Revista Brasileira de Computação Aplicada, Julho, 2021

DOI: 10.5335/rbca.v13i2.12466
Vol. 13, Nº 2, pp. 28–37
Homepage: seer.upf.br/index.php/rbca/index

ARTIGO ORIGINAL

Aplicação de classificador binário por RNC na detecção de acidentes de trânsito

Application of binary classifier by CNN in the detection of traffic accidents

Augusto Carvalho Soares [1] and Danilo César Pereira [1]

[1]FACOM – Universidade Federal de Uberlândia (UFU)

*augustocs@ufu.br; danilo.pereira@ufu.br

Recebido: 09/04/2021. **Revisado:** 18/05/2021. **Aceito:** 15/06/2021.

Resumo

A possibilidade de que haja veículos que trafeguem sem a necessidade de um condutor, isto é, veículos autônomos, é um vislumbre da ficção científica que nos últimos anos vem ganhando notoriedade por grandes fabricantes e pesquisadores. É imperativa a necessidade de desenvolvimento de sistemas que possam dotar veículos autônomos de capacidade para a identificação antecipada de colisão com precisão e segurança, sendo um importante tópico arduamente explorado por diversas áreas, entre as quais destacam-se a visão computacional e a inteligência artificial pelo vasto potencial que ambas têm apresentado quando combinadas. Partindo dessa necessidade, este trabalho teve por objetivo empregar técnicas de visão computacional e inteligência artificial para o processamento e classificação de imagens extraídas de clipes curtos contidos em vídeos gravados, obtendo assim uma classificação binária de uma dada situação que identifica a ocorrência ou não de um acidente. Foram avaliadas duas arquiteturas de redes neurais convolucionais: AlexNet e ResNet-50, para a rotulação dos momentos em um conjunto de 19 vídeos, totalizando 201 clips e 18.064 imagens analisadas em 30 épocas na fase de treinamento. A eficácia dos modelos foi avaliada considerando as medidas F_1 *score* e *Precision*. Os resultados foram apurados em duas condições distintas: sem melhoramentos aplicados às imagens e com melhoramentos como a equalização de histograma. Os resultados foram: AlexNet F_1 *score* médio de 91,5% contra 89,5% da ResNet-50 para o primeiro caso e AlexNet F_1 *score* médio de 88% contra 91,5% da ResNet-50 para o segundo.

Palavras-Chave: Acidente; AlexNet; Detecção; Rede Neural; ResNet-50

Abstract

The possibility that there are vehicles that travel without the need for a driver, that is, autonomous vehicles, is a glimpse of science fiction that in recent years has been gaining notoriety by major manufacturers and researchers. There is an imperative need to develop systems that can provide autonomous vehicles with the capacity for early collision identification with precision and safety, being an important topic hard explored by several areas, among which computer vision and artificial intelligence stand out. vast potential that both have presented when combined. Based on this need, this work aimed to employ computer vision and artificial intelligence techniques for the processing and classification of images extracted from short clips contained in recorded videos, thus obtaining a binary classification of a given situation that identifies the occurrence or not of an accident . Two architectures of convolutional neural networks were evaluated: AlexNet and ResNet-50, for the labeling of moments in a set of 19 videos, totaling 201 clips and 18,064 images analyzed in 30 epochs in the training phase. The effectiveness of the models was evaluated considering the measures F_1 *score* and *Precision*. The results were obtained in two different conditions: without improvements applied to the images and with improvements such as the histogram equalization. The results were: AlexNet F_1 *score* average of 91.5% against 89.5% of ResNet-50 for the first case and AlexNet F_1 *score* average of 88% against 91.5% of ResNet-50 for the second.

Keywords: AlexNet; Collision; Detect; Neural Network; ResNet-50

1 Introdução

Nos dias atuais a perspectiva de que veículos automotores sejam guiados de forma autônoma abre novas fronteiras e desafios. Certamente um dos maiores desafios é o desenvolvimento de sistemas capazes de conduzir os veículos de forma automática e segura. Neste contexto é necessário que tais sistemas provejam formas de controle e prevenção de colisões.

Com o uso de técnicas de Inteligência Artificial (IA), Redes Neurais (RN) e redes de aprendizado de máquina profundo, os sistemas podem ser treinados para tomar decisões e realizar ações de maneira automática, contribuindo assim para o desenvolvimento de veículos autoguiados.

O desenvolvimento de veículos autônomos vêm ganhando colaborações de diversas áreas com o intuito de criar tecnologias que possam ser aplicadas em escalas comerciais.

Entre as áreas de pesquisa que mais evoluíram, a Visão Computacional e Inteligência Artificial, merecem ser citadas já que ambas trabalham há certo tempo de forma conjunta para melhorar o processo de identificação de objetos, algo extremamente precioso ao se trabalhar com veículos autônomos.

Neste trabalho propõe-se um subsistema de auxílio para detecção de colisão de trânsito por meio de técnicas de processamento de imagens captadas por câmeras *onboard*, para isso foram aplicados métodos de classificação de imagens por redes neurais.

1.1 Uma Breve História

Em 1863, o famoso escritor Júlio Verne, vislumbrou uma Paris futura com veículos sem motorista, em que imaginava carruagens e trens movidos por infraestruturas pneumáticas e eletromagnéticas (Verne et al., 1996).

As primeiras ideias e tentativas de construção de veículos autônomos datam das décadas de 20 e 30 do século XX. Em 1939 Norman Bel Geddes apresentou sua visão de carros elétricos controlados por rádio, impulsionados por campos eletromagnéticos fornecidos por circuitos embutidos na estrada, em uma feira patrocinada pela General Motors.

Recentemente o primeiro grande evento cujo foco foi voltado a área ocorreu em 2004 por meio de uma competição criada pelo Departamento de Defesa dos Estados Unidos (DARPA), onde colocou-se à prova veículos autônomos oferecendo uma premiação de 1 milhão de dólares a equipe de entusiastas e/ou pesquisadores que conseguissem percorrer um trajeto de 228 km em até 10h no deserto de Mojave. Na ocasião, nenhuma equipe conseguiu chegar ao final. No ano seguinte, a mesma competição ofereceu como prêmio o valor de 2 milhões de dólares para a conclusão do percurso. Dessa vez, cinco equipes alcançaram a reta final e a equipe vitoriosa foi a da Universidade de Stanford com o robô Stanley (Bühler et al., 2009).

