

PONTIFÍCIA UNIVERSIDADE CATÓLICA DE SÃO PAULO
FACULDADE DE ESTUDOS INTERDISCIPLINARES
CURSO DE CIÊNCIA DE DADOS E INTELIGÊNCIA ARTIFICIAL

Ana Beatriz O. De Macedo

Bruna Bellini Faria

**APLICANDO A INTELIGÊNCIA ARTIFICIAL PARA ENRIQUECER OS
ESTÍMULOS VISUAIS NO DESENVOLVIMENTO COGNITIVO INFANTIL:
DETECÇÃO DE ANIMAIS EM VÍDEOS**

SÃO PAULO

2024

Ana Beatriz O. De Macedo

Bruna Bellini Faria

**APLICANDO A INTELIGÊNCIA ARTIFICIAL PARA ENRIQUECER OS
ESTÍMULOS VISUAIS NO DESENVOLVIMENTO COGNITIVO INFANTIL:
DETECÇÃO DE ANIMAIS EM VÍDEOS**

Trabalho de Conclusão de Curso apresentado à Faculdade de Estudos Interdisciplinares da Pontifícia Universidade Católica de São Paulo como requisito parcial para a obtenção do título de Bacharel em Ciência de Dados e Inteligência Artificial.

Orientador: Dr. Jefferson O. Silva

SÃO PAULO

2024

DEDICATÓRIA

Dedicamos este trabalho, primeiramente, aos nossos pais, por serem nossa base, por sempre nos apoiarem e ajudarem no que fosse preciso. Agradecemos aos nossos familiares, que sempre torceram por nós e se fizeram presentes em nossas vidas. Por último, agradecemos aos nossos amigos, que foram responsáveis por tornar tudo mais leve e fácil, que nos apoiaram e nos ajudaram a seguir em frente mesmo quando estava difícil. Sem vocês e sem o apoio e amor incondicional, não teríamos chegado até aqui. Agradecemos também ao nosso orientador, Prof. Dr. Jefferson O. Silva, por nos incentivar e apoiar.

"A análise de dados é o coração da transformação digital. É através dela que podemos tomar decisões baseadas em evidências e impulsionar a inovação."

Satya Nadella - CEO da Microsoft.

RESUMO

Este estudo explora a importância da diversidade de estímulos visuais no desenvolvimento cognitivo infantil, destacando como ela é fundamental para o crescimento saudável das crianças em seus primeiros estágios de vida. Um dos efeitos negativos da vida moderna é que muitas crianças não têm acesso a uma variedade adequada de estímulos visuais. A Inteligência Artificial (IA) pode oferecer soluções promissoras para esse problema. Apesar de a importância do tema ser reconhecida, poucas pesquisas têm investigado maneiras de integrar a IA para enriquecer esses estímulos. Este trabalho propõe um método para identificar em vídeos digitais, por meio da IA, elementos essenciais no desenvolvimento infantil. A pesquisa se concentra na implementação de um modelo que aprimora a rede neural You Only Look Once (YOLO), para identificar uma classe de animais: cachorros. Dessa forma, acreditamos contribuir para ampliar as opções de enriquecimento do desenvolvimento cognitivo das crianças nas fases mais cruciais de seu crescimento, até 2 anos. Além disso, desenvolvemos uma ferramenta com nosso modelo que oferece aplicações úteis para pais, educadores e desenvolvedores de conteúdo digital.

Palavras-chave: Desenvolvimento Cognitivo Infantil, Inteligência Artificial, Estímulo Visual.

ABSTRACT

This study explores the importance of visual diversity in children's cognitive development and how it is essential for healthy growth in the early years of life. One of the negative effects of modern life is that many children do not have access to an adequate variety of visual stimulation. Artificial Intelligence (AI) may offer promising solutions to this problem. Despite the recognized importance of this issue, few studies have investigated ways to integrate AI to enrich these stimuli. This work proposes a methodology to identify essential elements in digital videos for child development using AI. The research focuses on implementing a model that extends the You Only Look Once (YOLO) neural network to identify a class of animals: dogs. In this way, we aim to contribute to expanding the possibilities for enriching children's cognitive development during the most crucial stages of their growth, up to the age of 2. In addition, we have developed a tool with our model that offers useful applications for parents, educators, and digital content developers.

Keywords: Child cognitive development, artificial intelligence, visual stimulus.

SUMÁRIO

1. INTRODUÇÃO.....	8
2. TRABALHOS RELACIONADOS.....	11
3. MÉTODO.....	13
3.1 Arquitetura do Algoritmo de Visão Computacional.....	13
3.2 Identificação e Priorização das Categorias de Animais e Cachorro.....	14
3.3 Treinamento Inicial da Detecção de Cachorros.....	15
3.4 Coleta de Imagens Adicionais para o Retreinamento.....	15
3.5 Retreinamento do Algoritmo YOLO.....	16
3.6 Disponibilização do Modelo.....	16
4. RESULTADOS.....	17
5. DISCUSSÃO.....	35
6. LIMITAÇÕES.....	38
7. CONCLUSÃO.....	40
REFERÊNCIAS.....	43

1. INTRODUÇÃO

Nos últimos anos temos observado um crescimento significativo das plataformas de streaming e a proliferação do conteúdo digital voltado para crianças (Domoff, Borgen, Radesky, 2020). Portanto, a identificação e seleção de vídeos que ajudam no desenvolvimento cognitivo tornam-se questões cruciais. A falta de métodos eficazes para identificar elementos visuais que contribuem para o desenvolvimento cognitivo das crianças dificulta a implementação de soluções baseadas em Inteligência Artificial (IA) (American Academy of Pediatrics, 2017). Isso limita a oportunidade de aumentar o potencial de aprendizado em um ambiente digital, o que é particularmente importante, visto que muitos desses vídeos podem apresentar elementos que não auxiliam, da melhor maneira, o desenvolvimento cognitivo.

Embora a capacidade das crianças de reconhecer elementos seja uma habilidade inata, que se desenvolve naturalmente, essa habilidade pode ser constantemente aprimorada. No entanto, a ausência de exposição a uma diversidade de elementos pode resultar em dificuldades no reconhecimento de elementos fundamentais no ambiente em que vivem, levando a possíveis atrasos em seu desenvolvimento cognitivo.

Diante deste contexto, torna-se essencial utilizar esse tempo valioso com conteúdos que contribuam para o desenvolvimento das crianças. Portanto, mesmo com grande exposição, é fundamental trabalhar para garantir que essa exposição seja correta, ajudando a aumentar o desenvolvimento cognitivo em vez de diminuí-lo. Surpreendentemente, apesar das grandes possibilidades neste campo, percebe-se uma lacuna nos estudos recentes, destacando a necessidade de mais pesquisas e desenvolvimentos na área. Isso é fundamental para proporcionar acesso a conteúdo digital realmente proveitoso para o desenvolvimento infantil.

Estudos anteriores têm enfatizado a importância da exposição a elementos na infância (Flavell, 1999). Mas, muitas vezes, esses estudos não consideram como a diversidade dos elementos afeta o processo de reconhecimento das crianças. Mesmo que existam soluções, como programas educacionais, brinquedos educativos, literatura infantil e conteúdo digital, que buscam promover o desenvolvimento cognitivo das crianças por meio da exposição a elementos

diversos, elas nem sempre conseguem abordar de maneira completa a necessidade de uma ampla e correta variedade de estímulos visuais. Essa lacuna pode resultar em oportunidades perdidas para enriquecer o desenvolvimento cognitivo infantil.

Ademais, o estudo de Wang, Song e Krstic (2016) busca adquirir manualmente uma série de imagens de todos os tipos de animais, objetos e outros estímulos visuais para cobrir essa lacuna. No entanto, essa tarefa representa um desafio substancial e muitas vezes inviável. Isso ocorre porque temos grande diversidade do mundo real, e a coleta manual de imagens de todos estes elementos seria demorada e complexa. Além disso, a qualidade e autenticidade das imagens obtidas manualmente podem variar. Isso ressalta ainda mais a necessidade de soluções que possam automatizar e simplificar o processo de exposição a uma variedade significativa de estímulos visuais, a fim de enriquecer o desenvolvimento cognitivo das crianças de maneira eficaz e consistente.

A pesquisa em questão tem como foco a otimização do reconhecimento de elementos por crianças por meio da aplicação de visão computacional. O objetivo deste trabalho é implementar uma rede neural que seja capaz de reconhecer animais.

Aprimorar a qualidade do conteúdo audiovisual consumido por crianças, enquanto promovemos positivamente o desenvolvimento cognitivo delas. Nossa estratégia envolve a reconfiguração de uma rede neural, YOLO¹, especializada em identificar e classificar uma variedade de elementos vitais para o progresso cognitivo na infância. Além de uma interface prática e intuitiva destinada a auxiliar pais e educadores na escolha criteriosa de vídeos apropriados.

Nossa pesquisa adota um método organizado, iniciando com a revisão de estudos correlatos, abrangendo análises de elementos visuais em vídeos que demonstram impacto positivo no desenvolvimento cognitivo infantil. Além disso, exploramos investigações em visão computacional aplicada a vídeos no contexto do desenvolvimento infantil. Uma etapa fundamental do nosso estudo consiste na análise do treinamento de elementos pertinentes nas redes neurais YOLO, utilizando o conjunto de dados do *Common Objects in Context* (COCO)² pré-treinado para

¹ Link do YOLO: <https://arxiv.org/abs/1506.02640>

² Link do COCO: <https://cocodataset.org/>

avaliar sua eficácia e identificar classes já existentes. Posteriormente, concentramos nossos esforços no treinamento dessas redes para entender as classes com um desempenho inferior e, conseqüentemente, treiná-las novamente para terem uma melhor performance, com o objetivo de implementá-las em uma interface específica. Essa interface será capaz de analisar quadro a quadro os vídeos disponíveis no YouTube, identificando elementos considerados benéficos para o desenvolvimento cognitivo das crianças.

Com o uso dessa solução, podemos prever uma melhoria significativa na maneira como crianças até 2 anos e bebês consomem conteúdos digitais. Isso não apenas impactará positivamente o desenvolvimento cognitivo das crianças, mas também elevará a qualidade de sua experiência de aprendizado. Familiares, educadores, profissionais de saúde infantil e pesquisadores terão à disposição um sistema que auxilia na identificação de elementos visuais que promovam um crescimento saudável e o bem-estar das crianças, contribuindo assim para a melhora da cognição de crianças e bebês.

2. TRABALHOS RELACIONADOS

A pesquisa realizada por Salim *et al.* (2022) se baseia em estudos anteriores que examinaram a detecção de objetos como uma ferramenta para melhorar o aprendizado infantil. Esta pesquisa descreve o uso da arquitetura YOLOv3, com algumas modificações no modelo original. Este trabalho destaca a importância da detecção em tempo real e como ela pode ser usada em mídias educacionais para crianças. Em vez de focar apenas nas otimizações técnicas, nosso projeto busca ir além da simples detecção, analisando categorias específicas de elementos que podem estimular o desenvolvimento cognitivo das crianças.

