Análise do Erro Humano nos Rótulos de Bases de Dados de Detecção de Objetos

Daniel Almeida[1]
orcid.org/0009-0007-6541-4852

Arthur Silva[1]
orcid.org/0009-0002-0525-1354

Agostinho Freire[2]
orcid.org/0000-0002-6059-9014

Raphaell Maciel de Souza[1]
orcid.org/0000-0001-5826-6408

Bruno J. T. Fernandes[2]
orcid.org/0000-0002-6001-3925

Leandro H. de S. Silva[1]
orcid.org/0000-0001-6221-2250

[1] Instituto Federal de Educação Ciência e Tecnologia da Paraíba (IFPB), *Campus* Cajazeiras, Cajazeiras-PB, Brasil
E-mail: almeida.daniel@academico.ifpb.edu.br

[2] Escola Politécnica de Pernambuco, Universidade de Pernambuco, Recife, Brasil.

DOI: 10.25286/repa.v10i3.3122

Como citar este artigo pela NBR 6023/2018: Daniel Almeida; Arthur Silva; Agostinho Freire, Raphaell Maciel de Souza, Bruno J. T. Fernandes, Leandro H. de S. Silva. Análise do Erro Humano nos Rótulos de Bases de Dados de Detecção de Objetos. Revista de Engenharia e Pesquisa Aplicada, v.10, n. 3, p. 43-54, 2025,

RESUMO

A detecção de objetos é uma tarefa importante em visão computacional e a precisão dos modelos de detecção pode depender da qualidade das bases rotuladas. A anotação manual, sobretudo em cenas complexas com múltiplos objetos, é trabalhosa e propensa a erros, comprometendo a qualidade das bases de dados. Neste estudo, 30 participantes rotularam 30 imagens da base Pascal VOC distribuídas em três níveis de dificuldade. Os resultados indicaram médias de erros de 120 ± 15 (Nível 1), 200 ± 45 (Nível 2) e 450 ± 80 (Nível 3), enquanto a porção de rótulos considerados corretos caiu de 78 % no Nível 1 para 41 % no Nível 3. Esses achados evidenciam que a crescente complexidade das imagens eleva significativamente o número de erros de rotulagem e reduz a sobreposição média entre rótulos humanos e *ground-truth*.

PALAVRAS-CHAVE: Detecção de Objetos; Visão Computacional; Ruído de Anotação; Aprendizado de Máquina.

ABSTRACT

Object detection is an important task in computer vision, and the accuracy of detection models can depend on the quality of labeled datasets. Manual annotation, especially in complex scenes with multiple objects, is labor-intensive and prone to errors, compromising dataset quality. In this study, 30 participants annotated 30 images from the Pascal VOC dataset, distributed across three difficulty levels. The results showed mean errors of 120 ± 15 (Level 1), 200 ± 45 (Level 2), and 450 ± 80 (Level 3), while the proportion of labels considered correct dropped from 78% at Level 1 to 41% at Level 3. These findings demonstrate that increasing image complexity significantly raises labeling errors and reduces the average overlap between human annotations and ground truth.

KEY-WORDS: Object Detection; Computer Vision; Annotation Noise; Machine Learning.

1 INTRODUÇÃO

A detecção de objetos é uma tarefa da visão computacional que desempenha um papel fundamental em sistemas práticos, como a vigilância por vídeo, onde é essencial identificar e localizar objetos de interesse [1][2]. Além disso, sua relevância também se estende a tecnologias como os veículos autônomos, que dependem da detecção precisa de objetos no ambiente para operar com segurança e eficiência [3]. Outra aplicação relevante é o reconhecimento de expressões e detalhes faciais, fundamental em sistemas de segurança e interações humanas [4][5].

Um dos principais fatores que têm impulsionado o avanço efetivo da detecção de objetos é o progresso das técnicas de aprendizado profundo, com destaque para as redes neurais convolucionais [6], possibilitando a construção de modelos cada vez mais precisos e eficientes para tarefas de detecção de objetos, com aplicações nas áreas de diagnósticos [7], o funcionamento seguro de veículos e infraestruturas de transporte [8][9].

Tipicamente o paradigma da aprendizagem supervisionada é utilizado para treinamento dos modelos de detecção de objetos baseados em aprendizagem profunda [10]. Entretanto, bases de dados suficientemente grandes e completamente rotuladas para problemas de detecção de objetos são custosas e constantemente não estão disponíveis para determinadas aplicações [2]. Além disso, a anotação manual de dados, especialmente em tarefas que exigem detalhes minuciosos ou contêm múltiplos objetos em uma cena, pode ser uma tarefa demorada e propensa a erros humanos [11]. Essa dificuldade na rotulagem de dados pode afetar diretamente a eficácia dos modelos de aprendizado profundo, pois a qualidade do conjunto de treinamento é crucial para garantir que o modelo aprenda representações significativas e generalizáveis dos objetos em questão.

Diante desse contexto, este trabalho investiga quantitativamente como os erros de rotulagem humana impactam modelos de detecção de objetos. Para isso, 30 voluntários rotularam 30 imagens selecionadas da base Pascal VOC em três níveis de dificuldade (até 4, 4-10 e mais de 10 objetos). As discrepâncias entre rótulos humanos e ground-truth foram categorizadas via Intersection over Union (IoU) e analisadas estatisticamente. Ao estabelecer essa ponte direta entre a escassez de rótulos precisos — destacada na literatura — e um experimento controlado de rotulagem, demonstramos em que medida a complexidade da cena aumenta a incidência de erros e, por extensão, pode comprometer a robustez de sistemas de detecção treinados com esses dados.

2 TRABALHOS RELACIONADOS

Diversos estudos têm se debruçado sobre o impacto dos rótulos ruidosos em tarefas de detecção de objetos, investigando como esses erros de anotação afetam tanto a qualidade dos modelos quanto sua robustez. Adhikari et al. [12] exploram a sensibilidade das funções de perda em relação à ausência de rótulos em caixas delimitadoras. Ao comparar a perda de entropia cruzada com a perda focal, o estudo mostra que a escolha cuidadosa de hiperparâmetros pode minimizar os efeitos negativos de até 50% de rótulos ausentes, indicando que certos modelos de detecção podem ser ajustados para lidar com níveis significativos de ruído nos dados. Essa abordagem fornece uma análise detalhada sobre a relevância dos rótulos faltantes no desempenho do modelo, sugerindo estratégias para mitigar a degradação causada por erros de rotulagem.

