

Trabalho Final do Bootcamp

Definição do problema, compreensão da área de estudo, criação de dataset, proposta de novos métodos, fine tuning e avaliação de modelo deep learning.

Breno da Silva Lima

Bacharelado em Ciência da Computação

Universidade Tecnológica Federal do Paraná

Ponta Grossa, Paraná

brenolima@alunos.utfpr.edu.br

Resumo—Como trabalho final do Bootcamp de Machine Learning, este projeto visa consolidar os conhecimentos adquiridos ao longo das tarefas em um projeto próprio, explorando toda a pipeline de um problema de aprendizado de máquina. No artigo, identificou-se um problema real, onde a partir dele, uma solução foi proposta: a aplicação de OCR em hidrômetros para extrair a leitura de consumo de água. A área de estudo foi investigada em profundidade, e novos métodos foram propostos, resultando em melhorias significativas no modelo usado.

Palavras-Chave—OCR, hidrômetros, visão computacional.

I. INTRODUÇÃO

Atualmente na cidade de São Paulo as medições de consumo de água são realizadas pela Sabesp, a qual monitora o consumo de casas e condomínios. Desde o ano de 2021, assim como é para as casas, obrigatoriamente as leituras dos moradores de condomínios passaram ser realizadas individualmente, ou seja, cada morador deve receber sua conta, onde o consumo de outros moradores não influenciam em seu valor.

Devido ao grande número de pessoas habitando a cidade, conseguir mensurar o consumo de cada pessoa individualmente dentro dos condomínios pode se tornar uma atividade desafiadora. Por essa razão, a Sabesp passou a responsabilizar os próprios condomínios pela organização e leitura dos consumos, onde os mesmos geram relatórios mensais.

Esse cenário gera uma demanda grande de condomínios que procuram terceirizar esse trabalho, fazendo com que diferentes empresas sejam criadas propondo soluções diversas para conseguir realizar tais serviços. Hoje em dia existem soluções baseadas em hidrômetros inteligentes, os quais medem e enviam os dados para um servidor por si só. Há também soluções baseadas em protocolos bluetooth, fazendo com que os medidores armazenem as informações dos consumos, e uma vez que uma pessoa passe por perto com o celular, os dados sejam enviados e armazenados.

Contudo, ainda existem muitos cenários que não se adaptaram totalmente a essas novas tecnologias, possuindo hidrômetros do modelo padrão e resultando assim na única medição possível sendo a de uma pessoa realizá-la. O presente trabalho explora esse tema tendo consultado e acompanhado o caso de um funcionário que desempenha tal função, onde alguns problemas relacionados a essa abordagem foram encontrados.

Em cenários reais, a pessoa a medir esses hidrômetros opta pela realização da fotografia, onde em um mês de trabalho há por volta de seis mil fotos tiradas. Esse cenário gera uma carga de trabalho muito elevada, visto que a primeira parte do trabalho envolve tirar todas as fotos enquanto a pessoa está presente no condomínio, e a segunda rever foto a foto enquanto manualmente anota em alguma base dados o valor da leitura. A segunda parte do trabalho é levada para casa ou para o escritório após as leituras do dia, consumindo assim ainda mais horas de trabalho do funcionário.

Com o avanço de modelos de visão computacional nos dias atuais, há cada vez mais possibilidades sendo exploradas e analisada dentre os mais diversos tópicos, mas em específico o campo da Optical Character Recognition (OCR), que se propõe tratar da extração e interpretação de texto em imagens ou documentos. A OCR busca identificar padrões nos pixels da imagem e reconhecendo caracteres individuais, para então convertê-los em texto que pode ser pesquisável ou editável.

II. OBJETIVO

O presente trabalho explora a OCR aplicado à extração dos dígitos presentes nas fotografias de hidrômetros, conseguindo assim obter o valor da leitura de consumo, eliminando grande parte do trabalho da anotação manual. O objetivo é analisar e compreender o campo da OCR voltado a cenários semelhantes ao estudado, entendendo modelos e técnicas utilizadas, assim como propor novos métodos e ao fim treinar um modelo para realizar previsões.

