios, indistinguíveis de rostos verdadeiros.



Figura 20: Rostos gerados artificialmente.

<https://techxplore.com/news/2018-12-nvidia-face-making-approach-genuinely-gan-tastic.html>

O site:

<https://thisxdoesnotexist.com/>

gera versões fake de vários objetos, animais e rostos. Figura 21 mostra exemplos dessas imagens.



Figura 21: As imagens geradas artificialmente.

5) Aprender a jogar um jogo

O aprendizado por reforço é uma das sub-áreas do aprendizado de máquina que estuda como um agente inteligente deve realizar ações em um ambiente dinâmico para maximizar a recompensa. Aprendizado por reforço pode ser usada para fazer o computador aprender a jogar jogos, fazendo-o jogar várias vezes e informando-o se ganhou ou perdeu (figura 22).



Figura 22: Aprendizado por reforço pode ser usada para ensinar o computador a jogar jogos.

6) Notícia sobre “leitura da mente”

Uma notícia de 09/abril/2023

<https://universoracionalista.org/experiencia-incrivel-mostra-que-a-inteligencia-artificial-pode-ler-mentes-para-visualizar-nossos-pensamentos/>

e o artigo

[Chen2023] <https://arxiv.org/pdf/2211.06956.pdf>

dizem que o computador consegue “ler a mente”. Computador tenta “adivinhar”, a partir de sinais de fMRI (ressonância magnética funcional), a imagem que o paciente está vendo.

Em um experimento, a equipe coletou dados da atividade cerebral dos participantes via fMRI enquanto viam algumas imagens (figura 23). A partir dos sinais de fMRI, rede neural conseguiu “adivinhar” a imagem que os participantes estavam vendo.

Daqui a pouco, o computador pode conseguir decodificar os nossos sonhos.

As figuras 23 e 24 mostram a imagem que o paciente estava vendo e a correspondente imagem “adivinhada” pelo computador.

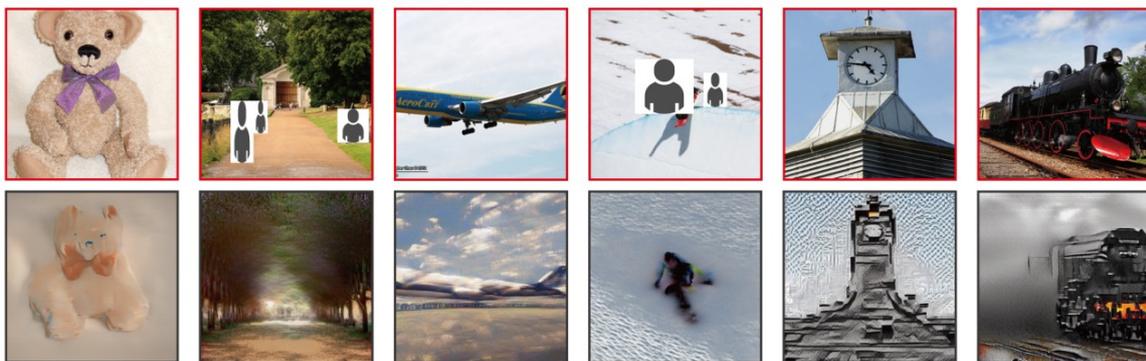


Figura 23: Na linha de cima, a imagem que o paciente estava vendo. Na linha de baixo, a imagem “adivinhada” pelo computador.