Júnior et al. [13] introduzem uma abordagem de auto aprendizado para a detecção de componentes eletrônicos em placas de circuito impresso (PCBs) recicladas, um domínio com poucos conjuntos de dados anotados. Utilizando uma técnica de pseudo-rótulos, eles empregam o modelo YOLOv5 para transferir o aprendizado de um conjunto de dados anotado (FICS-PCB) para outro parcialmente anotado (PCB-DSLR). Esse processo resulta na criação de rótulos adicionais para os componentes eletrônicos, permitindo que o modelo aluno ultrapasse o desempenho do modelo professor. O trabalho demonstra que mesmo em contextos com dados limitados e parcialmente anotados, é possível treinar modelos robustos, contribuindo diretamente para o avanço da reciclagem de resíduos eletrônicos e extração de materiais valiosos.

No estudo de Wang et al. [14], os desafios impostos por rótulos ruidosos em modelos de detecção de objetos ocultos são abordados por meio de um framework unificado. Esse framework envolve a modelagem de ruído de rótulos no nível de região e a correção dos rótulos durante o treinamento, utilizando uma perda ponderada por incerteza. Os experimentos conduzidos com imagens de ondas milimétricas, comumente usadas em cenários de segurança, demonstram que o framework proposto é eficaz na redução do impacto de anotações ruidosas, resultando em melhorias

significativas no desempenho da detecção. A abordagem sugere que a robustez a ruídos de rótulos pode ser aprimorada mesmo em aplicações de alta criticidade, como segurança.

Li et al. [15] focam na adaptação de domínio não supervisionada (UDA), especialmente em cenários onde não há acesso a dados rotulados do domínio de origem. A abordagem proposta, chamada de detecção de objetos adaptativa sem dados de origem (SFOD), utiliza uma rede pré-treinada para gerar pseudo-rótulos no domínio de destino, e uma métrica de auto-entropia (SED) é introduzida para avaliar a confiança dos pseudo-rótulos. Ao simular falsos negativos com técnicas de aumento de dados, os autores melhoram a robustez do modelo e demonstram que é possível obter resultados de ponta mesmo sem rótulos limpos e completos. Este trabalho amplia o conhecimento sobre o uso de dados ruidosos e reforça a importância de frameworks adaptáveis para cenários de adaptação de domínio.

Por fim, Zhou et al. [16] propõem uma abordagem para a detecção semissupervisionada de objetos, substituindo as pseudo-caixas por Pseudo-rótulos Densos (DPL). Essa técnica elimina a necessidade de pós-processamento complexo e ajustes finos de hiperparâmetros, que são frequentes em métodos de pseudo-caixas. A proposta utiliza DPL para manter informações ricas durante o processo de rotulagem, resultando em melhor desempenho em benchmarks como COCO [17] e VOC [18]. Esse trabalho representa um avanço significativo na área de detecção semi supervisionada, oferecendo uma alternativa mais eficiente e eficaz para a geração de pseudo-rótulos, o que facilita o treinamento de modelos em contextos de dados parcialmente anotados.

Esses estudos, em conjunto, demonstram a crescente importância de abordar o ruído nos rótulos, tanto em tarefas de detecção de objetos quanto em domínios correlatos. Eles fornecem insights valiosos sobre como modelos de aprendizado profundo podem ser ajustados ou treinados para lidar com ruído, destacando a necessidade de técnicas robustas e adaptativas para melhorar a precisão e confiabilidade dos sistemas de detecção.

3 METODOLOGIA

Para investigar o erro humano nos rótulos de bases de dados de detecção de objetos, esta pesquisa propôs um experimento onde 30 pessoas foram convidadas a rotular 30 imagens selecionadas da base de dados Pascal Visual Object Classes (VOC). A partir dos rótulos produzidos, é conduzida uma análise estatística dos erros cometidos, tomando como referência os rótulos originais na base de dados. Na Tabela 1, é apresentado o processo detalhado de realização do experimento.

Tabela 1: Etapas do experimento

Ordem	Etapa	Ação principal
1	Seleção de imagens	Escolha de 30 imagens da base Pascal VOC em três níveis de dificuldade
2	Configuração do projeto	Criação do dataset e das tarefas de rotulagem
3	Treinamento dos participantes	Tutorial rápido sobre uso da ferramenta e critérios de rotulagem
4	Rotulagem síncrona	Cada um dos 30 voluntários rotula as 30 imagens
5	Exportação dos rótulos	Download dos rótulos feitos pelos participantes
6	Cálculo do IoU	Comparação rótulos VOC × rótulos humanos; classificação dos erros
7	Análise estatística	Estatísticas descritivas, testes de normalidade, boxplots

Fonte: Os próprios autores (2025).

1. Planejamento do Experimento de coleta de rótulos

Os experimentos de coleta dos rótulos foram realizados de forma online síncrona, utilizando a plataforma online *Roboflow*[1]. Essa plataforma permite o cadastro das imagens selecionadas e a solicitação da tarefa de rotulagem para um usuário que possua cadastro na plataforma. A proposta de realização da coleta de dados ocorreu em grupos de forma síncrona, de forma a obter controle do tempo total de execução do experimento. Durante o experimento os participantes também responderam um formulário

[1] ROBOFLOW, Inc. Roboflow Universe: conjuntos de dados e modelos de visão computacional. Disponível em: https://universe.roboflow.com. Acesso em: 29 abr. 2025.

DOI: 10.25286/repa.v10i3.3122

contendo dados sobre idade, se já possuíam experiências na tarefa de rotular imagens e escolaridade.

O experimento foi realizado em três etapas: primeiro, os participantes foram instruídos a realizar cadastro na plataforma *Roboflow*; em seguida, foram treinados no uso da plataforma para a tarefa de rotulagem de objetos em imagens; e, por fim, cada participante foi designado para rotular trinta imagens previamente selecionadas. Importante destacar que todas as etapas ocorreram em um único encontro síncrono, todos os participantes forneceram rótulos para as mesmas imagens e, em nenhum momento, tiveram acesso aos rótulos produzidos por outros participantes ou aos rótulos originais da base de dados. Cabe ressaltar que não está no escopo deste projeto analisar os fatores que podem causar erros no processo de rotulagem, como, por exemplo, a fadiga dos participantes.