Já no ano de 2007, o DARPA resolveu convidar um seleto grupo de cientistas para uma competição que tinha como foco as rodovias em trajeto urbano. De forma mais específica, o objetivo dessa prova era percorrer um circuito de 90 km contendo diversos obstáculos encontrados diariamente nas rodovias americanas. A ideia da DARPA com esse evento foi demonstrar para as indústrias automobilísticas o potencial das tecnologias aplicadas nos veículos autônomos levando-se em consideração quesitos de segurança, estabilidade e praticidade.

1.2 O Desafio

Embora competições e premiações sejam importantes para o fomento de pesquisas em veículos autônomos, os ambientes em que ocorrem são na maioria das vezes, limitados e controlados.

Mesmo com o grande progresso que se obteve até o momento, alguns desafios ainda persistem quando se desenvolve um veículo autônomo. Um deles e talvez o mais importante é desenvolver sistemas que mediante as mais diversas situações encontradas nas vias de trânsito do mundo real, possam tomar a decisão segura e precisa do momento em que o veículo deve ser parado.

A tomada de decisão sobre a parada ou não do veículo é realizada por sistemas de inteligência artificial que dependem do pré-processamento de imagens adquiridas por sistemas de visão computacional acoplados ao veículo.

Segundo Thrun et al. (2006), para o sucesso de tais veículos autônomos, é imprescindível perceber e interagir com o tráfego em movimento, com precisão e livre de ocorrência de erros.

No entanto, para que a análise seja precisa é necessário treinar o sistema de inteligência artificial sobre cada cenário que um condutor poderia se deparar ao guiar um veículo no dia a dia. Porém , coletar cenas em que cada uma dessas situações ocorrem individualmente seria computacionalmente por demais custoso.

1.3 Revisão de trabalhos

Recentemente, com o surgimento de novas ferramentas computacionais que proporcionaram todo um novo universo de possibilidades em diversas áreas, inúmeros pesquisadores têm se debruçado sobre trabalhos que visam propor soluções que contornem os desafios de se construir um veículo autônomo.

A Industria automobilística já se move rumo às novas tecnologias, (Toyota, 2013, Volvo, 2013), grande parte das montadoras já tem projetos relacionados a autonomia de direção.

A título de exemplo, alguns trabalhos notáveis são os de: Fenton (1970), Aufrère et al. (2003), Carsten et al. (2012), de Winter et al. (2014), Liu et al. (2018), Gao et al. (2019).

Vários trabalhos têm surgido nos últimos anos aplicando visão computacional aliada a arquiteturas de redes neurais profundas, para detectar o momento de início de um evento adverso no trânsito de veículos automotores, chamado de "anomalia de tráfego".

Em seu livro intitulado *Introduction to AI Robotics*, Murphy (2018) apresenta uma pesquisa abrangente sobre algoritmos de inteligência artificial e organização de programação para sistemas robóticos tem sido um dos principais pilares de sistemas automatizados, em sua segunda edição traz notória e atualizada expansão.

Um trabalho de destaque que traz métodos de coleta

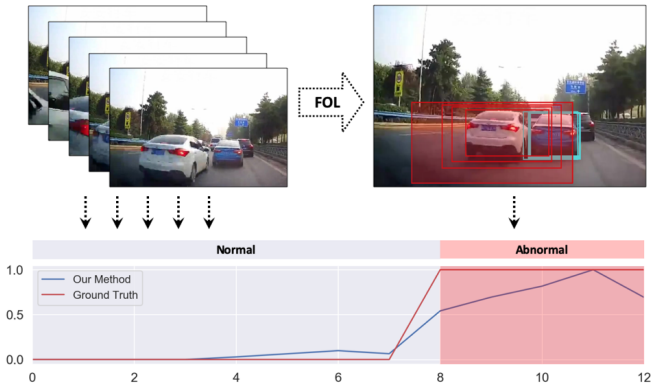

Figura 1: FOL (Localização Futura de Objetos)

de dados de sensores, algoritmos de cálculo de posicionamento e rotas é o de Thrun et al. (2006) *Stanley: The Robot that Won the DARPA Grand Challenge*, que é de leitura mandatória por pesquisadores em veículos autônomos, e já tendo sido prestigiado em milhares de citações.

Diferentemente do que ocorre na maioria dos trabalhos encontrados na literatura, onde a eficiência está baseada na quantidade de quilômetros que o veículo percorre, o que geralmente é pequena, alguns estudos defendem que veículos autônomos devem ser expostos a bilhões de quilômetros de rodagem ou a um grande conjunto de vídeos contendo todos os cenários possíveis de serem encontrados. Isso é defendido com o argumento de que o processo de avaliação da precisão e segurança do sistema é algo longo e demorado, portanto, não se pode chegar a essas afirmações sem as devidas provas. Por isso, uma alternativa é desenvolver modelos que consigam prever sinais de anomalia a partir de modelos treinados sob bases de dados contendo situações normais e anormais de dirigibilidade.

Como uma das mais bem-sucedidas estratégias para a detecção de anomalias de tráfego, é a análise de imagens de vídeos adquiridos por câmeras acopladas ao painel do carro, indica-se (Malla et al., 2019) com o trabalho *Future Object Localization* em que o sistema analisa a trajetória de um objeto, prevê sua localização futura e verifica se em determinado instante de tempo ocorre uma situação anormal, como ilustra a Fig. 1.

Diante dos desafios e estratégias de solução, este trabalho busca o objetivo de desenvolver um sistema capaz de identificar o momento em que uma colisão ocorre.

Para isso, foi utilizado como base o trabalho *Unsupervised Traffic Accident Detection in First-Person Videos* de Yao et al. (2019), que é um aprimoramento de seu artigo anterior *Egocentric Vision-based Future Vehicle Localization for Intelligent Driving Assistance Systems* (Yao et al., 2018), onde o autor emprega um conjunto de vídeos gravados em primeira pessoa, coletados de câmeras *on-board* em veículos e adquiridos da plataforma YouTube para avaliar seu sistema.