III. MONTAGEM DO DATASET

Presentes nas fotos tiradas pelos funcionários estão imagens retiradas dos blocos, de painéis digitais, fotos borradas ou totalmente pretas, as quais não oferecem nenhuma informação útil para o problema, e então, foram descartadas. Devido ao grande volume de fotos, e o trabalho de rotulação das mesmas, uma amostragem pequena foi escolhida, a fim de validar métodos e equiparar o modelo escolhido. Sendo assim, 200 fotos selecionadas para compor o dataset.

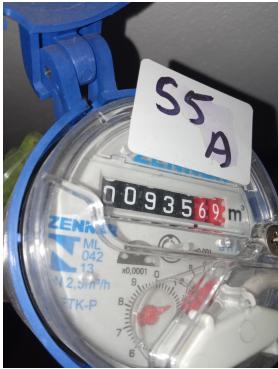
As imagens originais tiradas pelo celular possuem tamanho 3120x4160 para as imagens verticais, e 4160x3120 para as horizontais. Para que não seja usado uma grande quantidade

de recursos tanto para o armazenamento quanto para o treinamento da rede neural, as imagens foram redimensionadas para 640x640.

Por padrão as imagens são nomeadas com base nas datas que foram tiradas, fazendo com elas sejam ordenadas das mais até as menos recentes. Dessa forma, os dados ficam padronizados pelos condomínios e seus tipos de hidrômetros usados. Estarão presentes nos dados grandes sequências de um mesmo tipo de hidrômetros, podendo fazer com que, ao realizar um corte para dados de treino e teste, o modelo não tenha acesso a todos os tipos de hidrômetros no seu treinamento. Sendo assim, para alimentar essas imagens ao modelo, as imagens foram randomizadas e renomeadas, para que esteja presente nos dados de treino uma grande variedade de casos e assim aumentando a capacidade de generalização do modelo.

Para compor os dados de treino, validação e teste, as fotos foram distribuídas na seguinte ordem: 120 imagens para treino, 50 para validação e 30 para teste.

Na Figura 1 são mostrados alguns exemplos de fotos contidas dentro do dataset.



(a) Exemplo 1



(b) Exemplo 2

Figura 1: Fotos do dataset

IV. ANÁLISE DAS FOTOS

Através da observação das fotos contidas na base de dados, algumas características do problema podem ser notadas. Visto o funcionamento do hidrômetro de rotacionar o dígito conforme está havendo consumo da água, muitos casos podem possuir os dígitos sendo mostrados pela metade. Além disso, condições de luminosidade, texturas e padrões mudam para cada condomínio.

Percebe-se também a repetição das formas geométricas presentes no problema. Enquanto o hidrômetro apresenta uma forma circular, os dados dentro dele estão dispostos em uma forma de retângulo. Essas informações podem ser de utilidade para o modelo, visto que a grande maioria das fotos estão sempre rotacionadas.

Vale ressaltar que existem casos onde não se tem livre acesso aos hidrômetros, apenas o celular com a câmera consegue acessar, como forros, por exemplo. Logo, em uma possível aplicação real, a solução de limitar a área de funcionamento do

algoritmo para apenas um pequeno retângulo na câmera, e usar um modelo apenas para identificar dígitos não é totalmente possível, já que o usuário não poderá mirar o celular em todos os casos.

A. Fotos diversas

Junto às imagens mais comuns de hidrômetros presentes no dataset, existem também casos não tão ideais, com condições apresentadas na Figura 2.



(a) Exemplo 1



(b) Exemplo 2



(c) Exemplo 3



(d) Exemplo 4

Figura 2: Fotos

Apesar de serem casos extremos, eles demonstram que ruídos e alterações devem estar presentes nos dados, tanto naturalmente, quanto possivelmente artificialmente, fazendo com que o modelo possa generalizar cada vez mais para casos parecidos com esses.

