

Pontifícia Universidade Católica de São Paulo

**Curso de Bacharelado em Ciência de Dados e
Inteligência artificial**

Carlos Eduardo de Oliveira

João Pedro Taves Araujo

Thiago Rodrigues de Jesus Carvalho

**Aplicação de Redes Adversariais Generativas
para Aumento de Dados e Retreinamento do
YOLO V8n em Detecção de Semáforos**

Trabalho de Conclusão de Curso

2024

Orientador: Dr. Jefferson de Oliveira Silva

Aplicação de Redes Adversariais Generativas para Aumento de Dados e Retreinamento do YOLO V8n em Detecção de Semáforos

Resumo

Este estudo propõe uma abordagem para melhorar o desempenho do modelo YOLO V8n na identificação de objetos de trânsito, com um foco específico na detecção de semáforos, utilizando técnicas de Data Augmentation por meio da geração de imagens com Generative Adversarial Networks (GAN). O problema da pesquisa reside na busca por maneiras eficazes de aprimorar a precisão deste modelo em ambientes de tráfego complexos. A solução proposta envolve a geração de imagens de semáforos por meio de GANs e a subsequente reconfiguração do modelo YOLO V8n com essas imagens criadas. A integração dessas imagens geradas no processo de treinamento visa melhorar a capacidade do modelo YOLO V8n de identificar com precisão semáforos em diversas condições. A avaliação do desempenho do modelo YOLO V8n antes e depois do reajuste permite verificar se a solução proposta consegue mitigar os problemas de detecção identificados.

Palavras-chave

Generative Adversarial Networks (GAN), YOLO V8n, Data Augmentation, Objetos de Trânsito, Semáforo, Segurança Viária.

Introdução

As Redes Generativas Adversariais (GANs) têm sido amplamente utilizadas na geração de imagens artificiais, "fornecendo meios para criar dados sintéticos que mimetizam dados reais com alta fidelidade" (Goodfellow et al., 2014). Este trabalho propõe explorar o potencial das GANs na geração de imagens para treinamento de algoritmos de detecção de objetos, utilizando o Roboflow como ferramenta de gerenciamento de imagens dos conjuntos de dados. Ao gerar imagens sintéticas, pretendemos investigar como esses dados influenciam a eficácia do detector de objetos, contribuindo para o avanço da pesquisa em visão computacional (Radford et al., 2015).

A geração de imagens artificiais e a detecção de objetos são áreas cruciais em visão computacional e inteligência artificial. Com a crescente demanda por sistemas de detecção de objetos precisos e robustos, a utilização de GANs para gerar conjuntos de dados sintéticos torna-se uma abordagem promissora (Zhu et al., 2017). Esta pesquisa é relevante porque aborda a questão da generalização do detector de objetos em diferentes conjuntos de dados, incluindo dados sintéticos. Além disso, ela pode contribuir para a redução da dependência de conjuntos de dados reais, muitas vezes limitados em tamanho e variedade, para o treinamento de modelos de detecção de objetos.

A relevância desse problema é evidenciada pela necessidade de uma identificação precisa de objetos de trânsito para garantir a segurança viária e prevenir acidentes. Neste contexto, Shrivastava et al. (2017) destacam a eficácia do treinamento com imagens sintéticas e não supervisionadas, uma metodologia que pode aprimorar significativamente a capacidade de generalização dos modelos de detecção. Esse

aprimoramento é crucial para lidar com as condições variadas e imprevisíveis encontradas em ambientes de tráfego complexos, onde a precisão na detecção de semáforos pode diretamente influenciar a redução de riscos de colisões e violações das normas de trânsito.

O objetivo deste estudo é investigar a eficácia de um detector de objetos ao utilizar imagens geradas por GANs como dados de entrada. Para isso, planejamos treinar o detector de objetos com conjuntos de dados contendo imagens reais e imagens sintéticas geradas por GANs. Posteriormente, iremos avaliar o desempenho do detector no conjunto de dados reais e comparar os resultados, analisando a capacidade de generalização do modelo em relação a objetos reais e sintéticos.

A justificativa para este estudo reside na necessidade de desenvolver métodos eficazes para a geração de dados sintéticos que possam ser usados para treinar e avaliar algoritmos de visão computacional, como os detectores de objetos (Antoniou et al., 2017). A utilização de imagens geradas por GANs oferece uma alternativa viável para superar limitações associadas à disponibilidade e variedade de conjuntos de dados reais. Além disso, entender como os modelos se comportam em dados sintéticos pode revelar insights valiosos sobre suas capacidades de generalização e robustez. Este estudo contribuirá para avançar o campo da visão computacional, fornecendo uma abordagem inovadora para testar e melhorar algoritmos de detecção de objetos.