Faz-se menção a Yang et al. (2018), que estudou a detecção de anomalias em seu trabalho: *Traffic Anomaly Detection and Prediction Based on SDN-Enabled ICN*.

Aprofundando um pouco mais as discussões acerca da detecção de anomalias, Yao et al. (2020) explora as características de porque, quando e como essas anomalias acontecem em seu mais novo trabalho *When, Where, and What? A New Dataset for Anomaly Detection in Driving Videos*, em que analisa um novo conjunto de dados.

2 Metodologia

A ideia inicial foi usar como base um trabalho que explorasse o uso das tecnologias de redes neurais artificiais (RNAs), que são métodos e técnicas empregadas para que um programa de computador trabalhe mais ou menos na mesma mecânica de funcionamento do cérebro humano, simulando assim características humanas como o reconhecimento de padrões e objetos bem como aprendizado e classificação.

O objetivo principal então foi empregar alguns algoritmos de redes neurais e avaliar se seriam melhores na tarefa de classificar uma anomalia de trânsito, quando algum tipo de melhoramento é aplicado nas imagens. Para isso o mesmo grupo de imagens foi submetido à classificação em duas fases: A (sem melhoramentos), B (melhoramentos aplicados).

Neste modelo, a base de dados foi decomposta em várias cenas de colisão para posteriormente serem novamente divididas em sequências de quadros. Esse volume de imagens, e então foi repassado aos modelos de redes neurais AlexNet e ResNet-50 a fim de treiná-los e assim as saídas fornecidas na etapa de classificação foram utilizadas para remontar os videoclipes e demarcar os quadros identificados como colisão, melhor visualizado pela Fig. 2.

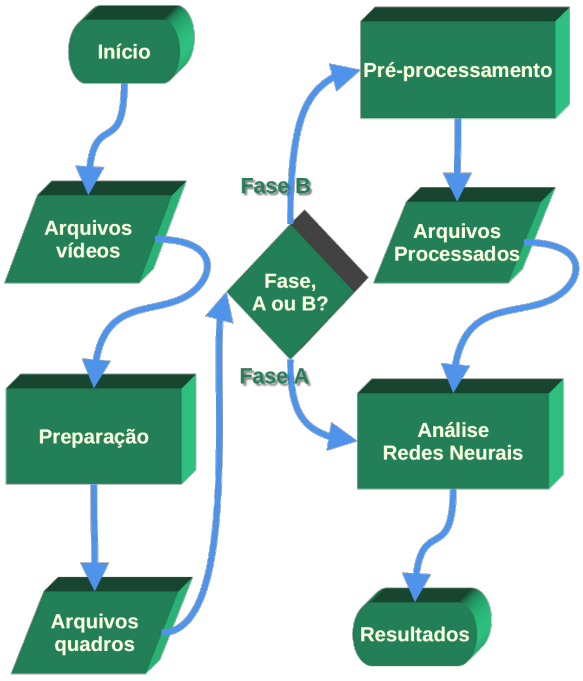

Figura 2: Fluxo das etapas

Conforme destacado anteriormente, o grande problema quando se desenvolve sistemas para veículos autônomos está no fato de que não é possível expô-lo às mais diversas situações encontradas diariamente nas rodovias e tão pouco encontrar vídeos suficientes que retratem cada uma delas.

Para tarefas de classificação em redes neurais, uma etapa desgastante é a rotulação manual do conjunto de treinamento, para que o classificador possa "treinar" e "aprender" corretamente como rotular cada instância a ele apresentada.

Sendo assim este trabalho utiliza o conjunto de vídeos disponibilizados por Yao et al. (2019), podendo ser baixado publicamente por meio do serviço de multimídia YouTube (os *hashes* identificadores dos vídeos estão no repositório do projeto). Um grande ganho ao utilizar esse conjunto está no fato de que também foi disponibilizado pelo autor, arquivos com as demarcações e rótulos de classificação realizadas manualmente para cada quadro dos vídeos.

Esses dados são valiosos, visto que as redes neurais enquadram-se no grupo de algoritmos supervisionados e que por isso, necessitam de dados rotulados para aprenderem. Os arquivos do projeto original estão disponíveis publicamente no repositório do autor [1].

2.1 Dataset e espaço amostral

No conjunto de dados selecionado pelo autor existem 204 vídeos e 1.500 *short clips*. Estes são pequenas partes do vídeo que mostram uma situação de anomalia que pode ou não ser um acidente de trânsito.

Todavia, destaca-se que neste trabalho foram selecionados apenas 19 vídeos com um total de 201 *short clips*. Essa diminuição da quantidade de dados foi necessária para que os recursos computacionais disponíveis possibilitassem a execução das redes neurais, uma vez que a execução de tarefas de aprendizado de máquina requerem um *hardware* consideravelmente poderoso ou a utilização de modernas placas de processamento gráfico (GPU).

A primeira etapa foi extrair todas as imagens dos vídeos (quadro a quadro), para este processo criou-se um *shell script* que automatizou chamadas da ferramenta FFmpeg, usando amostragem de extração com a frequência de 10 fps (*frames* por segundo), gerando 18.064 imagens.

Em seguida o universo amostral de imagens foi dividido em 3 grupos distintos para as fases de treinamento, validação e testes, tal separação é importante para a correta aferição dos resultados.

O percentual que cada grupo representará do universo amostral é objeto de uma escolha sensível e crítica, que influencia fortemente os resultados.

Embora não se tenha um consenso sobre esses valores, é aconselhável reservar a maior parte para a fase de treinamento, uma vez que quanto mais se treina, mais se aprende.