2. Escolha da Base de Dados e Seleção das Imagens

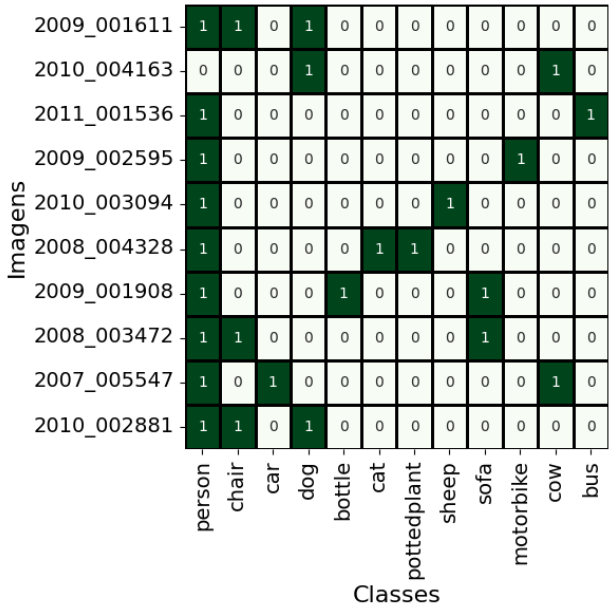
Para o desenvolvimento deste projeto, foi utilizada a base de dados Pascal Visual Object Classes (VOC), com mais de 17 milhões de imagens. Os rótulos da base de dados foram considerados como *groundtruth* e foram confrontados com os rótulos realizados pelas pessoas. Dessa forma, foram escolhidas 30 imagens, divididas em 3 diferentes níveis de dificuldade. Os níveis de dificuldade no experimento são definidos de acordo com a complexidade das cenas e a quantidade de objetos presentes nas imagens. Através da tabela 2, é possível verificar a comparação entre os diferentes níveis de dificuldade. O Nível 1 apresenta até 4 objetos na cena, onde os objetos ocupam mais de 20% da área da imagem e há, no máximo, 3 classes de objetos por imagem. No Nível 2, o número de objetos aumenta para entre 4 e 10, com uma representação de área menor, entre 10% e 20%, e de 3 a 5 classes diferentes de objetos. Já no Nível 3, a dificuldade é elevada com mais de 10 objetos por cena, ocupando menos de 10% da área da imagem, e mais de 5 classes de objetos. Essas características tornam as imagens progressivamente mais difíceis de rotular (Tabela 1). Nas Figuras 1, 2 e 3 é possível visualizar as classes presentes nas imagens em níveis de dificuldades respectivos, onde as classes presentes são determinadas pelo o valor 1 e as ausentes com o valor 0.

Atributo	Nível 1	Nível 2	Nível 3
N° de objetos por imagem	Até 4	4 - 10	> 10
Área Média Ocupada Por Objetos	> 20% da imagem	10% - 20%	< 10%
N° de Classes Distintas	≤ 3	3 - 5	> 5

Tabela 2: Tabela comparativa dos níveis de dificuldade.

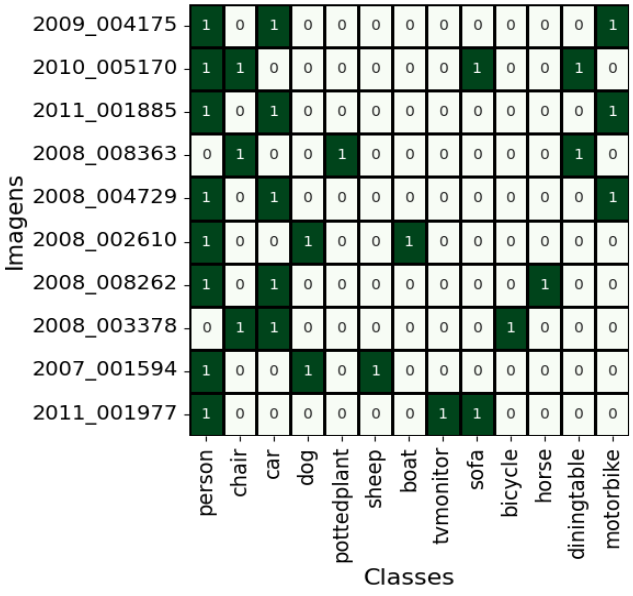
Fonte: Os próprios autores (2025).

Figura 1: Nível de dificuldade 1 das imagens selecionadas.



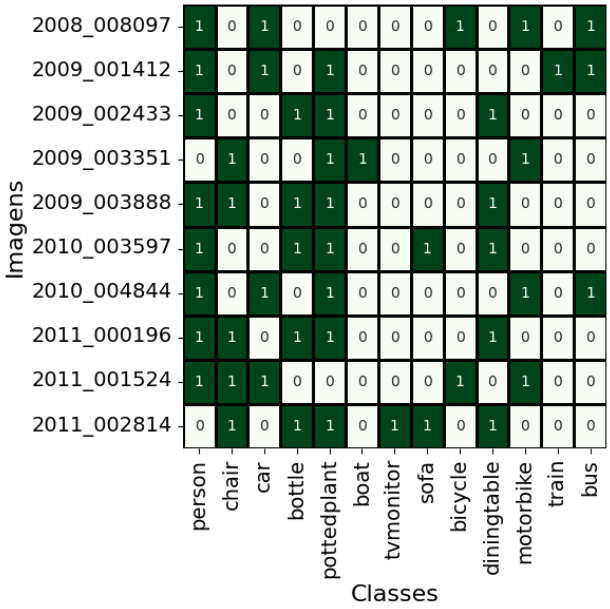
Fonte: Os próprios autores (2024).

Figura 2: Nível de dificuldade 2 das imagens selecionadas.



Fonte: Os próprios autores (2024).

Figura 3: Nível de dificuldade 3 das imagens selecionadas.



Fonte: Os próprios autores (2024).

3. Taxonomia dos erros

Para identificar os erros, foi implementado o cálculo do IoU (Intersection over Union). A escolha do IoU como métrica neste estudo se justifica por sua capacidade de quantificar com precisão o grau de sobreposição entre duas regiões delimitadas, sendo particularmente útil na comparação entre rótulos originais e rótulos produzidos por humanos.

O IoU é definido como a razão entre a área de interseção e a área de união de duas caixas delimitadoras (BB) [19]:

$$IoU = \frac{Área(BBorig \cap BBhum)}{Área(BBorig \cup BBhum)}$$

onde BBorig é o rótulo da base VOC e BBhum o rótulo fornecido pelo participante. Na Tabela 3, é possível identificar os Limiares de IoU utilizados neste trabalho.

A partir do IoU entre todos os rótulos, foi possível categorizar os tipos de erros/ruídos de anotação:

- Erro tipo #1: Acontece quando um objeto presente nos rótulos originais da imagem (Figura 4) não é rotulado pelo humano ou quando o humano rotula um objeto que originalmente não encontra-se listado nos rótulos originais. Distinguimos esses dois subtipos de erros da seguinte forma:
    - Erro tipo 1.1 - Há um objeto nos rótulos originais que não foi identificado pelo humano (Figura 5);
    - Erro tipo 1.2 - O humano rotulou um objeto que não está contido nas anotações originais da base de dados (Figura 6).
- Erro tipo #2: Confusão de identidade, onde a classe de um objeto é trocada. Nesse caso, um objeto rotulado pelo humano aparece com alta semelhança (IoU alto) a um objeto contido nos rótulos originais da imagens, entretanto, há uma diferença na classe do objeto (Figura 7).
- Erro tipo #3: Esse erro é a distorção na localização do objeto, sendo constituído pela diferença entre as coordenadas do objeto nos rótulos originais e as coordenadas do objeto identificada pelo humano. Dado que é improvável que o humano marque exatamente as mesmas coordenadas do rótulo original, esse tipo de erro é sempre esperado em alguma magnitude (Figura 8).