Enquanto Rakison e Butterworth (1998) se concentram na manipulação tangível de objetos físicos, nossa pesquisa adota uma abordagem diferente ao incorporar tecnologias de Inteligência Artificial para investigar interações visuais com elementos digitais. Eles abordam o uso de partes de objetos na categorização inicial por bebês, enquanto nosso estudo se fundamenta na detecção de elementos. Nós queremos entender como a exposição a diferentes categorias de elementos em ambientes digitais pode influenciar o desenvolvimento cognitivo infantil.

Ademais, o estudo de Pereira *et al.* (2008) examina a exploração ativa de elementos por crianças em idade pré-escolar. Embora compartilhemos a inquietação sobre como as interações com elementos impactam o desenvolvimento cognitivo, nossa abordagem diverge ao integrar a detecção de elementos em vídeos infantis como metodologia de análise. Nosso objetivo é identificar quais categorias específicas de elementos apresentam maior propensão para o aprendizado cognitivo.

Em síntese, este projeto se destaca por seu foco na detecção de elementos, utilizando tecnologias de Inteligência Artificial, e pela integração de *insights* provenientes de estudos anteriores para orientar a seleção criteriosa de elementos em ambientes digitais de aprendizado infantil. A convergência dessas perspectivas proporciona uma base sólida para investigar a otimização da detecção de elementos, visando promover o desenvolvimento cognitivo de crianças em contextos educacionais digitais.

Analisando as pesquisas anteriores, ressalta-se a ausência de abordagens semelhantes até o momento, destacando a originalidade e relevância deste projeto para o avanço do entendimento sobre o impacto em um contexto moderno. Nesse cenário, desde o nascimento, as crianças são expostas à tecnologia por meio de vídeos no YouTube e outras plataformas de *streaming*. O objetivo consiste em contribuir para a compreensão de como esse meio digital específico pode ser otimizado para auxiliar no desenvolvimento cognitivo infantil.

3. MÉTODO

Nesta seção, apresentamos o método desta pesquisa. A Figura 1 ilustra o método que utilizamos para estruturar e solucionar o problema da pesquisa.

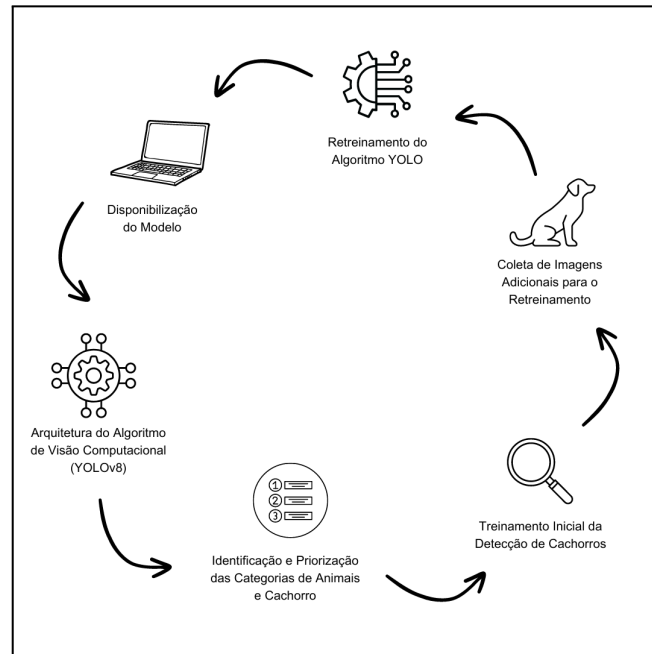


Figura 1 - Ilustração do método

Fonte: Elaborado pelas autoras (2024)

Nas próximas seções descrevemos em detalhes cada passo do método.³

3.1 Arquitetura do Algoritmo de Visão Computacional

Utilizamos o algoritmo YOLO devido à sua eficiência como um algoritmo de visão computacional avançado para treinar um modelo de reconhecimento de elementos relacionados ao estímulo visual infantil. Mais especificamente, o YOLO é um método de detecção de objetos de passada única que utiliza uma rede neural convolucional (CNN) como extrator de características em imagens e vídeos. Escolhemos este método baseado na alta precisão e rapidez para detecção de elementos diversos, além do potencial de aprimoramento em certas aplicações. Utilizamos a oitava e mais atual versão do YOLO à época da escrita desse trabalho: YOLOv8.

³ Link para o Google Drive contendo toda a implementação da metodologia:
<https://drive.google.com/drive/folders/1BcIIsrOT9ALukV7PUG9L67WK4FOftwy>

3.2 Identificação e Priorização das Categorias de Animais e Cachorro

A decisão de priorizar a detecção de animais antes de brinquedos e objetos domiciliares foi fundamentada em duas razões principais. Primeiramente, brinquedos e objetos domiciliares abrangem uma variedade muito ampla e diversificada de itens. Essa complexidade torna a identificação precisa e a categorização desses elementos um desafio significativo para modelos de Inteligência Artificial, como o YOLOv8, especialmente considerando as limitações de tempo e recursos de um trabalho de conclusão de curso.

Além disso, a escolha de priorizar a detecção de animais é respaldada por pesquisas que indicam a importância da exposição visual a animais no desenvolvimento cognitivo infantil. Estudos de Chouinard, Harris e Maratsos (2007), bem como de Bambach et al. (2020), demonstraram que a visualização e observação de animais permitem que as crianças estabeleçam associações rápidas e significativas, oferecendo uma base sólida para o aprimoramento cognitivo inicial.

Iniciamos nossa pesquisa de aprimoramento das classes específicas de animais no dataset do COCO, que contém 10 classes de animais (pássaros, cachorros, ovelhas, elefantes, zebras, gatos, cavalos, vacas, ursos e girafas) já pré treinadas pelo YOLO no seu modelo base.

Para melhorar a precisão das detecções, coletamos 200 novas imagens que não fazem parte do dataset COCO, 20 imagens para cada uma das 10 classes de animais já presentes no COCO. A seleção dessas imagens foi feita com o objetivo de garantir uma variedade de cores, ângulos e proximidades, proporcionando uma representação abrangente e diversificada.

Para realizar estas previsões, configuramos o modelo de modo que rejeitasse qualquer classificação que não se enquadra nas 10 classes de animais definidas pelo COCO. Estabelecemos um limite mínimo de confiança de 50% e um limite máximo de 2 detecções por foto, para garantir precisão e relevância nas detecções.

Após essa análise decidimos focar especificamente na classe de cachorros. Embora a classe de cachorros já apresentasse uma média de confiança de 89%, percebemos que havia margem para melhoria, especialmente em relação ao reconhecimento de diferentes raças e variações dentro da classe.

3.3 Treinamento Inicial da Detecção de Cachorros

Para treinar o YOLOv8, e conseguirmos atingir um treinamento semelhante do Ultralytics, carregamos o modelo pré-treinado. Utilizamos a biblioteca FiftyOne para baixar todas as imagens de cachorros dos conjuntos de treino e validação do COCO, totalizando 4385 imagens. Essa abordagem visou reduzir o tempo de treinamento e o processamento do modelo, além de minimizar a necessidade de recursos computacionais avançados.

A preparação das imagens no Roboflow e a escolha do modelo YOLOv8m para o retreinamento foram baseadas na sua complexidade moderada, que inclui 25,9 milhões de parâmetros e executa 78,9 bilhões de operações de ponto flutuante (FLOPs). O processo de retreinamento foi conduzido ao longo de 100 épocas, consideradas apropriadas devido ao volume moderado de novas imagens por classe. Essa quantidade de épocas foi selecionada para otimizar o aprendizado sem ocorrer sobreajuste.

Para o treinamento de um modelo de detecção de elementos com YOLOv8, utilizamos os principais argumentos padrão do YOLO, exceto pelo número de épocas, que foi ajustado para 100.

3.4 Coleta de Imagens Adicionais para o Retreinamento

Para aprimorar a detecção de cachorros, identificamos áreas específicas que necessitavam de melhorias, como a identificação de raças menos comuns e a detecção de expressões variadas. Observamos que o modelo apresentava menor confiabilidade na identificação de raças como o Spitz Alemão e dificuldades com imagens de baixa resolução ou com expressões variadas.

Adicionamos 1026 novas imagens não presentes no dataset COCO, coletadas de fontes como Roboflow, Kaggle e Google, garantindo a conformidade com os direitos autorais. Essas imagens foram divididas em categorias estratégicas para abordar lacunas identificadas no treinamento anterior, como emoções caninas, raças raras e cenários variados, além de variedade em cores, ângulos e proximidades, enriquecendo o conjunto de dados e melhorando a representatividade

e diversidade dos dados de treinamento. Utilizamos o Roboflow para organizar e preparar as imagens, tornando-as compatíveis para o treinamento com o YOLOv8.⁴

3.5 Retreinamento do Algoritmo YOLO

Para o aprimoramento da classe "cachorro", iniciamos um novo treinamento básico utilizando apenas as novas imagens, sem incluir as 4 mil imagens do dataset COCO. Essa abordagem evitou o sobreajuste, pois os pesos utilizados eram do modelo treinado anteriormente com o dataset completo do COCO. Focamos em novas imagens para que o modelo aprendesse novas informações sem a influência repetida das imagens antigas.

Utilizamos todas as novas imagens anotadas através da ferramenta Roboflow e aumentamos o número de épocas para 200, garantindo tempo suficiente para o modelo se ajustar às novas imagens e melhorar sua capacidade de detecção. Implementamos um parâmetro de paciência de 50 e aumentamos o tamanho do lote de 16 para 32.

Com esses ajustes, o modelo foi treinado mantendo os demais parâmetros em default. A metodologia final garantiu um treinamento eficaz, resultando em melhorias substanciais na precisão e robustez da detecção de cachorros.

3.6 Disponibilização do Modelo

Por fim, desenvolvemos uma aplicação online utilizando a plataforma Streamlit, permitindo aos usuários inserir URLs de vídeos do YouTube e visualizar as principais estatísticas relacionadas à detecção de elementos. Esta aplicação exibe métricas como a média de detecções por frame, o número máximo de detecções em um único frame, e o total de detecções ao longo do vídeo, proporcionando uma interface prática para pais, educadores e pesquisadores.

⁴ Link do YOLOv8: <https://roboflow.com/model/yolov8>

4. RESULTADOS

Com base na aplicação do método, foi determinado que o treinamento inicial se concentraria nas categorias relacionadas a animais. Importante destacar que os animais desempenham um papel significativo no desenvolvimento cognitivo das crianças, mesmo quando a interação ocorre apenas por meio da observação visual. Os estudos demonstraram que a visualização e observação de animais permitem que as crianças estabeleçam associações rápidas e significativas, oferecendo uma base sólida para o aprimoramento cognitivo inicial (Chouinard, Harris e Maratsos, 2007; Bambach et al., 2020).

Essa decisão foi respaldada também pelo desempenho superior dessas categorias durante as fases de testes iniciais, além da comum utilização de imagens de animais em estudos científicos para detecção e da diversidade e complexidade visual para estudos. Durante as fases experimentais, ainda que algumas classes específicas não estivessem diretamente representadas no dataset COCO, as iterações do modelo YOLO exibiram tendências de correlacionar entidades visualmente similares a categorias conhecidas. Esta tendência é ilustrada na associação de lobos e raposas à categoria 'Cachorro', evidenciando a capacidade do modelo de generalizar a partir de características visuais compartilhadas.