V. TRABALHOS RELACIONADOS

Visto que boa parte das soluções relacionada à visão computacional e redes neurais possuem um caráter experimental no que se diz respeito a sua arquitetura, parâmetros e hiperparâmetros, uma pesquisa literária foi feita a fim de compreender abordagens semelhantes ao objetivo deste trabalho, para ao fim explorar quais modelos são mais utilizados e quais outros processamentos podem ser aplicados.

Primeiro, definiu-se uma string de busca S1, responsável por representar o assunto buscado. Em seguida, foi pesquisado

artigos dentre as bases de artigo da IEEE e WoS. Nenhum filtro foi adicionado às pesquisas.

S1: ("computer vision"OR "machine learning"OR "deep learning"OR "neural network"OR "deep neural network"OR "CNN"OR "convolutional neural network"OR "digit recognition") AND ("hydrometer"OR "water meter"OR "water reading"OR "water consumption"OR "water monitoring")

Foram retornados 43 artigos científicos, dos quais apenas 6 se alinhavam com a proposta da pesquisa, sendo esses:

- Application of Deep Residual Neural Network to Water Meter Reading Recognition [1];
- Research on water meter reading recognition based on deep learning [2];
- Automatic wheel-type water meter digit reading recognition based on deep learning [3];
- Automatic Water Meter Reading Development Based On CNN and LoRaWAN [4];
- Key point localization and recurrent neural network based water meter reading recognition [5];
- Evaluation of Recognition of Water-meter Digits with ApplicationPrograms, APIs, and Machine Learning Algorithms [6].

Após a seleção e leitura dos artigos, mais informações sobre a área de estudo foram anotadas. A Tabela 1 mostra de forma resumida como cada artigo aborda o tema, enquanto a Tabela 2 mostra informações gerais dos resultados de cada um. Também foi possível entender os seguintes tópicos.

A. Modelos One-Stage e Two-Stages

Existem dois tipos de modelos implementados para a proposta desse trabalho. O primeiro é o de One-Stage, o qual converte o problema de posicionamento do quadro alvo em um problema de regressão, e em seguida realiza a classificação de objetos. Ou seja, não há o uso de propostas de região pré-geradas (caixas delimitadoras de objetos candidatos). Alguns modelos desse tipo é o YOLO e SSD.

Two-Stages se refere a um algoritmo que primeiro gera uma série de quadros candidatos como amostras por um algoritmo e, em seguida, classifica as amostras por meio de uma rede neural convolucional. Algoritmos comuns incluem R-CNN, Fast R-CNN, Faster e etc.

O trade-off presente entre os dois modelo é que enquanto modelos One-Stage são mais rápidos, ele tendem a possuir uma performance pior do que modelos Two-Stages.

B. Distribuição Long Tail

Os hidrômetros vem com a leitura zerada de fábrica, e é necessário muito tempo de consumo até que todos os seus dígitos saiam de zero, por essa razão um comportamento esperado no dataset é uma distribuição Long Tail, na qual os dígitos mais próximo de zero possuem mais instâncias do que os dígitos próximo a nove.

Visto isso, o modelo terá mais recursos para aprender aquelas classes de dígitos mais próximos de zero, enquanto o dígito nove terá uma performance pior de classificação, por exemplo.

Esse problema é abordado por Zhu et al.[4] onde através do uso da Mix Suppression, cada dígito no hidrômetro pode ser alterado trocando ele por outros dígitos da mesma imagem, copiando e colando um dígito no outro. Ou seja, o dataset é balanceado realizando a substituição dos dígitos, onde em uma imagem de hidrômetro poderia possuir o dígito zero, ele é futuramente substituído pelo dígito seis da mesma imagem, por exemplo.

VI. ANALISANDO OS DADOS

Dentro das 200 fotos rotuladas pelo dataset, a distribuição na Figura 3 de dados pode ser encontrada.