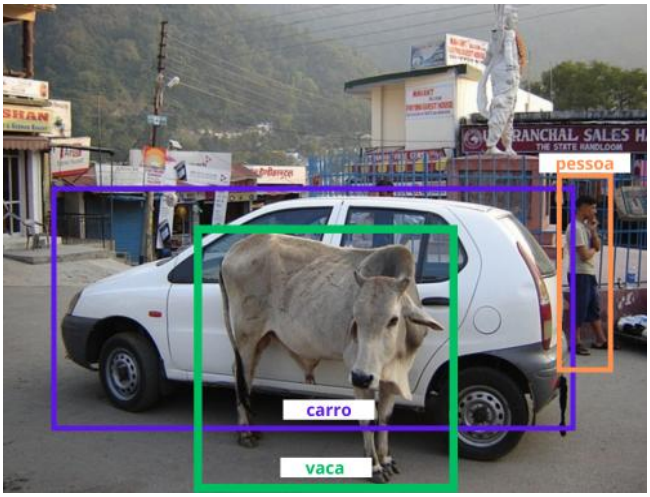
Tabela 3: Faixas de Intersection over Union (IoU) adotadas e respectiva associação aos tipos de erro de rotulagem

Faixa de IoU	Interpretação	Relacionamento com a taxonomia de erros
IoU <0,10	ausente/excedente	Erro 1.1 (objeto ausente) ou 1.2 (objeto extra)
0,10≤ IoU <0,50	localização grosseiramente imprecisa	Erro 3 (distorção de localização)
IoU ≥ 0,50 e classe diferente	alta sobreposição, classe trocada	Erro 2 (confusão de identidade)
IoU ≥ 0,50 e classe correta	rótulo considerado correto	-

Fonte: Os próprios autores (2025).

Figura 4: Imagem com rótulos originais.

DOI: 10.25286/repa.v10i3.3122

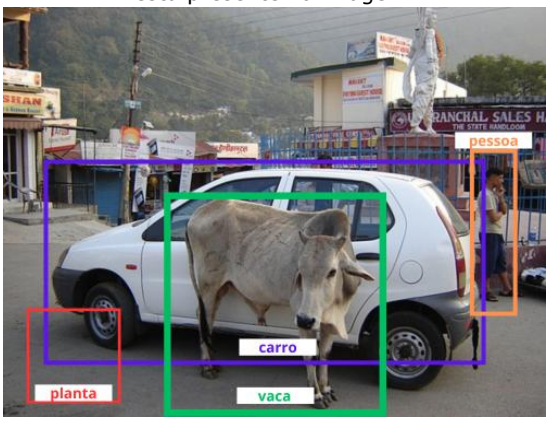


Fonte: Os próprios autores (2024).

Figura 5: Imagem com exemplo do erro de rotulagem tipo 1.1, onde o humano esqueceu de rotular o carro que está presente.

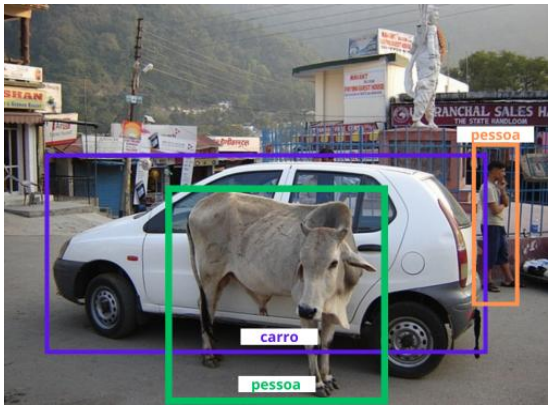
Fonte: Os próprios autores (2024).

Figura 6: Imagem com exemplo do erro de rotulagem tipo 1.2, onde o humano rotulou uma planta que não está presente na imagem.



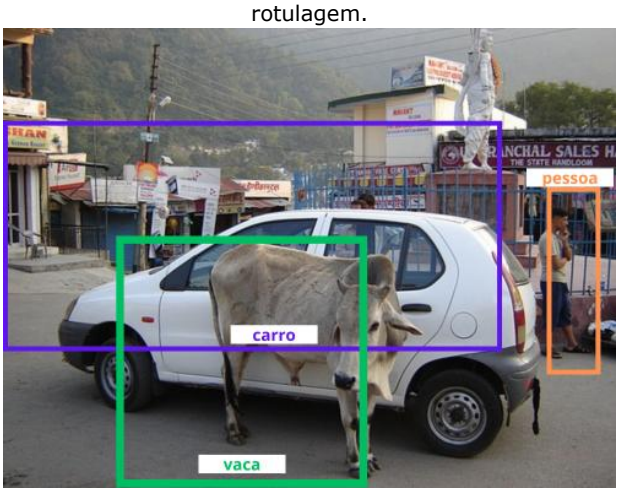
Fonte: Os próprios autores (2024).

Figura 7: Imagem com exemplo do erro de rotulagem tipo 2, onde uma vaca foi confundida com uma pessoa, havendo assim uma confusão de classes.



Fonte: Os próprios autores (2024).

Figura 8: Imagem com exemplo do erro de rotulagem tipo 3, onde há uma distorção de localização da rotulagem.



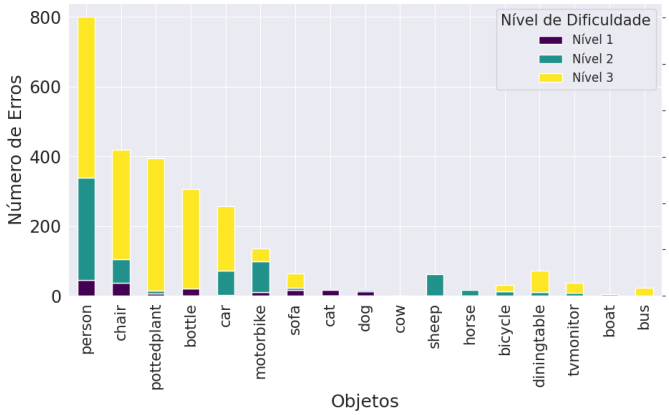
Fonte: Os próprios autores (2024).

