



**PONTIFÍCIA UNIVERSIDADE CATÓLICA DO PARANÁ
PRÓ-REITORIA DE GRADUAÇÃO, PESQUISA E PÓS-GRADUAÇÃO**

**Programa Institucional de Bolsas de Iniciação Científica
PIBIC**

RELATÓRIO FINAL

CLASSIFICAÇÃO DE VAGAS DE ESTACIONAMENTO UTILIZANDO FEWSHOT LEARNING E CONTINUAL LEARNING

Andre Gustavo Hochuli

CURITIBA

2024

MATEUS ANTONIO DANIEL
ANDRE GUSTAVO HOCHULI

CIBERSEGURANÇA - PUCPR

**CLASSIFICAÇÃO DE VAGAS DE ESTACIONAMENTO UTILIZANDO FEWSHOT
LEARNING E CONTINUAL LEARNING**

Relatório Final apresentado ao Programa Institucional de Bolsas de Iniciação Científica, Pró-Reitoria de Pesquisa e Pós-Graduação da Pontifícia Universidade Católica do Paraná, e órgãos de fomento, sob orientação do **Prof. Andre G. Hochuli**.

CURITIBA

SUMÁRIO

RESUMO	i
LISTA DE FIGURAS	i
LISTA DE TABELAS	i
LISTA DE ANEXOS	i
1 INTRODUÇÃO	1
1.1 PROBLEMATIZAÇÃO	
1.2 ESTADO DA ARTE	
1.3 MOTIVAÇÃO AO PROJETO	
2 OBJETIVO	2
3 MATERIAIS E MÉTODO	3
3.1 DATASET.....	
3.2 PREPARAÇÃO DE DADOS.....	
3.3 MODELO.....	
3.4 TREINAMENTO.....	
3.5 INFERÊNCIA.....	
4 RESULTADOS	4
5 DISCUSSÃO	5
6 CONCLUSÃO	6
REFERÊNCIAS	7
ANEXOS	8

RESUMO

A busca por vagas de estacionamento em centros urbanos densamente povoados é um problema recorrente que exige soluções inovadoras. Este estudo utiliza técnicas de inteligência artificial para classificar vagas de estacionamento como ocupadas ou livres, aproveitando o dataset PKLot. O trabalho avalia redes siamesas treinadas em diferentes sub-datasets do PKLot (UFPR04, UFPR05 e PUC), com volumes de treino de 2.000, 5.000 e 10.000 imagens, e testa a predição com poucos pares de teste de outros sub-datasets. Resultados indicam que a rede treinada com UFPR04 apresenta alta acurácia e precisão, com leve queda ao ser testada em PUC e maior variabilidade com UFPR05, especialmente com menores volumes de treino. A rede treinada com PUC mantém desempenho excepcional no próprio sub-dataset, mas enfrenta desafios ao generalizar para UFPR05. Por outro lado, a rede treinada com UFPR05 mostra robustez ao classificar seu próprio sub-dataset, mas apresenta performance inferior ao testar com UFPR04 e PUC, evidenciando a influência das características intrínsecas dos sub-datasets na capacidade de generalização. Aumentar o volume de dados de treino geralmente melhora a performance, mas é importante cuidar com o super ajuste aos dados de treino na rede. Esses achados ressaltam a necessidade de métodos que exijam menos dados para treinamento eficiente, proporcionando insights valiosos para aprimorar sistemas de monitoramento de estacionamentos.

LISTA DE FIGURAS

Figura 1. Exemplo de arquitetura de uma Rede Neural Convolucional.....	0
Figura 2. Exemplo de aumentos de dados sintéticos	0
Figura 3. Imagens do dataset PKLot	0
Figura 4. PKLot dataset imagens de exemplo	0
Figura 5. Pares positivos e negativos	0
Figura 6. Entrada de nova imagem	0
Figura 7. Referências fixas de cada classe.....	0
Figura 8. Pares formados pela rede.....	0
Figura 9. Exemplo de rede siamesa simples.....	0
Figura 10. Arquitetura do modelo	0
Figura 11. Entrada de nova imagem teste.....	0
Figura 12. Pares de teste formados.....	0

LISTA DE TABELAS

Tabela 1. Resultados da rede treinada no dataset UFPR05	0
Tabela 2. Resultados da rede treinada no dataset UFPR04.....	0
Tabela 3. Resultados da rede treinada no dataset PUC	0

1. INTRODUÇÃO

1.1. PROBLEMATIZAÇÃO

A busca por vagas de estacionamento em centros urbanos de alta densidade tem se mostrado um problema recorrente no dia a dia da população. Nesse sentido, cada vez mais se vê necessário inovar os ambientes e sistemas urbanos, a fim de auxiliar os motoristas a estacionarem seus carros de forma rápida e conveniente.

Nesse contexto, aproveitando o avanço da inteligência artificial, várias ideias e soluções de cidade inteligente têm sido propostas[1]. Nesse sentido, as abordagens baseadas em imagem para classificação de vagas de estacionamento utilizando técnicas de aprendizado de máquina têm se mostrado uma excelente alternativa, devido à sua vantagem de custo e infraestrutura em relação a outros modelos[2], ademais de sua ampla possibilidade de modelos e abordagens. A escassez de dados, particularmente de imagens, é uma problemática recorrente em modelos de inteligência artificial, representando um desafio significativo durante o processo de treinamento do modelo.