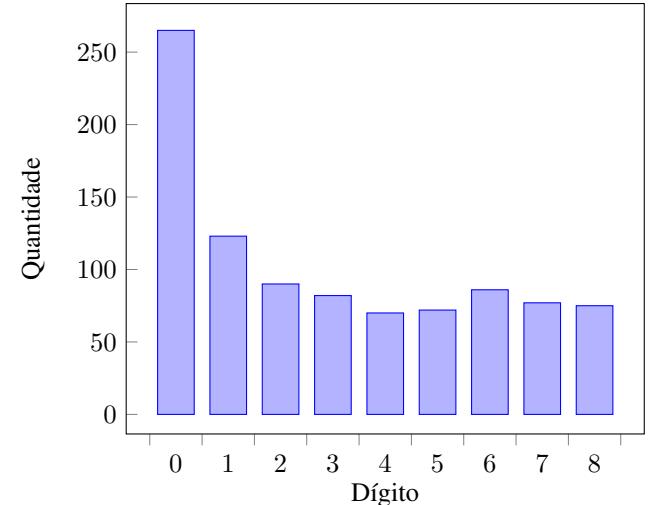


Figura 3: Distribuição dos dígitos

Juntamente ao conceito de Mix Suppression usado por [4], fez-se uso do mesmo agora desenvolvido e aplicado à imagens de hidrômetros rotacionados. Cada dígito presente em uma imagem possui uma chance de ser selecionado para ser trocado, e uma vez isso acontecendo, os dígitos restantes compõem uma espécie de roleta, onde um ponteiro aleatório é rodado selecionando o dígito que será usado na substituição. Ou seja, o número mais presente nos dados tem mais chances de sofrer a troca, e o números menos presente tem mais chance de ser escolhido para substituir. A Figura 4 mostra o uso do método desenvolvido.

Após a aplicação desse método, a distribuição dos números no dataset passou a ser o da Figura 5.

Ainda é possível perceber certa variação na quantidade dos números, porém, ao utilizar do método proposto, conforme o dataset expande de tamanho, é esperado cada vez mais que os dados se mantenham mais平衡ados.

VII. ESCOLHA DO MODELO

Em relação aos resultado obtidos na tabela III, é importante ressaltar que dado o fato de cada artigo trabalhar com seu próprio dataset, e existir uma grande variação de tipos de hidrômetros no mundo, cada um com sua condições de sujeira e iluminação, há viés presente no resultados de cada artigo. Aqueles que modelos performaram melhor, podem possuir um



(a) Exemplo 1



(b) Exemplo 2



(c) Exemplo 3



(d) Exemplo 4

Figura 4: Variação de Mix Suppression desenvolvida

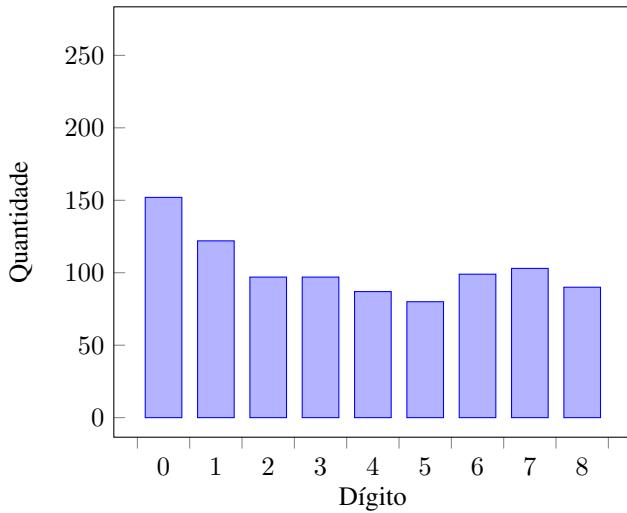


Figura 5: Distribuição dos dígitos平衡ados

tipo de hidrômetro só, sem sujeira, iluminado e com dígitos sempre aparentes, facilitando assim a aplicação de um modelo. Logo, há de se ter um cuidado para não relacionar o melhor modelo de um artigo como melhor modelo para qualquer outro.

Dentre os modelos encontrados na pesquisa literária presentes na tabela III, o modelo YOLO, junto a outros métodos, obteve um resultado constantemente satisfatório. Sendo assim, visto que em uma possível aplicação no mundo real a performance e velocidade do modelo são características importantes, foi escolhido o YOLOv8m-OBB como modelo

de implementação do presente artigo.