Ao final da etapa de preparação dos dados, as imagens ficaram distribuídas em:

- Treinamento: 11.745 imagens (65%)
- Validação: 3.609 imagens (20%)
- Teste: 2.710 imagens (15%)

O tamanho dos arquivos separados entre imagens e rótulos para cada subconjunto de treino, validação e teste é apresentado a seguir:

- Images_train.npy (6,8 GB)
- Images_val.npy (2,1 GB)
- Images_test.npy (1,6 GB)
- Rotulos_train.npy (92 KB)
- Rotulos_val.npy (29 KB)
- Rotulos_test.npy (22 KB)

2.2 Pré-processamento e Redes Neurais

Analisando o aspecto das imagens, propõe-se uma mudança do espaço de cores de RGB para o YCbCr como indicado por Szeliski (2011) aliada a um processo de equalização de histograma a fim de prover equilíbrio aos canais de cores visto em Li et al. (2014) para tentar reduzir a relação sinal x ruído da base de dados e aumentar a taxa de classificação. Isto foi feito para todo *dataset* em uma etapa distinta, para avaliação da eficácia da proposta.

A próxima etapa do trabalho foi a implementação das arquiteturas de redes neurais. Essas redes agrupam-se na categoria de algoritmos de classificação supervisionados onde o processo de aprendizagem se alicerça sobre um conjunto de dados previamente rotulados para posteriormente realizar-se a classificação de dados desconhecidos.

As redes neurais convolucionais (CNNs) são um subgrupo de redes neurais criadas por LeCun and Bengio (1995) com a proposta de aplicar operações lineares e não lineares por meio de filtros convolucionais dispostos entre as camadas da rede neural, modelo visto na Fig. 3.

[1] https://github.com/MoonBlvd/tad-IROS2019/

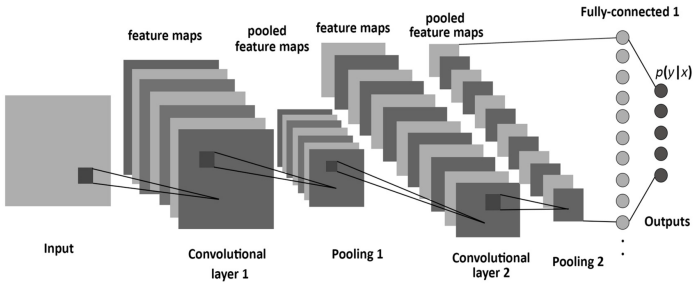

Figura 3: Funcionamento de uma rede neural convolucional

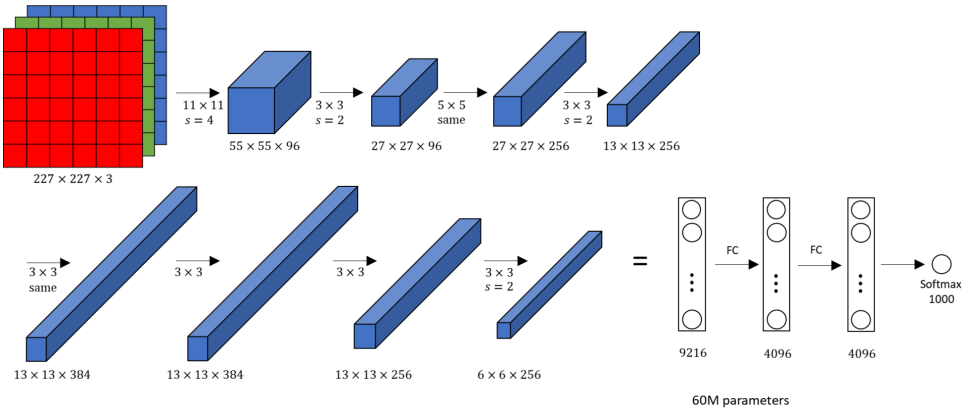

Figura 4: Arquitetura da rede neural AlexNet

Esses modelos de redes recentemente vêm ganhando destaque nas etapas de processamento e análise de imagens devido a capacidade de extrair características e avaliar nativamente as informações espaciais.

A rede neural parte de um conjunto de neurônios artificiais agrupados em diversas camadas com ou sem subníveis, e instanciados com pesos aleatórios que são ajustados a cada época de treinamento com base na retro propagação do erro calculado entre a saída da rede e valor do rótulo que deveria ser obtido.

Este trabalho empregou duas arquiteturas de redes neurais convolucionais tradicionais: AlexNet e ResNet-50.

Devido ao grande volume de dados que cada vídeo gera, optou-se por descartar o processo de aumento de dados, ou seja, processo que aplica operações de rotação, translação, espelhamento, entre outras ao conjunto de imagens para aumentar a diversidade de dados e submeter as redes a dados desprovidos de padronização a fim de evitar o super aprendizado.

2.3 AlexNet

A AlexNet é uma rede neural convolucional que foi desenvolvida por Alex Krizhevsky em colaboração com Ilya Sutskever e Geoffrey Hinton. Ela é muito popular e já foi citada por mais de 70.000 vezes (dados do Google Scholar), servido de base para várias outras implementações.

A estrutura da rede é formada por uma camada com filtros de convolução mais uma operação para diminuição do número de parâmetros através da escolha do maior valor (*max polling*) dentro de uma janela de vizinhança pré-definida. O mesmo conjunto de operações é realizado na segunda, terceira, quarta e quinta camada, alterando-se apenas o tamanho dos filtros de convolução e tamanho das janelas do *max polling* como observado em Alom et al. (2018).

Em seguida, duas camadas com todos os neurônios conectados (*Dropout*) são usadas para forçar os neurônios a não dependerem da presença de outros e dessa forma, tanto minimizar o *overfitting*, quanto tornarem a rede neural mais robusta para a perda de qualquer informação pontual. Por último é empregado uma função de ativação sigmóide (*Softmax*) que transforma as saídas dos neurônios da última camada de *Dropout* em valores de classes variando 0 ou 1, ou seja, para uma classificação binária, ilustrado na Fig. 4.

A saída da última camada totalmente conectada é alimentada a um *softmax* de 1000 vias que produz uma distribuição nas 1000 etiquetas de classe. Esta rede maximiza o objetivo de regressão logística multinomial, que é equivalente a maximizar a média entre os casos de treinamento do log-probabilidade do rótulo correto sob a distribuição de predição (Krizhevsky et al., 2017).