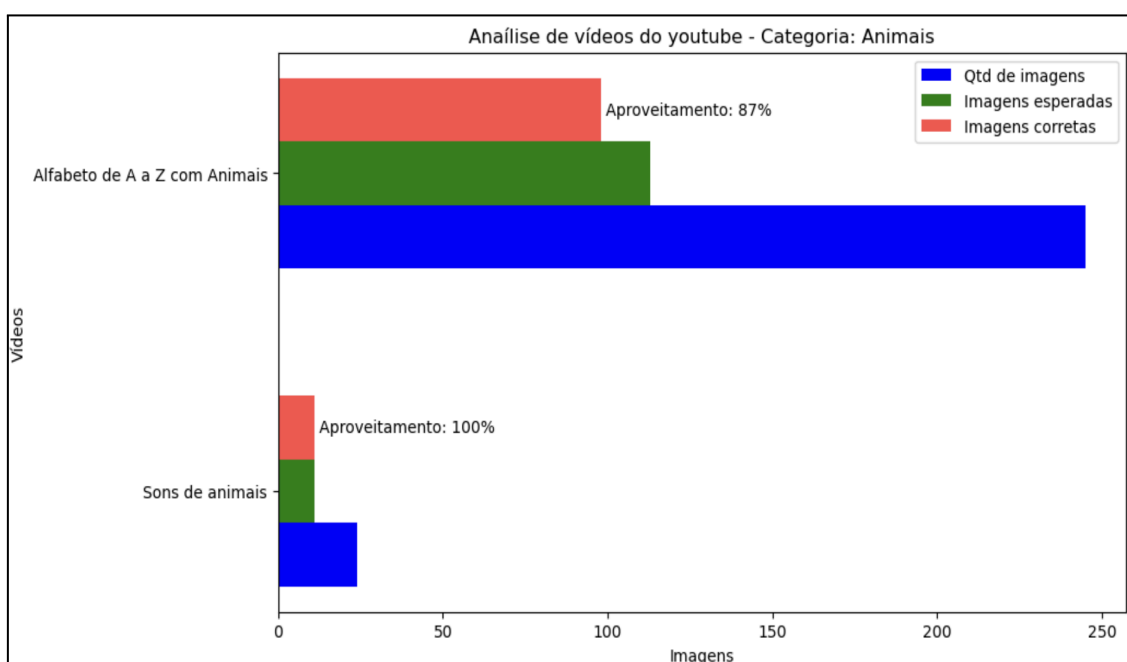


Figura 2 - Análise de vídeos do YouTube da categoria Animais

Fonte: Elaborado pelas autoras (2024)

Os resultados, sumarizados no Figura 2, exibem a competência do modelo em distinguir e corretamente classificar imagens de animais. A métrica de porcentagem de aproveitamento, definida como a proporção entre detecções corretas e o número total de imagens onde detecções eram antecipadas, conforme a equação 1, reflete a precisão do modelo em relação às expectativas fundamentadas na presença de classes no *dataset* COCO. O desempenho superior em imagens de animais, com um aproveitamento de 100% para o Vídeo 1, KidsTV (2016) e 87% para o Vídeo 2, Aprenda Brincando (2019), sugere uma robustez particular do modelo para esta categoria, apesar das limitações impostas pela ausência de classes específicas. Tal resultado reforça a pertinência da escolha de animais como foco inicial, tendo em vista o potencial de generalização do modelo para classes similares e sua aplicabilidade em cenários reais de detecção.

$$Aproveitamento\% = \frac{\text{Imagens corretamente identificadas}}{\text{Total de imagens esperadas}} \times 100 \quad (1)$$

Os resultados da terceira análise de vídeo com o uso do YOLOv8 também revelaram padrões similares aos observados anteriormente nas categorias de objetos de casa, brinquedos e animais, reforçando a relevância da diversidade de elementos no desenvolvimento cognitivo infantil. Na análise, a "média de confiança" refere-se à precisão com que o modelo de Inteligência Artificial identifica e classifica corretamente um objeto dentro de uma categoria específica. Esta métrica é crucial para avaliar a eficácia do modelo em reconhecer e entender diferentes elementos.

Segmento	Média de confiança
Objetos da casa	0,52
Brinquedos	0,46
Animais	0,50

Figura 3. Resultados de média de confiança por categoria

Visto os resultados acima, a decisão de priorizar a detecção de animais antes de brinquedos e objetos domiciliares se fundamenta em duas razões principais. Primeiramente, brinquedos e objetos domiciliares abrangem uma variedade muito

ampla e diversificada de itens, podendo existir milhões de variações distintas em termos de cores, formas e formatos. Essa complexidade torna a identificação precisa e a categorização desses elementos um desafio significativo para modelos de Inteligência Artificial, como o YOLOv8, especialmente considerando as limitações de tempo e recursos de um trabalho de conclusão de curso.

Além disso, a escolha de priorizar a detecção de animais é respaldada pela pesquisa de **Chouinard**, Harris e Maratsos (2007), mencionada na seção anterior do trabalho. Este estudo ressalta a importância da exposição visual a animais no desenvolvimento cognitivo infantil, indicando que a interação com imagens de animais pode contribuir significativamente para o processo de aprendizagem e formação de associações cognitivas nas crianças.

Com isso, iniciamos nossa pesquisa e aprimoramento das classes específicas de animais no dataset do COCO. Atualmente, ele inclui 10 classes de animais já treinados, como pássaros, cachorros, ovelhas, elefantes, zebras, gatos, cavalos, vacas, ursos e girafas. Nosso objetivo é melhorar o desempenho na previsão de uma dessas classes, e também explorar a adição de uma nova classe ao modelo. Observamos um desempenho variado entre essas classes, o que nos levou a investigar estratégias de aprimoramento contínuo.

Primeiramente, coletamos 200 imagens que não fazem parte do dataset do COCO, com o objetivo de avaliar a média de confiança obtida, assim como na previsão de vídeos para objetos de casa, brinquedos e animais anteriormente. Incluímos 20 imagens adicionais para cada uma das 10 classes já presentes no *dataset* COCO. A escolha dessas novas classes se baseia em estudos, que demonstram benefícios no desenvolvimento cognitivo das crianças quando expostas a vídeos desses animais (Guellai *et al.*, 2022), seja para despertar a curiosidade e o interesse em aprender mais sobre a vida, hábitos e os habitats ou expandir o conhecimento das crianças sobre comportamentos animais para aspectos importantes no desenvolvimento emocional e social, além de estimular a imaginação.

Cada conjunto de 20 imagens para cada classe foi selecionado com cuidado, visando uma variedade de cores, ângulos e proximidades para garantir uma representação abrangente. Nosso objetivo foi incluir tanto imagens dos animais

individualmente quanto em grupo, proporcionando uma diversidade de contextos. Optamos por imagens simples e identificáveis, tanto para facilitar a análise computacional quanto para a percepção humana. Esse cuidado meticuloso permitiu-nos escolher cada classe de maneira precisa para futuras melhorias e refinamentos.

Para realizar estas previsões e conseguir o cenário mais equilibrado para as detecções das distintas classes, configuramos o modelo de modo que rejeitasse qualquer classificação que não se enquadra nas 10 classes de animais definidas pelo YOLO. Isso nos permitiu focar exclusivamente na análise das classes de animais. Além disso, estabelecemos um limite mínimo de confiança de 50% para evitar que o modelo detectasse classes muito distintas da realidade. Outro aspecto foi a imposição de um limite máximo de 2 detecções por foto. Isso se mostrou necessário porque, mesmo com a montagem dos *datasets*, encontramos casos em que uma classe continha imagens com vários animais em diferentes distâncias. Essa variação compromete a confiabilidade geral da detecção da classe. Limitar o número de detecções por foto permitiu que o modelo concentrasse nas detecções mais relevantes, melhorando a precisão e a classificação.

Classe de animal	Média de confiança	Aproveitamento
Passaro (<i>Bird</i>)	0,89	100%
Urso (<i>Bear</i>)	0,91	90,91%
Gato (<i>Cat</i>)	0,91	95,45%
Vaca (<i>Cow</i>)	0,90	93,94%
Cachorro (<i>Dog</i>)	0,89	100%
Elefante (<i>Elefant</i>)	0,92	96,77%
Girafa (<i>Giraffe</i>)	0,94	96,77%
Cavalo (<i>Horse</i>)	0,91	100%
Ovelha (<i>Sheep</i>)	0,90	100%
Zebra (<i>Zebra</i>)	0,94	100%

Figura 4. Resultados de média de confiança por classe de animal

Assim, o foco na classe cachorros durante o retreinamento, apesar de já apresentar uma média de confiança de 89%, se dá pela ainda existente margem para melhoria se comparado às outras classes, especialmente porque certas raças de animais não foram adequadamente reconhecidas pelo modelo original do COCO. Além disso, consideramos necessário expandir e diversificar o conjunto de dados com imagens que incluam diferentes ângulos, cores e cenários com múltiplos animais. Essa estratégia visa melhorar significativamente o desempenho da detecção da classe de cachorros, garantindo uma classificação mais precisa e robusta.

Para treinar o YOLOv8 com novas classes sem perder as antigas, é necessário combinar o *dataset* antigo com o novo, modificar a configuração do modelo para incluir todas as classes, carregar o modelo pré-treinado e então treiná-lo novamente com os dados combinados. Com isso, utilizamos a biblioteca FiftyOne⁵ para baixar todas as imagens de cachorros dos conjuntos de treino e validação do COCO (4385 imagens). Optamos por selecionar apenas as imagens de cachorros para otimizar nosso treinamento, utilizando uma quantidade menor de imagens e focando em uma única classe. Essa abordagem visa reduzir o tempo de treinamento e o processamento do modelo, além de minimizar a necessidade de recursos computacionais avançados. Essa escolha é motivada pelas limitações de recursos de processamento disponíveis, permitindo um treinamento eficiente e viável dentro das condições estabelecidas.

A preparação das imagens no Roboflow e a escolha do modelo YOLOv8m para o retreinamento foram baseadas na sua complexidade moderada, que inclui 25,9 milhões de parâmetros e executa 78,9 bilhões de operações de ponto flutuante (FLOPs). Essa característica permite uma rápida execução em GPUs de alto desempenho. O processo de retreinamento foi conduzido ao longo de 100 épocas, consideradas apropriadas devido ao volume moderado de novas imagens por classe. Essa quantidade de épocas foi selecionada para otimizar o aprendizado sem incorrer em sobreajuste.

⁵ Link para FiftyOne: https://docs.voxel51.com/getting_started/install.html

Para o treinamento de um modelo de detecção de elementos com YOLOv8, foram utilizados os principais argumentos padrão do YOLO, exceto pelo número de épocas. Esses argumentos incluem: a tarefa de detecção (*task: detect*), modo de treinamento (*mode: train*), modelo pré-treinado *yolov8m.pt*, dados de treinamento especificados em *dog-detection-3/data.yaml*, 100 épocas de treinamento (*epoch: 100*), paciência de 50 épocas (*patience: 50*), tamanho de lote de 16 (*batch: 16*), resolução das imagens de 640 pixels (*imgsz: 640*), uso de pesos pré-treinados (*pretrained: true*), otimizador automático definido pelas necessidades específicas do treinamento e as características do conjunto de dados e do modelo (*optimizer: AdamW*), taxa de aprendizado inicial de 0,01 (*lr0: 0,01*), momentum de 0,937 (*momentum: 0,937*), decaimento de peso de 0,0005 (*weight_decay: 0,0005*), uso de AMP (*Mixed Precision*) (*amp: true*), e a métrica IoU (*Intersection over Union*) configurada para 0,7 (*iou: 0,7*).