1.2. ESTADO DA ARTE

Na análise do estado da arte, dois principais trabalhos, [3] - [4], foram tomados como base da pesquisa, os quais estão resumidos abaixo.

O artigo [3] busca descobrir os desafios na criação de um modelo global para a classificação de vagas de estacionamento, ou seja, classificar se as vagas de estacionamento estão livres ou ocupadas em um determinado estacionamento alvo sem a necessidade de amostras de treinamento rotuladas do estacionamento alvo. Para este fim, foram definidas duas perguntas:

P1) Quão precisos são os modelos de aprendizagem profunda existentes quando aplicados a um cenário de conjunto de dados cruzados?

P2) Em relação às diferentes arquiteturas e estratégias de conjunto, qual estrutura é a mais adequada para cenários de conjuntos de dados cruzados? Após diversos experimentos testando diferentes conjuntos de dados, modelos e arquiteturas, incluindo estratégias de fusão e métodos de conjunto, foi atingido um resultado médio de 95%. Porém, a arquitetura MobileNetV3 obteve os melhores resultados, mostrando-se a mais adequada, juntamente com o conjunto de dados PKLoT, que mostrou melhor diversidade de dados para treinar uma estrutura generalista.

Já o segundo estudo base [4] tem como foco investigar e elucidar qual tipo de anotação (polígonos, caixas delimitadoras e quadrados de tamanho fixo) oferece o melhor

compromisso entre a precisão da anotação artesanal e o desempenho do modelo, juntamente com o número de amostras anotadas necessárias para ajustar um modelo pré-treinado em um estacionamento alvo. Para esse fim, foram definidas duas perguntas:

P1) Uma abordagem relativamente barata para demarcar as posições dos lugares de estacionamento, como caixas delimitadoras, pode fornecer bons resultados? P2)

Considerando um modelo pré-treinado, quantas amostras rotuladas do estacionamento alvo são necessárias para ajustar o modelo?

Ao final dos experimentos, concluiu-se que a anotação de caixa delimitadora representa o melhor compromisso entre o esforço de anotação e a precisão, levando a resultados semelhantes aos do polígono, e que com apenas um pequeno conjunto composto por 1.000 amostras, podemos ajustar satisfatoriamente um modelo CNN pré-treinado para um novo domínio alvo, podendo atingir até 97% de desempenho.

1.3. MOTIVAÇÃO AO PROJETO

A aquisição e preparação de dados, especialmente imagens, juntamente com sua classificação e anotação para o treinamento de modelos em projetos de inteligência artificial, representam desafios recorrentes. Esses desafios surgem devido à sua complexidade, custo, esforço ou escassez, o que pode dificultar a viabilidade prática desses conjuntos de dados para implantações no mundo real. Não é de surpreender que diversas abordagens no estado da arte estejam em busca de soluções e métodos para superar essas dificuldades.

2. OBJETIVOS

Esta pesquisa tem como objetivo investigar e analisar os resultados da classificação de vagas de estacionamento - determinando se uma vaga está livre ou ocupada - utilizando um conjunto restrito de amostras rotuladas. Para isso, foram testadas abordagens, métodos de redes Siamesas para Few-shot, além de técnicas de aumento de dados sintéticos. Esse enfoque se deve ao fato de que um dos principais obstáculos enfrentados em projetos de inteligência artificial é a necessidade de um grande volume de amostras para treinamento do modelo, como discutido anteriormente. Para nossos experimentos, utilizamos o conjunto de dados PKLot [5], reconhecido como um dos maiores e mais abrangentes disponíveis para a classificação de estacionamentos no estado da arte. Como evidenciado em trabalhos anteriores [3], o PKLot destaca-se pela sua diversidade de dados, o que o torna ideal para treinar uma estrutura generalista, por isso será utilizado neste trabalho.

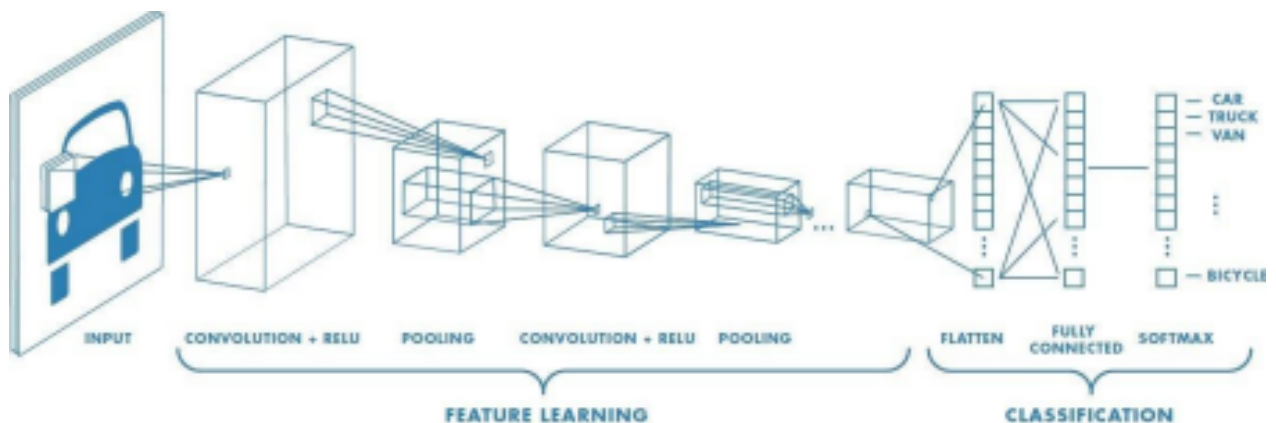


Figura 1. Exemplo de arquitetura de uma Rede Neural Convolucional.

