



damarals/scd24-visao-computacional

ENSINANDO COMPUTADORES

A VEREM: UMA INTRODUÇÃO À VISÃO COMPUTACIONAL

Por Daniel Amaral

íris | Laboratório de
Inovação e Dados

 **CEARÁ**
GOVERNO DO ESTADO

 UNIVERSIDADE
FEDERAL DO CEARÁ

2024 | **SCD**
I Semana Acadêmica
da Ciência de Dados



AGENDA

1. Motivação e Fundamentos

Por que ensinar máquinas a verem é difícil? Uma analogia com nosso próprio sistema visual e como transformamos pixels em compreensão.

2. Da Imagem à Informação

Como computadores "veem" imagens? Entendendo a transformação de luz em números e de números em significado.

3. Três Tarefas da Visão Computacional

Classificação, detecção e segmentação: do "que objeto é esse?" ao "qual sua forma exata?". Métricas e aplicações práticas.

4. Demonstrações e Possibilidades

Teoria em prática: vendo cada tipo de tarefa em ação e suas aplicações no mundo real.



ANTES QUE ALGUÉM PERGUNTE...

Não Vamos Ver 🤔

- Como treinar modelos do zero
- Ajuste fino (fine-tuning)
- Otimização de hiperparâmetros
- Diagnóstico de problemas
- Arquiteturas avançadas
- Deploy em produção (MLOps)



MOTIVAÇÃO

Se computadores são tão "inteligentes", por que
é tão **difícil fazê-los ver?**





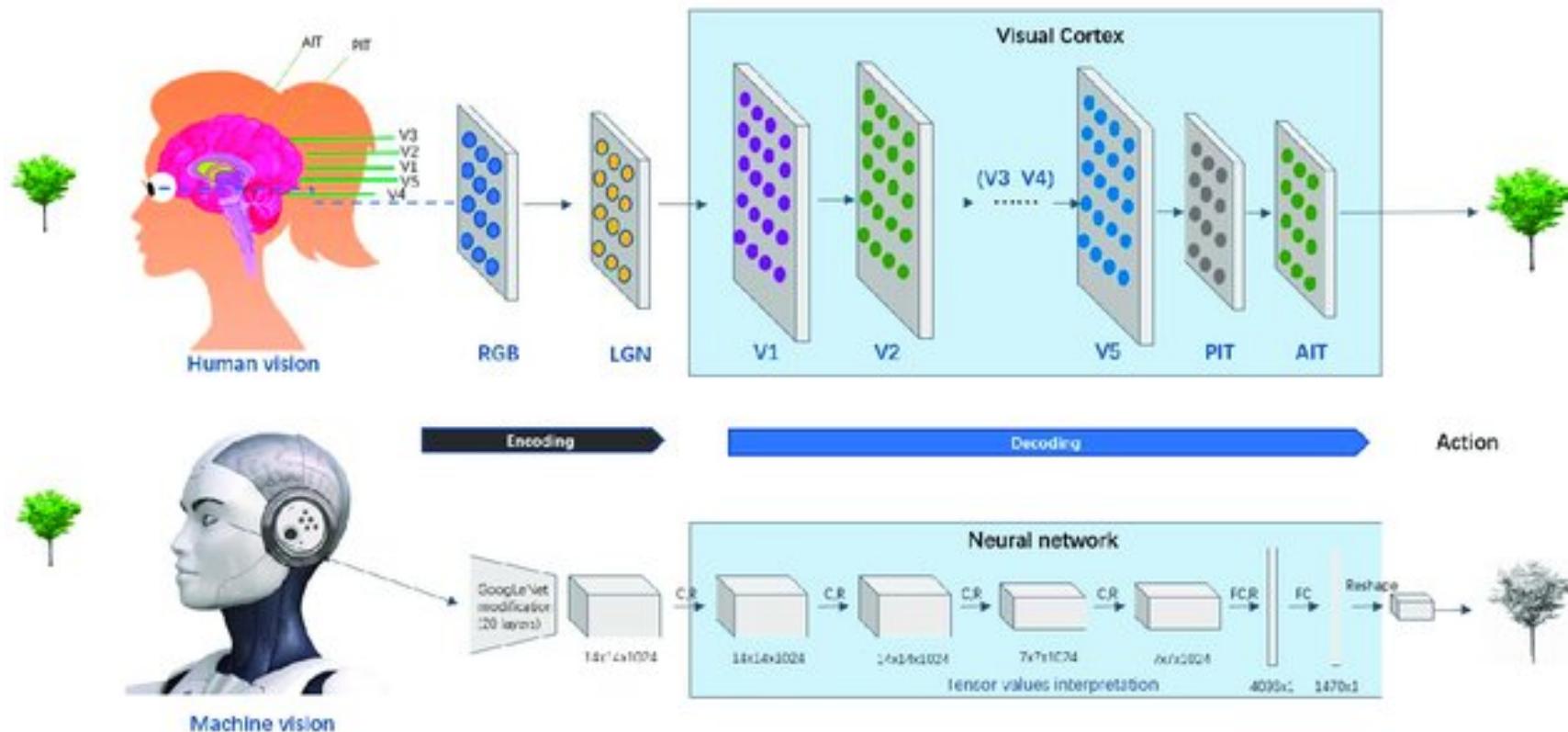
MOTIVAÇÃO

Se computadores são tão "inteligentes", por que é tão **difícil fazê-los ver?**





MOTIVAÇÃO



Fonte: Zhang, H., & Lee, S. (2022).



