**PONTIFÍCIA UNIVERSIDADE CATÓLICA DE MINAS GERAIS  
NÚCLEO DE EDUCAÇÃO A DISTÂNCIA**

**Pós-graduação *Lato Sensu* em Inteligência Artificial e Aprendizado de Máquina**

**André Luís Baptista da Silva**

**USO DE VISÃO COMPUTACIONAL PARA DETECÇÃO DE VAGAS EM ESTACIONAMENTOS**

Macaé/RJ - Brasil

2021

**André Luís Baptista da Silva**

**USO DE VISÃO COMPUTACIONAL PARA DETECÇÃO DE VAGAS EM ESTACIONAMENTOS**

Trabalho de Conclusão de Curso apresentado ao Curso de Especialização em Inteligência Artificial e Aprendizado de Máquina como requisito parcial à obtenção do título de especialista.

Macaé/RJ - Brasil

2021 **SUMÁRIO**

[1. Introdução 4](#_Toc63612071)

[1.1. Contextualização 4](#_Toc63612072)

[1.2. O problema proposto 5](#_Toc63612073)

[2. Coleta de Dados 6](#_Toc63612074)

[3. Processamento/Tratamento de Dados 7](#_Toc63612075)

[4. Análise e Exploração dos Dados 8](#_Toc63612076)

[4.1 Obtendo e pré-processando dados 8](#_Toc63612077)

[4.2 Usando redes neurais convolucionais como geradores de vetores de caracterísiticas 8](#_Toc63612078)

[4.3 Cálculo de similaridades usando diferença dos vetores de características 8](#_Toc63612079)

[4.4 Usando listas de top-k mais semelhantes 9](#_Toc63612080)

[5. Criação de Modelos de Machine Learning 10](#_Toc63612081)

[6. Apresentação dos Resultados 11](#_Toc63612082)

[7. Links 15](#_Toc63612083)

[REFERÊNCIAS 16](#_Toc63612084)

# 1. Introdução

## 1.1. Contextualização

Uma aplicação da visão computacional com inúmeras possibilidades é a busca por imagens semelhantes. Neste trabalho apresento o uso de uma rede neural convolucional (convolutional neural network - CNN) pré treinada para identificação de vagas vazias e ocupadas em estacionamentos.

## 1.2. O problema proposto

A automatização de processos em geral é uma tendência mundial e irreversível. Existe enorme demanda por automatização de estacionamentos de diversos tipos, como supermercados, shoppings, empresas, etc. Uma abordagem tradicional é o uso de sensores de presença individuais para cada vaga, com custo e complexidade consideráveis, conforme figura:



No presente trabalho apresentaremos um sistema de identificação de vagas baseado em visão computacional (machine vision), que oferece vantagens importantes em custo e complexidade de instalação e manutenção.

A base de dados em que o trabalho se baseia é o dataset CNRPARK+EXT (AMATO et al., 2015), composto de aproximadamente 150.000 patches (pedaços) de imagens de estacionamentos. Estas imagens foram coletadas entre julho de 2015 e fevereiro de 2016, sob várias condições meteorológicas e ângulos de visão, com diferentes iluminações e oclusões, no estacionamento da CNR Research Area em Pisa, Itália.

O objetivo é comprovar a viabilidade do processo, em termos de acurácia (detectar vagas vazias e ocupadas corretamente) e performance (conseguir executar a detecção em curto espaço de tempo).

# 2. Coleta de Dados

O dataset escolhido para análise, CNRPARK+EXT (AMATO et al., 2015) é subdividido em dois: o original (CNR-PARK, menor, com 242 frames, coletado em apenas dois dias de julho de 2015) e o posterior (CNR-EXT, maior, com 4.081 frames, coletado em 23 dias entre novembro de 2015 e fevereiro de 2016, com grande variedade climática). Cada frame é subdividido em vários patches (segmentos de imagem) que representam uma vaga cada.

O subset CNR-PARK é composto por duas câmeras (A e B), e os arquivos seguem o seguinte padrão de nomenclatura:

<CAMERA>/<CLASSE>/YYYYMMDD\_HHMM\_<NUM\_SLOT>.jpg

Onde:

* <CAMERA> pode ser A ou B;
* <CLASSE> pode ser “free” ou “busy”, repectivamente “livre” ou “ocupada”;
* YYYYMMDD\_HHMM\_ representa a data e hora da captura;
* <NUM\_SLOT> é um número identificador da vaga para aquela câmera.