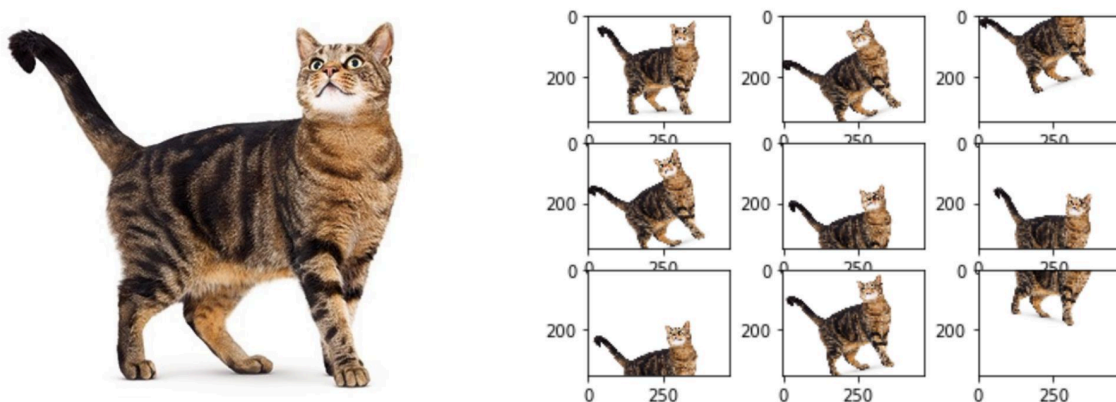


Figura 2. Exemplo de aumento de dados sintéticos. Aonde uma única imagem pode se tornar nove imagens diferentes.

3. MATERIAIS E MÉTODO

Os materiais empregados nesta pesquisa consistem em um computador de alta performance equipado com uma GPU robusta, juntamente com um sistema operacional GNU/Linux e acesso à internet. Além disso, incorporamos as amostras de dados provenientes da PKLot[5], cujo papel fundamental é destacado em nossas análises. Quanto aos métodos adotados, estes abrangem uma abordagem meticulosa que não se limita à análise dos artigos disponíveis no estado da arte, sejam eles recomendados pelo orientador ou selecionados pelo aluno. Também é explorado outros recursos, tanto aqueles fornecidos pelo orientador quanto os encontrados por meio de pesquisas na internet. Esses materiais são estudados e testados, englobando tanto os modelos propostos quanto os resultados obtidos. O objetivo é alcançar insights e promover o contínuo aprimoramento do trabalho desenvolvido.

3.1. DATASET

No presente trabalho, utilizamos o dataset PKLot [5], que é amplamente reconhecido por sua abrangência e complexidade. Este dataset compreende três cenários distintos: dois ângulos de câmera para o estacionamento da UFPR e um ângulo de câmera para o estacionamento da PUCPR (conforme ilustrado na Figura 3). As imagens para cada um desses cenários foram coletadas ao longo de aproximadamente trinta dias, com um intervalo de cinco minutos entre as capturas. Essa metodologia permitiu a coleta de uma ampla variedade de condições, incluindo variações nas posições das câmeras, mudanças no fundo, diferentes condições climáticas e a presença de obstruções.

O dataset PKLot [5] destaca-se como um dos maiores e mais abrangentes conjuntos de dados disponíveis para a classificação de estacionamentos no estado da arte. Com imagens capturadas ao longo de cerca de três meses, com um intervalo consistente de cinco minutos entre cada imagem, o dataset fornece um total impressionante de 12.417 vagas de estacionamento e aproximadamente 700.000 amostras anotadas. Essas amostras estão distribuídas em três cenários distintos, denominados UFPR04, UFPR05 e PUCPR, cada um oferecendo uma visão única e detalhada dos diferentes ambientes de estacionamento. Exemplos das imagens incluídas no dataset PKLot são apresentados na Figura 4, ilustrando a diversidade e a riqueza dos dados coletados.



Figura 3. Imagens do dataset PKLot. Este compreende 3 cenários nomeados a) UFPR04, b) UFPR05 e c) PUCPR.



Figura 4. PKLot dataset imagens de exemplo.

3.2. PREPARAÇÃO DOS DADOS

As classes em um dataset referem-se às diferentes categorias ou rótulos que as imagens podem ter. Em nosso dataset (PKLot), as classes podem ser "vaga ocupada" e "vaga vazia". Cada imagem no dataset pertence a uma dessas classes. Redes siamesas, em particular, apresentam uma abordagem única, onde a necessidade de pares de imagens, tanto positivos quanto negativos é fundamental. Pares positivos consistem em imagens da mesma classe (por exemplo, duas imagens de vagas ocupadas), enquanto pares negativos são formados por imagens de classes diferentes (por exemplo, uma imagem de vaga ocupada e uma imagem de vaga vazia).



Figura 5. Pares positivos e negativos.

Ao comparar diretamente pares de imagens, a rede neural aprende a discernir as diferenças e semelhanças de forma mais eficaz, o que pode resultar em um desempenho superior em tarefas de classificação com dados limitados. No entanto, para garantir a eficácia do modelo, é crucial equilibrar o número de exemplos em cada classe. Esse equilíbrio não só previne o viés durante o treinamento, mas também assegura que a rede desenvolva uma capacidade robusta de generalização.