4 RESULTADOS

Conduzindo uma exploração inicial dos erros do tipo 1 e 2, a Figura 9 apresenta a soma dos erros por classe e por nível das imagens, totalizando 2 660 falhas de rotulagem. Observa-se que o objeto "person" concentra 800 erros (30,1 % de todos os erros), com clara predominância no Nível 3. Em seguida aparecem "chair" com 419 erros (15,8 %) e "pottedplant" com 395 erros (14,9 %), igualmente influenciados pela complexidade do Nível 3. Esse padrão confirma que as cenas mais densas multiplicam a dificuldade de anotação. No Nível 2, "motorbike" registra 135 erros (5,1 %) e "car" 258 erros (9,7 %), evidenciando que, mesmo em dificuldade intermediária, esses objetos permanecem desafiadores. Já no Nível 1, embora o volume global de erros seja reduzido, "person" (46) e "chair" (38) ainda lideram, sugerindo características visuais que confundem tanto anotadores humanos quanto algoritmos de detecção. Essas cinco classes principais — "person", "chair", "pottedplant", "car" e

"motorbike" — somam 75,6 % de todos os erros, indicando que a maior parte dos deslizes se concentra nelas. Em contraste, objetos como "sheep" (63 erros; 2,4 %), "diningtable" (72; 2,7 %) e "sofa" (64; 2,4 %) têm participação modesta, enquanto "tvmonitor" (38; 1,4 %), "bicycle" (32; 1,2 %), "bus" (24; 0,9 %), "horse" (16; 0,6 %), "boat" (5; 0,2 %) e "cow" (1; 0,04 %) apresentam proporções inferiores a 3 %. Tais resultados sugerem que esses objetos são mais facilmente identificáveis ou aparecem com menor frequência nas imagens analisadas.

Figura 9: Número de erros por objetos em diferentes níveis de dificuldade.



Fonte: Os próprios autores (2024).

1. Caracterização da População do Experimento

Com base nos dados dos participantes, as idades variam de 18 a 34 anos, sendo majoritariamente entre 18 e 23 anos. No que diz respeito à escolaridade, o grupo é composto por 1 pessoa com doutorado, 1 com ensino superior completo, 2 com mestrado e 26 estão cursando graduação. Além disso, a partir dos dados coletados, apurou-se que 22 participantes afirmaram não possuir nenhuma experiência com rotulagem de objetos ou áreas correlatas, enquanto 8 participantes confirmaram ter experiência nesse campo. A maioria das experiências relatadas pelos participantes é de natureza institucional, abrangendo projetos de pesquisa, projetos de extensão, atividades de estágio, estudos individuais em inteligência artificial, além da escrita de artigos e trabalho de doutorado na área de visão computacional.

Ademais, também foi possível comparar a experiência dos usuários pela quantidade de erros obtidos por eles no experimento. Observa-se, na Figura 10, que o grupo com experiência ("Sim") tende a cometer menos erros, com uma mediana ligeiramente abaixo de 275 erros. A distribuição dos dados para esse grupo é mais compacta, indicando uma menor variação na quantidade de erros. Além disso, há um único outlier (321 erros) no grupo dos que têm experiência, que pode representar um caso isolado de alguém que, apesar da experiência, teve um desempenho significativamente pior que os demais. A cauda inferior para esse grupo indica que algumas pessoas cometeram erros próximos de 255, o que pode ser considerado um desempenho melhor do que a média.

Por outro lado, o grupo sem experiência ("Não") apresenta uma mediana mais alta, em torno de 290 erros, sugerindo que, em média, essas pessoas cometeram mais erros. A variação dentro desse grupo é maior, conforme indicado pelo tamanho maior do intervalo interquartil. Isso reflete que as pessoas sem experiência apresentaram um desempenho mais inconsistente, com alguns participantes cometendo erros na faixa dos 260, enquanto outros chegaram a cometer cerca de 310 erros. Entretanto, devido a diferença no tamanho das amostras dos indivíduos com e sem experiência, com a maioria dos participantes da pesquisa sendo constituída de pessoas sem experiência, não é possível traçar conclusões com significância estatística sobre a relação entre experiência e quantidade de erros.

2. Análise Distributiva dos Dados

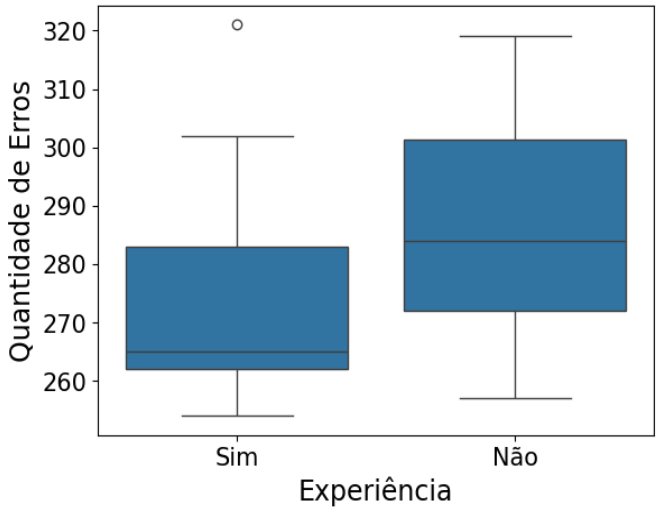
Em seguida, foi possível analisar o comportamento dos ruídos nos rótulos conforme a progressão no nível das imagens (Figura 11). Os dados mostram que, o Nível 1 apresenta a menor variação de erros, com o total de erros cometidos pelos participantes entre aproximadamente 100 e 150 erros, e uma mediana de cerca de 120 erros (somando-se todos os tipos de erro). O Nível 2 possui uma variação maior, com a quantidade de erros variando entre aproximadamente 100 e 400. A mediana está em torno de 200, sugerindo que metade das imagens deste nível tem menos de 200 erros e a outra metade mais de 200. O Nível 3 mostra a maior variação e quantidade de erros, com a mediana em torno de 450 e uma extensão de erros que vai de cerca de 350 a 550. Notavelmente, há um outlier acima de 700, indicando uma imagem com um número de erros significativamente maior que as outras. Assim, à medida que o nível aumenta de 1 para 3, a quantidade de erros e a variabilidade

DOI: 10.25286/repa.v10i3.3122

entre as imagens também aumentam, com o Nível 3 apresentando maior dispersão e um outlier significativo.

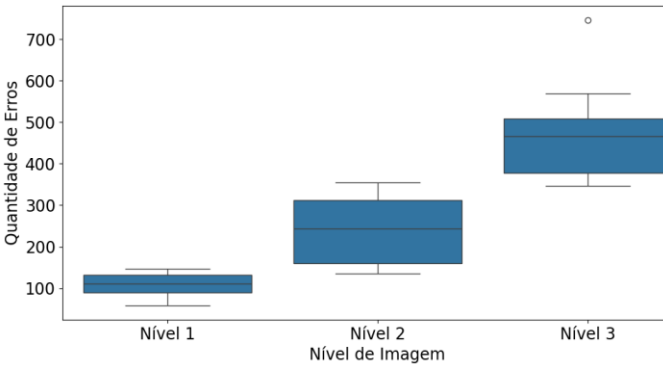
Através da Tabela 4, verifica-se, para o Erro #1, uma tendência clara de aumento no número de erros à medida que a dificuldade das imagens progride, com uma média de 9,21 erros no Nível 1, que sobe para 95,10 no Nível 3. O desvio padrão também cresce, indo de 5,21 no nível mais simples para 18,60 no mais complexo, indicando maior variação nos erros conforme a dificuldade aumenta. Isso sugere que os participantes cometeram mais erros e apresentaram maior inconsistência à medida que enfrentavam imagens mais difíceis.