Já o segundo subset, CNR-EXT, é composto por 9 câmeras (1 a 9), e as imagens seguem este padrão de nomenclatura:

PATCHES/<CLIMA>/<DATA\_CAPTURA>/camera<NUM\_CAM>/<W\_ID>\_< DATA\_CAPTURA >\_<HORA\_CAPTURA>\_C0<NUM\_CAM>\_<NUM\_SLOT>.jpg

* <CLIMA> pode ser “sunny”, “overcast” ou “rainy”, respectivamente: “ensolarado”, “nublado” ou “chuvoso”;
* <DATA\_CAPTURA> é a data de captura da imagem no formato YYYY-MM-DD;
* <NUM\_CAM> é o número identificador da câmera, entre 1 e 9;
* <W\_ID> é o identificador do tipo de clima, “S”, “O” ou “R”;
* <HORA\_CAPTURA> é a hora da captura da imagem no formato HH.MM, com horas de 0 a 23;
* <NUM\_SLOT> é o identificador global de cada vaga, pode ser usado para identificar unicamante daterminada vaga no dataset CNR-EXT.

# 3. Processamento/Tratamento de Dados

Para o presente trabalho foi escolhida a câmera “A” do subset “CNR-PARK”, que possui um total de 6.171 patches (pedaços de frames, cada um representando uma vaga individual), sendo 3.621 vagas ocupadas e 2.550 vagas vazias. Através de um pequeno utilitário Console Application feito em C# (disponível na seção Links), foram sorteados 100 patches (50 vagas ocupadas e 50 vazias) para servirem como dados de treinamento e referência posterior. Essas 100 imagens foram copiadas para um diretório específico (“A - Imagens tratadas – treinamento”), enquanto todas as outras 6.071 imagens foram copiadas para o diretório “A - Imagens tratadas - busca”.

Todas essas imagens de patches de vagas tiveram os nomes prefixados com “busy\_” ou “free\_” para facilitar a análise do resultado.

# 4. Análise e Exploração dos Dados

## 4.1 Obtendo e pré-processando dados

Seguindo a abordagem vista em (KORZEC, 2020), usamos uma rede neural convolucional pré-treinada chamada resnet18 (PYTORCH TEAM, 2017) incluída na biblioteca PyTorch, ela requer como entrada imagens RGB de resolução 224x224 pixels, com valores resultantes no intervalo [0, 1]. As imagens são normalizadas para a média [0,485, 0,224, 0,225] e desvio padrão [0,229, 0,224, 0,225], então são todas redimensionadas para este formato, cuidando para que não se perca informação de orientação da imagem. Estas ponderações foram obtidas através do treinamento da resnet18 com a ImageNet (DENG et al., 2009), em caso de treinamento com outros datasets, os valores serão outros.

## 4.2 Usando redes neurais convolucionais como geradores de vetores de caracterísiticas

Métodos tradicionais de classificação por características normalmente envolvem busca por bordas (edge detection) em diferentes direções (orientações) e densidades locais. Ao invés de fazer todo este processo trabalhoso e maçante, já existem modelos pré treinados em bases de imagens gigantescas. A Imagenet, usada neste caso pela resnet18, por exemplo possui mais de 14 milhões de imagens. Não é necessário o entendimento do processo por trás, funcionando como uma abordagem caixa-preta. Só é necessário saber que imagens semelhantes geram entradas semelhantes na tabela de vetores de características.

Quando usamos uma rede pré treinada, como a resnet18, todos os parâmetros já estão fixos, não sendo necessário se preocupar com eles. Apenas deve-se pré processar as imagens para se adequar ao formato de entrada usado pela rede, como visto anteriormente.

## 4.3 Cálculo de similaridades usando diferença dos vetores de características

Para o cálculo do nível de similaridade entre as imagens uma abordagem simples e eficiente é o cálculo da similaridade de cosseno para todos os pares de imagens, conforme demonstrado no artigo por (KORZEC, 2020).

A similaridade de cosseno é uma medida de comparação da orientação de dois vetores não nulos, independentemente da magnitude deles. Dois vetores com a mesma orientação possuem uma similaridade de cosseno igual a 1, dois vetores distantes por 90º tem a similaridade igual a zero, e dois vetores diametralmente opostos possuem a similaridade igual a -1, conforme (SINGHAL, 2001).

O cálculo de similaridade é um tópico de estudo muito frequente na área correlata de sistemas de recomendações. Posteriormente poderiam ser avaliadas outras métricas de semelhança, em busca de resultados melhores em precisão e/ou performance.

## 4.4 Usando listas de top-k mais semelhantes

De posse dos índices de similaridade entre todos os pares de imagens, dispostos em forma de matriz, basta selecionarmos as top-k mais semelhantes, com k=4 no nosso caso.

# 5. Criação de Modelos de Machine Learning

Para este trabalho foi desenvolvido um script em Pyhton, usando a ferramenta Jupyter Notebook, rodando localmente (o Google Colab não é tão prático quando se trata de uma quantidade grande de imagens).