MOTIVAÇÃO

Voltando ao carro...

Você consegue identificar a **marca** abaixo?





MOTIVAÇÃO

Voltando ao carro...

Você consegue identificar a **marca** abaixo?



Você **aprendeu!**



IMAGENS SÃO NÚMEROS

 Pixel

Cada **ponto** da imagem é um número (ou três!)

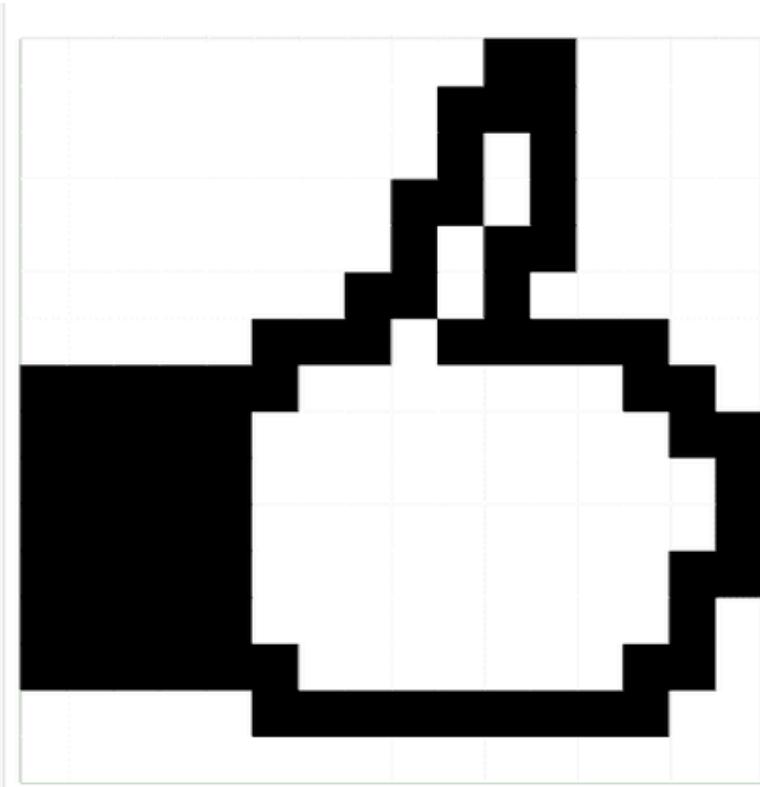


IMAGENS SÃO NÚMEROS

Cada **ponto** da imagem é um número (ou três!)

0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 0 1 1 1 0 0 0 0
0 0 0 0 0 0 0 0 0 0 0 0 1 0 1 0 0 0 0
0 0 0 0 0 0 0 0 0 0 0 1 1 0 1 0 0 0 0
0 0 0 0 0 0 0 0 0 0 1 0 1 1 0 0 0 0 0
0 0 0 0 0 0 0 0 0 1 1 0 1 0 0 0 0 0 0
0 0 0 0 0 0 0 0 1 1 0 1 0 0 0 0 0 0 0
0 0 0 0 0 0 0 1 1 1 0 1 1 1 1 1 0 0 0
1 1 1 1 1 1 0 0 0 0 0 0 0 0 0 1 1 0 0
1 1 1 1 1 0 0 0 0 0 0 0 0 0 0 0 1 1 0
1 1 1 1 1 0 0 0 0 0 0 0 0 0 0 0 0 1 0
1 1 1 1 1 0 0 0 0 0 0 0 0 0 0 0 0 0 1
1 1 1 1 1 0 0 0 0 0 0 0 0 0 0 0 0 1 0
1 1 1 1 1 0 0 0 0 0 0 0 0 0 0 0 0 1 0
0 0 0 0 0 1 1 1 1 1 1 1 1 1 0 0 0 0
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0

Pixel





IMAGENS SÃO NÚMEROS

Pixel

Cada **ponto** da imagem é um número (ou três!)