Metodologia

O método deste estudo está dividido em três etapas principais. Em todas as etapas de treinamento e teste dos modelos, empregamos um sistema contendo uma GPU

GTX 1660 para aceleração, 8GB de RAM DDR4 e um CPU Ryzen 5 1600, com 6 núcleos e 12 threads.

Parte 1: Preparação do Dataset

Inicialmente, foram coletados em torno de 5 mil frames de trânsito das ruas e avenidas brasileiras. Utilizamos o modelo YOLO V8m para identificar objetos de semáforos nessas imagens. O YOLO V8m, ao identificar um objeto, desenha um retângulo ao redor do mesmo, chamado de "bounding box", que delimita a área onde o objeto foi detectado. Foi desenvolvido um algoritmo capaz de identificar a bounding box em cada frame e recortar os objetos, salvando-os em uma pasta.

Os objetos detectados são recortados e salvos em duas versões: original e com ruído adicionado. O ruído é gerado de forma uniforme, aplicando pequenas variações nos pixels das imagens para simular condições adversas de captura. As imagens com ruído são originadas das imagens normais, resultando em uma distribuição de 50% imagens originais e 50% imagens com ruído. O conjunto de dados resultante foi organizado exclusivamente para semáforos, com um total de 3334 imagens de tamanho 100x100.



Figura 1: fluxo de identificação e recorte de semáforo pelo YOLO V8m

Este diagrama ilustra o processo da identificação e recorte dos objetos pelo YOLO V8.

Parte 2: Geração de Imagens com GAN

Para a geração de imagens sintéticas, utilizamos a arquitetura Deep Convolutional Generative Adversarial Network (DCGAN), baseada no trabalho de Radford, Metz e Chintala (2015). A rede é composta por duas partes principais: o gerador, que cria as imagens, e o discriminador, que avalia sua autenticidade. Ambos os componentes da rede foram treinados utilizando a função de perda Binary Cross-Entropy Loss (BCELoss), que mede a distância entre as probabilidades previstas e os valores reais, ideal para este tipo de tarefa de classificação binária.

Hiperparâmetros:

- **Taxa de Aprendizado (Learning Rate, lr):** Hiperparâmetros: Taxa de Aprendizado (Learning Rate, lr): Foi definida como 0.0002 para ambos, gerador e discriminador, equilibrando a estabilidade e eficácia do treinamento das redes, conforme recomendado por Radford et al. (2015).
- **Otimizador:** Utilizamos o Adam, conhecido por sua eficiência em tarefas que envolvem GANs, com os parâmetros beta de (0.5, 0.999). Esse otimizador ajuda a ajustar a taxa de aprendizado dinamicamente, o que é crucial para manter o equilíbrio entre o gerador e o discriminador durante o treinamento.
- **Número de Épocas:** Cada treinamento foi realizado até que a geração das imagens alcançasse uma qualidade aceitável visualmente e conforme medidas de validação intermediárias.

Os semáforos recortados da Parte 1 serviram como entrada para esta rede. Inicialmente, concentrarmos nossos esforços em gerar imagens de semáforos, dada sua relevância e desafio particular em termos de variação em iluminação e perspectiva nas cenas de trânsito. As imagens geradas foram avaliadas iterativamente para garantir que a qualidade e a diversidade estivessem alinhadas com os requisitos de um dataset robusto para treinamento de visão computacional.

Este processo resultou na geração de 2000 novas imagens de semáforos com tamanho 64x64, as quais foram cuidadosamente inspecionadas e validadas antes de serem adicionadas ao conjunto de dados de treinamento para o retreinamento do YOLO V8n. Ao final da seleção, 500 imagens foram mantidas. A utilização destas imagens busca não apenas aumentar a quantidade de dados disponíveis, mas também introduzir uma variabilidade controlada que pode ser crucial para melhorar a generalização do modelo em cenários reais.

Parte 3: Retreinamento do YOLO V8n

O método proposto permite investigar se as imagens geradas por GANs podem melhorar a precisão e a capacidade de generalização do modelo YOLO V8n na identificação de semáforos em cenários diversos.

O método proposto explora se as imagens sintéticas geradas por GANs podem aprimorar a precisão e a capacidade de generalização do modelo YOLO V8 na identificação de semáforos em diversos cenários urbanos. Após gerar as imagens sintéticas, utilizamos a plataforma Roboflow para processá-las, identificando e etiquetando corretamente os semáforos com a label 'traffic-light'. Durante o pré-processamento no Roboflow, implementamos técnicas de aumento de dados

que incluíram o redimensionamento das imagens para um formato padrão de 128x128 pixels usando um método de stretch. Adicionalmente, aplicamos técnicas adversariais conforme descrito por Behpour et al. (2019), que incluíram flip horizontal e vertical, rotações de 90 graus tanto no sentido horário quanto no anti-horário, e ajustes no brilho com redução e aumento de 15%, visando replicar o ganho de até 16% em performance observado pelos autores. Essas modificações são cruciais para garantir que as imagens geradas sejam efetivamente utilizadas no treinamento do modelo, mantendo consistência nas anotações e introduzindo variação controlada que enriquece o conjunto de dados de treino.