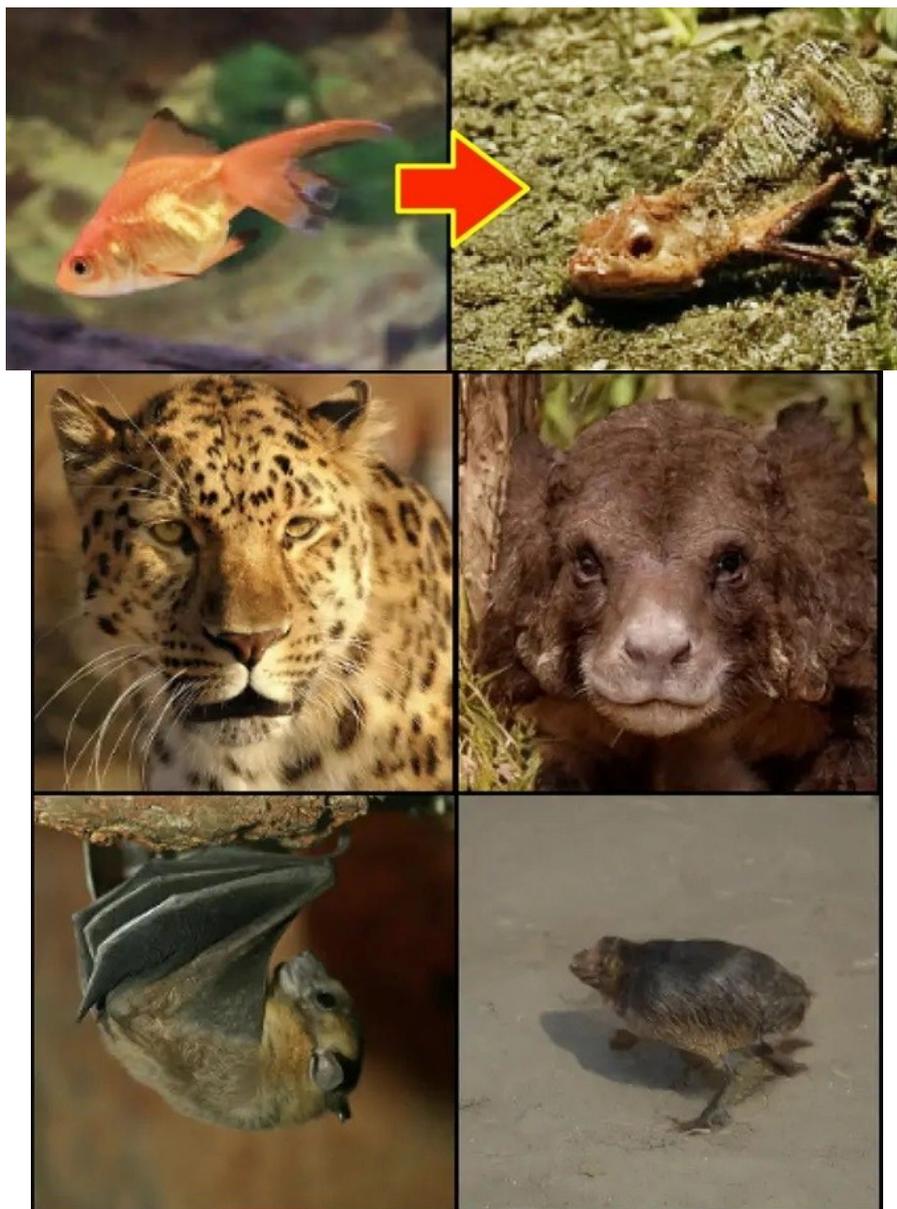


Figura 24: Esquerda, a imagem que o paciente estava vendo. Direita, a imagem “adivinhada” pelo computador.

7) Deep fake

“Deep fake” usa IA para trocar o rosto de pessoas em vídeos, sincronizar movimentos labiais, expressões e demais detalhes.



Figura 25: Deep fake usa IA para trocar o rosto de pessoas.

8) IA gera imagens e vídeos a partir da descrição em português/inglês

Pedindo para chatgpt.com:

“Crie uma imagem realista de um marciano andando numa moto na avenida paulista.”

“Crie uma imagem realista de um gatinho fofo brincando com um bebê chimpanzé.”

“Gere uma imagem realista de cachorro de óculos de sol andando de moto na avenida paulista.”



chargpt



Poe.com

VI. Diferença filosófica entre IA e inteligência humana

Há um artigo interessante [SuperInteressante] que considera 6 “mitos” da inteligência artificial. Em resumo, esse artigo diz que há diferenças essenciais (isto é, substanciais, fundamentais) entre IA e a mente humana, e essa diferença não é apenas de grau. IA não consegue pensar da mesma forma como nós raciocinamos.

Esse artigo diz, citando o livro [Larson2021]: “Vamos supor que você está tentando ensinar o robô a reconhecer gatos (...). Para fazer isso, você o alimenta com fotos de gato (...) e ele vai aprendendo a identificar os bichanos. Mas você não pode ensinar coisas que você mesmo não sabe. (...) Quando Copérnico postulou que a Terra girava em torno do Sol, e não o contrário, ele ignorou montanhas de dados e evidências acumuladas através dos séculos. IA, se existisse no século 16, não teria ajudado Copérnico – pelo contrário, diria que ele estava errado, já que ela se basearia em dados do modelo geocêntrico, que era o aceito na época. IA é uma excelente ferramenta para investigar melhor o que já conhecemos. Mas não tem, por definição, a capacidade de pensar de formas totalmente novas.”

Na sua opinião, o desenvolvimento “progressivo” das atuais técnicas de Inteligência Artificial (isto é, sem que aconteça nenhuma inovação completamente imprevisível) fará com que um dia o computador alcance a inteligência dos seres humanos?

Referências

[Larson2021] Erik J. Larson, *The Myth of Artificial Intelligence - Why Computers Can't Think the Way We Do*, Belknap Press 2021.

[Superinteressante] B. Garattini, Os 6 grandes mitos da inteligência artificial <https://super.abril.com.br/especiais/os-6-grandes-mitos-da-inteligencia-artificial/>

[Zhang2023] Aston Zhang et al., Dive into Deep Learning, 2023. <https://d2l.ai/>

[PSI3471 aula 1 – pular até aqui]

[PSI3471 aula 1 parte 1 – fim]