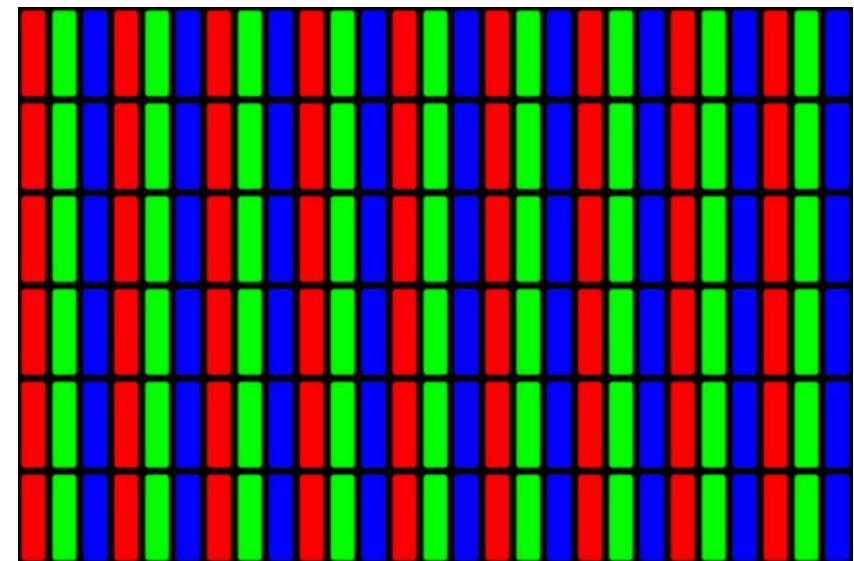
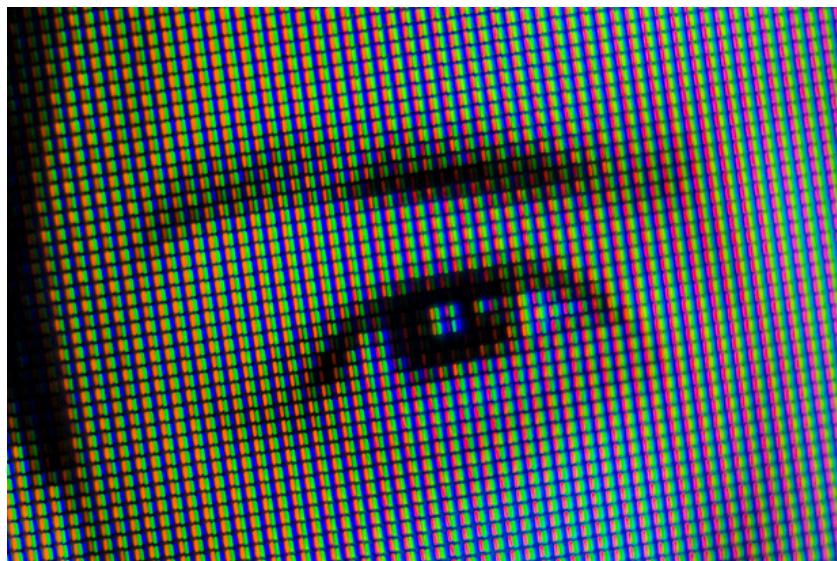


Pixel Defeituoso
(Dead Pixel)



IMAGENS SÃO NÚMEROS

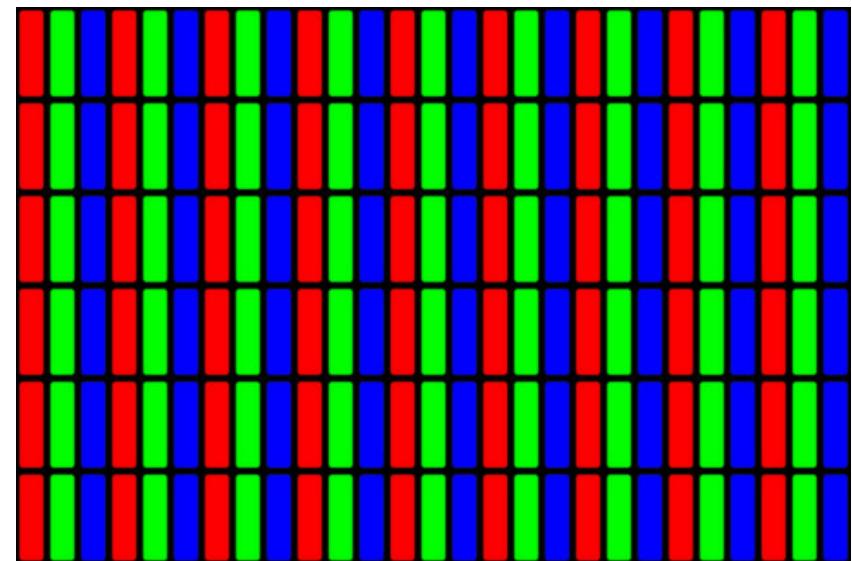
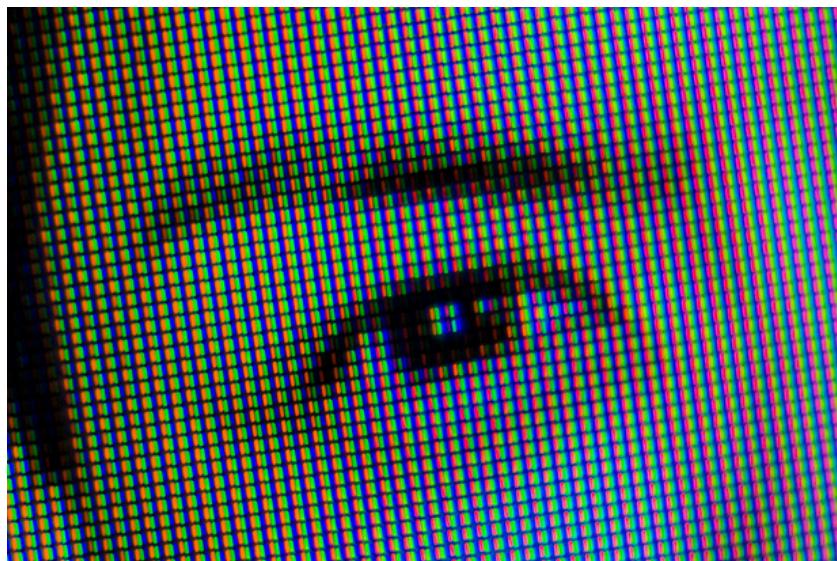
Cada **pixel** da imagem é um número (ou **três!**)