Uma vez etiquetadas e pré-processadas, as imagens foram combinadas ao COCO dataset, um dos datasets mais abrangentes e utilizados na área de visão computacional. Este procedimento visou não apenas expandir a quantidade de dados disponíveis, mas também preservar a base de conhecimento pré-existente do YOLO V8n, assegurando que o modelo continue a ser eficaz na identificação de uma ampla gama de objetos além dos semáforos. Ao final deste processo, contávamos com aproximadamente 120 mil imagens de treino e 5.000 imagens de validação. Para todas as etapas, utilizamos o mesmo sistema descrito para a DCGAN.

Retreinamento do Modelo: O retreinamento do modelo YOLO V8n foi realizado com os seguintes hiperparâmetros:

- **Número de Épocas:** 10, uma escolha influenciada pela limitação computacional disponível.
- **Taxa de Aprendizado Inicial (lr_0):** 0.001, para promover uma convergência rápida e eficaz sem causar instabilidade durante o treinamento.

- **Freeze Layers (freeze):** 10, o que significa que os primeiros 10 layers do modelo foram congelados durante o treinamento inicial, focando o ajuste nos layers superiores para refinamento fino.
- **Automatic Mixed Precision (amp):** false, optamos por não utilizar a precisão mista automática devido à nossa configuração de hardware e necessidades de treinamento específicas.

Este processo de retreinamento foi meticulosamente monitorado para garantir que o modelo não apenas aprendesse a identificar semáforos com maior precisão, mas também mantivesse sua performance em relação a outros objetos de trânsito e elementos urbanos. Utilizamos métricas de avaliação padrão, como precisão e recall, para comparar o desempenho do modelo antes e após o retreinamento, visando documentar quaisquer melhorias significativas ou ajustes necessários para futuras iterações de treinamento.

Para medir a performance do modelo retreinado, comparamos seu desempenho antes e depois do treinamento em imagens de teste. Utilizamos duas métricas principais: a quantidade de objetos identificados e a precisão das detecções, verificando se os objetos foram identificados corretamente.

RESULTADOS

1: Imagens Geradas

Nesta seção discutimos os resultados obtidos a partir das imagens geradas pela Deep Convolutional Generative Adversarial Network (DCGAN), uma técnica baseada nos princípios estabelecidos por Radford et al. (2015). A qualidade das

imagens é avaliada com base em critérios visuais qualitativos, com o objetivo de determinar sua adequação para o treinamento de modelos de detecção de objetos.

O código fonte deste projeto está disponível no [GitHub](#).

1.1: Qualidade Visual das Imagens Geradas

As imagens geradas foram analisadas visualmente para verificar sua clareza, realismo e fidelidade em relação aos semáforos reais. Diversas amostras das imagens geradas são apresentadas, destacando tanto sucessos quanto falhas do modelo. Como podemos ver na Figura 2, foram identificados casos onde a GAN conseguiu replicar com certa fidelidade características dos semáforos, enquanto em outros casos, artefatos como borrões, formas distorcidas ou outros formatos não identificáveis foram observados (Goodfellow et al., 2014). Podemos destacar também o baixo volume de variedade nas amostras geradas pelos melhores pesos da geradora, indicando uma carência de mais dados e iterações do fluxo de treinamento.