Durante o desenvolvimento do modelo de detecção de cachorros utilizando YOLOv8, após 100 épocas de treinamento com o otimizador AdamW, alcançamos resultados promissores, conforme evidenciado por várias métricas e visualizações gráficas.

A Figura 5 apresenta a curva de F1-Confiança, onde observamos que a métrica F1, que é a média harmônica entre precisão e *recall*, alcança um valor elevado. Este valor se mantém alto em diferentes níveis de confiança, indicando um equilíbrio excelente entre a capacidade do modelo de detectar corretamente os cachorros (*recall*) e de precisar corretamente essas detecções (precisão).

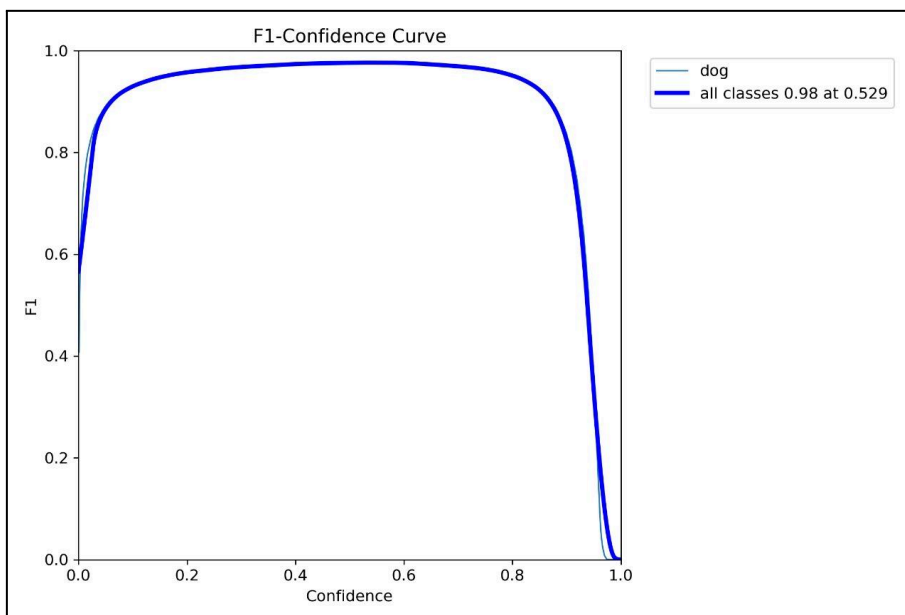


Figura 5 - Curva de confiança do F1 no treinamento inicial da classe de cachorro com YOLOv8

Fonte: Elaborado pelas autoras (2024)

A Figura 6 exibe várias curvas de perdas e métricas durante o treinamento e validação. Observamos que as curvas de *recall* e precisão indicam uma performance robusta do modelo. Mesmo para valores altos de confiança, o modelo mantém uma taxa elevada de *recall* e precisão, sugerindo que ele é eficaz na detecção de cachorros com poucos falsos negativos e positivos. Isso demonstra a habilidade do modelo em identificar corretamente as instâncias de cachorros nas imagens, mesmo sob critérios rigorosos de confiança.

As curvas de perda, tanto para *box_loss*, *cls_loss* e *dfl_loss*, mostram um decréscimo consistente e convergem para valores baixos. Especificamente, na melhor época de treinamento (época 100), os valores de perda foram os seguintes: *train/box_loss*: 0,44931, *train/cls_loss*: 0,26949, e *train/dfl_loss*: 0,9211. Esses valores indicam que o modelo está aprendendo de forma eficiente e ajustando bem os parâmetros durante o treinamento. As perdas na validação também foram baixas, com *val/box_loss*: 0,41403, *val/cls_loss*: 0,22523, e *val/dfl_loss*: 0,85282, sugerindo uma boa generalização do modelo para dados não vistos.

Além disso, as métricas de precisão e *recall* do modelo são muito elevadas. A precisão (*metrics/precision(B)*) alcançou 0,98865 e o *recall* (*metrics/recall(B)*) foi de 0,96425 na melhor época. Esses valores altos indicam que o modelo é capaz de

identificar corretamente a maioria das instâncias de cachorros, minimizando tanto falsos positivos quanto falsos negativos. As métricas mAP50 e mAP50-95 são cruciais para avaliar modelos de detecção de elementos. O mAP (*Mean Average Precision*) avalia a precisão média em diferentes limiares de IoU (*Intersection over Union*), indicando a sobreposição entre as áreas prevista e real do objeto. Embora o padrão comum seja um IoU de 0,5, neste projeto utilizamos um IoU de 0,7 como critério principal, exigindo maior precisão nas detecções. O mAP50-95 expande essa análise para IoUs de 0,5 a 0,95, oferecendo uma visão detalhada da precisão do modelo em níveis variados de exigência. Com um mAP50 de 98,8% e um mAP50-95 de 90,7%, os resultados demonstram ótima precisão e robustez do modelo em detectar cachorros de forma eficaz.

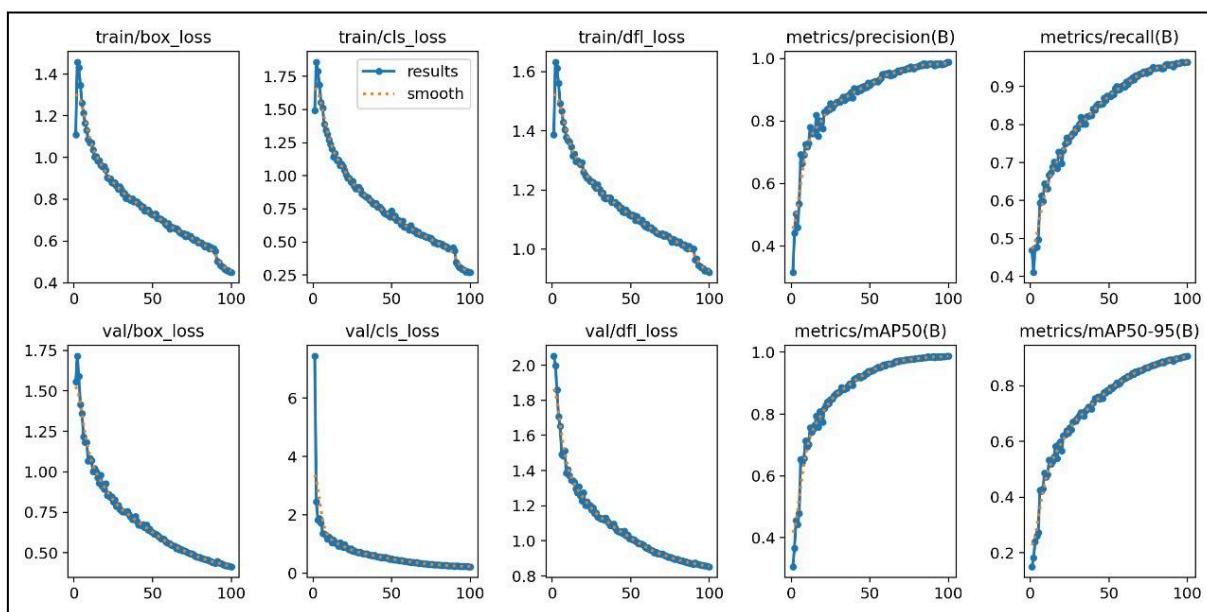


Figura 6 - Métricas de avaliação do treinamento inicial da classe de cachorro com YOLOv8

Fonte: Elaborado pelas autoras (2024)

A comparação desses resultados com os obtidos utilizando o YOLO já treinado pela Ultralytics⁶ mostra similaridades, como podemos observar na Figura 7 e 8. Na previsão da classe do cachorro com YOLOv8, alcançamos inicialmente uma média de confiança de 0,89 e um aproveitamento das imagens de 100%. Nosso novo treinamento, embora tenha registrado uma média de confiança menor de 0,88, manteve um aproveitamento das imagens de 100%. Este desempenho, alcançado apesar de termos recursos e tempo de treinamento mais limitados que os

⁶ link do Ultralytics: <https://www.ultralytics.com/>

disponíveis para o treinamento original do YOLO, é considerado um resultado ótimo e demonstra a eficácia do nosso treinamento especializado. Esse resultado reforça a competência do nosso modelo para competir com soluções pré-existentes, com uma eficiência similar na identificação de cachorros, e destaca as áreas para futuras melhorias, especialmente na otimização da confiança nas previsões enquanto mantemos um alto nível de aproveitamento das imagens.

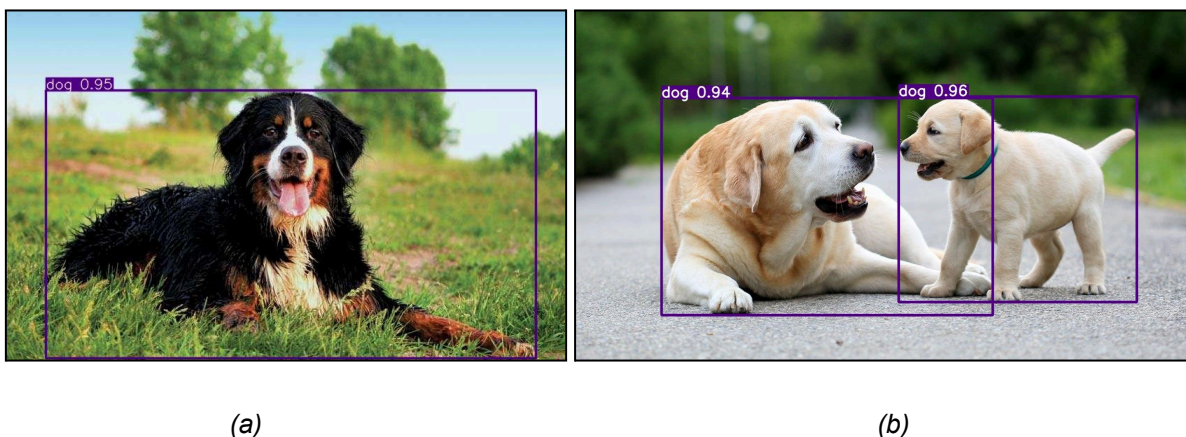


Figura 7. (a) Resultado previsão com YOLOv8 cachorro na grama; (b) Resultado previsão com YOLOv8 dois cachorros no asfalto