IMAGENS SÃO NÚMEROS

Cada **pixel** da imagem é um número (ou **três!**)



**Red Green Blue
RGB**



IMAGENS SÃO NÚMEROS

Utilizando 8 bits (binário), podemos representar 256 valores diferentes de intensidades para cada **canal**:

$$2^8 = 256$$

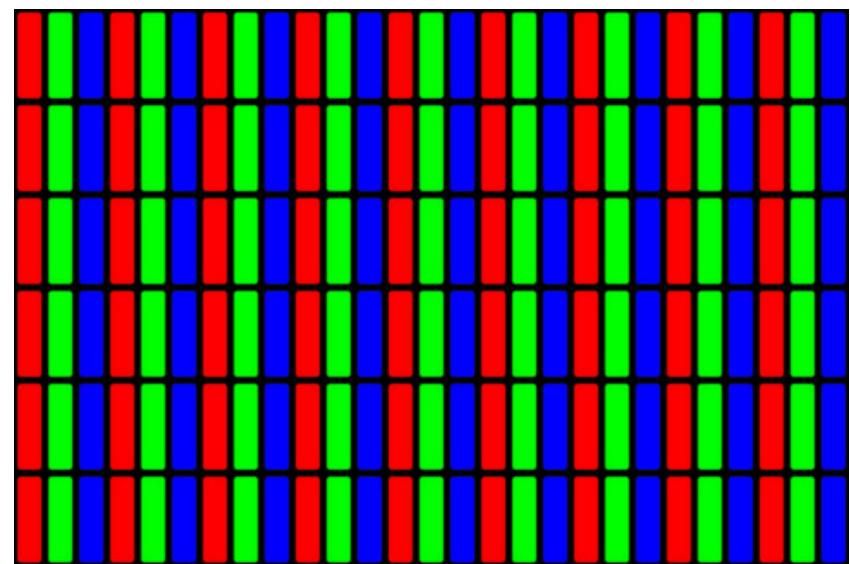


$(256, 256, 256)$

$(0, 0, 0) \rightarrow (255, 255, 255)$