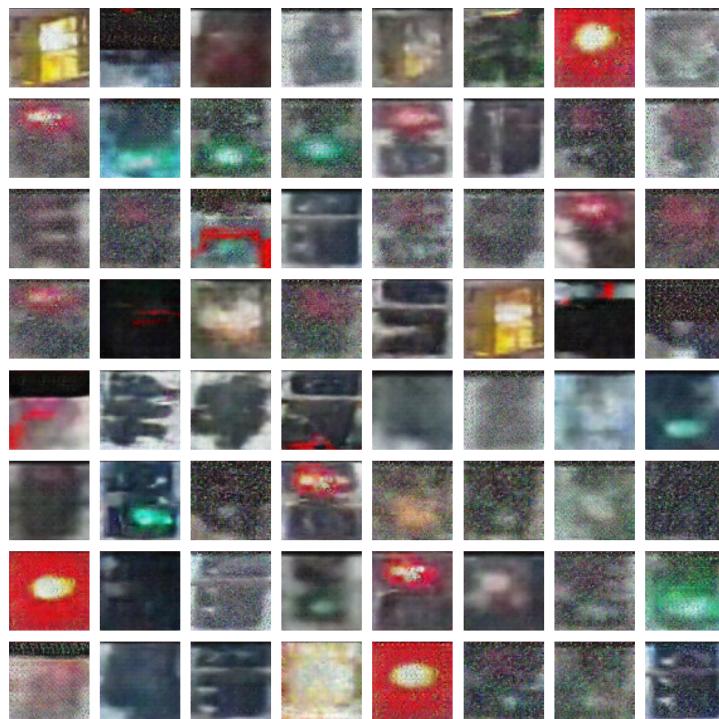


Figura 2: Exemplo de batch gerado pela GAN

1.2: Análise Quantitativa

Mudando o foco para a progressão dos valores de perda ao longo das épocas (Figura 3), podemos perceber que ainda existe uma notável discrepância entre os valores da rede geradora em relação à discriminadora. Destaca-se, especialmente, os baixíssimos valores atingidos pela rede discriminadora, de maneira basicamente constante ao longo das épocas, enquanto que a rede geradora tem valores bem mais altos e instáveis ao longo do treinamento (Goodfellow et al., 2014). Tal comportamento indica que está ocorrendo um desequilíbrio entre os dois modelos e levanta ainda mais evidências de que, em futuras iterações do projeto, seria de suma importância aumentar o volume de dados coletados para treinamento. No entanto, podemos notar que as redes tiveram certo sucesso em gerar imagens utilizáveis em nosso estudo. Vale destacar que devido à natureza do código de treinamento da GAN, optamos por fazer a média das perdas por iteração no

treinamento, pois o número de iterações atinge números maiores que o número de épocas ao final do treinamento.

Vale destacar que devido à natureza do código de treinamento da GAN, optamos por fazer a média das perdas por iteração no treinamento, pois o número de iterações atinge números maiores que o número de épocas ao final do treinamento.

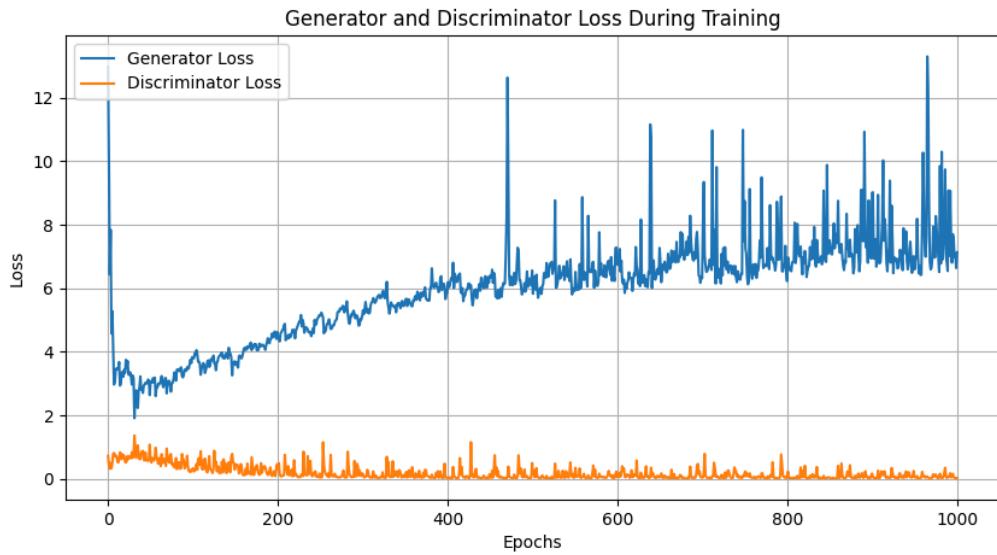


Figura 3: Evolução da perda das redes geradora (azul) e discriminadora (laranja)

2: Previsões do Modelo Retreinado

Nesta seção, exploramos os resultados alcançados pelo modelo base (YOLO V8n) após ser retreinado com um conjunto de dados enriquecido, que inclui tanto imagens originais do dataset COCO 2017, quanto imagens sintéticas geradas pela GAN. Essa abordagem visa avaliar a eficácia do modelo em integrar e aprender a

partir de dados sintéticos adicionais no contexto de detecção de objetos (Shrivastava et al., 2017).

2.1: Avaliação Quantitativa

Considerando a análise da curva F1-Confidence (Figura 4) do modelo após o retreinamento, observamos que o escore máximo de F1 para todas as classes (representado pela linha azul) é relativamente baixo. Além disso, as curvas F1 das classes individuais exibem uma ampla variação no desempenho. Esses resultados sugerem que o modelo enfrenta dificuldades em identificar as classes de maneira precisa e consistente. Essa performance insatisfatória pode ser parcialmente atribuída ao curto período de treinamento, que se limitou a apenas 10 épocas devido a restrições computacionais do projeto. Portanto, um número maior de épocas de treinamento poderia ser crucial para permitir que o modelo desenvolva uma compreensão mais aprofundada das características das classes, potencialmente melhorando sua precisão geral (Goodfellow et al., 2014).