O fato de se tratar da versão Oriented Bounding Box (OBB) se dá pelo fato do dataset estar rotulado usando esse padrão de dados, ou seja, retângulos rotacionados representando cada dígito e sua classe.

VIII. FINE TUNING

Para fins de comparação, o YOLOv8m-OBB foi treinado em ambos dataset, o primeiro desbalanceado, e o segundo com a variação do Mix Suppression aplicada e dados balanceados. A prática de usar um modelo previamente construído e treinado, onde a partir de sua estrutura propor ele para um novo problema, se denomina fine tuning. O modelo foi treinado por 100 épocas e com batch de 8.

IX. RESULTADOS

A tabela I mostra que pequenas melhorias nos resultados foram encontradas quando os dados foram balanceados pelo método desenvolvido. A baixa melhora é possivelmente proveniente do baixo número de imagens no conjunto de dados. Contudo, o método proposto se prova eficaz e traz resultados satisfatórios, podendo ser aplicado de conjunto dados maiores gerando mais impacto.

Tabela I: Resultados

	P (%)	R (%)	mAP50 (%)	mAP50-95 (%)
Não balanceado	0.81315	0.73608	0.81599	0.67963
Balanceado	0.83994	0.7735	0.84313	0.70714

X. CONCLUSÃO

Neste trabalho, foi possível compreender uma nova área da visão computacional: a aplicação de OCR em hidrômetros. Características presentes na natureza do problema foram descobertas e analisadas. Também foi possível realizar uma pesquisa literária com o objetivo de explorar métodos e modelos usados pela comunidade acadêmica, assim como também propor uma variação de um método existente, resultando em melhorias significativas para o modelo YOLOv8m-OBB escolhido. Este estudo contribui para o avanço e aprimoramento da aplicação de OCR em hidrômetros, oferecendo insights para futuras pesquisas e desenvolvimentos nessa área específica da visão computacional.

Tabela II: Observações resumidas dos artigos

Artigo	Observação
Application of Deep Residual Neural Network to Water Meter Reading Recognition	Primeiro usa ResNet-101 para feature extraction; em seguida, usa a rede RPN para gerar regiões de ROI; em seguida, divide o ROI em nove regiões, gera o espectro de recursos de cada região e, finalmente, realiza a detecção de classificação por votação.
Research on water meter reading recognition based on deep learning	Há dois planos propostos, o plano A utiliza uma sequência de passos que inclui o ajuste do tamanho da imagem de entrada, a identificação dos números nos medidores de água e a saída dos números após a ordenação. Por outro lado, o Plano B envolve a identificação do retângulo preto, o corte das imagens como entrada e a saída dos números após a ordenação.
Automatic wheel-type water meter digit reading recognition based on deep learning	Mix suppression foi usado para expandir o conjunto de dados, melhorar a distribuição de classes e aumentar o desempenho da detecção de pequenos objetos. Junto foi implementado attention mode BSAM para orientar a atenção da CNN para a área de concentração de objetos e melhorar ainda mais o desempenho do modelo de detecção de objetos
Automatic Water Meter Reading Development Based On CNN and LoRaWAN	Não há menção explícita sobre as métricas de desempenho ou acurácia dos modelos. O artigo se concentra na tecnologia IoT, redes LoRaWAN, redes neurais convolucionais (CNN) e sua aplicação no desenvolvimento de um sistema inteligente para a leitura automática de medidores de água analógicos.
Key point localization and recurrent neural network based water meter reading recognition	Localização de pontos-chave e método de reconhecimento de leitura de hidrômetro baseado em rede neural recorrente. três módulos principais, incluindo localização de pontos-chave na área digital, correção e segmentação de imagens distorcidas, reconhecimento e transcrição da sequência de leitura de hidrômetros.
Evaluation of Recognition of Water-meter Digits with Application Programs, APIs, and Machine Learning Algorithms	Visto o contexto de detecção de dígitos em hidrômetros, compara correção de 5 programas/APIs, Anyline, Line, Google Vision, Microsoft Azure Computer Vision, e Naver (Clovar) com soluções como KNN, SVM, CNNs, AutoML, Tesseract e etc. O artigo mostrou que a arquitetura baseada em CNN pode fornecer resultados de reconhecimento comparáveis a programas comerciais e APIs. Fotos classificadas em três classes: reta, 90 graus para direita, ponta cabeça