Para a formação dos pares positivos e negativos, foram estabelecidas três imagens de referência fixas de cada classe, gerando seis imagens fixas. Isso resulta na formação de seis pares para cada nova imagem do dataset de treino, sempre sendo três pares positivos e três pares negativos. Um exemplo de como funciona este fluxo pode ser visto a seguir, onde na Figura 6, apresenta-se um exemplo de uma nova imagem de entrada para formação de pares com as referências fixas, conforme ilustrado na Figura 7, e o resultado desta formação de pares na Figura 8. A definição de referências fixas visa auxiliar o modelo a treinar de forma mais confiável e estável, além de aumentar a quantidade de dados, visto que cada imagem gera seis pares. Como o modelo calcula a distância entre as imagens, a utilização de pares aleatórios poderia comprometer essa estabilidade, resultando em perda de credibilidade dos resultados.

Imagem do dataset de treino



Referências fixas:

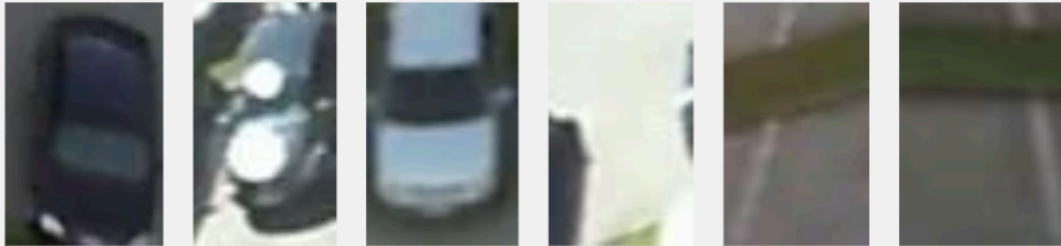


Figura 6. Entrada de nova imagem.

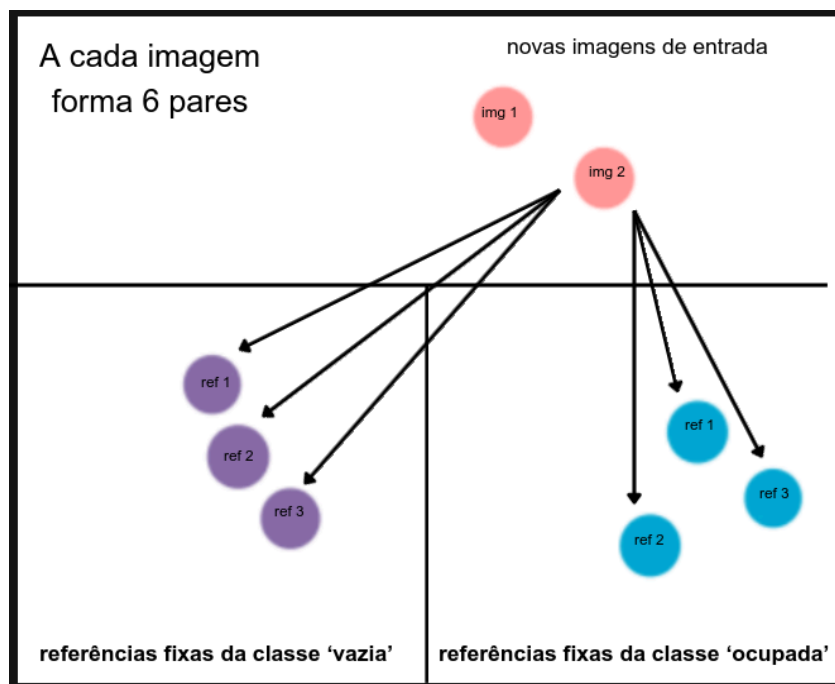


Figura 7. Referências fixas de cada classe.

Exemplo dataset de Treino

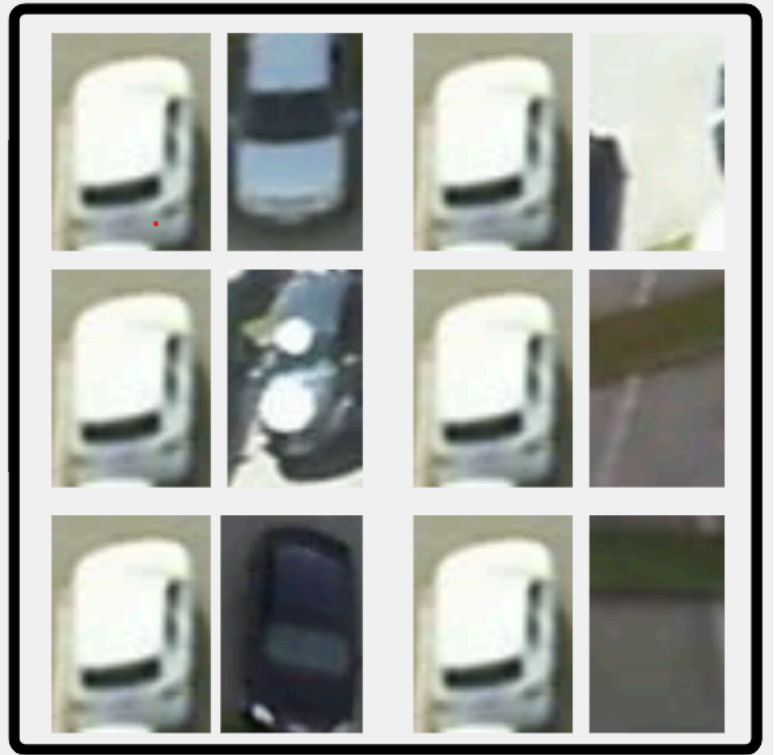


Figura 8. Pares formados pela rede.

