

VISÃO COMPUTACIONAL

02.1 – Classificação de Imagens

POTENCIAL DE UMA IMAGEM

- Uma ideia complexa pode ser (potencialmente) mais facilmente expressa através de uma imagem.
- Analisar gráficos de valorização de investimentos, jogadas, evolução de uma pandemia, entre outras coisas são tarefas rotineiras para as pessoas nos dias atuais.
- Enquanto que é natural para nós, humanos, interpretamos uma imagem, para uma máquina a tarefa não é tão simples.
 - *Tudo que o computador “enxerga” é uma matriz numérica. Não existe uma relação direta do significado agregado a imagem que já existe em nós.*

CATEGORIZAÇÃO DE IMAGENS

- Uma das primeiras tarefas que pode ser feita no contexto de visão computacional é a categorização de uma imagem.
 - *A tarefa consiste em utilizar visão computacional juntamente com aprendizado de máquina para tentar extrair o significado (principal) de uma imagem.*
 - *Dentre os diversos sentidos que 'significado' pode trazer, a maneira mais simples de se realizar essa tarefa é atribuir um rótulo para o que a imagem contém.*
- Certamente, a categorização de imagens é um dos tópicos mais evidentes no momento.
 - *Grandes empresas vem investindo fortemente identificação automática de imagens.*

CATEGORIZAÇÃO DE IMAGENS

- Existem algumas maneiras de se realizar a categorização de imagens.
 - *Classificação de imagens*
 - *Agrupamento de imagens*
- O primeiro desses tópicos se refere ao ato de atribuir um rótulo (categorias) previamente existente a uma imagem.
 - *Já sabemos do que se trata.*
- O segundo, se refere a organização das imagens por semelhança em grupos distintos (categorias), mas sem rótulos predefinidos.
 - *Não sabemos ainda do que se trata cada um dos grupos.*

CLASSIFICAÇÃO DE IMAGENS

- A classificação de imagens consiste na tarefa de atribuir um rótulo a uma imagem considerando um conjunto de rótulos já pré-existentes.
 - *Esse conjunto de rótulos indica (potencialmente) o significado de uma imagem.*
 - *Por exemplo, o conjunto de rótulos pode ser {cachorro, gato, vaca} que indica o animal presente na imagem. Ou seja, estamos classificando imagens que contém (potencialmente) animais e queremos saber qual imagem contém cada animal.*



GAP SEMÂNTICO

- Se olharmos as duas imagens, conseguimos facilmente dizer que uma delas contém uma vaga e a outra, um gato.
- No entanto, para o computador, tudo o que ele enxerga são duas matrizes (ou seis) numéricas com os valores dos pixels.
 - *Sem significado semântico*



GAP SEMÂNTICO

- A diferença que existe entre como os humanos percebem o conteúdo de uma imagem e a maneira como uma imagem pode ser representada para que o computador possa entender é o que estamos chamando de *gap* semântico.
- Enquanto que nós temos facilidade para interpretar a imagem e distinguir dois animais diferentes, os computadores não sabe nem que existem animais lá.

DESCRIÇÃO DE UMA IMAGEM

- Considere a imagem da praia ao lado. Se quisermos descrever essa imagem, podemos ir por alguns caminhos:
 - *Espacial: temos o céu ao topo da imagem com a areia e o oceano na parte de baixo.*
 - *Cores: o céu é azul claro, o mar tem uma tonalidade esverdeada e a areia é amarela.*
 - *Textura: o céu possui um paleta uniforme enquanto que a areia possui granularidades diferentes.*
- Como codificamos todas essas informações de uma maneira que o computador possa entender?



Fonte da imagem: <https://www.guiadasemana.com.br/turismo/galeria/praias-para-conhecer-no-litoral-norte-de-sao-paulo>