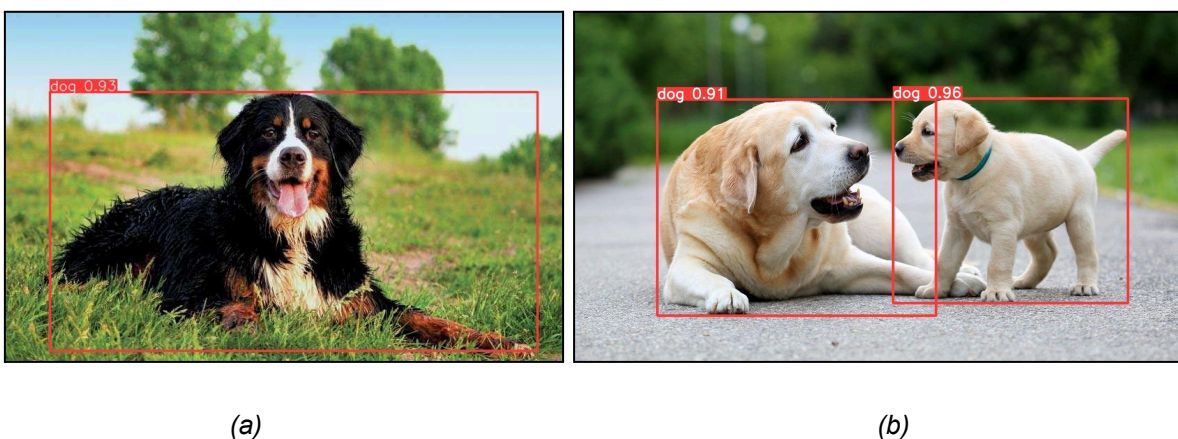


Figura 8. (a) Resultado com treinamento YOLOV8 cachorro na grama; (b) Resultado treinamento com YOLOV8 dois cachorros no asfalto

Assim, como mencionado anteriormente, apesar dos resultados serem bons, ainda existem oportunidades significativas para aprimoramentos. A análise de um conjunto de 20 imagens utilizadas em nossas previsões destaca algumas áreas específicas para melhoria. Por exemplo, observamos que o modelo apresenta menor confiabilidade na identificação da raça Spitz Alemão em comparação com outras raças, como pode ser visto na Figura 9. Isso sugere a necessidade de enriquecer

nosso conjunto de dados com mais variações dessa e de outras raças menos comuns.

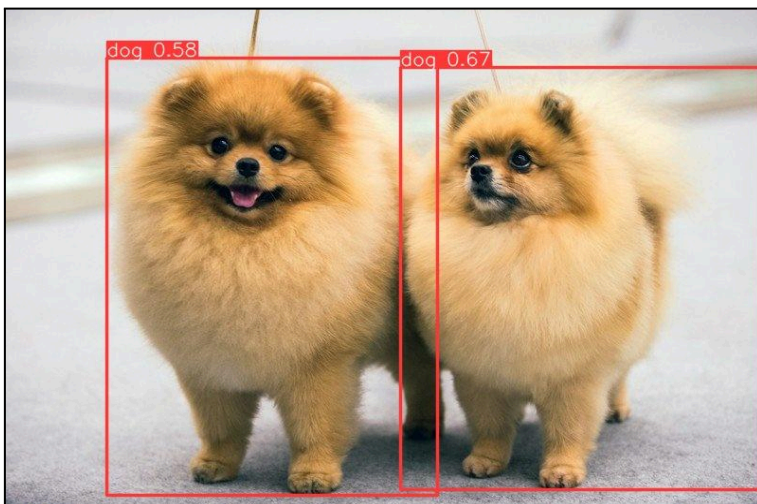


Figura 9. Resultado da previsão de imagens de Spitz Alemão do primeiro treinamento

Além disso, em situações com imagens de resolução mais baixa e onde os cachorros apresentam expressões mais variadas, como visto na Figura 10, destaca-se a importância de incluir imagens com essas características durante o treinamento para melhorar a robustez do modelo.



Figura 10. Resultado da previsão de imagens de cachorro sorrindo do primeiro treinamento

Para enfrentar esses desafios, considerar ajustes finos nas configurações do otimizador e nos hiperparâmetros pode ser benéfico. Tais ajustes poderiam refinar a capacidade do modelo de ajustar-se a variações sutis nas imagens, aprimorando a

precisão geral nas detecções. Em suma, esses passos não só aumentarão a precisão em condições desafiadoras mas também ampliarão a generalização do modelo para identificar corretamente cachorros em um cenário mais abrangente de situações e condições visuais.

Para o novo treinamento do nosso modelo YOLOv8, escolhemos manter o otimizador AdamW, uma decisão estratégica baseada em sua eficácia comprovada em treinamentos anteriores e sua adequação para tarefas complexas de visão computacional como a detecção de cachorros nas imagens. O otimizador AdamW é uma variante do tradicional Adam, conhecido por combinar as vantagens do *RMSprop* e do *Stochastic Gradient Descent com Momentum*. O AdamW modifica a forma como as correções de peso são aplicadas, integrando a decaída do peso diretamente no cálculo dos gradientes, o que ajuda a prevenir o ajuste excessivo durante o treinamento — uma preocupação particularmente relevante dado nosso amplo conjunto de dados de 5411 imagens.

Optamos por uma divisão de 90% para treinamento e 10% para validação para aumentar o uso dos dados disponíveis, assegurando uma quantidade substancial de exemplos para o aprendizado do modelo, enquanto ainda reservamos uma parcela significativa para validação e ajustes finos. Essa proporção apoia a eficiência do AdamW, facilitando uma convergência mais rápida e estável, essencial para lidar com a variação natural encontrada em uma base de dados com múltiplas representações de cachorros.

Além disso, adicionamos 1026 novas imagens, que não estavam presentes no *dataset* COCO, para enriquecer o conjunto de dados. Essas imagens foram cuidadosamente coletadas de diferentes fontes, como Roboflow, Kaggle e imagens publicadas no Google, respeitando rigorosamente os direitos autorais e garantindo que não sejam utilizadas para divulgação ou fins lucrativos. Essas novas imagens foram estrategicamente selecionadas para abordar lacunas específicas identificadas no treinamento anterior, garantindo uma melhor representatividade e diversidade dos dados.

Essas imagens foram divididas em categorias específicas: duas delas focam em capturar emoções de cachorros, variando entre raiva e felicidade, para melhorar a precisão do modelo na detecção de expressões caninas intensas. Outras quatro

categorias são dedicadas a raças de cachorros raras ou pouco representadas no dataset COCO, como Poodle, Spitz Alemão, Schnauzer e Chow Chow. A inclusão dessas raças visa aprimorar a acurácia do modelo na identificação dessas raças e enriquecer o conteúdo digital para o público infantil, promovendo o aprendizado sobre diversidade, características, temperamentos e cuidados específicos de cada raça.

Também optamos por abranger alguns cenários variados, como múltiplos cachorros, cachorros em camas, cachorros agachados, cachorros segurando ou cercados por brinquedos, cachorros deitados em diferentes cenários, e cachorros com roupas. Isso permitiu corrigir detecções de *bounding box* e melhorar a confiança do modelo em situações diversas. Essas adições foram estratégicas para melhorar a robustez e a precisão do modelo, garantindo uma melhor performance na detecção de emoções, raças raras e situações complexas envolvendo cachorros.

Começamos com um novo treinamento básico, utilizando apenas as novas imagens, sem as 4 mil imagens do dataset COCO. Decidimos por essa abordagem porque os pesos que estávamos utilizando já eram do modelo treinado anteriormente. Ao manter a mesma classe no treino e treinar o modelo exclusivamente com as novas imagens, conseguimos evitar o sobreajuste que poderia ocorrer se o modelo revisse as mesmas imagens novamente. Essa estratégia se mostrou eficaz, pois permitiu que o modelo se concentrasse em aprender as novas informações sem ser influenciado repetidamente pelas imagens antigas.

Para realizar o aprimoramento da classe "cachorro", utilizamos o arquivo de melhores pesos do primeiro treinamento, que foi similar ao treinado pela Ultralytics com o *dataset* completo do COCO. Utilizamos todas as novas imagens anotadas através da ferramenta Roboflow para este treinamento. Decidimos passar o número de épocas para 200 para garantir que o modelo tivesse tempo suficiente para ajustar-se às novas imagens e melhorar sua capacidade de detecção.

Essa decisão foi baseada na complexidade adicional trazida pela inclusão das novas raças e situações específicas, que requerem mais tempo de treinamento para alcançar uma convergência estável e eficaz. Implementamos um parâmetro de *patience* de 50, permitindo que o treinamento fosse interrompido automaticamente

se não houvesse um melhor resultado nas métricas de mAP50 e mAP50-95 nas últimas 50 épocas. Com essa configuração, conseguimos garantir que o treinamento não continuasse indefinidamente sem ganhos substanciais, economizando tempo e recursos computacionais.

Antes de encontrar a melhor configuração, tentamos algumas outras abordagens de aprimoramento, nas quais vimos melhoria muito pequena ou nenhuma na confiança das detecções das imagens, principalmente nas antigas, cujos modelos eram insuficientes para permitir que se adaptasse e identificasse novos conhecimentos de forma eficaz.

Com todos esses testes, treinamos o modelo com 200 épocas, mantendo todos os parâmetros em *default* e fazendo uma única mudança: aumentar nosso tamanho do lote (*Batch Size*) de 16 para 32. Aumentar o *batch size* melhorou o desempenho do modelo porque permitiu que o modelo atualizasse seus parâmetros com uma quantidade maior de dados em cada iteração. Isso pode resultar em um gradiente de atualização mais estável e representativo, reduzindo a variabilidade entre as atualizações e permitindo que o modelo mudasse de forma mais eficiente. Com um tamanho de lote maior, o modelo exigiu mais memória da GPU, o que resultou em um tempo de treinamento maior, mas possibilitou melhorias significativas na qualidade das detecções.

Concluindo este treinamento, já conseguimos ver uma melhora nas métricas de avaliação. Observando primeiramente o gráfico da métrica F1, podemos notar que, neste modelo aprimorado, a curva é mais plana e mantém valores elevados em uma gama maior de níveis de confiança (Figura 11). Isso indica que o modelo consegue manter a performance de maneira consistente e robusta, mesmo sob critérios rigorosos.

