

<https://eventos.utfpr.edu.br//sicite/sicite2019>

## Aprendizado profundo para reconhecimento de marca e modelo de automóveis brasileiros

### Deep learning for brazilian car make and model recognition

#### RESUMO

**Lucas Augusto Albini**

[lucasalbini@alunos.utfpr.edu.br](mailto:lucasalbini@alunos.utfpr.edu.br)  
Laboratório de Bioinformática e Inteligência Computacional, Universidade Tecnológica Federal do Paraná - UTFPR, Curitiba, Paraná, Brasil

**Matheus Gutoski**

[matheusgutoski@alunos.utfpr.edu.br](mailto:matheusgutoski@alunos.utfpr.edu.br)  
Laboratório de Bioinformática e Inteligência Computacional, Universidade Tecnológica Federal do Paraná - UTFPR, Curitiba, Paraná, Brasil

**Heitor Silvério Lopes**

[hslopes@utfpr.edu.br](mailto:hslopes@utfpr.edu.br)  
Laboratório de Bioinformática e Inteligência Computacional, Universidade Tecnológica Federal do Paraná - UTFPR, Curitiba, Paraná, Brasil

A classificação de marcas e modelos de carros é uma questão frequentemente discutida na literatura devido às suas diversas aplicações em segurança, controle de tráfego e planejamento urbano, especialmente no contexto de cidades inteligentes. Atualmente, os métodos de aprendizado profundo são o estado da arte para classificação de imagens e vídeos. Este trabalho apresenta um método para classificar carros no nível de marca e modelo de forma simples e eficaz, utilizando a aprendizagem profunda. Para realizar essa tarefa, a rede neural *Inception-v3* foi utilizada para treinar e avaliar o modelo. Outro objetivo deste trabalho é criar um conjunto de dados de imagens de carros produzidos pela indústria brasileira, para que o trabalho seja contextualizado nacionalmente. Nossos resultados mostraram que a abordagem proposta foi muito bem sucedida para fins de classificação e incentiva o desenvolvimento futuro.

**PALAVRAS-CHAVE:** Visão computacional. Aprendizagem profunda. Classificação de carros.

#### ABSTRACT

Car make and model classification is an issue frequently discussed in the literature due to its several applications in security, traffic control, and urban planning, especially in the context of smart cities. Currently, deep learning methods are the state-of-the-art for image and video classification. This work presents a method for classifying cars at the level of make and model in a simple and effective way using deep learning. To accomplish this task, the Inception-v3 neural network was used to train and evaluate the model. Another objective of this work is to create a dataset of images of cars produced by the Brazilian industry, so that the work is contextualized nationally. Our results showed that the proposed approach was very successful for classification purposes and encourages further development.

**KEYWORDS:** Computer Vision. Deep Learning. Car Classification.

**Recebido:** 19 ago. 2019.

**Aprovado:** 01 out. 2019.

**Direito autorial:** Este trabalho está licenciado sob os termos da Licença Creative Commons-Atribuição 4.0 Internacional.



## INTRODUÇÃO

As cidades inteligentes são aquelas que utilizam intensivamente a tecnologia para levar o bem-estar à sua população, melhorando a segurança e a mobilidade, favorecendo o crescimento econômico e a sustentabilidade. Em grandes cidades brasileiras existe uma crescente preocupação com mobilidade e segurança, principalmente devido ao crescente número de acidentes de trânsito e roubos de carros.

Ao longo dos anos, o número de câmeras de segurança nas cidades cresceu exponencialmente, sejam elas públicas ou privadas que estão espalhadas por todos os lugares. No entanto, não existem pessoas suficientes para monitorá-las em tempo real. Esse fato leva a sistemas puramente reativos, de tal forma que as câmeras de vigilância são úteis apenas para visualizar eventos passados. Esta é uma das motivações para a criação de meios de análise automática de imagens e vídeos de câmeras de segurança, de modo a contornar a necessidade de especialistas humanos.