Para maximizar a eficiência do aprendizado no dataset PKLot, que possui as classes "vaga ocupada" e "vaga vazia," foram utilizadas redes siamesas, formando pares citados anteriormente que são passados à rede. Como sempre são formados seis pares para cada imagem do dataset de treino, caso o dataset de treino possua 1.000 imagens, serão formados 6.000 pares. Essa estratégia não só aumenta a quantidade de dados, mas também previne vieses e assegura uma capacidade robusta de generalização, crucial para o desempenho eficaz da rede neural em tarefas de classificação com dados limitados.

3.3. MODELO

Neste trabalho foi utilizado o modelo de redes siamesas, as quais são uma arquitetura particular de redes neurais projetadas para aprender a distinguir entre duas entradas, usualmente imagens, ao invés de simplesmente classificá-las em uma única categoria. Essas redes foram inicialmente propostas por Bromley e LeCun (1993) em seu trabalho "Signature Verification using a Siamese Time Delay Neural Network"[6]. Desde então, têm sido amplamente utilizadas em diversas aplicações, como verificação de assinaturas, reconhecimento facial e detecção de anomalias entre outras diversas

aplicações onde é difícil conseguir dados.

Uma rede siamesa consiste em duas sub-redes idênticas que compartilham os mesmos parâmetros e pesos. Cada sub-rede processa uma das entradas e extrai um conjunto de características. As saídas dessas sub-redes são então comparadas utilizando uma métrica de distância, como a distância Euclidiana, para determinar o grau de similaridade entre as entradas, exemplo abaixo na figura 9 . O aprendizado ocorre ajustando os pesos das sub-redes de forma a minimizar a diferença entre pares de entradas semelhantes (pares positivos) e maximizar a diferença entre pares de entradas diferentes (pares negativos).

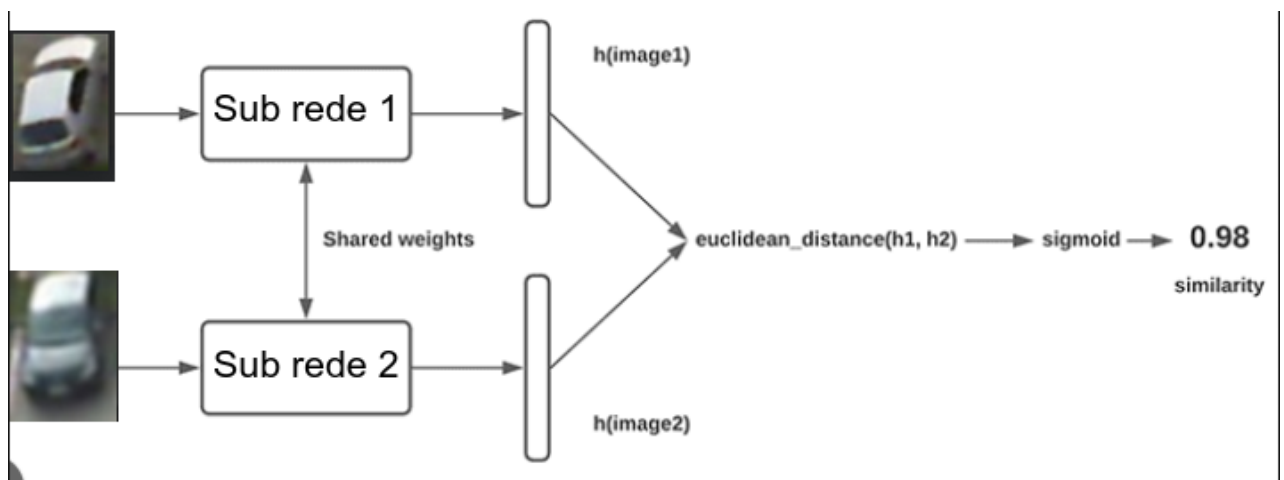


Figura 9. Exemplo de rede siamesa simples.

Este modelo foi projetado para maximizar a eficiência do aprendizado com dados limitados, utilizando uma estratégia de comparação direta de pares de imagens, o que assegura um treinamento mais confiável e estável, além de aumentar a robustez do modelo em tarefas de classificação. A seguir, explicarei a arquitetura detalhada da nossa rede siamesa, baseada no artigo [7].

A rede recebe imagens de entrada com dimensões de 105x105 pixels. A primeira camada convolucional aplica 64 filtros de tamanho 10x10, seguidos pela função de ativação ReLU. Em seguida, uma operação de max-pooling com tamanho de 2x2 é realizada. A segunda camada convolucional utiliza 128 filtros de tamanho 7x7, também seguidos pela função de ativação ReLU, e novamente é aplicada uma operação de max-pooling com tamanho de 2x2.