DESCRIÇÃO DE UMA IMAGEM

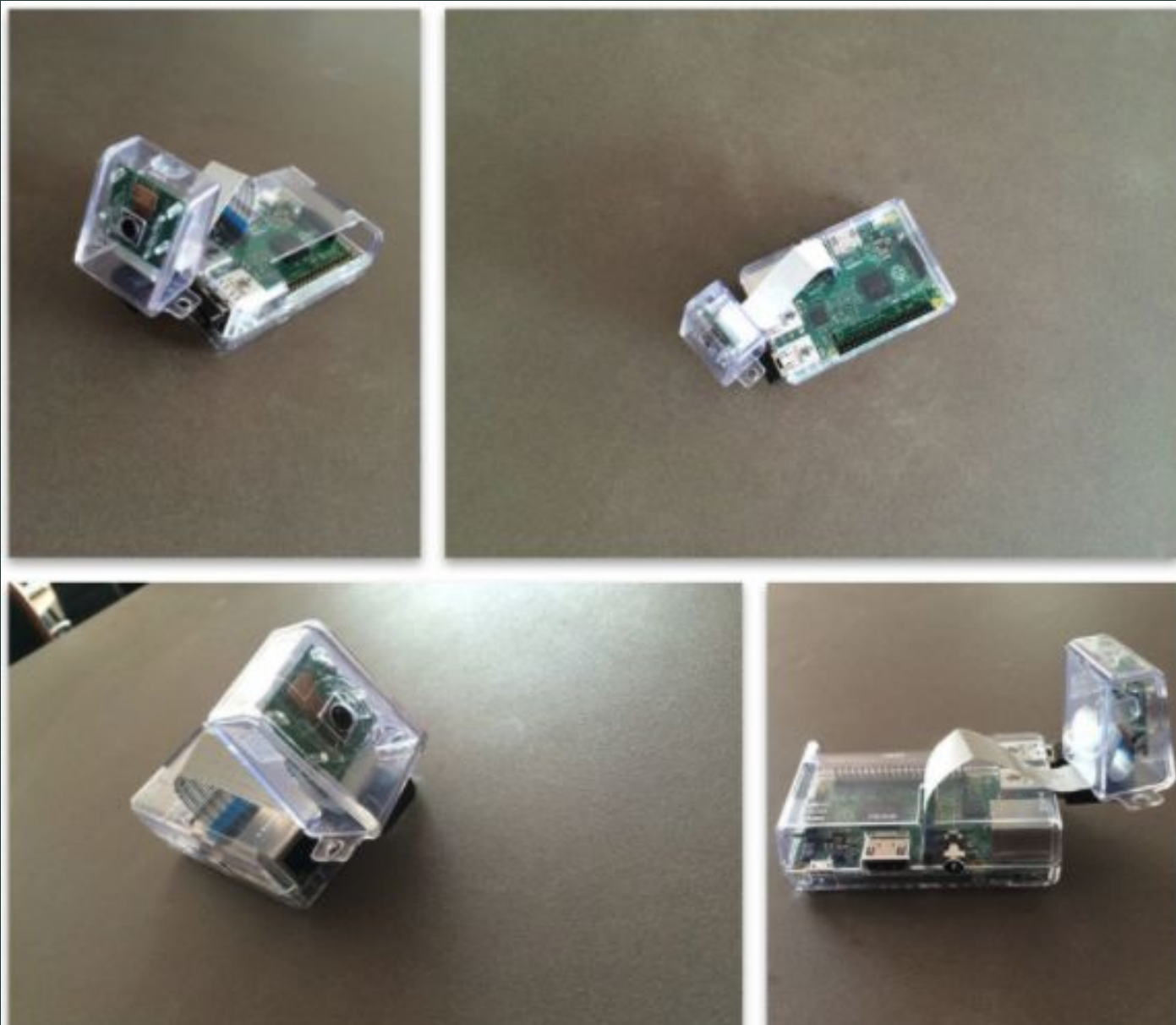
- Uma das respostas pode estar nas várias técnicas de extrair descritores de imagens.
- Utilizando os descritores, nós podemos, de fato, extrair características e quantificar regiões da imagem.
 - *Mais recentemente, o uso de Deep Learning se tornou a maneira mais robusta para classificar imagens.*
 - *Alguns descritores são usados para codificar informações espaciais, enquanto que outros são usados para codificar informações de cores ou texturas.*
- Com base nesses descritores, podemos utilizar aprendizado de máquina para ensinar o computador a entender a composição de características que cada imagem é formada.
- *Deep Learning* vem sendo bastante utilizado por conseguir extrair características e aprender ao mesmo tempo.

DESAFIOS

Temos ainda outros desafios para enfrentar quando estamos trabalhando com imagens.

O primeiro deles se refere a variação das imagens, ainda que dentro de uma mesma categoria. As variações mais comuns são:

- *Orientação*
- *Escala*
- *Deformação*
- *Oclusões*
- *Iluminação*
- *Excesso de informação de fundo*
- *Variação intra-classe*



ORIENTAÇÃO

- Quando mudamos o ponto de vista do observador, temos uma variação na orientação do objeto.
- Se o ponto de vista do observador permanecer, o objeto pode ainda ser rotacionado.
- Independentemente da sua orientação, o objeto é exatamente o mesmo.



ESCALA

- Muitas vezes, o mesmo objeto pode ser representado de tamanhos diferentes. Mas, tecnicamente, ainda é o mesmo objeto (ou pertencem a mesma categoria).
- A mudança de escala também pode ser provocada por mudança na distância entre a câmera e o objeto.



DEFORMAÇÃO

- Uma das variações mais difíceis de tratar se dá pela deformação.
- A deformação ocorre quando um objeto aparece de forma distorcida na imagem.

OCLUSÃO



- Os sistemas de visão computacional deve estar preparados para tratar oclusões, que são imagens onde uma grande porção do objeto a ser detectado está “escondida” do ponto de vista da captura.
- Nas imagens ao lado, o mesmo animal pode ser visto de corpo inteiro (esquerda) e somente uma parte da cabeça (direita).

ILUMINAÇÃO

Diferenças de iluminação podem alterar drasticamente uma imagem.





EXCESSO DE INFORMAÇÃO DE FUNDO

- Essa imagem de 'onde está wally?' representa bem a ideia de excesso de informação no fundo da imagem.
- O ruído gerado por ela dificulta e muito identificar um objeto de interesse.

VARIAÇÃO INTRA-CLASSE

- Existem muitas maneiras de representar um objeto no mundo real.
- Um bom algoritmo deve ser capaz de suportar essas diferentes representações.



DESAFIOS – PARTE 2

- Gerenciar uma dessas variações é uma tarefa desafiadora. No entanto, em cenários reais, o mais comum é ter que lidar com a maioria, senão todas elas ao mesmo tempo.
- Para minimizar os impactos negativos de possíveis falhas, devemos sempre ter em mente:
 - *Qual o objetivo principal do projeto?*
 - *Quais suposições e considerações eu tenho que fazer?*
 - *Que limite de erro é tolerável?*
 - *Que tipo de erro é aceitável?*
- Ou seja, defina bem o escopo do seu projeto de visão computacional.

Referências

Richard Szeliski. Computer Vision: Algorithms and Applications. 2nd Edition. 2021.
<http://szeliski.org/Book/>

Adrian Rosebrock. PyImageSearch Gurus Course. Disponível em:
<https://customers.pyimagesearch.com/>