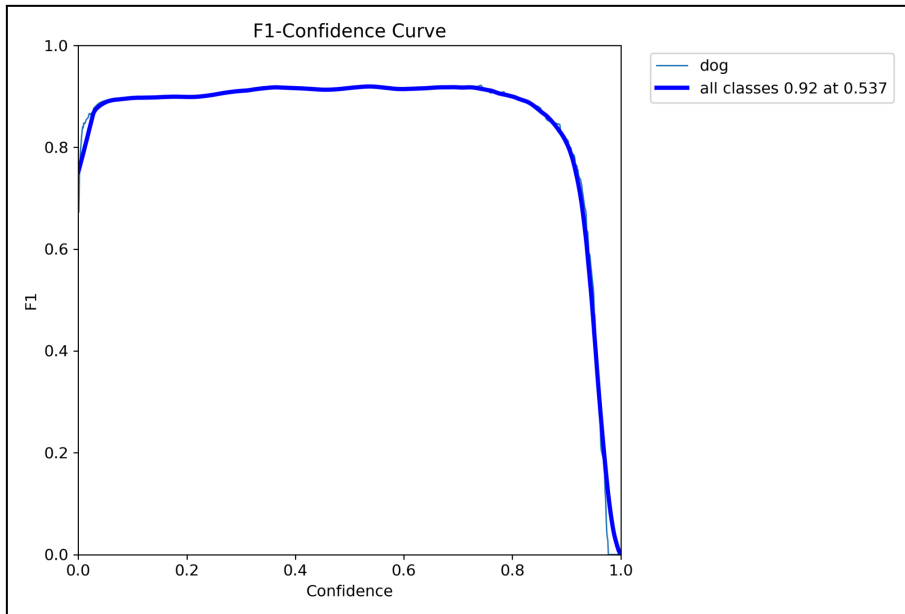


Figura 11 - Curva de confiança do F1 no retreinamento da classe de cachorro com YOLOv8

Fonte: Elaborado pelas autoras (2024)

As curvas de perda no treinamento exibem uma diminuição consistente e rápida, convergindo para valores ainda mais baixos de maneira mais estável (Figura 12). Especificamente, a perda de box (*train/box_loss*) atinge um valor de 0,22198, a perda de classificação (*train/cls_loss*) chega a 0,16489, e a perda de dfl (*train/dfl_loss*) é de 0,87971 na melhor época de treinamento (época 198). Esses valores baixos indicam que o modelo está aprendendo de forma eficiente e ajustando bem os parâmetros durante o treinamento.

Além disso, as perdas na validação também mostram uma diminuição significativa. A perda de box (*val/box_loss*) é de 0,6213, a perda de classificação (*val/cls_loss*) é de 0,38038, e a perda de dfl (*val/dfl_loss*) é de 1,1864. Essas reduções nas perdas de validação sugerem uma melhor generalização do modelo e maior eficiência no aprendizado.

As curvas de precisão (*metrics/precision(B)*) e *recall* (*metrics/recall(B)*) do modelo aprimorado apresentam valores mais altos e menores flutuações. Especificamente, a precisão atinge 0,94256 e o *recall* chega a 0,89304 na melhor época. Isso indica que o modelo é mais robusto e confiável, sendo capaz de detectar corretamente os verdadeiros positivos e mantendo uma taxa de acertos acima de 80%.

As métricas de mAP50 e mAP50-95 demonstram melhorias significativas, com valores mais elevados e variações menores. O mAP50 atinge 0,96433 e o mAP50-95 chega a 0,82627, reforçando a precisão e robustez do modelo. Esses valores indicam que o modelo consegue manter uma boa precisão na detecção de cachorros em diferentes limiares de IoU (*Intersection over Union*).

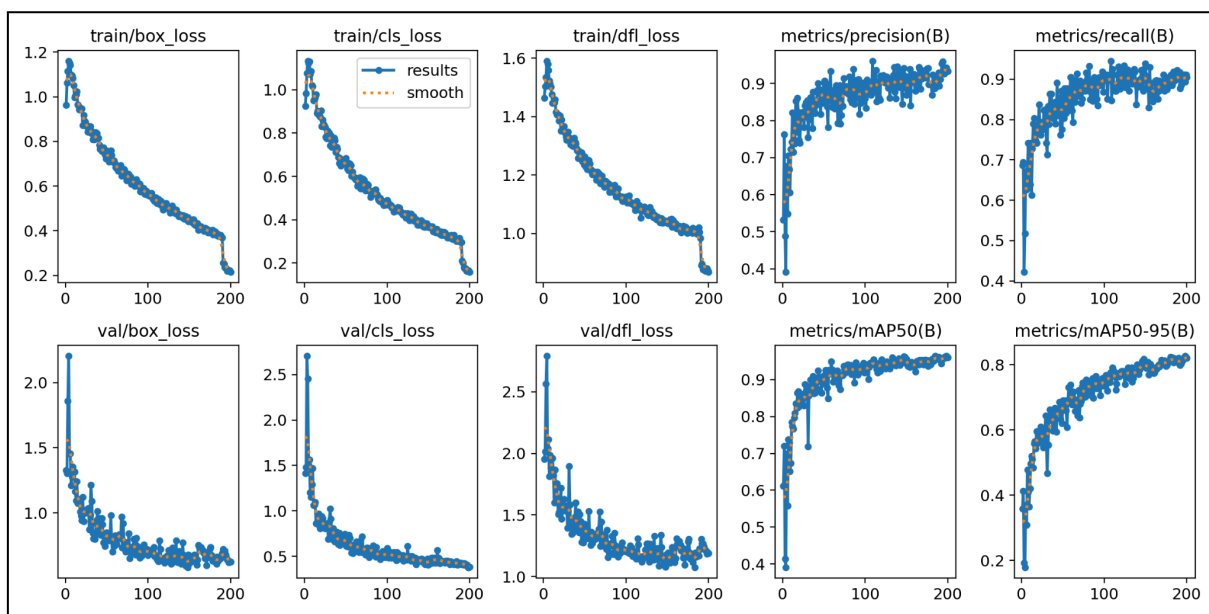


Figura 12 - Métricas de avaliação do retreinamento da classe de cachorro com YOLOv8

Fonte: Elaborado pelas autoras (2024)

Comparando os gráficos, o modelo aprimorado apresenta uma melhoria significativa em termos de eficiência do aprendizado, estabilidade nas perdas de validação e robustez nas métricas de precisão e *recall*. As curvas de mAP50 e mAP50-95 reforçam a boa precisão e robustez do modelo aprimorado. A curva F1-Confiança indica um excelente equilíbrio entre precisão e *recall* em diversos níveis de confiança, confirmando a eficácia do modelo YOLOv8 na detecção de cachorros e evidenciando uma performance robusta e estável durante o processo de treinamento e validação.

Quando comparamos os resultados do modelo aprimorado com os do primeiro treinamento, observamos melhorias significativas. No primeiro treinamento, os valores de perda durante o treinamento foram maiores: a perda de box (*train/box_loss*) era de 0,44931, a perda de classificação (*train/cls_loss*) era de

0,26949, e a perda de dfl (*train/dfl_loss*) era de 0,9211. No treinamento aprimorado, esses valores foram reduzidos para 0,22198, 0,16489, e 0,87971, respectivamente.

As perdas na validação também mostraram uma redução no modelo aprimorado: a perda de *box* (*val/box_loss*) caiu de 0,41403 para 0,6213; a perda de classificação (*val/cls_loss*) subiu de 0,22523 para 0,38038; e a perda de dfl (*val/dfl_loss*) aumentou de 0,85282 para 1,1864.

Apesar das perdas na validação terem mostrado uma piora, as métricas de precisão e *recall* do modelo aprimorado se destacaram por serem mais estáveis e robustas. No primeiro treinamento, a precisão (*metrics/precision(B)*) era de 0,98865 e o *recall* (*metrics/recall(B)*) era de 0,96425. Embora os valores de precisão e *recall* do modelo aprimorado seja um pouco mais baixo, a estabilidade e a consistência são melhoradas, conforme indicado pelas menores flutuações nos gráficos.

Adicionalmente, as métricas de mAP50 e mAP50-95 no primeiro treinamento eram de 0,98766 e 0,90678, respectivamente. No modelo aprimorado, embora esses valores sejam menores, com mAP50 de 0,96433 e mAP50-95 de 0,82627, a curva F1-Confiança sugere um desempenho mais equilibrado e robusto em diversos níveis de confiança.

Essas comparações demonstram que, apesar de um aumento nas perdas de validação, o modelo aprimorado oferece maior estabilidade e robustez nas métricas de avaliação, evidenciando um progresso significativo em relação ao primeiro treinamento.

Além das melhorias nas métricas de avaliação mencionadas anteriormente, conseguimos aumentar nossa média de confiança de 88% para 93,6%. Esse aumento demonstra que conseguimos ajustar as imagens que estavam com baixa confiança para serem melhor detectadas pelo modelo, resultando em níveis de confiança mais elevados, como demonstrado nas figuras 13 e 14. A média de confiança foi calculada com base nas mesmas imagens de teste utilizadas anteriormente, com um máximo de duas detecções por imagem.

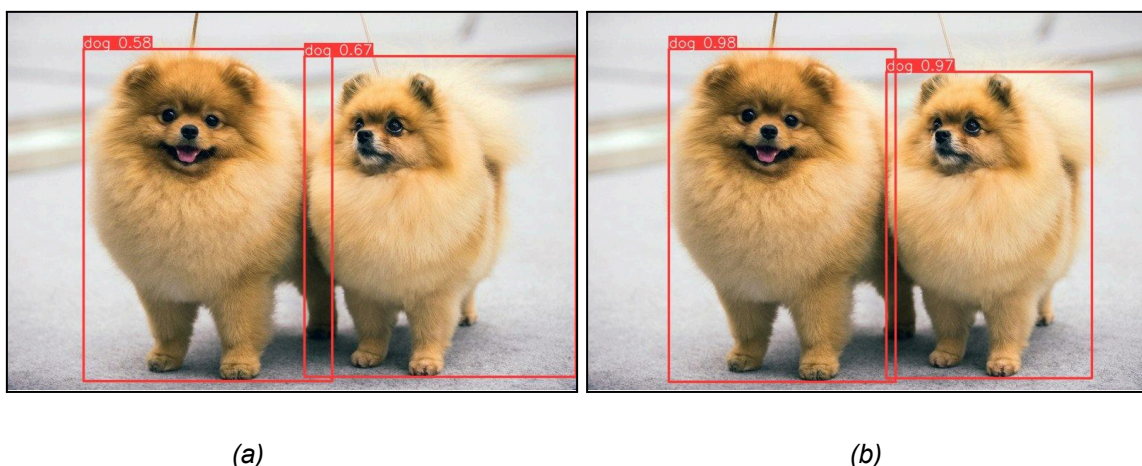


Figura 13. (a) Resultado primeiro treinamento com YOLOv8 dois Spitz Alemão; (b) Resultado retreinamento com YOLOv8 dois Spitz Alemão.



Figura 14. (a) Resultado primeiro treinamento YOLOv8 cachorro sorrindo; (b) Resultado retreinamento YOLOv8 cachorro sorrindo.

Esses resultados comprovam que é possível aprimorar o treinamento de uma classe específica utilizando o YOLOv8. Através da adição de imagens estrategicamente selecionadas e um ajuste cuidadoso dos parâmetros de treinamento, conseguimos assim não apenas melhorar a precisão do modelo na detecção de cachorros, mas também aumentar significativamente a confiança nas detecções realizadas. Como demonstrado nas Figuras 15 e 16, conseguimos manter grande confiança nas imagens que já apresentavam detecções precisas, além de aumentar a detecção e a confiança em outras imagens que anteriormente tinham baixa confiança. Isso evidencia a eficácia do YOLOv8 em permitir refinamentos contínuos e específicos, promovendo um desempenho superior do modelo em situações complexas.



Figura 15. (a) Resultado primeiro treinamento YOLOv8 cachorro na cama com brinquedos; (b) Resultado retreinamento YOLOv8 cachorro na cama com brinquedos.



Figura 16. (a) Resultado primeiro treinamento com YOLOv8 vários cachorros; (b) Resultado retreinamento com YOLOv8 vários cachorros.