A terceira camada convolucional aplica 128 filtros de tamanho 4x4, seguidos pela função de ativação ReLU, e é acompanhada por uma terceira operação de max-pooling com tamanho de 2x2. A quarta camada convolucional utiliza 256 filtros de tamanho 4x4, seguidos pela função de ativação ReLU.

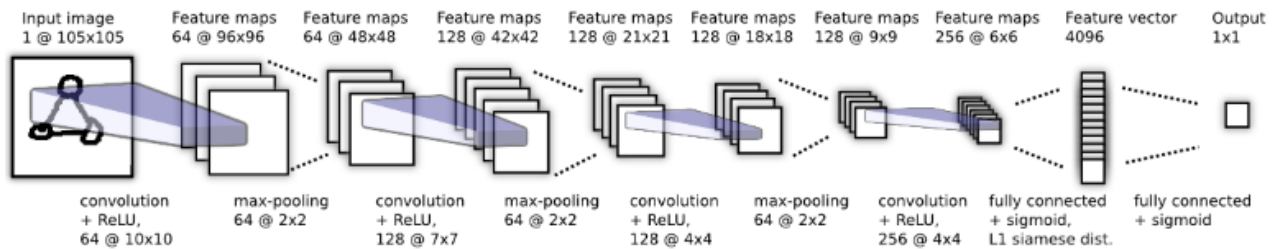


Figura 10. Melhor arquitetura convolucional selecionada para a tarefa de verificação. A rede siamesa gêmea não é representada, mas se une imediatamente após a camada totalmente conectada de 4096 unidades, onde a distância L1 entre os vetores é calculada componente por componente.

Após as camadas convolucionais, as saídas são achatadas em um vetor de 4096 características, seguido por uma função de ativação sigmoidal. Em seguida, a camada de comparação calcula a distância L1 componente a componente entre os vetores de características das duas sub-redes. Finalmente, uma camada totalmente conectada com função de ativação sigmoidal é usada para discriminar entre as classes "vaga ocupada" e "vaga vazia".

3.4. TREINAMENTO

Para uma avaliação completa do desempenho e das comparações de precisão, realizei o treinamento da rede em diferentes estacionamentos dentro do conjunto de dados PKLot. Especificamente, treinei as redes nos sub-datasets UFPR04, UFPR05 e PUC. Para cada um desses sub-datasets, conduzir experimentos com três diferentes volumes de dados de treinamento: 2.000, 5.000 e 10.000 imagens. O objetivo foi analisar a relação entre o aumento do volume de dados e o desempenho da rede.

A análise também incluiu o tempo de treinamento e a convergência dos modelos, observando como o aumento do volume de dados afetava a eficiência do treinamento. Esse estudo abrangente proporciona uma visão clara de como a rede se comporta em

diferentes cenários e com diferentes quantidades de dados, oferecendo insights valiosos para futuras melhorias e aplicações práticas em sistemas de monitoramento de estacionamentos.

3.5. INFERÊNCIA

Para a realização dos testes de avaliação, foi utilizado um método similar ao de treinamento, onde pares de imagens são passados pela rede. As características das imagens são extraídas e a distância euclidiana entre elas é calculada para determinar se pertencem à mesma classe ou não. No entanto, a preparação do dataset de teste é feita de maneira um pouco diferente.

Primeiramente, uma imagem fixa é escolhida do dataset de teste. Esta imagem servirá como referência para formar pares com todas as outras imagens do dataset de teste. Cada imagem do dataset de teste é então pareada com a imagem fixa, criando um conjunto de pares de teste. Por exemplo, se o dataset de teste contém 1.000 imagens, serão formados 1.000 pares, todos contendo a imagem fixa como um dos elementos do par. Isto está representado na Figura 11, onde é possível ver a imagem fixa e as imagens do dataset de teste (representadas apenas quatro delas na imagem), e na Figura 12 pode-se ver um exemplo dos pares de teste formados.

Esses pares de imagens são então passados pela rede siamesa. As duas sub-redes idênticas processam cada imagem do par separadamente, extraindo um conjunto de características para cada uma. As saídas das sub-redes, que são vetores de características, são então comparadas. Esta comparação é feita utilizando a métrica de distância euclidiana, que mede a similaridade entre os dois vetores de características. Finalmente, a precisão da rede é avaliada contando quantas vezes a classificação está correta. Esta avaliação é feita mesmo quando as imagens do par pertencem a sub-datasets diferentes do PKLot, testando a capacidade de generalização da rede.

Imagem fixa



Imagens aleatorias do dataset de teste:



Figura 11. Entrada de nova imagem teste.

Exemplo dataset de test

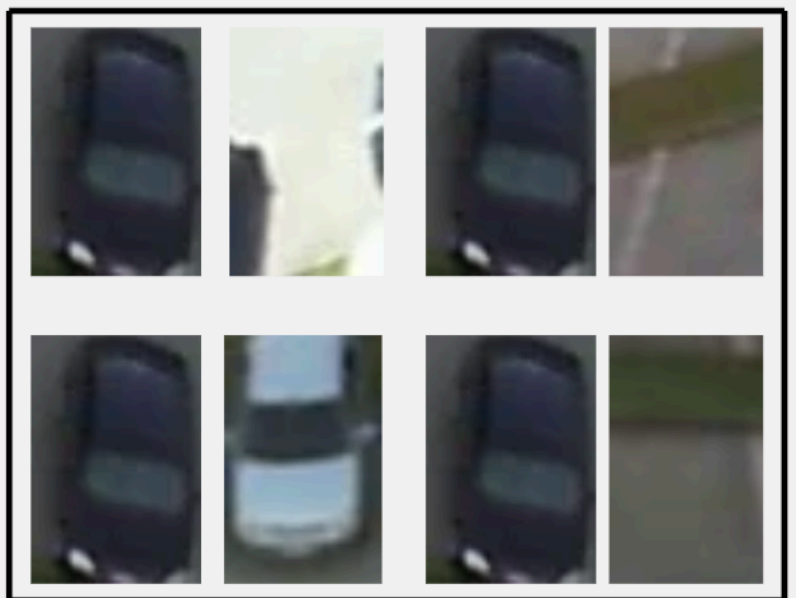


Figura 12. Pares de teste formados.