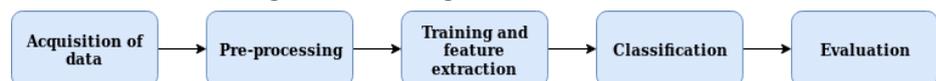
Um sistema capaz de reconhecer a marca e o modelo de veículos em câmeras de vigilância pode ser de grande utilidade para a segurança pública. Pode ser usado, por exemplo, para prevenir roubos ou aumentar a prontidão policial, além de ser útil para controle de tráfego e planejamento urbano.

Este trabalho propõe um método para classificar marca e modelo de carro a partir de imagens brutas. Para treinar o classificador, foi criado um novo conjunto de dados, utilizando imagens de veículos produzidos recentemente no Brasil. Este conjunto de dados inclui imagens de vários modelos e marcas, com uma grande diversidade de ângulos e planos de fundo.

## METODOLOGIA

Neste trabalho, o método a seguir foi utilizado: a primeira etapa foi criar um conjunto de dados e pré-processar imagens brutas. O próximo passo foi treinar uma rede neural convolucional a partir de um modelo de pré-treinado (Transferência de aprendizado). Finalmente, imagens de teste foram usadas para avaliar o desempenho de classificação. A Figura 1 mostra o fluxograma do método proposto.

Figura 1 – Abordagem do Problema.



Fonte: Própria.

### AQUISIÇÃO DOS DADOS

Para capturar imagens de veículos, um código escrito em Python foi usado para a raspagem da web. Imagens foram retiradas de vários sites, incluindo sites de fabricantes, sites de vendas online, google imagens, entre outros. A anotação de marca e modelo foi feita manualmente.

### PRÉ-PROCESSAMENTO

O pré-processamento de dados foi realizado em várias etapas. Primeiro, todos os formatos de imagem incompatíveis (como GIF, TIFF, PNG, entre outros) foram removidos da coleção de dados inicial. Imagens corrompidas devido a falhas na raspagem dos dados também foram eliminadas.

Devido à natureza do problema, existe uma variabilidade muito grande do fundo das imagens dentro dos dados adquiridos. Para evitar falhas nos resultados e facilitar a limpeza dos dados, foi realizado um procedimento de recorte para reduzir o fundo e fazer com que o veículo ocupasse a maior parte da imagem. Para fazer isso, usamos o YOLO-v3 [Redmon e Farhadi 2018], pré-treinado com o *ImageNet dataset* [Russakovsky et al. 2015], para encontrar uma caixa delimitadora que inclua o veículo na imagem. Embora o YOLO-v3 não seja mais o mais preciso, é muito rápido e confiável para a identificação de objetos em tempo real. Portanto, usando as coordenadas de caixa delimitadora, a imagem é recortada, como mostrado na Figura 2. As imagens brutas não foram padronizadas em tamanho, e usamos o *Keras* para redimensionar imagens para a entrada padrão da rede, neste caso, para três canais RGB com 224x224 pixels.

O próximo passo foi organizar o conjunto de dados em classes. As imagens foram agrupadas em classes que correspondem aos modelos dos carros. Nesse ponto, o conjunto de dados foi dividido aleatoriamente em conjuntos de treinamento e teste, com uma proporção de 80% e 20%, respectivamente.

Como etapa final, foi implementado um código para aumento de dados. Com operações simples são aplicadas transformações nas imagens, gerando imagens novas. As transformações utilizadas, com respectivas faixas, são mostradas na Tabela 1. Além disso, cada transformação pode ser vista individualmente na Figura 5.

Tabela 1- Transformações para aumento dos dados.