O código foi baseado no artigo (KORZEC, 2020), com as devidas adaptações necessárias à diferente finalidade. Originalmente era um sistema de recomendação de produtos baseado na semelhança de imagens, ou seja, é possível um pré processamento e deixar as k imagens mais semelhantes já preparadas. No caso de um sistema de detecção de vagas em tempo real, existe a necessidade de trabalhar com imagens inéditas o tempo todo.

O script pode ser encontrado na seção Links.

# 6. Apresentação dos Resultados

Num primeiro momento foi feita uma comparação de cada uma das 100 imagens do dataset de treinamento com as demais (excluindo a própria imagem). O resultado foi este:

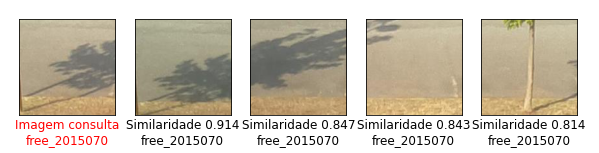
Total: 100

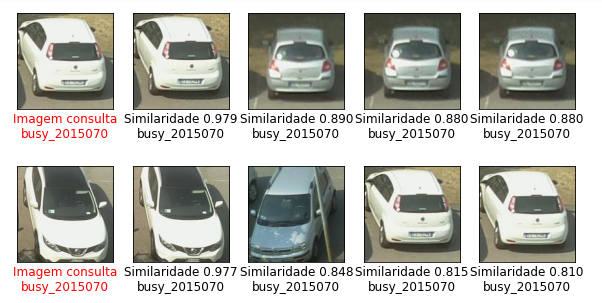
Taxa de acerto da imagem mais semelhante: 100.0%

Taxa de acerto das 4 imagens mais semelhantes: 94.5%

A taxa de acerto da primeira imagem (ou seja, a mais semelhante) foi de 100%. No acumulado das quatro imagens mais semelhantes houve uma taxa de acerto de 94,5%. No exemplos a seguir o nome do arquivo de imagem foi truncado em 12 caracteres. Segue abaixo uma amostra do resultado de busca de algumas imagens:









Posteriormente foi feita a análise com o dataset de busca, mais próximo do caso de uso real, composto por 6.071 imagens, com o seguinte resultado:

Total: 6071

Taxa de acerto da imagem mais semelhante: 98.17163564486904%

Taxa de acerto das 4 imagens mais semelhantes: 96.70564981057487%

Tempo total: 43min 52s

Tempo por imagem: 2,31s

Alguns exemplos de imagens que foram equivocadamente avaliadas, numa primeira análise os erros são bem mais comuns em vagas vazias do que em vagas ocupadas:











É possível notar que o erro se dá, normalmente, na imagem mais semelhante, sendo possível de tratar com algum simples algoritmo quando as quatro imagens mais semelhantes indicarem resultados diferentes entre si, aplicando pesos para cada uma das quatro imagens mais semelhantes.

# 7. Links

* Repositório contendo o script, imagens, testes e o utilitário usado para o pré tratamento das imagens: <https://github.com/albs-br/tcc-machine-vision>
* Vídeo com demonstração do script em Python testado contra a base de treinamento com 100 imagens de vagas, metade delas ocupadas e metade vazia: <https://youtu.be/CW7c21cET5M>

# REFERÊNCIAS

KORZEC, Maciej D. Effortlessly Recommending Similar Images. *In:* **Towards Data Science**. [*S. l.*], 27 jun. 2020. Disponível em: <https://towardsdatascience.com/effortlessly-recommending-similar-images-b65aff6aabfb>. Acesso em: 11 jan. 2021.

AMATO, Giuseppe *et al.* **CNRPark+EXT**: A Dataset for Visual Occupancy Detection of Parking Lots. [*S. l.*], jul. 2015. Disponível em: <http://cnrpark.it/>. Acesso em: 4 jan. 2021.

PYTORCH TEAM. **RESNET**: Deep Residual Networks Pre-Trained on ImageNet. [*S. l.*], 2017. Disponível em: <https://pytorch.org/hub/pytorch_vision_resnet/>. Acesso em: 13 jan. 2021.

SINGHAL, Amit. **Modern Information Retrieval: A Brief Overview. Bulletin of the IEEE Computer Society Technical Committee on Data Engineering**, [*s. l.*], v. 24, n. 4, p. 35-43, 2001. Disponível em: <http://sites.computer.org/debull/A01dec/A01DEC-CD.pdf>. Acesso em: 5 fev. 2021.

DENG, J. et al. ImageNet: A Large-Scale Hierarchical Image Database. **IEEE Computer Vision and Pattern Recognition (CVPR)**, [*s. l.*], 2009. Disponível em: <http://www.image-net.org/papers/imagenet_cvpr09.pdf>. Acesso em: 11 jan. 2021.