Preto

Branco

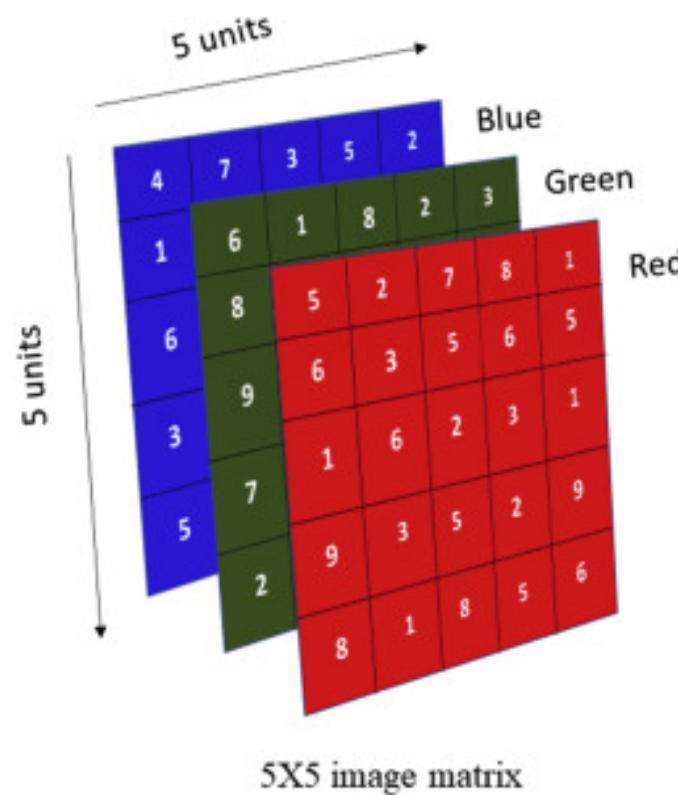


**Red Green Blue
RGB**



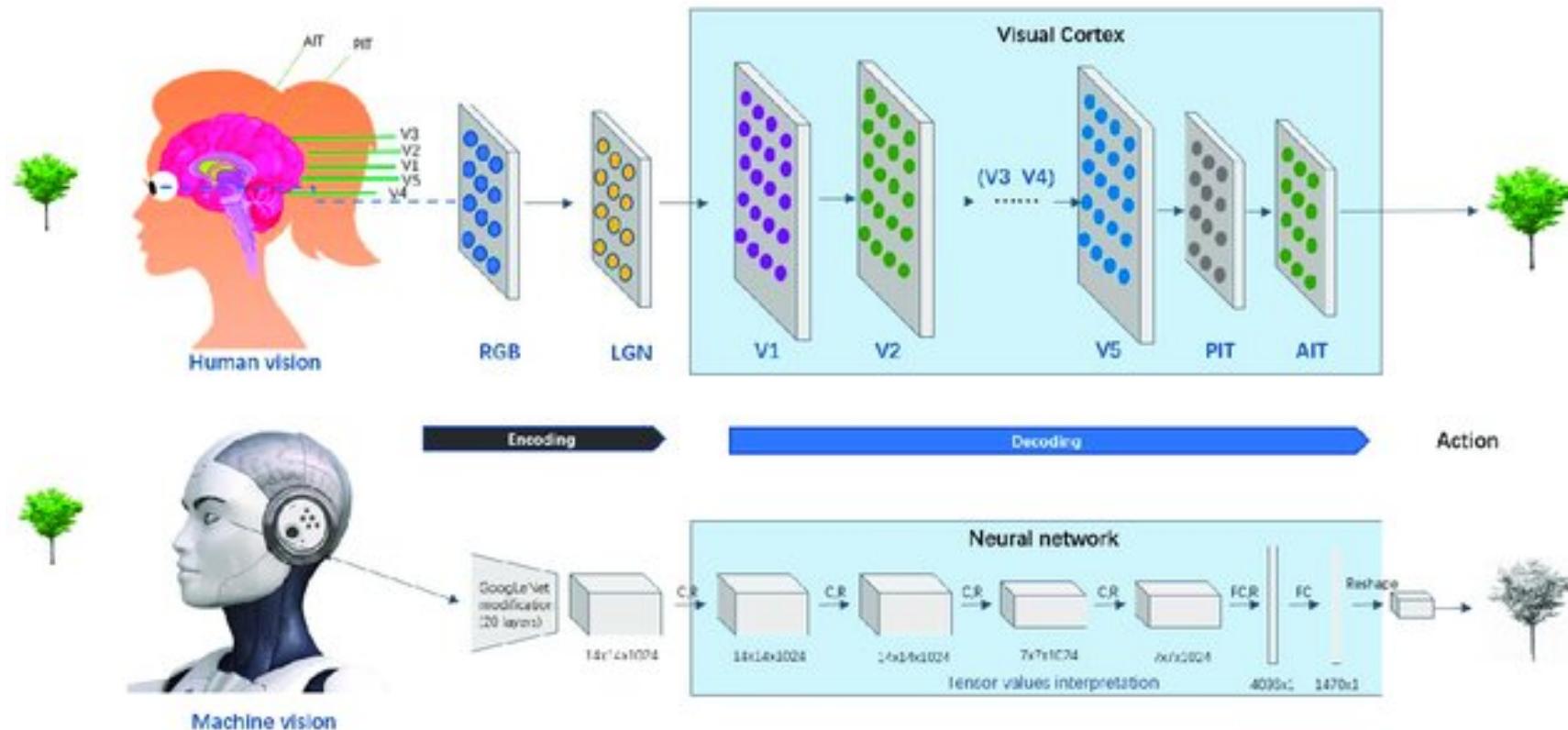
IMAGENS SÃO NÚMEROS

No final das contas, temos uma **matriz 3d RGB** de intensidades:





DO PIXEL À INTELIGÊNCIA

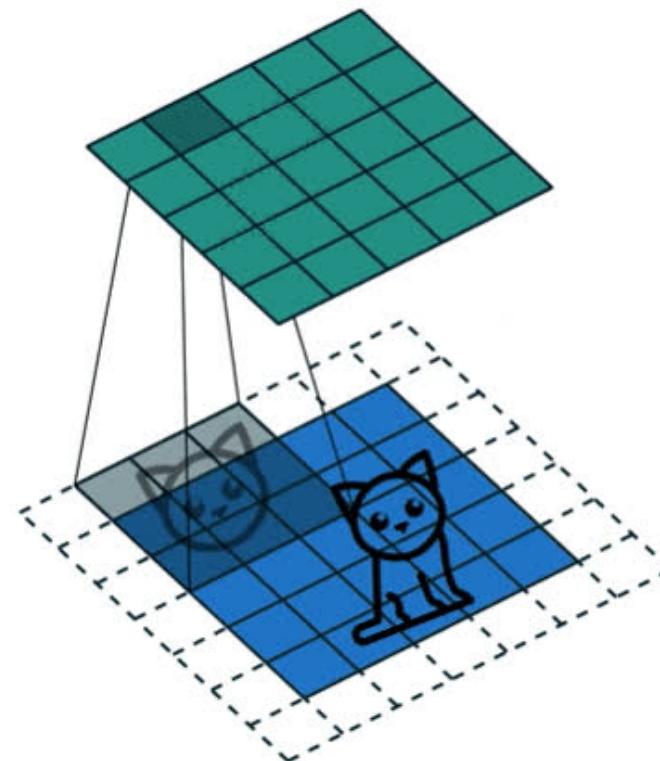


Fonte: Zhang, H., & Lee, S. (2022).