2.4 ResNet-50

Outra arquitetura clássica de redes neurais artificiais profundas é a ResNet-50, foi proposta por He et al. (2015), quando venceu uma importante competição, a *ImageNet Large Scale Visual Recognition Challenge* (ILSVRC), e é um dos mais influentes modelos de aprendizado profundo.

Esta arquitetura foi revolucionária, pois, não só obteve as melhores taxas de acerto em tarefa de classificação na base de dados ImageNet, mas também demonstrou solução para o problema do esvanecimento de gradiente.

A ResNet-50 é uma arquitetura de rede residual, possui 49 camadas com filtros de convolução e 1 camada totalmente conectada. Tem-se ao final de cada camada de convolução um processo de retro propagação do erro com o repasse de parte dessas informações para as camadas posteriores (*residual blocks*), melhor compreendido na Fig. 5.

As ResNets referem-se a redes neurais em que as conexões ignoradas ou conexões residuais fazem parte da arquitetura da rede. Essas conexões de salto permitem que as informações de gradiente passem pelas camadas, criando "rodovias" de informações, onde a saída de uma camada ou de uma ativação anterior é adicionada à saída de uma camada mais profunda. Isso permite que as informações das partes anteriores da rede sejam passadas para as partes mais profundas da rede, ajudando a manter a propagação do sinal mesmo em redes mais profundas. Pular conexões são um componente crítico que permitiu o treinamento bem-sucedido de redes neurais mais profundas (He et al., 2015).

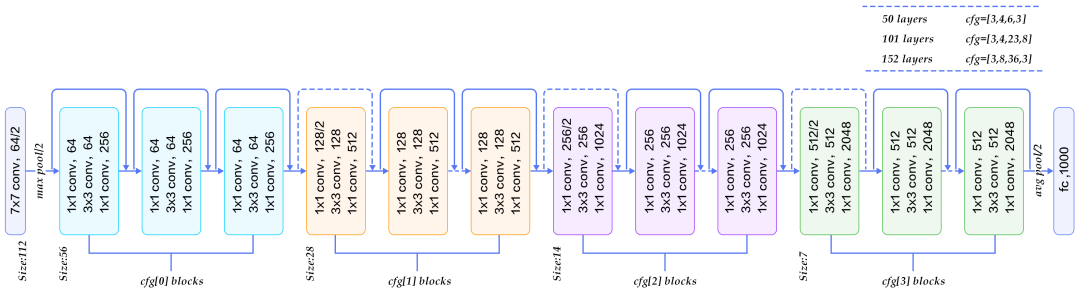

Figura 5: Arquitetura de rede neural residual - ResNet

Estes "saltos" entre as camadas de neurônios Fig. 6, permitiu o desenvolvimento de redes com profundidades maiores, sem que o excessivo número de camadas causasse problemas de perda de desempenho e acurácia.

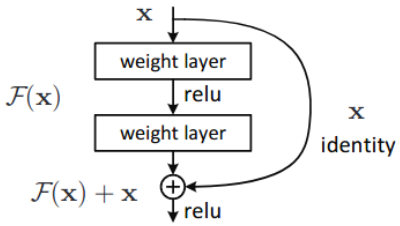

Figura 6: Salto de camadas

2.5 Ambiente de execução

Os códigos desenvolvidos neste trabalho foram executados em um computador Desktop com processador Intel(R) Core(TM) i7-8700 CPU @ 3.20 GHz, memória RAM de 16 GB, placa gráfica (GPU) NVIDIA GP106 [GeForce GTX 1060 6 GB], disco rígido do tipo SSD com taxas de leitura e escritas sequenciais de 550 e 490 MBps, respectivamente.

Todas as execuções ocorreram por meio da GPU e por isso, foi necessário definir os parâmetros: quantidade de épocas de treinamento e tamanho do *batch*. O primeiro valor é responsável por informar a rede neural a quantidade de iterações que serão conduzidas para o treinamento da rede; o segundo corresponde ao tamanho dos subconjuntos de treinamentos manipulados pela rede neural a cada época de treinamento. Ambos os parâmetros dependem da quantidade de memória disponível na placa gráfica, que nesse caso foram respectivamente de 30 e 32.

É importante ressaltar que havia a possibilidade de produzir os mesmos testes utilizando o processador e a memória RAM, o que aumentaria o tamanho dos subconjuntos e da quantidade de iterações. No entanto, quando se opta por esse caminho, o tempo para execução é automaticamente expandido de maneira exponencial. Sendo assim, não seria viável dado o volume de dados a ser processado e o espaço de tempo gasto.

A última etapa do trabalho foi a reconstrução dos *short clips* a partir das imagens rotuladas pelas redes neurais. As redes neurais geram um arquivo Python (.npy) com todos os rótulos para as imagens de um *short clip* passados às camadas de entrada. A partir dos rótulos realizou-se a demarcação das imagens classificadas como colisão pela rede neural e remontou-se o vídeo com os mesmos parâmetros utilizados na divisão, ou seja, 10 *frames* por segundos como taxa de amostragem.

2.6 Medidas de validação

As medidas de validação foram tomadas em duas fases, quando se avaliou o desempenho e a eficácia dos modelos, a primeira fase quando não houve modificações nas imagens, e a segunda quando foram aplicadas técnicas de melhoramento de imagens.

As analises comparativas entre as redes neurais foram realizadas a partir dos resultados extraídos das etapas de treinamento, validação e testes, gerando a matriz de confusão, que compara o número de positivos reais com o número de positivos preditos, assim como o número negativos reais e o número de negativos preditos.

Com isso pode-se calcular os valores das medidas *Precision* e *Recall*, Eq. (1).

$$precision = \frac{TP}{TP + FN}$$
$$recall = \frac{TP}{TP + FN} \quad (1)$$

Neste trabalho, optou-se por utilizar o F_1 *score*, medida já consagrada para avaliação de acertos em tarefas de classificação por redes neurais convolucionais. Que é calculado com base nos valores de *Precision* e *Recall*, sua definição equacional é vista em Eq. (2).

$$F_1 score = 2. \frac{precision \,.\, recall}{precision + recall} \quad (2)$$

Outras medidas também foram extraídas em duas fases distintas, para avaliar as diferenças de eficácia das arquiteturas de rede neural, como a sensibilidade a perda e a especificidade. Que serviram de balizamento na avaliação da eficácia de classificação das redes neurais aqui estudadas, como a sensibilidade, a especificidade e a acurácia.