Figura 10: Boxplot de experiência por quantidade de erros.



Fonte: Os próprios autores (2024).

Figura 11: Boxplot do total de erros por nível de imagem.



Fonte: Os próprios autores (2024).

O Erro #2 tem médias bem mais baixas, sendo o menos frequente, com uma média máxima de 2,86 no Nível 3 e desvios padrões relativamente menores. Por fim, o Erro #3 segue uma trajetória semelhante ao Erro #1, com um número de erros que vai de 26,10 no Nível 1 para 93,28 no Nível 3, e um desvio padrão que cresce significativamente, refletindo uma variabilidade maior em níveis mais complexos. Dessa forma, percebe-se que tanto a média quanto o desvio padrão indicam que os erros aumentam substancialmente com a dificuldade das imagens.

Já na Tabela 5, apresentam-se os valores de significância (p) obtidos em dois testes de normalidade — Shapiro-Wilk e Kolmogorov-Smirnov (KS) — empregados especificamente para verificar se a distribuição dos erros cometidos na rotulagem segue ou não um padrão normal em cada nível de dificuldade das imagens. No teste Shapiro-Wilk, os valores de p para os Níveis 1 e 2 (0,0704 e 0,0513, respectivamente) estão muito próximos do limite de significância de 0,05, sugerindo que os dados nestes níveis podem não seguir exatamente uma distribuição normal, mas ainda assim não se afastam drasticamente dela. Já no Nível 3, o valor de 0,2915 indica que a normalidade não pode ser rejeitada, apontando para uma maior adequação a uma distribuição normal em níveis mais complexos. Os resultados do teste KS são ainda mais robustos, com todos os níveis apresentando valores de p acima de 0,28, e o Nível 3, em particular, mostrando um valor de 0,955, que fortemente confirma a normalidade dos dados. Assim, destaca-se que, especialmente nos níveis mais difíceis, neste experimento, os ruídos cometidos pelos participantes tendem a seguir uma distribuição normal.

Tabela 4: Média e desvio padrão da quantidade de erros por Nível.

Tipo de Erro	Nível 1	Nível 2	Nível 3	Todos os níveis
Erro #1	9,21 (5,21)	36,00 (16,72)	95,10 (18,60)	140 (38,5)
Erro #2	1,85 (0,93)	1,68 (0,68)	2,86 (1,83)	6,03 (2,31)

Fonte: Os próprios autores (2024).

Tabela 5: Testes de Normalização.

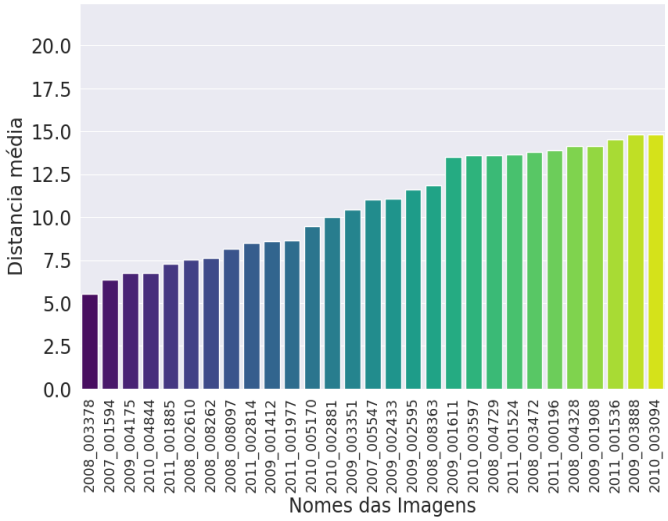
Teste Shapiro-Wilk p-value			Teste Kolmogorov-Smirnov p-value		
Nível 1	Nível 2	Nível 3	Nível 1	Nível 2	Nível 3
0.0704	0.0513	0.2915	0.408	0.281	0.955

Fonte: Os próprios autores (2024).

Para analisar os erros de deslocamento da caixa delimitadora, analisamos os rótulos que os participantes designaram corretamente. Sendo assim, primeiramente, observa-se a distância entre o centro do rótulo original e o centro do rótulo fornecido pelo participante. No gráfico da Figura 11, a imagem 2008_003378 apresenta a menor diferença de erro

(somando-se todos os desvios de centro de todos os rótulos e todos os participantes), enquanto a imagem 2010_004163, representada na Figura 12, apresenta a maior diferença. A imagem da Figura 12 é de Nível 1, possuindo apenas dois grandes objetos na cena. Sendo assim, essa quantidade de ruído não parece estar relacionada à dificuldade da imagem. Entretanto, percebe-se que a distância média entre o centro das caixas delimitadoras originais e as caixas delimitadoras designadas pelos participantes diferem em no máximo pouco mais do que 20 pixels, sendo inferior a 15 pixels para todas as classes, com exceção da classe 'cow', como pode ser visualizado na Figura 13.

Figura 11: Distância média entre o centro das caixas delimitadoras de todas as imagens.

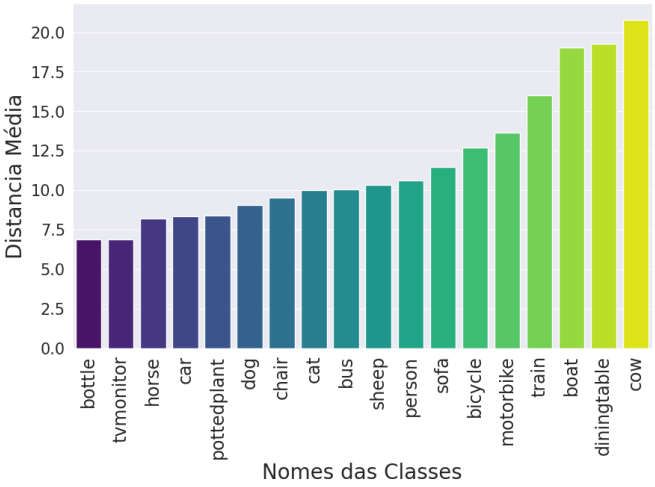


Fonte: Os próprios autores (2024).

Figura 12: Imagem de nível 1, 2010_004163.