DO PIXEL À INTELIGÊNCIA

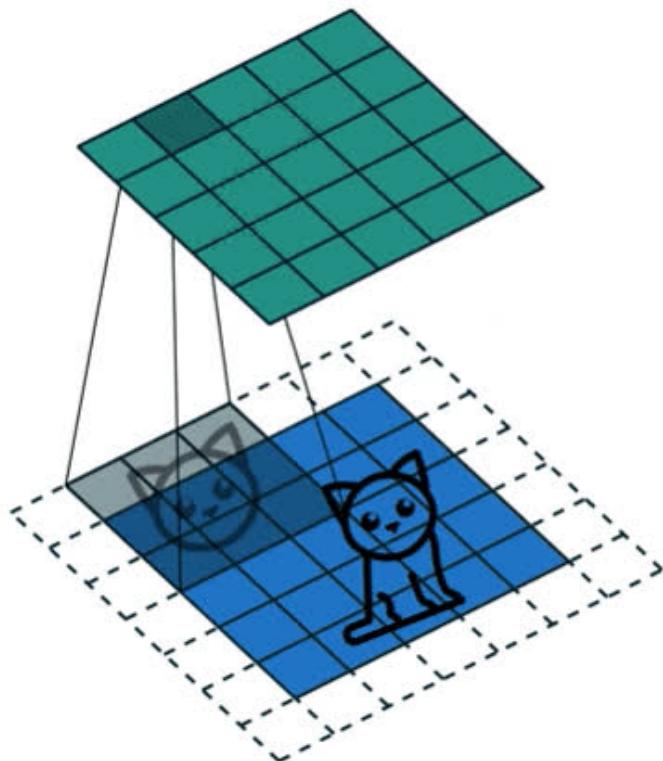
Por que "convolução" é importante:





DO PIXEL À INTELIGÊNCIA

Por que "convolução" é importante:



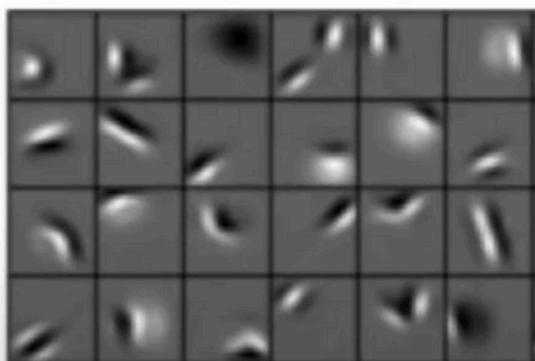
"É como passar uma lupa por toda a imagem, procurando padrões específicos"



DO PIXEL À INTELIGÊNCIA

A combinação de convoluções nos gera hierarquia de características:

Low level features



Edges, dark spots

Mid level features



Eyes, ears, nose

High level features

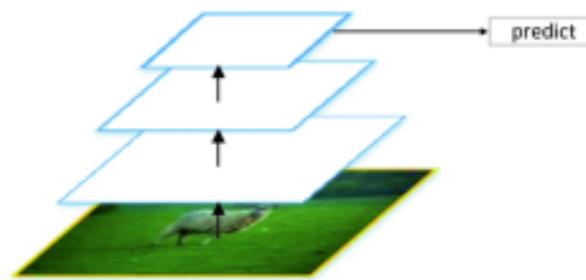


Facial structure

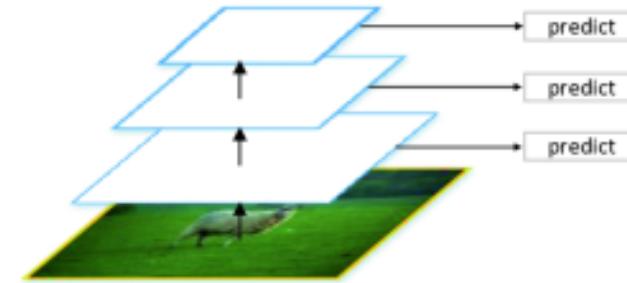


DO PIXEL À INTELIGÊNCIA

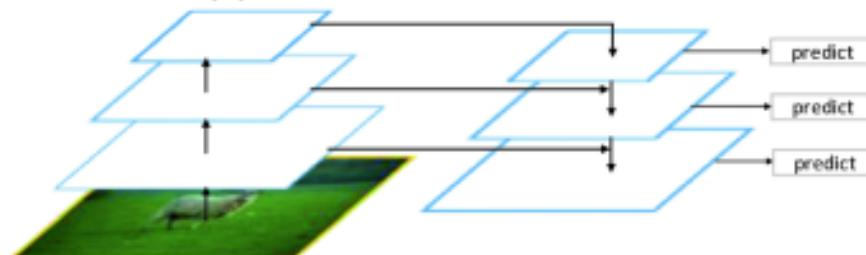
A hierarquia pode ser facilmente vista através da arquitetura comum nas redes neurais atuais:



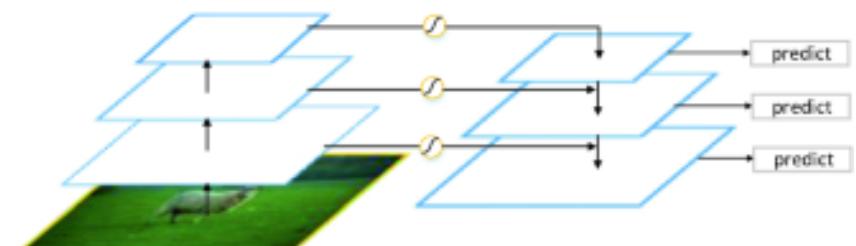
(a) YOLO



(b) SSD



(c) FPN

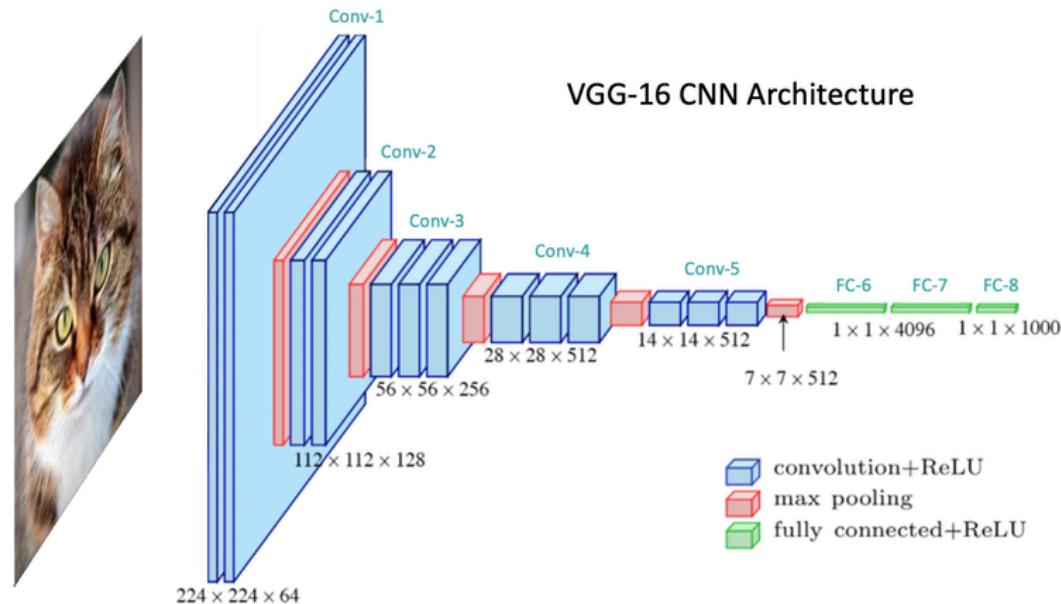


(d) GFPN



DO PIXEL À INTELIGÊNCIA

Uma rede neural convolucional (CNN) é como um detetive juntando pistas: **pequenas evidências formam a conclusão final**





TRÊS TAREFAS DA VISÃO COMPUTACIONAL

- Classificação: "É ou Não É?"
- Detecção: "Onde Está?"
- Segmentação: "Que Forma Tem?"



TRÊS TAREFAS DA VISÃO COMPUTACIONAL

Classificação:

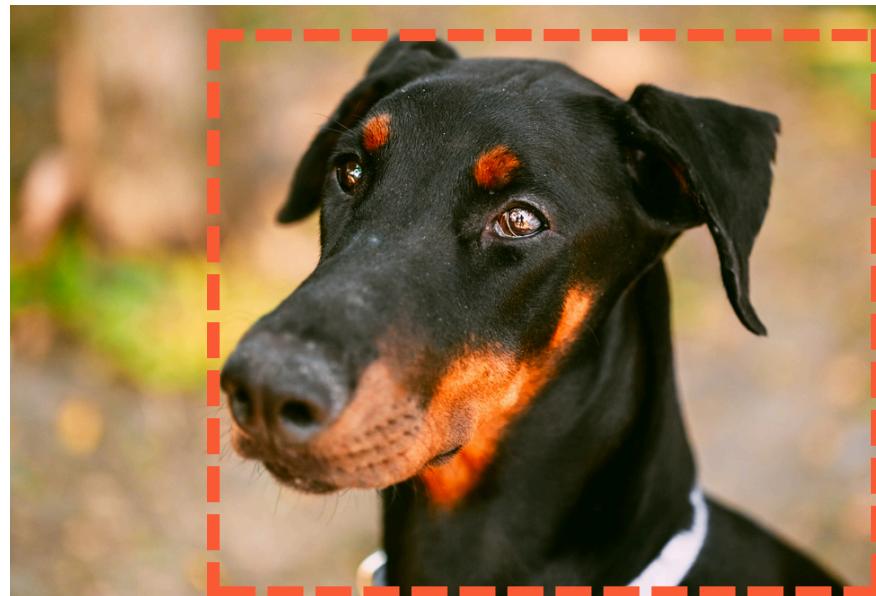


É um cachorro



TRÊS TAREFAS DA VISÃO COMPUTACIONAL

Detecção:



Caixa Delimitadora
(Bounding Box, BBox)