3 Resultados

Com os dados pré-fixados de 30 épocas e 32 imagens como *batch*, o tempo consumido pelas redes neurais na etapa de treinamento foram de aproximadamente 86 minutos para a rede neural convolucional AlexNet e 67 minutos para a rede neural convolucional ResNet-50 para cada fase.

A outra comparação realizada foi em relação a taxa de erro e acurácia de ambos os modelos durante o processo de treinamento. Para isso, coletou-se informações a cada época sobre o erro de aprendizado juntamente com a taxa de acerto (acurácia).

Estas informações serviram para elaborar os gráficos de desempenho de aprendizado de cada rede neural. O primeiro gráfico corresponde aos dados da AlexNet Fig. 7, e o segundo da ResNet-50 Fig. 8.

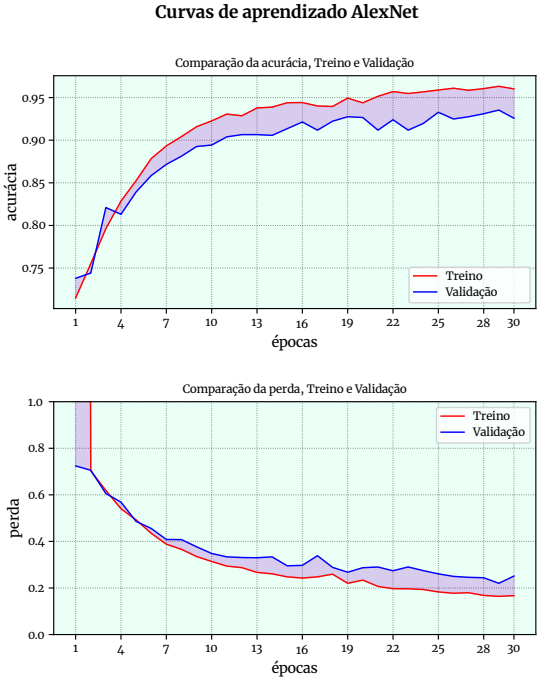

Figura 7: Gráficos de perda e acurácia para AlexNet

Os dados gerados pela rede neural convolucional AlexNet mostram que ao iniciar o processo de treinamento, a quantidade de rótulos classificados erroneamente é grande, pois não se sabe a priori quais os pesos corretos que devem ser atribuídos aos neurônios para se obter uma classificação exata. No entanto, com o passar das épocas e o aprendizado consequente das características do que é um acidente, a rede passa a diminuir o erro no processo ao mesmo que se aumenta a taxa de classificação, chegando ao final das 30 épocas com uma perda pouco acima de 23% e 93% de precisão.

Considerando os mesmos parâmetros, a rede neural convolucional ResNet-50 apresentou uma acurácia de 94% contra uma taxa de perda de 25%.

Também elaborou-se a tabela de precisão, *recall* e F_1

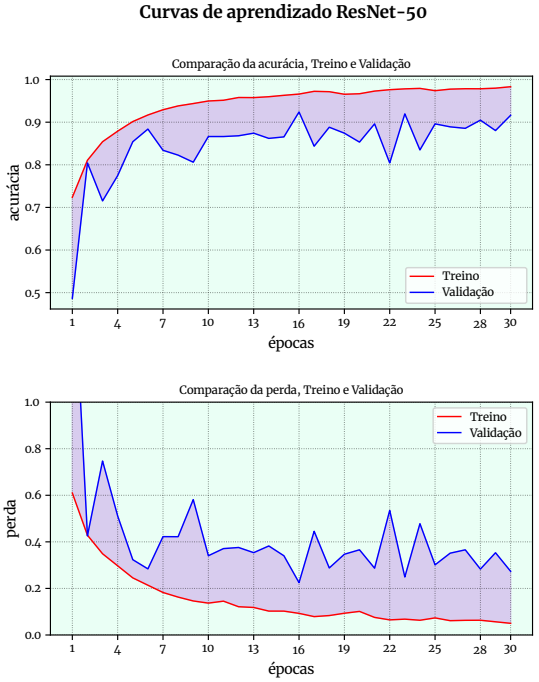

Figura 8: Gráficos de perda e acurácia para ResNet-50

score para cada rede neural. Essa tabela é gerada na etapa de testes da rede neural, que é sucessiva a etapa de treinamento. Nesse processo um conjunto de dados diferente das etapas de treino e validação foi informado a rede neural e ela se incumbiu de classificar as imagens com base no que aprendeu. Após isso, comparou-se com o rótulo correto e computou-se os valores de verdadeiro positivo, falso positivo, verdadeiro negativo e falso negativo para finalmente obter-se os valores das medidas de validação.

Tabela 1: Relatório de Classificação - AlexNet

		precision	recall	f1-score
Fase A	Imagens Puras			
	Classe normal	0.93	0.97	0.95
	Classe acidente	0.92	0.84	0.88
Fase B	Imagens Process.			
	Classe normal	0.90	0.98	0.94
	Classe acidente	0.94	0.76	0.84

Pela análise da Tabela 1 vê-se que, na fase A, em que a rede neural AlexNet classificou imagens sem melhoramentos (Imagens Puras) a precisão foi de 93% para as imagens que não continham colisão (Classe normal) contra 92% para as imagens que continham colisão (Classe acidente), ou seja, daquelas imagens que a rede neural classificou como colisão, quantas efetivamente eram dessa classe.

Já a taxa de *recall* ficou em 97% em casos que não houve colisão contra 84% nos casos em que as colisões de fato ocorreram, esta é uma medida que avalia quando realmente é da classe colisão X, o quão frequente ela classifica como X.

Por último, a medida F_1 *score* avalia o equilíbrio entre uma boa precisão com bom *recall*. Os valores obtidos foram de 95% para imagens sem colisão e 88% para imagens com colisão.