Arguments	Transformation	Range
Rotation	Rotate the image up to a limit of 40 degrees	40
Width Shift	Move the image horizontally to a fraction of total width	0.2
Height Shift	Move the image vertically to a fraction of total height	0.2
Shear	Turn the rectangular image into a parallelogrammed image with a transformation matrix.	0.2
Zoom	Zoom image within the range	0.8-1.0
Channel Shift	Changes the colors schemes given an intensity	30

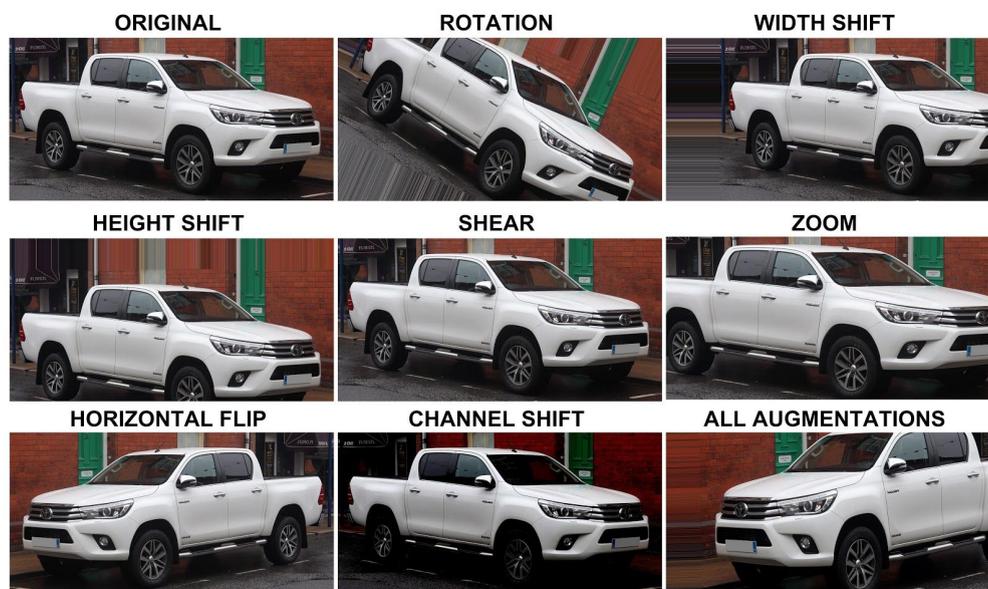
Fonte: Própria

Figura 2 – Recorte do automóvel



Fonte: Própria.

Figura 3 – Aumento de dados



Fonte: Própria.

#### TREINAMENTO DO MODELO E CLASSIFICAÇÃO DOS DADOS

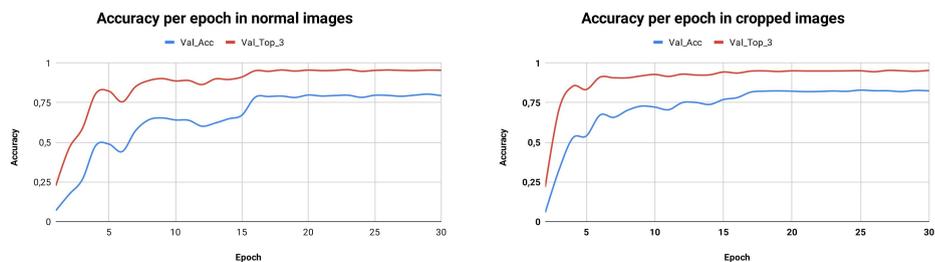
Em vez de treinar nossa rede de classificação de marcas e modelos a partir do zero, nós usamos a rede neural Inception-v3 [Szegedy et al. 2015] pré-treinado em ImageNet como ponto de partida.

Considerando que o problema abordado aqui é o aprendizado supervisionado, o modelo CNN treinado foi usado para classificação de modo a encontrar a classe à qual a imagem pertence. Usando o procedimento de aumento de dados mencionado anteriormente, nosso conjunto de dados foi balanceado no nível de imagens por modelo, ou seja, todas as classes têm a mesma quantidade de amostras. Portanto, a avaliação foi feita usando a precisão no conjunto de testes. Calculamos a precisão geral e a precisão para as três melhores classes classificadas.