4. RESULTADOS

Os resultados da pesquisa foram organizados em três tabelas principais, cada uma representando o desempenho da rede em um sub-dataset específico do PKLot: UFPR04, UFPR05 e PUC. Cada tabela contém resultados de testes com diferentes tamanhos de conjuntos de dados de treino (2.000, 5.000 e 10.000 imagens) e outras variações.

A Tabela 1 apresenta os resultados dos testes, onde a rede foi treinada utilizando o sub-dataset UFPR04. Os testes foram realizados com três volumes de dados de treinamento distintos: 2.000, 5.000 e 10.000 imagens, nos três sub-datasets (UFPR04, UFPR05 e PUC). Cada cor na tabela representa os testes realizados contra um sub-dataset específico. Por exemplo, na cor amarela temos os resultados de todos os testes realizados para analisar o desempenho da rede treinada com 2.000, 5.000 e 10.000 imagens, ao classificar imagens do sub-dataset PUC. Esta abordagem permite analisar como a rede lida com novos dados não vistos antes, avaliando sua capacidade de generalização e precisão na classificação.

REDE TREINADA NO CONJUNTO DE DADOS UFPR04					
REDE TREINADA N* IMAGENS	Estacionamento testado	Num. pares para teste	Acuracia	Brier Score	Precisão
2.000	UFPR05	128	0.9375	0.625	0.9118
	UFPR04	128	0.9844	0.0156	0.9844
	PUC	128	0.9531	0.0469	0.9531
5.000	UFPR05	500	0.9140	0.0860	0.9259
	UFPR04	500	0.9900	0.0100	0.9842
	PUC	500	0.9760	0.0240	0.9612
10.000	UFPR05	1.000	0.9430	0.0570	0.9955
	UFPR04	1.000	0.9930	0.0070	0.9901
	PUC	1.000	0.9640	0.0360	0.9895

Tabela 1. Apresentação dos testes realizados na rede treinada com o sub-dataset UFPR04, com o objetivo de elucidar qual a melhor opção entre o números de dados e acurácia.

Como visto nesta Tabela, são analisados 3 parâmetros para avaliar o desempenho, os quais irei explicar agora (o número de imagens a esquerda se refere às imagens do dataset, as quais formam pares com as referências fixas, logo o número de pares de treinamento da rede é (número de imagens * 6)). Acurácia analisa a proporção de previsões corretas feitas pelo modelo em relação ao total de previsões, resumidamente mede quantas vezes o modelo acertou em todas as suas previsões. Já a precisão analisa a proporção de previsões positivas corretas em relação ao total de previsões positivas feitas pelo modelo, ou seja, mede quantas das previsões positivas do modelo estavam realmente corretas. e Por último a Brier Score, na qual avalia a precisão das previsões probabilísticas do modelo, medindo o quão perto as previsões estão dos resultados reais. Medindo o quão confiável o modelo é quando diz a probabilidade de algo acontecer, o quanto menor melhor, pois indica que o modelo está classificando bem.

REDE TREINADA NO CONJUNTO DE DADOS UFPR05					
REDE TREINADA	Estacionamento	Num. pares	Acurácia	Brier Score	Precisão
N* IMAGENS	testado	para teste			
2.000	UFPR05	128	0.9688	0.0312	0.9545
	UFPR04	128	0.9375	0.0625	0.9242
	PUC	128	0.8984	0.1016	0.9636
5.000	UFPR05	500	0.9780	0.0220	0.9703
	UFPR04	500	0.8740	0.1260	0.8006
	PUC	500	0.8900	0.1100	0.8197
10.000	UFPR05	1.000	0.9900	0.0100	0.9880
	UFPR04	1.000	0.9380	0.0620	0.9228
	PUC	1.000	0.8060	0.1940	0.7217

Tabela 2. Apresentação dos testes realizados na rede treinada com o sub-dataset UFPR05, com o objetivo de elucidar qual a melhor opção entre o números de dados e acurácia.

REDE TREINADA NO CONJUNTO DE DADOS PUC					
REDE TREINADA	Estacionamento	Num. pares	Acuracia	Brier Score	Precisão
N* IMAGENS	testado	para teste			
2.000	UFPR05	128	0.8594	0.7805	0.1406
	UFPR04	128	0.9297	0.0703	0.9825
	PUC	128	0.9922	0.0078	0.9900
5.000	UFPR05	500	0.9040	0.0960	0.9120
	UFPR04	500	0.9520	0.0480	0.9871
	PUC	500	0.9940	0.0060	0.9708
10.000	UFPR05	1.000	0.9250	0.0750	0.9953
	UFPR04	1.000	0.9710	0.0290	0.9555
	PUC	1.000	0.9990	0.0010	0.9805

Tabela 3. Apresentação dos testes realizados na rede treinada com o sub-dataset PUC, com o objetivo de elucidar qual a melhor opção entre o números de dados e acurácia.