Na fase B, em que a rede neural AlexNet classificou imagens com melhoramentos (Imagens Process.), os valores das medidas de precisão, *recall* e F_1 *score* foram respectivamente 90%, 98% e 94% para a classificação de imagens onde não houve acidente, e 94%, 76% e 84% para imagens onde o acidente realmente ocorreu.

O mesmo conjunto de imagens foi submetido aos testes da rede neural ResNet-50 nos mesmos termos e parâmetros que a primeira, com os seguintes resultados:

Tabela 2: Relatório de Classificação - ResNet-50

		precision	recall	f1-score
Fase A	Imagens Puras			
	Classe normal	0.94	0.94	0.94
	Classe acidente	0.85	0.86	0.85
Fase B	Imagens Process.			
	Classe normal	0.94	0.96	0.95
	Classe acidente	0.90	0.87	0.88

Observa-se na Tabela 2 que, a rede neural ResNet-50 na fase A (Imagens Puras), alcançou valores de precisão, *recall* e F_1 *score* de 94% quando não houve acidente, e 85%, 86% e 85% respectivamente, quando houve. Na fase B (Imagens Process.), os valores alcançados de precisão, *recall* e F_1 *score* foram de 94%, 96% e 95% respectivamente quando não houve acidente, e 90%, 87% e 88%, quando houve.

4 Discussões

Pela comparação dos resultados obtidos nas etapas A onde as imagens não sofreram alterações e B onde foram realizadas alterações como a mudança do espaço de cores de RGB para YCbCr e equalização de histograma. Comprova-se mínima diferença nas taxas de acerto de classificação pelas arquiteturas de redes neurais aqui avaliadas.

Contudo nota-se que a rede neural ResNet-50 pode-se beneficiar das técnicas de melhoramento de imagens aqui propostas.

Aponta-se ainda a superioridade do processo de aprendizado da rede neural AlexNet em comparação a ResNet-50 (neste *dataset* específico), pela menor variação de sua curva de aprendizado no conjunto de validação, claramente visto na Fig. 7.

Os resultados demonstram que a rede neural convolucional ResNet-50 tem melhor capacidade (ainda que discreta) de classificar imagens contendo colisão e imagens normais.

A justificativa para isso é que a arquitetura ResNet-50 tem uma quantidade muito maior de camadas do que a AlexNet (50 contra 6) e consequentemente mais neurônios para ajudarem no processo de tomada de decisão.

Mesmo trabalhando com um subconjunto de dados de treinamento pequeno, e a quantidade de iterações não sendo altos, a rede ResNet-50 tem uma leve superioridade em relação à taxa de classificação (90,25% AlexNet, 90,50% ResNet-50 de F_1 *score* médio) conforme os resultados apresentados nas Tabela 1 e Tabela 2.

A Fig. 9 ilustra a rotulação feita pelas redes neurais analisadas, observam-se duas situações onde foi obtido sucesso ao classificar o quadro como acidente(9a,9c) e outras duas em que foi observado erro de classificação (9b,9d).

(a) acerto **(b)** erro

(c) acerto **(d)** erro

Figura 9: Quadros Rotulados pelas redes neurais.

5 Conclusão

Esse trabalho teve como objetivo realizar a comparação de dois modelos tradicionais de redes neurais convolucionais (AlexNet e ResNet-50) para a demarcação dos momentos em que veículos sofrem algum tipo de acidente. Para essa tarefa, utilizou-se como base de dados vídeos gravados em primeira pessoa a bordo de veículos.

Com as configurações discorridas ao longo do trabalho, chegou-se à conclusão de que a rede neural convolucional ResNet-50 possui maior taxa de acurácia média do que a AlexNet.

Nos resultados ainda percebe-se uma diferença da taxa de classificação entre as diferentes classes de imagens. Isso ocorre devido as classes estarem desbalanceadas, implicando no maior aprendizado de uma classe em detrimento da outra.

Pelos resultados obtidos, nenhum dos modelos alcançou valores máximos (100%) e nem mínimos (0%), tanto visto pelos gráficos quanto nas matrizes de confusão, sendo assim um indício de que pode não ter ocorrido super ou sub treinamento.

Como trabalhos futuros sugere-se que novas comparações sejam realizadas aumentando-se o espaço amostral (número de vídeos utilizados) como treinamento, validação e testes.

Também recomenda-se aumentar o número de épocas de treinamento e o tamanho do subconjunto para treinamento das épocas (*batchs*), bem como um melhor ajuste dos hiper-parâmetros.

Os códigos desenvolvidos para este trabalho encontram-se disponíveis de forma livre e gratuita sob licença GPL e podem ser obtidos em https://github.com/augustocsoa/SSADAT.

Referências

Alom, M. Z., Taha, T. M., Yakopcic, C., Westberg, S., Sidike, P., Nasrin, M. S., Van Esesn, B. C., Awwal, A. A. S. and Asari, V. K. (2018). The history began from AlexNet: A comprehensive survey on deep learning approaches, *arXiv:1803.01164 [cs]* . Disponível em https://arxiv.org/abs/1803.01164v2.

Aufrère, R., Gowdy, J., Mertz, C., Thorpe, C., Wang, C.-C. and Yata, T. (2003). Perception for collision avoidance and autonomous driving, *Mechatronics* **13**: 1149–1161. https://doi.org/10.1016/S0957-4158(03)00047-3.

Bühler, M., USA and Advanced Research Projects Agency (2009). *The DARPA Urban Challenge autonomous vehicles in city traffic*, Springer. https://doi.org/10.1007/978-3-642-03991-1.

Carsten, O., Lai, F. C. H., Barnard, Y., Jamson, A. H. and Merat, N. (2012). Control task substitution in semi-automated driving: Does it matter what aspects are automated?, *Human Factors: The Journal of the Human Factors and Ergonomics Society* **54**: 747–761. https://doi.org/10.1177%2F0018720812460246.

de Winter, J. C., Happee, R., Martens, M. H. and Stanton, N. A. (2014). Effects of adaptive cruise control and highly automated driving on workload and situation awareness: A review of the empirical evidence, *Transportation Research Part F: Traffic Psychology and Behaviour* **27**: 196–217. https://doi.org/10.1016/j.trf.2014.06.016.