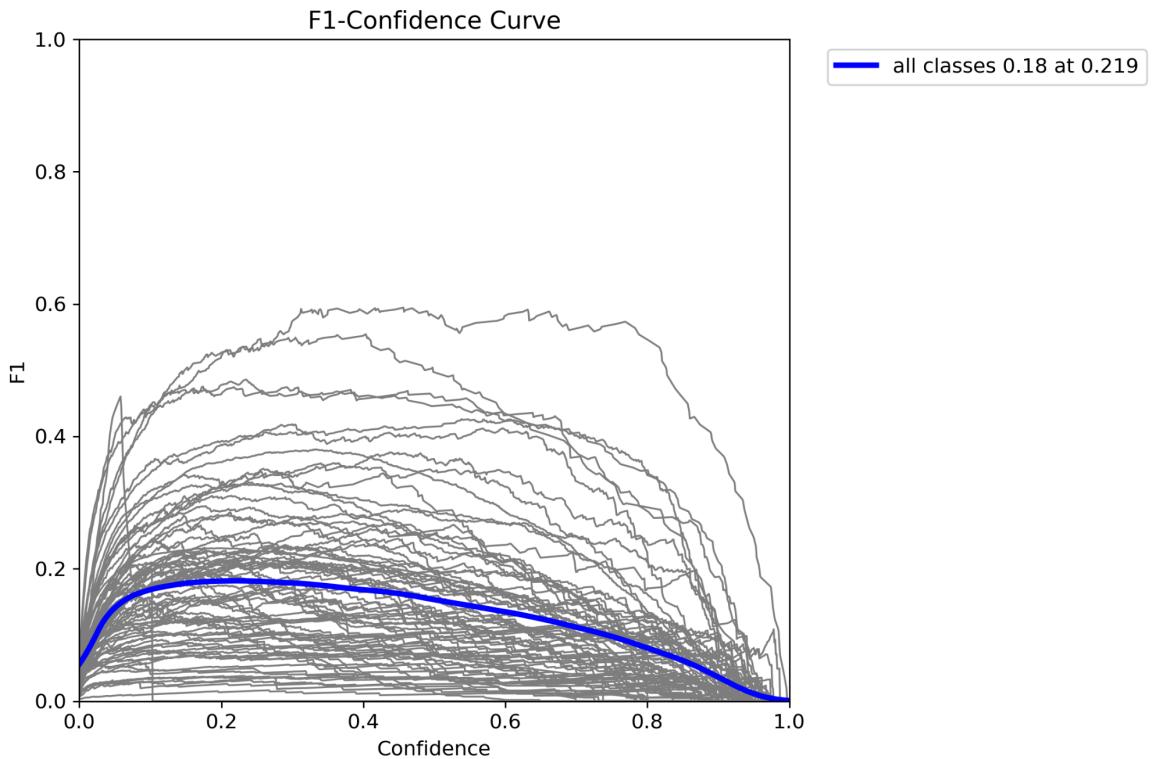


Figura 4: Curva F1 versus Confiança

Ademais, a partir da análise da curva de Precision-Confidence (Figura 5), podemos observar que o modelo demonstra uma melhoria progressiva na precisão à medida que o limiar de confiança aumenta, atingindo uma precisão notavelmente alta perto do limiar máximo de 1.0. Isso indica que o modelo é bastante preciso nas detecções que ele classifica como positivas com alta confiança. No entanto, essa alta precisão em limiares elevados também sugere que o modelo pode estar sendo excessivamente conservador, possivelmente omitindo detecções válidas com confianças mais baixas (Zhu et al., 2017). Esta tendência, embora positiva em termos de minimizar falsos positivos, pode limitar a abrangência do modelo, reduzindo o recall e, por consequência, impactando o equilíbrio geral entre precisão e recall necessário para uma detecção de objetos eficaz em cenários do mundo real. Considerando novamente que o treinamento foi realizado em apenas 10 épocas, é possível que uma extensão no número de épocas possa ajudar a suavizar

essa curva de precisão, permitindo que o modelo mantenha uma alta precisão, mas com um limiar de confiança menos restritivo. Isso pode potencialmente aumentar a sensibilidade do modelo em detectar objetos com uma variedade maior de confianças, melhorando a utilidade prática do modelo em aplicações reais.