Tabela III: Resultados da pesquisa da literatura

Artigo	Dataset próprio	Tamanho dataset	Métodos de box extraction usados	Arquiteturas para target-detection usadas	Avaliação do(s) modelo(s)	
					Acc (%)	mAP (%)
Application of Deep Residual Neural Network to Water Meter Reading Recognition	Sim	160.785 imagens (15.000 treino e 145.785 teste)	RPN	R-FCN	81.79	
Research on water meter reading recognition based on deep learning	Sim	4000 imagens originais 14.400 com expansão (12.960 treino e 1.440 teste)	RPN	SSD, YOLOv3 e Faster R-CNN	Plano A SSD: 78.63 YOLOv3: 87.30 Faster R-CNN: 82.96 Plano B SSD: 80.42 YOLOv3: 90.61 Faster R-CNN: 81.53	Plano A SSD: 75.67 YOLOv3: 78.10 Faster R-CNN: 77.07 Plano B SSD: 78.61 YOLOv3: 82.02 Faster R-CNN: 75.46
Automatic wheel-type water meter digit reading recognition based on deep learning	Sim	1277 imagens (700 treino e 577 teste)		Yolov4-tiny (com Mix suppression e BSAM)		97.2
Automatic Water Meter Reading Development Based On CNN and LoRaWAN				FCSRN		
Key point localization and recurrent neural network based water meter reading recognition	Sim	45861 (43.681 treino e 2.000 teste)	Stacked hourglass networks	CRNN (com correção de distorção)	95.3	
Evaluation of Recognition of Water-meter Digits with Application Programs, APIs, and Machine Learning Algorithms	Dataset próprio, MNIST, EMNIST e SVHN	96 imagens originais, futuramente misturado com outros datasets		Anyline, Line, Google Vision, Microsoft Azure Computer Vision, Naver (Clovar), AutoML, Tesseract, KNN, SVM eCNNs	(Top 3 modelos de cada classe de fotos) Retas Anyline: 96.88 Naver OCR: 92.19 GoogLeNet: 86.72 90 graus para direita Naver OCR: 87.5 GoogLeNet: 78.91 Automl: 75 Ponta cabeça Naver OCR: 85.16 GoogLeNet: 74.22 Microsoft Azure: 69.53	

Células vazias representam dados não informados.

REFERÊNCIAS

- [1] Y. Peng and Z. Chen, "Application of Deep Residual Neural Network to Water Meter Reading Recognition," 2020 IEEE International Conference on Artificial Intelligence and Computer Applications (ICAICA), Dalian, China, 2020, pp. 774-777, doi: 10.1109/ICAICA50127.2020.9182460.
- [2] Liang, Y., Liao, Y., Li, S. et al. Research on water meter reading recognition based on deep learning. *Sci Rep* 12, 12861 (2022). <https://doi.org/10.1038/s41598-022-17255-3>
- [3] Zhu, J., et al. "Automatic wheel-type water meter digit reading recognition based on deep learning," in *Journal of Electronic Imaging*, vol. 31, pp. 023023, 2022.
- [4] Bangkit, H., et al, "Automatic Water Meter Reading Development Based On CNN and LoRaWAN," in Book title is required!, 2023, pp. 212-215.
- [5] Jiguang Zhang, undefined., et al. "Key point localization and recurrent neural network based water meter reading recognition," in *Displays*, vol. 74, pp. 102222, 2022.
- [6] Eurviriyanukul, K., et al, "Evaluation of Recognition of Water-meter Digits with Application Programs, APIs, and Machine Learning Algorithms," in 2020 8th International Electrical Engineering Congress (iEECON), 2020, pp. 1-4.