## RESULTADOS E DISCUSSÃO

Dois experimentos foram feitos: usando as imagens originais (com o fundo) e usando as imagens cortadas. Isso foi feito para avaliar quanto o fundo atrapalha na classificação do carro. A evolução da precisão é mostrada na Figura 4.

Figura 4 -Gráfico de Acurácia por época.



Fonte: Própria

Os resultados finais dos dois experimentos são mostrados na Tabela 2. Percebemos que o ganho de acurácia ao cortar os carros nas imagens era muito pequeno, o que sugere que a CNN foi capaz de extrair características relevantes do objeto com uma influência muito pequena do fundo. No entanto, houve uma melhora significativa comparando a precisão da classe superior e a precisão com base nas três principais classes.

Tabela 2 – Acurácia média.

	Normal Images	Cropped Images
<b>Accuracy</b>	79,40%	82,36%
<b>Top-3 accuracy</b>	95,43%	94,87%

Fonte: Própria.

Isto ocorreu devido ao fato da rede confundir modelos que são bastante semelhantes. Isso pode ser analisado posteriormente, observando-se a matriz de confusão da Tabela 3, para as classes em que ocorreram os piores e melhores resultados.

Tabela 3 – Matriz de confusão.

	<b>HB20</b>	<b>HB20S</b>		<b>Amarok</b>	<b>Captur</b>
<b>HB20</b>	138	144		276	0
<b>HB20S</b>	165	98		0	257

Fonte: Própria.

## CONCLUSÕES

O reconhecimento do modelo e da marca do carro em imagens ou vídeos é um problema não trivial, pois alguns modelos não mostram diferenças aparentes um do outro quando vistos de certos ângulos da visão. Este trabalho apresentou uma abordagem ligeiramente diferente daqueles da literatura, onde as imagens são apenas de um ângulo específico e de baixa diversidade.

O modelo proposto foi capaz de alcançar bons resultados, desde que o conjunto de dados tivesse uma grande diversidade de imagens. O uso do aumento de dados foi muito importante para as classes de equilíbrio e levou o classificador a altas precisões.

A parte mais desafiadora deste trabalho foi construir e rotular o banco de dados. Portanto, esta é uma contribuição significativa para outros pesquisadores interessados no reconhecimento de marca e modelo de carro.

## AGRADECIMENTOS

Agradeço a todos os colegas de laboratório pelo apoio no desenvolvimento desta pesquisa, assim como meu Professor orientador Heitor Silvério Lopes pela oportunidade oferecida. Também agradeço ao CNPQ pelo apoio financeiro concedido.

## REFERÊNCIAS

AQUINO, N. M. R., GUTOSKI, M., HATTORI, L. T., AND LOPES, H. S. (2017). **THE EFFECT OF DATA AUGMENTATION ON THE PERFORMANCE OF CONVOLUTIONAL NEURAL NETWORKS**. IN PROC XIII BRAZILIAN CONGRESS ON COMPUTATIONAL INTELLIGENCE(CBIC). ABRICOM.

GUTOSKI, M. (2018). **LEARNING AND TRANSFER OF FEATURE EXTRACTORS FOR AUTOMATIC ANOMALY DETECTION IN SURVEILLANCE VIDEOS**. MSc. DISSERTATION, GRADUATE PROGRAM IN ENGINEERING AND COMPUTER SCIENCE, FEDERAL UNIVERSITY OF TECHNOLOGY PARANÁ – UTFPR.

TAFAZZOLI, F., NISHIYAMA, K., AND FRIGUI, H. (2017). **A LARGE AND DIVERSE DATASET FOR IMPROVED VEHICLE MAKE AND MODEL RECOGNITION**. IN PROCEEDINGS OF THE IEEE CONFERENCE ON COMPUTER VISION AND PATTERN RECOGNITION (CVPR).