Fenton, R. (1970). Automatic vehicle guidance and control—a state of the art survey, *IEEE Transactions on Vehicular Technology* **19**: 153–161. https://doi.org/10.1109/T-VT.1970.23443.

FFmpeg Developers (2020). *FFmpeg: Multimedia Framework, able to encode, decode and transcode*. Available at https://www.ffmpeg.org/.

Gao, M., Xu, M., Davis, L. S., Socher, R. and Xiong, C. (2019). StartNet: Online detection of action start in untrimmed videos, *arXiv:1903.09868 [cs]* . Disponível em http://arxiv.org/abs/1903.09868.

He, K., Zhang, X., Ren, S. and Sun, J. (2015). Deep residual learning for image recognition, *arXiv:1512.03385 [cs]* . Disponível em https://arxiv.org/abs/1512.03385v1.

Kaehler, A. and Bradski, G. R. (2017). *Learning OpenCV 3: computer vision in C++ with the OpenCV library*, first edition, second release edn, O'Reilly Media, Sebastopol, CA.

Keras (2020). *Keras Documentation: Keras API reference*. Available at https://keras.io/api.

Krizhevsky, A., Sutskever, I. and Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks, *Advances in neural information processing systems* **25**: 1097–1105. https://proceedings.neurips.cc/paper/2012/file/c399862d3b9d6b76c8436e924a68c45b-Paper.pdf.

Krizhevsky, A., Sutskever, I. and Hinton, G. E. (2017). ImageNet classification with deep convolutional neural networks, *Communications of the ACM* **60**(6): 84–90. https://doi.org/10.1145/3065386.

LeCun, Y. and Bengio, Y. (1995). Convolutional networks for images, speech, and time-series, *in* M. A. Arbib (ed.), *The Handbook of Brain theory and Neural Networks*, MIT Press.

Li, Z.-N., Drew, M. S. and Liu, J. (2014). *Fundamentals of Multimedia*, Springer International Publishing.

Liu, W., Luo, W., Lian, D. and Gao, S. (2018). Future frame prediction for anomaly detection - a new baseline, *arXiv:1712.09867 [cs]* . Disponível em http://arxiv.org/abs/1712.09867.

Malla, S., Dwivedi, I., Dariush, B. and Choi, C. (2019). NEMO: Future Object Localization Using Noisy Ego Priors, *arXiv:1909.08150 [cs, eess]* . Disponível em https://arxiv.org/abs/1909.08150v1.

Murphy, R. (2018). *Introduction to AI Robotics*, The MIT Press, Cambridge, MA.

OpenCV (2020). *OpenCV 4.4.0 documentation*. Available at https://docs.opencv.org/4.4.0/.

Python (2020). *Python3 3.8.6 online documentation*. Available at https://docs.python.org/3.8/.

Skodras, A., Christopoulos, C. and Ebrahimi, T. (2001). The JPEG 2000 still image compression standard, *IEEE Signal Processing Magazine* **18**(5): 36–58. http://ieeexplore.ieee.org/document/952804/.

Szeliski, R. (2011). *Computer vision: algorithms and applications*, Springer. https://doi.org/10.1007/978-1-84882-935-0.

TensorFlow (2020). *TensorFlow API Documentation v2.3.0*. Available at https://www.tensorflow.org/api_docs?hl=pt-br.

Thrun, S., Montemerlo, M., Dahlkamp, H., Stavens, D., Aron, A., Diebel, J., Fong, P., Gale, J., Halpenny, M., Hoffmann, G., Lau, K., Oakley, C., Palatucci, M., Pratt, V., Stang, P., Strohband, S., Dupont, C., Jendrossek, L.-E., Koelen, C., Markey, C., Rummel, C., van Niekerk, J., Jensen, E., Alessandrini, P., Bradski, G., Davies, B., Ettinger, S., Kaehler, A., Nefian, A. and Mahoney, P. (2006). Stanley: The robot that won the DARPA grand challenge, *Journal of Field Robotics* **23**(9): 661–692. https://doi.org/10.1002/rob.20147.

Toyota (2013). 2013 consumer electronics show – toyota motor corp. and lexus advance active safety research vehicle, *2013 Consumer Electronics Show* . Disponível em https://pressroom.lexus.com/2013-toyota-lexus-consumer-electronics-show-mark-templin-jan7/.

Verne, J., Howard, R., Weber, E. and Wenngren, A. (1996). *Paris in the Twentieth Century*, Random House.

Volvo (2013). Volvo cars reveals world-class safety and support features to be introduced in the all-new xc90 in 2014, *Volvo Cars Global Newsroom* . Disponível em https://www.media.volvocars.com/global/en-gb/media/pressreleases/49875/volvo-cars-reveals-world-class-safety-and-support-features-to-be-introduced-in-the-all-new-xc90-in-2/.

Yang, F., Jiang, Y., Pan, T. and E., X. (2018). Traffic anomaly detection and prediction based on SDN-enabled ICN, *2018 IEEE International Conference on Communications Workshops (ICC Workshops)*, IEEE, pp. 1–5. https://doi.org/10.1109/ICCW.2018.8403693.

Yao, Y., Wang, X., Xu, M., Pu, Z., Atkins, E. and Crandall, D. (2020). When, where, and what? a new dataset for anomaly detection in driving videos, *arXiv:2004.03044 [cs]* . Disponível em https://arxiv.org/abs/2004.03044v1.

Yao, Y., Xu, M., Choi, C., Crandall, D. J., Atkins, E. M. and Dariush, B. (2018). Egocentric vision-based future vehicle localization for intelligent driving assistance systems, *arXiv:1809.07408 [cs]* . Disponível em https://arxiv.org/abs/1809.07408v1.

Yao, Y., Xu, M., Wang, Y., Crandall, D. J. and Atkins, E. M. (2019). Unsupervised traffic accident detection in first-person videos, *2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, IEEE, pp. 273–280. https://doi.org/10.1109/IROS40897.2019.8967556.