Fonte: (VOC2012).

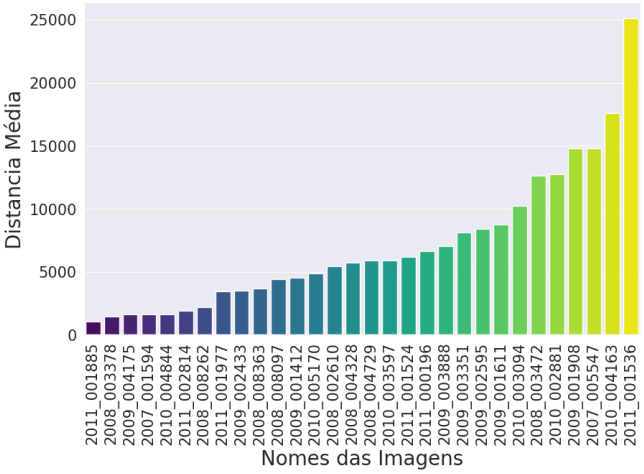
Figura 13: Distância média de todas as classes.



Fonte: Os próprios autores (2024).

Realizando as mesmas análises para a diferença entre a área do rótulo original e a área do rótulo fornecido pelo participante (Figura 14), observamos que, ao considerar a diferença média das áreas por imagem, a imagem com a menor diferença de área é a 2011_001885, enquanto a imagem com a maior diferença de área é a 2011_001536 (Figura 15). Portanto, podemos concluir que as imagens de nível 1, em média, exibem uma maior diferença de área, por normalmente apresentarem objetos maiores. Por fim, ao analisar a diferença média de área dos objetos (Figura 16), concluímos que o objeto com a maior diferença de área é o barco, presente na imagem com a maior média de diferença de área entre as imagens de nível 3. Em contraste, os objetos com as menores diferenças de área, assim como na análise da distância do centro, são a garrafa/copo e a TV.

Figura 14: Diferença de área média de todas as imagens.



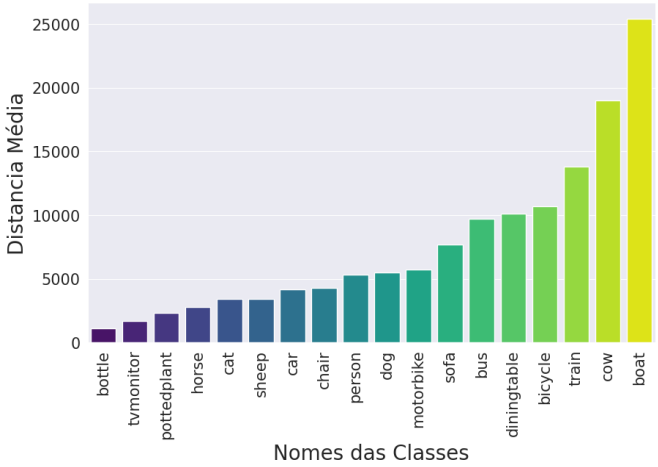
Fonte: Os próprios autores (2024).

Figura 15: Imagem de nível 1, 2011_001536.

DOI: 10.25286/repa.v10i3.3122

Fonte: (VOC2012).

Figura 16: Diferença de área média de todos os objetos.



Fonte: Os próprios autores (2024).

5 CONCLUSÃO

Os resultados deste estudo evidenciam de forma clara que a complexidade das imagens tem um impacto significativo na precisão da rotulagem de objetos, conforme revelado pelas médias e desvios-padrão apresentados na Tabela 4. O aumento no número de objetos, a diversidade de classes e a diminuição da área ocupada por esses objetos estão diretamente relacionados com o aumento dos erros cometidos pelos participantes. Esses achados reforçam a importância de levar em conta a complexidade das cenas ao desenvolver sistemas de rotulagem automatizada, oferecendo uma base sólida para ajustes em algoritmos de visão computacional.

A análise dos testes de normalidade, descritos na Tabela 5, revela que, apesar das variações nos dados, especialmente em cenários de maior dificuldade, os erros seguem uma distribuição normal. Isso valida o uso de análises estatísticas robustas para compreender o comportamento do ruído na rotulagem. Essa descoberta não apenas fundamenta a relevância dos resultados, como também abre caminho para que métodos de mitigação mais eficazes possam ser aplicados em situações desafiadoras.

Além disso, a possibilidade de simular o ruído de rotulagem humana em bases de dados amplamente utilizadas, como a Pascal VOC, oferece uma contribuição prática significativa. Isso permitirá que detectores de objetos sejam avaliados em cenários mais realistas, onde o ruído de rotulagem ocorre naturalmente, e não apenas em contextos idealizados. Avaliar modelos na presença de ruído real fornece insights críticos sobre o impacto desse tipo de erro no desempenho dos algoritmos.

Sugere-se, como linha de pesquisa futura, aplicar a metodologia proposta a domínios de alta relevância prática — por exemplo, detecção de lesões em imagens médicas ou monitoramento urbano por câmeras de segurança — a fim de quantificar como o ruído de rotulagem afeta sistemas operando em ambientes reais. Estudos de caso nesses contextos poderão avaliar se a perda de desempenho observada nos experimentos controlados se mantém, além de identificar requisitos específicos de robustez e estratégias de pré-processamento (p. ex., filtros de consistência inter-anotadores) que minimizem falhas críticas em aplicações sensíveis.

Importa salientar que está fora do escopo desta pesquisa discutir detalhadamente os motivos dos erros observados nos rótulos — por exemplo, fatores humanos específicos ou fadiga dos participantes. Todavia, recomendam-se, como trabalho futuro, estratégias de verificação dupla (double check) que comparem as discrepâncias entre rótulos gerados por diferentes anotadores, visando minimizar tais erros.

Por fim, a pesquisa fornece subsídios importantes para o aprimoramento de sistemas de visão computacional, indicando que a compreensão das dificuldades enfrentadas por rotuladores humanos pode guiar o desenvolvimento de algoritmos mais robustos. Ao incorporar cenários de rotulagem complexos em fases de treinamento e validação, futuros sistemas de detecção poderão se tornar mais resilientes ao ruído e, consequentemente, mais eficazes em aplicações do mundo real, especialmente em contextos com imagens de alta complexidade e diversificação de objetos.

Essa investigação representa um avanço importante para o campo da detecção de objetos, não apenas demonstrando os efeitos do ruído na rotulagem humana, mas também propondo estratégias concretas para mitigar seu impacto nos sistemas de aprendizado profundo.

6 REFERÊNCIAS BIBLIOGRÁFICAS

[1] DEL-BLANCO, C.; JAUREGUIZAR, F.; GARCIA, N. **An efficient multiple object detection and tracking framework for automatic counting and video surveillance applications.** IEEE Transactions on Consumer Electronics, v. 58, n. 3, p. 857–862, ago. 2012. Acesso em: Acesso em: 20 ago. 2023.