Para concluir a seção de resultados, podemos afirmar que é possível aprimorar uma classe utilizando o YOLO. Devemos sempre considerar o estudo das lacunas existentes no conjunto de dados, reunir uma base de dados nova, consistente e bem trabalhada. Além disso, é importante priorizar o treinamento sem muitos ajustes de parâmetros, especialmente para o YOLO, que já vem bem ajustado. Focar no aumento das épocas, fazer ajustes na taxa de aprendizado e no tamanho do lote, e realizar diversos testes até encontrar o melhor resultado são práticas muito recomendadas.

5. DISCUSSÃO

A discussão revela a amplitude e profundidade do impacto potencial desta pesquisa, destacando tanto as conquistas quanto as áreas para melhoria e expansão futura. Ao continuar a refinar e validar essas abordagens, podemos esperar contribuições significativas para o campo da educação infantil e além.

Portanto, a pesquisa continua a buscar soluções que otimizem o processo de reconhecimento de elementos nas fases iniciais da infância, promovendo um desenvolvimento cognitivo mais completo por meio da exposição a uma ampla variedade de elementos, incluindo objetos de casa, brinquedos e animais. No entanto, é crucial reconhecer que a ausência dessa exposição pode gerar consequências significativas. Crianças que não têm a oportunidade de interagir com uma diversidade de elementos e pessoas podem enfrentar desafios na formação de suas habilidades cognitivas, levando a possíveis defasagens em seu desenvolvimento. Problemas como a dificuldade em distinguir elementos-chave no ambiente ou atrasos no aprendizado de conceitos fundamentais, como cores e formas, podem surgir. Diante deste cenário, surge a necessidade premente de encontrar soluções eficazes para otimizar o processo de reconhecimento de elementos nas fases iniciais da infância.

A fase inicial desta pesquisa surgiu da constatação de um problema crucial no desenvolvimento cognitivo das crianças: a necessidade de uma diversidade de elementos para aprimorar o processo de reconhecimento. Até então, as soluções existentes pareciam inadequadas, deixando lacunas significativas nesse processo crucial. A ausência de uma variedade suficiente de estímulos visuais pode resultar em um desenvolvimento subótimo nas fases iniciais da infância. Isso levanta a questão: como podemos oferecer um espectro mais amplo de estímulos de maneira eficaz e personalizada para cada criança?

Nossa proposta surge da ideia de treinar uma rede neural altamente específica, capaz de identificar e classificar uma ampla gama de elementos de interesse para o desenvolvimento cognitivo infantil. Ao fazer isso, visamos preencher uma lacuna crítica na abordagem atual, abrindo portas para um crescimento mais robusto e saudável nas capacidades cognitivas das crianças. Além de sua

relevância individual, esta pesquisa também traz implicações de grande alcance para um público não técnico. É uma iniciativa que pode moldar de forma tangível o desenvolvimento e a educação de crianças em suas fases mais cruciais. Ao demonstrar o impacto direto que este trabalho pode ter na vida de tantas famílias e comunidades, buscamos envolver um público amplo e diversificado na discussão e aplicação desses resultados.

Nesta busca, é importante explorar e compreender os fundamentos do processo. Isso nos conduziu a uma extensa pesquisa sobre trabalhos relacionados, especialmente no contexto de crianças com menos de 5 anos. A análise detalhada de como elementos afetam positivamente o desenvolvimento cognitivo foi o primeiro passo, seguido pela coleta e análise de vídeos para avaliar a eficácia da tecnologia YOLO na detecção desses elementos. Uma constatação intrigante surgiu durante este processo: para animais, mesmo detecções incorretas estavam sempre associadas a alguma classe pré-existente. Essa observação nos impulsionou a refletir sobre a necessidade de um retreinamento inicial, especialmente no caso de animais, dada a sua eficácia no auxílio do desenvolvimento cognitivo infantil.

Adicionalmente, é importante destacar o impacto significativo que o aumento de 5,6% na média de confiança nas 20 imagens de teste representa para nosso modelo. Esse incremento evidencia não apenas a eficácia do treinamento, mas também como o YOLO pode ser uma ferramenta valiosa para o aprimoramento de classes específicas, como no nosso caso, a detecção de cachorros. No entanto, essas melhorias também ressaltam o potencial não totalmente explorado das soluções de código aberto. Embora essas ferramentas sejam poderosas, suas limitações atuais indicam que há espaço para avanços ainda maiores, especialmente quando aplicadas a contextos específicos e desafiadores como a detecção em vídeo de diversas classes visando o desenvolvimento cognitivo infantil.

Além dos aspectos técnicos, é importante considerar o contexto social e educacional em que essas tecnologias serão aplicadas. A adoção de ferramentas de reconhecimento de elementos em vídeos educacionais pode transformar a forma como pais e educadores selecionam e utilizam conteúdos digitais. A capacidade de identificar e classificar elementos automaticamente oferece um meio eficaz de garantir que as crianças estejam expostas a materiais que promovam seu desenvolvimento cognitivo de maneira mais equilibrada e diversificada.

Finalmente, esta pesquisa também abre caminho para futuros estudos que podem explorar a integração de tecnologias de visão computacional com outras ferramentas educacionais e interativas, criando um ecossistema de aprendizado mais holístico e adaptativo. A combinação de detecção de elementos com *feedback* adaptativo, por exemplo, poderia personalizar ainda mais a experiência de aprendizado das crianças, respondendo diretamente às suas necessidades e progressos individuais.

6. LIMITAÇÕES

Nossa abordagem, embora bem-sucedida para a detecção de animais, pode não ser diretamente aplicável a todas as classes do YOLO em uma escala mais ampla. A eficácia do método pode requerer ajustes significativos para lidar com um conjunto mais abrangente de classes distintas.

Para garantir a viabilidade do projeto, optamos por utilizar uma quantidade limitada de imagens para formar o conjunto de treinamento. É importante destacar que o desempenho do modelo pode variar caso o conjunto de treinamento seja expandido. A inclusão de mais imagens e classes pode melhorar a robustez e a precisão do modelo, mas também exigiria mais recursos computacionais e tempo de processamento.

Embora nosso estudo sugira uma possível melhora na cognição infantil, é importante dizer que não conduzimos uma análise sistemática nem realizamos testes aprofundados para validar essa hipótese. Portanto, são necessárias pesquisas adicionais para confirmar ou refutar os resultados obtidos até o momento. Estudos empíricos com crianças, monitorando seu desenvolvimento cognitivo ao longo do tempo, seriam essenciais para validar a eficácia da tecnologia proposta.

O treinamento inicial foi realizado de maneira semelhante ao do modelo YOLO oficial, mas devido a limitações de hardware, não foi possível replicar exatamente o mesmo procedimento. Essas restrições não impedem que o modelo seja posteriormente aprimorado com base no que foi construído até agora. Nosso aproveitamento e métricas foram baseados em um treinamento com 5411 imagens para uma única classe, e poderiam ser significativamente melhorados com mais de 200 épocas de treinamento e um maior número de imagens.

Além disso, vale mencionar que a aplicação final no servidor se limita a uma única classe de detecção de animais, o que representa uma limitação que pode ser superada com futuras criações e expansão de novas classes do modelo. Outra limitação a ser considerada é a dependência de vídeos que representam situações do cotidiano real das crianças; vídeos animados apresentaram desafios que precisam ser abordados em pesquisas futuras.

Em suma, as limitações deste estudo apontam para a necessidade de mais dados, recursos computacionais avançados e validações empíricas para expandir e confirmar a eficácia das soluções propostas. Continuar a investigar e aprimorar o modelo permitirá uma aplicação mais ampla e eficaz da tecnologia de reconhecimento de elementos no desenvolvimento cognitivo infantil.

7. CONCLUSÃO

Este estudo representa uma tentativa significativa de abordar o problema crucial do desenvolvimento cognitivo infantil em um mundo cada vez mais dominado pela tecnologia e pelos meios digitais. Reconhecendo a importância da exposição a uma ampla variedade de elementos para o aprimoramento cognitivo das crianças, nossa pesquisa propôs uma metodologia de otimização no reconhecimento de elementos em vídeos, especialmente em ambientes digitais ricos em conteúdo de *streaming*, como os disponíveis no YouTube. Dada a análise abrangente conduzida neste estudo sobre a otimização do reconhecimento de elementos para crianças nas primeiras fases do desenvolvimento cognitivo, várias conclusões e recomendações podem ser destacadas.

Primeiramente, a pesquisa demonstrou a importância da diversidade de estímulos visuais no crescimento cognitivo infantil, especialmente em ambientes digitais ricos em conteúdo de *streaming*. A implementação da rede neural YOLO oferece uma solução viável para identificar eficientemente uma ampla gama de elementos em vídeos, contribuindo assim para o desenvolvimento e a educação de crianças nas fases mais cruciais de seu crescimento.

Os resultados dos testes de detecção realizados com vídeos selecionados destacaram a eficácia da YOLO na identificação de elementos do mundo real, como animais, brinquedos e objetos domésticos. No entanto, foram observados desafios na associação de elementos em desenhos animados, ressaltando a necessidade de focar em vídeos que representam situações do cotidiano real das crianças. Além disso, durante a análise mais aprofundada das classes, o retreinamento da rede neural, especialmente para classes específicas como animais, mostrou-se crucial para melhorar a precisão e a consistência das detecções.

A etapa de retreinamento revelou um aumento na média de confiança de 5,6%, de 88% para 93,6%, demonstrando que é possível aprimorar significativamente a eficácia do modelo com ajustes apropriados no conjunto de dados e parâmetros de treinamento. Esses resultados comprovam que a tecnologia de visão computacional pode ser uma ferramenta valiosa para o aprimoramento de

classes específicas, promovendo um desempenho superior do modelo em situações complexas.

O retreinamento da classe "cachorro" foi apenas o início do que este trabalho pode se tornar quando aplicamos a metodologia a outras classes que podem ou não estar presentes no COCO, que auxiliam no desenvolvimento cognitivo de crianças. O potencial de expandir essa abordagem para incluir uma variedade ainda maior de classes de elementos é vasto, prometendo uma aplicação mais ampla e um impacto ainda maior no enriquecimento dos estímulos visuais oferecidos às crianças.

Apesar dos resultados positivos, a pesquisa encontrou algumas limitações. A quantidade limitada de imagens no conjunto de treinamento pode influenciar o desempenho do modelo. Além disso, a análise sistemática dos benefícios cognitivos não foi realizada, sugerindo a necessidade de pesquisas adicionais para validar as hipóteses apresentadas. O treinamento inicial foi realizado de maneira semelhante ao do modelo YOLO oficial, mas devido a limitações de hardware, não foi possível replicar exatamente o mesmo procedimento. Essas restrições não impedem que o modelo seja posteriormente aprimorado com base no que foi construído até agora.

As recomendações para futuras pesquisas incluem o aprimoramento contínuo da tecnologia de detecção de elementos, especialmente em relação a elementos específicos que apresentaram desempenho abaixo do ideal, como os animais. Também é fundamental considerar o impacto a longo prazo do uso dessas tecnologias no desenvolvimento cognitivo das crianças e explorar maneiras de integrar essas soluções de forma responsável e eficaz em ambientes educacionais e de entretenimento infantil.