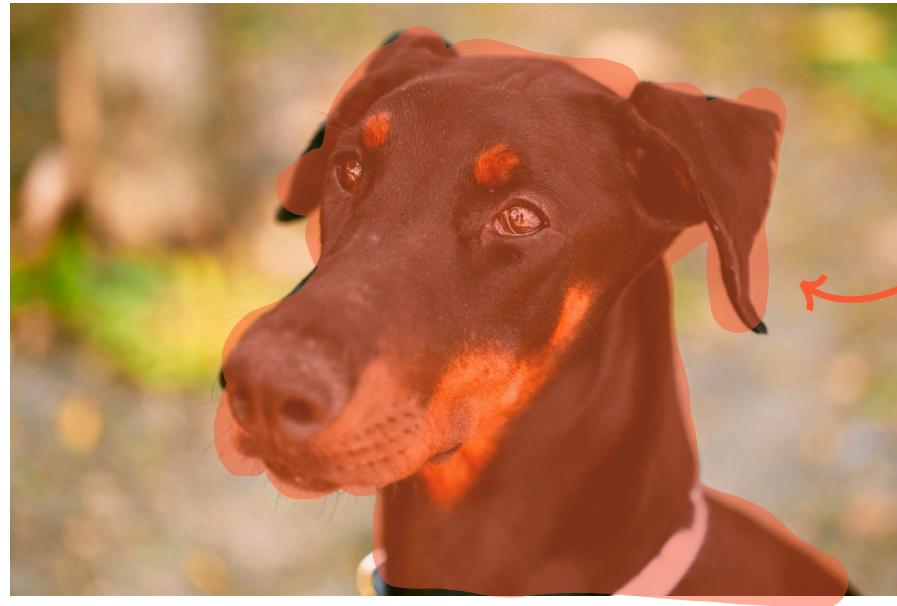


É um cachorro e
está no retângulo destacado



TRÊS TAREFAS DA VISÃO COMPUTACIONAL

Segmentação:



Máscara
(Mask)



É um cachorro e
está "pintado" na imagem



damarals/scd24-visao-computacional

TECNOLOGIAS

Como podemos implementar esses métodos?



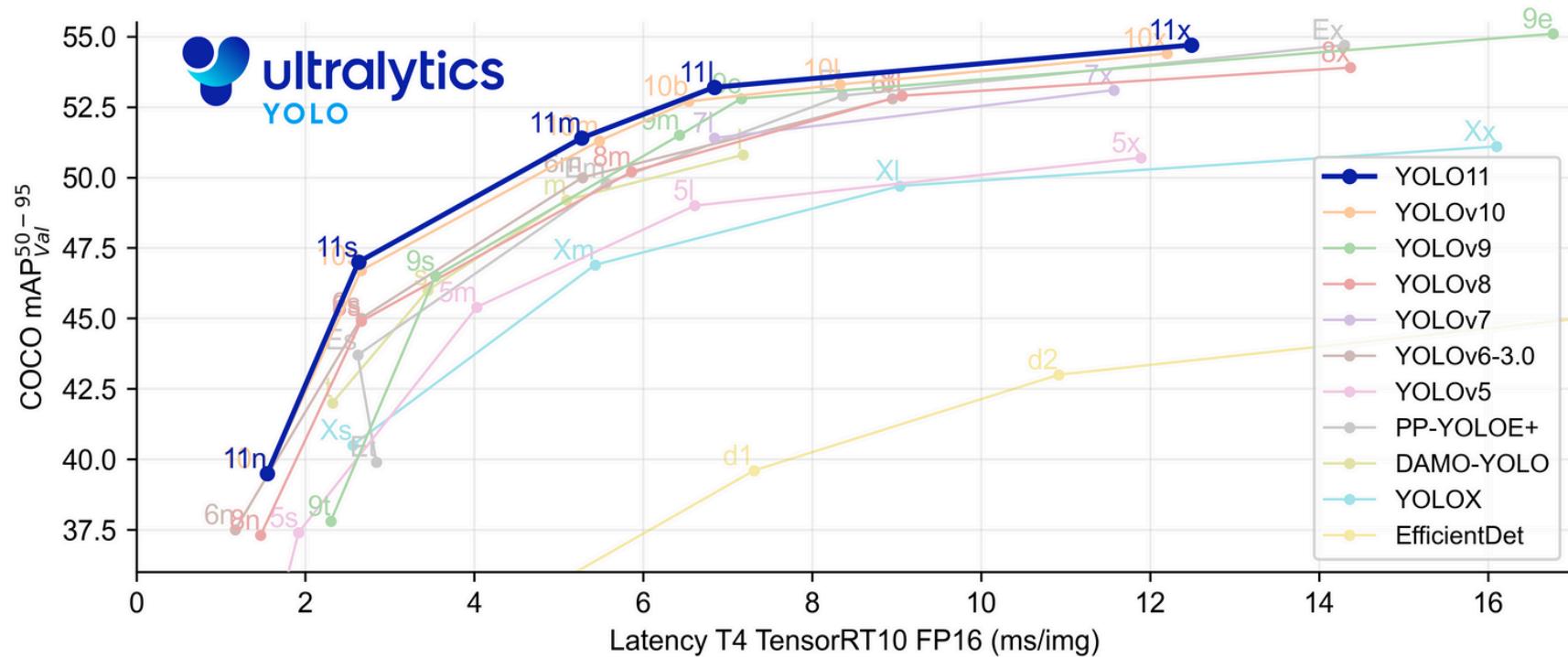
TensorFlow





TECNOLOGIAS

O que vamos utilizar agora:





damarals/scd24-visao-computacional

MATERIAIS

Repositório do **Slide** e do **Notebook**:



damarals/scd24-visao-computacional



damarals/scd24-visao-computacional

OBRIGADO!

Perguntas?

 | Laboratório de
Inovação e Dados

 **CEARÁ**
GOVERNO DO ESTADO

 UNIVERSIDADE
FEDERAL DO CEARÁ

20
24 | SCD
I Semana Acadêmica
de Ciência de Dados