[2] CHEN, Bo-Hao; SHI, Ling-Feng ; KE, Xiao, **A Robust Moving Object Detection in Multi-Scenario Big Data for Video Surveillance**. IEEE Transactions on Circuits and Systems for Video Technology, v. 29, n. 4, p. 982–995, 2019. Acesso em: 20 ago. 2023.

[3] ZHAO, Xiangmo; SUN, Pengpeng; XU, Zhigang; et al. **Fusion of 3D LIDAR and Camera Data for Object Detection in Autonomous Vehicle Applications**. IEEE Sensors Journal, p. 1–1, 2020. Disponível em: <https://ieeexplore.ieee.org/document/8957313>. Acesso em: 20 ago. 2023.

[4] HAPPY, S. L. ; ROUTRAY, Aurobinda. **Automatic facial expression recognition using features of salient facial patches**. IEEE Transactions on Affective Computing, v. 6, n. 1, p. 1–12, 2013. Disponível em: <https://ieeexplore.ieee.org/document/6998925>. Acesso em: 20 ago. 2023.

[5] SHIN, Jongju ; KIM, Daijin. **Hybrid Approach for Facial Feature Detection and Tracking under Occlusion**. IEEE Signal Processing Letters, v. 21, n. 12, p. 1486–1490, 2014. Disponível em: <https://ieeexplore.ieee.org/document/6855350>. Acesso em: 25 ago. 2023.

[6] LI, Zewen; LIU, Fan; YANG, Wenjie; et al. **A Survey of Convolutional Neural Networks: Analysis, Applications, and Prospects**. IEEE Transactions on Neural Networks and Learning Systems, v. 33, n. 12, p. 1–21, 2021. Disponível em: <https://ieeexplore.ieee.org/document/9451544>. Acesso em: 24 ago. 2023.

[7] WANG, Xiaosong; PENG, Yifan; LU, Le; et al. **ChestX-Ray8: Hospital-Scale Chest X-Ray Database and Benchmarks on Weakly-Supervised Classification and Localization of Common Thorax Diseases.** 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017. Disponível em: <https://arxiv.org/abs/1705.02315>. Acesso em: 20 ago. 2023.

[8] BAI, Zhengwei; WU, Guoyuan; QI, Xuewei; et al. **Infrastructure-Based Object Detection and Tracking for Cooperative Driving Automation: A Survey.** arXiv:2201.11871 [cs, eess], 2022. Disponível em: <https://arxiv.org/abs/2201.11871>. Acesso em: 20 ago. 2023.

[9] SONG, Huansheng; LIANG, Haoxiang; LI, Huaiyu; et al. **Vision-based vehicle detection and counting system using deep learning in highway scenes**. European Transport Research Review, v. 11, n. 1, 2019. Disponível em: <https://etrr.springeropen.com/articles/10.1186/s12544-019-0390-4>. Acesso em: 20 ago. 2023.

[10] ZAIDI, S. S. A.; ANSARi, M. S.; ASLAM, A.; KANWAL, N.; ASGHAR, M.; LEE, b. **A Survey of Modern Deep Learning based Object Detection Models.** arXiv:2104.11892 [cs, eess], 12 maio 2021. Disponível em: <https://arxiv.org/abs/2104.11892>. Acesso em: 29 Out. 2024.

[11] PÖLLABAUER, T. et al. **Fast Training Data Acquisition for Object Detection and Segmentation using Black Screen Luminance Keying.** Disponível em: <https://arxiv.org/abs/2405.07653>. Acesso em: 30 out. 2024.

[12] B. Adhikari, J. Peltom¨aki, S. B. Germi, E. Rahtu, H. Huttunen, **Effect of label noise on robustness of deep neural network object detectors**. International Conference on Computer Safety, Reliability, and Security, Springer, 2021, pp. 239–250.

[13] A. A. J´unior, L. H. d. S. Silva, B. J. Fernandes, G. O. Azevedo, S. C. Oliveira, **Learning from pseudo-labels: Self-training electronic components detector for waste printed circuit boards**. 2022 35th SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI), Vol. 1, IEEE, 2022, pp. 252–257.

[14] C. Wang, J. Shi, C. Tao, F. Xu, X. Tang, L. Li, Y. Zhou, B. Tian, S. Wei, X. Zhang, **Multitype label noise modeling and uncertainty-weighted label correction for concealed object detection**. IEEE Transactions on Instrumentation and Measurement 72 (2023) 1–12.

DOI: 10.25286/repa.v10i3.3122

[15] X. Li, W. Chen, D. Xie, S. Yang, P. Yuan, S. Pu Y. Zhuang, **A free lunch for unsupervised domain adap-
tive object detection without source data**. Proceed ings of the AAAI Conference on Artificial Intelligence, Vol. 35, 2021, pp. 8474–8481.

[16] H. Zhou, Z. Ge, S. Liu, W. Mao, Z. Li, H. Yu, J. Sun, **Dense teacher: Dense pseudo-labels for semi-supervised object detection**. European Conference on Computer Vision, Springer, 2022, pp. 35–50.

[17] LIN, Tsung-Yi; MAIRE, Michael; BELONGIE, Serge; BOURDEV, Lubomir; GIRSHICK, Ross; HAYS, James; PERONA, Pietro; RAMANAN, Deva; DOLLÁR, Piotr; ZITNICK, C. Lawrence. **Microsoft COCO: Common Objects in Context. In: FLEET, David;** PAJDLA, Tomas; SCHIELE, Bernt; TUYTELAARS, Tinne (eds.). Computer Vision – ECCV 2014. Cham: Springer, 2014. p. 740–755. (Lecture Notes in Computer Science, v. 8693). Disponível em: https://doi.org/10.1007/978-3-319-10602-1_48. Acesso em: 29 abr. 2025.

[18] EVERINGHAM, Mark; VAN GOOL, Luc; WILLIAMS, Christopher K. I.; WINN, John; ZISSERMAN, Andrew. T**he PASCAL Visual Object Classes (VOC) Challenge.** International Journal of Computer Vision, Dordrecht, v. 88, n. 2, p. 303–338, jun. 2010. Disponível em: https://doi.org/10.1007/s11263-009-0275-4. Acesso em: 29 abr. 2025.

[19] REZATOFIGHI, Hamid; TSOI, Nathan; GWAK, JunYoung; SADEGHIAN, Amir; REID, Ian; SAVARESE, Silvio. **Generalized Intersection over Union: A Metric and a Loss for Bounding Box Regression**. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2019. p. 658–666. Disponível em: https://doi.org/10.48550/arXiv.1902.09630. Acesso em: 29 abr. 2025.