Este estudo oferece *insights* sobre como a tecnologia de visão computacional pode ser aplicada para enriquecer o desenvolvimento cognitivo infantil em ambientes digitais, destacando a importância da diversidade de estímulos visuais e da otimização contínua das técnicas de detecção de elementos. Essas descobertas têm o potencial de informar e influenciar práticas educacionais e de entretenimento para crianças em todo o mundo, promovendo um crescimento cognitivo mais saudável e completo nas fases iniciais da infância.

A aplicação da rede neural YOLO demonstrou ser uma solução viável para identificar uma ampla gama de elementos em vídeos, contribuindo para a educação

e desenvolvimento das crianças. Com a continuação e refinamento desta pesquisa, espera-se que mais classes sejam integradas e a precisão das detecções continue a melhorar, oferecendo ferramentas cada vez mais eficazes para pais e educadores.

Em suma, a otimização do reconhecimento de elementos em vídeos é uma área promissora que pode transformar significativamente a maneira como o conteúdo digital é utilizado para apoiar o desenvolvimento cognitivo infantil, proporcionando um ambiente digital mais enriquecedor e responsável.

REFERÊNCIAS

AMERICAN ACADEMY OF PEDIATRICS. **Digital Screen Media and Cognitive Development**. *Pediatrics*, v. 140, n. Suppl 2, p. S57-S61, 2017. Disponível em: https://publications.aap.org/pediatrics/article/140/Supplement_2/S57/34173/Digital-Screen-Media-and-Cognitive-Development. Acesso em: 16 jun. 2024.

APRENDA BRINCANDO. **Alfabeto de A a Z com Animais** – Aprenda Brincando. YouTube, 2019. Disponível em: <https://www.youtube.com/watch?v=e6WMtMaG4tY>. Acesso em: 15 set. 2023.

BAMBACH, S.; CRANDALL, D. J.; SMITH, L. B.; YU, C. **Active Viewing in Toddlers Facilitates Visual Object Learning**: An Egocentric Vision Approach. *Journal of Child Development*. School of Informatics and Computing, 2020. Disponível em: https://cogdev.siteshost.iu.edu/labwork/cogsci16_toys.pdf. Acesso em: 30 nov. 2023.

BANERJEE, I. S. **Animal Image Dataset - 90 Different Animals**. Kaggle, 2021. Disponível em: <https://www.kaggle.com/datasets/iamsouravbanerjee/animal-image-dataset-90-different-animals?resource=download>. Acesso em: 1 maio 2024.

BRASIL, Victor. **50 ANIMAIS - Nome e Som de 50 Animais**. YouTube, 2020. Disponível em: <https://www.youtube.com/watch?v=Meu9L5fQGG8>. Acesso em: 15 set. 2023.

BRASIL, Victor. **Vídeos de animais para a família** - sons de vaca, gato, elefante, girafa, pato. YouTube, 2022. Disponível em: <https://www.youtube.com/watch?v=cmqsSxpJZFY>. Acesso em: 15 set. 2023.

CANAL DO SH. **Aprenda as cores e meios de transporte com caminhão e carros coloridos para crianças**. YouTube, 2017. Disponível em: https://www.youtube.com/watch?v=MDVbkgFjN_8. Acesso em: 15 set. 2023.

CHOUINARD, M. M.; HARRIS, P. L.; MARATSOS, M. P. **Children's Questions: A Mechanism for Cognitive Development**. *Monographs of the Society for Research in Child Development*, v. 72, p. 1-129, 2007.

CKN. **Giant Color Toy Building Blocks House Pretend Play Fun CKN**. YouTube, 2019. Disponível em: <https://www.youtube.com/watch?v=3aWwSwkJDew>. Acesso em: 15 set. 2023.

DHARAHAS CREATION. **How to make a Lorry with Building Blocks**. YouTube, 2021. Disponível em: <https://www.youtube.com/watch?v=t0qaJ-69EKg&list=PL3RK-YxKgGTdIAE-Ky65oX3sUr4jN-kdO>. Acesso em: 15 set. 2023.

DOMOFF, S.; BORGEM, A.; RADESKY, J. **Interactional Theory of Childhood Problematic Media Use**. Human Behavior and Emerging Technologies, Michigan, v. 2, p. 343-354, set. 2020.

DUMUID, D. Screen Time in Early Childhood. **The Lancet Child & Adolescent Health**, 2020. DOI: 10.1016/S2352-4642(20)30005-5. Disponível em: [https://www.thelancet.com/journals/lanchi/article/PIIS2352-4642\(20\)30005-5/fulltext](https://www.thelancet.com/journals/lanchi/article/PIIS2352-4642(20)30005-5/fulltext). Acesso em: 26 nov. 2023.

ESTAÇÃO EDUCAÇÃO. **Alfabeto das FRUTAS em Português ABC Completo** (Com Alguns Animais Comidas e Plantas). YouTube, 2018. Disponível em: <https://www.youtube.com/watch?v=b9yHPm44i2w>. Acesso em: 15 set. 2023.

FABRICANDO MÚSICA. **Que som é esse Sons dos Meios de Transporte**. YouTube, 2021. Disponível em: <https://www.youtube.com/watch?v=i1HWpH5ODMg>. Acesso em: 15 set. 2023.

FLAVELL, J. H. **Cognitive Development: Children's Knowledge About the Mind**. Annual Review of Psychology, v. 50, p. 21-45, fev. 1999. DOI: 10.1146/annurev.psych.50.1.21. Disponível em: <https://www.annualreviews.org/doi/full/10.1146/annurev.psych.50.1.21>. Acesso em: 26 nov. 2023.

GLAM, Inc. **100 Years of Toys ★ Glamcom**. YouTube, 2016. Disponível em: <https://www.youtube.com/watch?v=EDAPaEVr1Hk>. Acesso em: 15 set. 2023.

GUELLAI, B.; SOMOGYI, E.; ESSEILY, R.; CHOPIN, A. **Effects of Screen Exposure on Young Children's Cognitive Development: A Review**. Frontiers in Psychology, 2022; 13. DOI: 10.3389/fpsyg.2022.923370. Disponível em:

<https://www.frontiersin.org/articles/10.3389/fpsyg.2022.923370/full>. Acesso em: 1 maio 2024.

HEALEY, A.; MENDELSON, A. **Selecting Appropriate Toys for Young Children in the Digital Era**. Journal of Child Development. From The American Academy of Pediatrics, 2019. Disponível em: <https://publications.aap.org/pediatrics/article/143/1/e20183348/37330/Selecting-Appropriate-Toys-for-Young-Children-in?autologincheck=redirectedf>. Acesso em: 30 nov. 2023.

HIHO KIDS. **Crystal's House | HiHo Cribs | HiHo Kids**. YouTube, 2018. Disponível em: <https://www.youtube.com/watch?v=5qIRpO7mJmM>. Acesso em: 15 set. 2023.

HIHO KIDS. **Twin House Tour! Ethan and Helena's Crib | HiHo Cribs | HiHo Kids**. YouTube, 2019. Disponível em: <https://www.youtube.com/watch?v=va4B6EXCqqo>. Acesso em: 15 set. 2023.

HOUDI TV. **Aprende os Veículos com os Sons - Português** (Portugal). YouTube, 2017. Disponível em: <https://www.youtube.com/watch?v=etQICtOhdLU>. Acesso em: 15 set. 2023.

HUSTON, A. C.; WRIGHT, J. C.; RICE, M. L.; RERKMEN, D.; ST. PETERS, M. **Development of Television Viewing Patterns in Early Childhood: A Longitudinal Investigation**. The American Psychological Association, Inc, 1990. Disponível em: <http://cdmc.georgetown.edu/wp-content/uploads/2017/10/1990-03x.pdf>. Acesso em: 30 nov. 2023.

IDIOMA BRASIL. **Português ao Vivo - Meios de transporte**. YouTube, 2015. Disponível em: <https://www.youtube.com/watch?v=7u9Sd2a3ZVA>. Acesso em: 15 set. 2023.

KIDSTV. **Sons de animais | Aprender sons de animais em português**. YouTube, 2016. Disponível em: <https://www.youtube.com/watch?v=NBbICffDAEQ>. Acesso em: 15 set. 2023.

KYLIE JENNER SNAPCHATS SONGS. **Kim Kardashian Gives a Tour of her Kids' Playroom in her House**. YouTube, 2020. Disponível em: <https://www.youtube.com/watch?v=6wZVs5e0mUQ>. Acesso em: 15 set. 2023.

NY & NA Art. **ABC Food Flashcard Read and Write** – Learning about Food Vocabulary – Food Names for Kids. YouTube, 2021. Disponível em: <https://www.youtube.com/watch?v=GnXJZq3k-fYk>. Acesso em: 15 set. 2023.

PEREIRA, A. F.; JAMES, K. H.; JONES, S. S.; SMITH, L. B. **Active Object Exploration in Toddlers and its Role in Visual Object Recognition**. Academia Edu, 2008. Disponível em: <https://api.semanticscholar.org/CorpusID:16760214>. Acesso em: 30 nov. 2023.

QUIZ BLITZ. **Guess the Fruit in 3 Seconds 52 Different Types of Fruit**. YouTube, 2023. Disponível em: https://www.youtube.com/watch?v=4_XyvLFDJBs. Acesso em: 15 set. 2023.

RAKINSON, D. H.; BUTTERWORTH, G. E. **Infants' use of object parts in early categorization**. Developmental Psychology, v. 34, p. 49-62, 1998. DOI: 10.1037/0012-1649.34.1.49. Disponível em: <https://psycnet.apa.org/record/1997-42746-004>. Acesso em: 29 nov. 2023.

SALIM, M. P.; ONG, J.; EDBERT, I. S.; SURHATONO, D. **Object Detection for Child Learning Media**. 8th International Conference on Science and Technology (ICST), 2022. Disponível em: <https://ieeexplore.ieee.org/abstract/document/10136287/keywords#keywords>. Acesso em: 29 nov. 2023.

TAVEIRA, Arnaldo. **Aprenda os nomes dos animais** - Professor Arnaldo. YouTube, 2021. Disponível em: <https://www.youtube.com/watch?v=w3u2iA0e7kl>. Acesso em: 15 set. 2023.

THE LABRANT FAM. **LaBrant Fam Official Tennessee House Tour (Part 2)**. YouTube, 2023. Disponível em: <https://www.youtube.com/watch?v=FoxsbjKT2yA>. Acesso em: 15 set. 2023.

THE LABRANT FAM. **The LaBrant Fam Official Tennessee House Tour**. YouTube, 2023. Disponível em: <https://www.youtube.com/watch?v=lmnsNb97Y7U>. Acesso em: 15 set. 2023.

WANG, Y.; SONG, Y.; KRSTIC, M. **Collectively Rotating Formation and Containment Deployment of Multiagent Systems: A Polar Coordinate-Based Finite Time Approach**. IEEE Transactions on Cybernetics, v. 47, dez. 2016. DOI:

10.1109/TCYB.2016.2624307. Disponível em:
<https://ieeexplore.ieee.org/document/7782778>. Acesso em: 28 nov. 2023.

WATCHSTATS. **Toys Evolution | 9000BC - 2020**. YouTube, 2020. Disponível em:
<https://www.youtube.com/watch?v=AtdboPVtU6s>. Acesso em: 15 set. 2023.