5. DISCUSSÃO

Os resultados apresentados nas Tabelas 1, 2 e 3 mostram o desempenho da rede treinada nos sub-datasets UFPR04, UFPR05 e PUC. Cada tabela destaca a acurácia, precisão e Brier Score para diferentes tamanhos de conjuntos de dados de treino (2.000, 5.000 e 10.000 imagens).

A rede treinada com UFPR04 apresenta excelente performance ao classificar imagens do próprio sub-dataset, com acurácia e precisão acima de 0.98. Testes com o sub-dataset PUC mostram uma leve queda na performance, mas ainda se mantêm altos. Entretanto, ao testar com UFPR05, a rede enfrenta mais desafios, com quedas mais significativas, especialmente com menores volumes de treino. Já a rede treinada com PUC mostra uma performance excepcional ao classificar imagens do próprio sub-dataset. No entanto, ao ser testada com UFPR04, há uma pequena diminuição na performance. Já com UFPR05, a rede enfrenta dificuldades, principalmente com 2.000 imagens de treino, sugerindo desafios de generalização para imagens com características diferentes. E por fim, a rede treinada com UFPR05 apresenta robustez ao classificar imagens do próprio sub-dataset. Contudo, ao ser testada com UFPR04, a performance é inferior, indicando desafios na generalização. Com PUC, os resultados são mistos, mostrando dificuldades de generalização para conjuntos de dados diferentes.

A capacidade de generalização da rede é influenciada pelo sub-dataset de treinamento e pelo número de imagens utilizadas. Aumentar o número de imagens de treino geralmente melhora a performance, mas a extensão dessa melhoria varia. Um possível ajuste significativo com 10.000 imagens pode levar ao sobreajuste devido à formação de múltiplos pares, o que pode ser considerado para ajustes futuros sem afirmar conclusões definitiva

6. CONCLUSÃO

Esta pesquisa teve como objetivo avaliar o desempenho de uma rede treinada utilizando diferentes sub-datasets do PKLot (UFPR04, UFPR05 e PUC) e analisar sua capacidade de generalização e precisão na classificação de imagens de estacionamento. Os resultados indicam que a performance da rede é altamente dependente do sub-dataset de treinamento e do volume de dados utilizados para o treinamento. Por isso é importante estar sempre avaliando métodos que precisem de menos dados.

A rede treinada com UFPR04 apresentou alta acurácia e precisão ao classificar imagens do próprio sub-dataset, com leve queda de desempenho ao ser testada com PUC e maior variabilidade com UFPR05, especialmente em volumes menores de treino. A rede treinada com PUC manteve desempenho excepcional no próprio sub-dataset. A rede treinada com UFPR05 foi robusta ao classificar imagens do próprio sub-dataset, mas apresentou performance inferior ao ser testada com UFPR04 e PUC, indicando que a variabilidade entre os sub-datasets influencia significativamente a capacidade de generalização da rede.

A capacidade de generalização da rede é fortemente influenciada pelo dataset de treinamento e pelo número de imagens utilizadas. Aumentar o volume de dados de treino geralmente melhora a performance, mas a extensão dessa melhoria varia. Um possível ajuste significativo com 10.000 imagens pode levar ao sobreajuste devido à formação de múltiplos pares, o que deve ser considerado para ajustes futuros

REFERÊNCIAS

- [1] S. Nosratabadi, A. Mosavi, R. Keivani, S. Ardabili, and F. Aram, "State of the art survey of deep learning and machine learning models for smart cities and urban sustainability," in International Conference on Global Research and Education. Springer, 2019, pp. 228–238
- [2] V. Paidi, H. Fleyeh, J. Hakansson, and R. G. Nyberg, "Smart parking ° sensors, technologies and applications for open parking lots: a review," IET Intelligent Transport Systems, vol. 12, no. 8, pp. 735–741, 2018
- [3] Hochuli, A.G., Barddal, J.P., Palhano, G.C., Mendes, L.M., & Almeida, P.R. (2023). Deep Single Models vs. Ensembles: Insights for a Fast Deployment of Parking Monitoring Systems. ArXiv, abs/2309.16495.
- [4] A. G. Hochuli, A. S. Britto, P. R. L. de Almeida, W. B. S. Alves and F. M. C. Cagni, "Evaluation of Different Annotation Strategies for Deployment of Parking Spaces Classification Systems," 2022 International Joint Conference on Neural Networks (IJCNN), Padua, Italy, 2022, pp. 1-8, doi: 10.1109/IJCNN55064.2022.9892783.
- [5] P. R. Almeida, L. S. Oliveira, A. S. Britto Jr, E. J. Silva Jr, and A. L. Koerich, "Pklot – um conjunto de dados robusto para classificação de estacionamentos", Especialista Sistemas com Aplicações, vol. 42, não. 4937–4949, 2015.
- [6] Bromley, J., Guyon, I., LeCun, Y., Sackinger, E., & Shah, R. (1993). Signature Verification using a Siamese Time Delay Neural Network. In Advances in Neural Information Processing Systems (pp. 737-744).
- [7] Koch, G., Zemel, R. & Salakhutdinov, R. (2015). Siamese Neural Networks for One-shot Image Recogniti