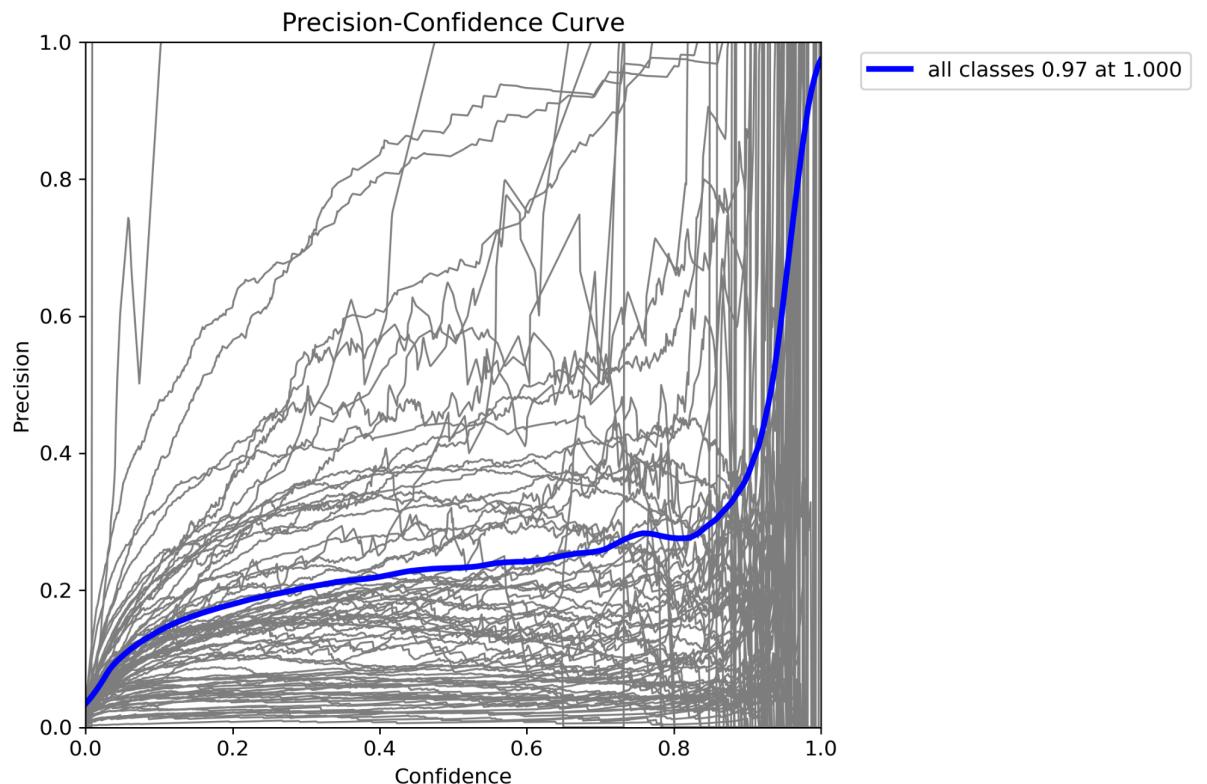


Figura 5: Curva Precisão versus Confiança

Número de Semáforos Identificados: Ambos os modelos identificaram um total de 40 semáforos no conjunto de dados de teste. Este resultado sugere que a adição de imagens geradas por GAN ao dataset de treinamento não afetou a capacidade de ambos os modelos de identificar a presença de semáforos em termos de quantidade detectada.

Média de Confiança das Detecções:

- **Modelo 1:** A média de confiança nas detecções de semáforos pelo primeiro modelo foi de 0.334875. Este valor reflete a confiança geral do modelo nas suas previsões quando treinado com o conjunto de dados COCO original, indicando um nível moderado de certeza nas suas detecções.
- **Modelo 2:** Para o segundo modelo, que foi retreinado com a inclusão de imagens sintéticas, a média de confiança foi ligeiramente inferior, situando-se em 0.308104. Isso pode indicar que, embora as imagens geradas por GAN possam ter introduzido uma variabilidade útil, elas também podem ter introduzido um nível de incerteza que afetou levemente a confiança do modelo nas suas previsões de semáforos.

Implicações e Considerações Futuras: A similaridade no número de detecções, mas a ligeira variação na confiança entre os dois modelos, sugere que as imagens adicionadas pelo processo GAN podem ter impactado sutilezas na forma como o modelo percebe e classifica semáforos. É importante considerar que a integração de dados sintéticos pode exigir ajustes adicionais nos parâmetros de treinamento ou mesmo uma revisão na qualidade e diversidade das imagens geradas para otimizar o equilíbrio entre quantidade e confiança das detecções.

Em futuras iterações do projeto, seria prudente explorar não apenas o aumento no volume de dados, mas também aprimorar a qualidade das imagens GAN para garantir que elas complementam de maneira mais eficaz o conjunto de dados de treinamento, potencialmente melhorando a precisão e a confiança do modelo em suas previsões sem comprometer a performance geral.

2.2: Análise de Casos Específicos

Para discutir a performance geral do modelo retreinado, selecionamos alguns frames específicos que ilustram as particularidades, desafios apresentados nas discussões quantitativas. Em cada uma das imagens apresentadas, o modelo 1 representa o modelo YOLO V8n base e o modelo 2 representa o modelo retreinado.

2.2.1: Análise de Casos Específicos - Frame 1

Como podemos observar a partir das imagens, o modelo retreinado (figura 7) apresentou melhora na detecção dos semáforos, detectando os dois objetos de interesse diretamente à frente da câmera no frame. No entanto, falhou em detectar o semáforo que se encontra no canto superior esquerdo da imagem, que foi detectado pelo modelo base (Figura 6). Tal fato sugere que as imagens geradas pela GAN estão adicionando conhecimento ao modelo, mas ainda são necessárias mais épocas de treinamento para a base de conhecimento ser replicada totalmente.

Modelo 1

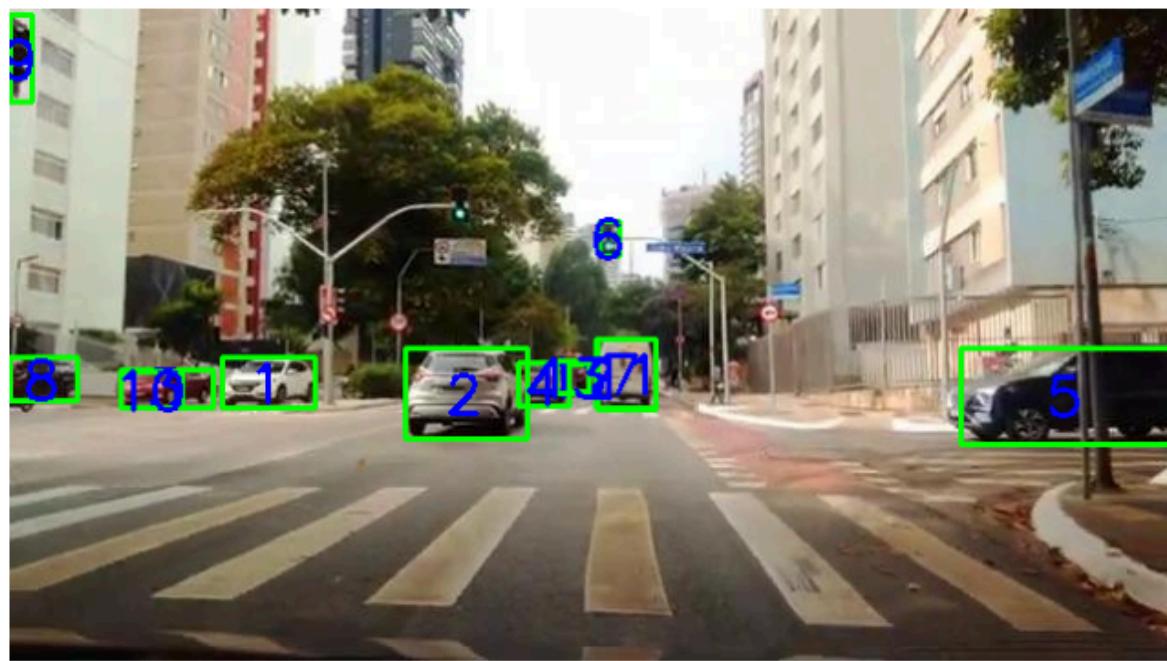


Figura 6: Predições do modelo base em primeira imagem de análise

Modelo 2

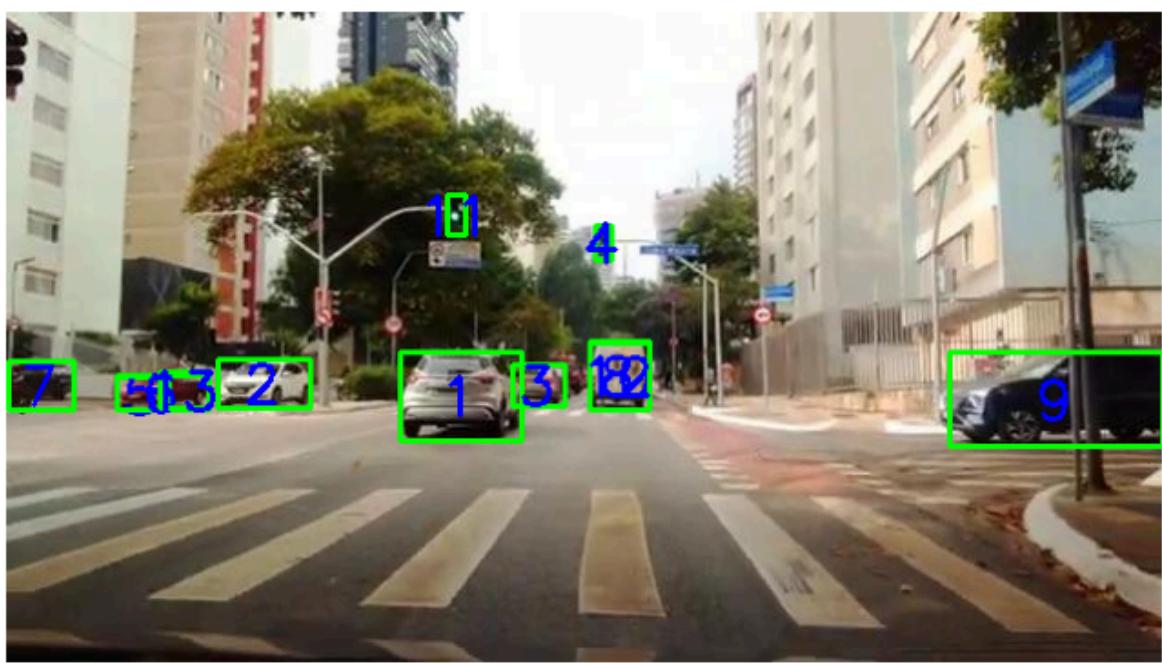


Figura 7: Predições do modelo retreinado em primeira imagem de análise

2.2.2: Análise de Casos Específicos - Frame 2

No segundo caso, observamos um dos desafios enfrentados na presente iteração do projeto. Como podemos ver, tanto na Figura 8 quanto na Figura 9, os modelos não foram capazes de identificar os semáforos diretamente à frente na imagem. Isso indica que o modelo não aprendeu o suficiente para identificar o objeto de interesse nesse caso. Tal fato reforça a hipótese de que o modelo deve ser retreinado em mais épocas e que devemos adicionar mais variedade ao conjunto de imagens geradas.

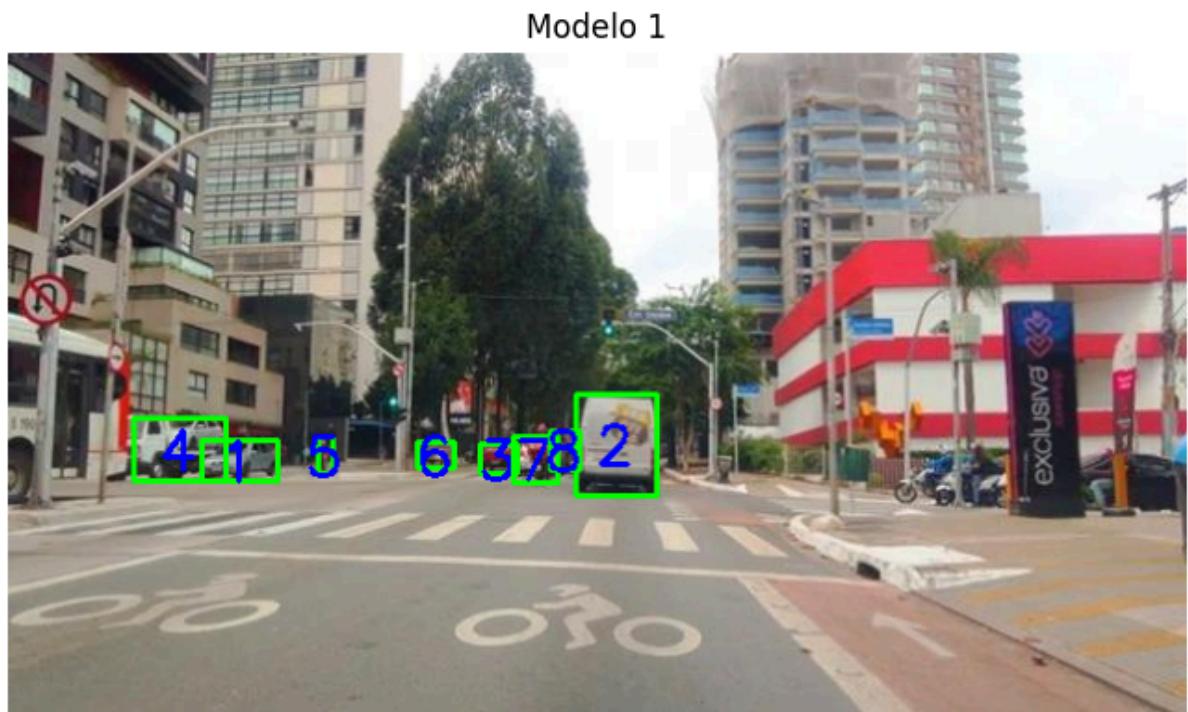


Figura 8: Predições do modelo base na segunda imagem de análise

Modelo 2

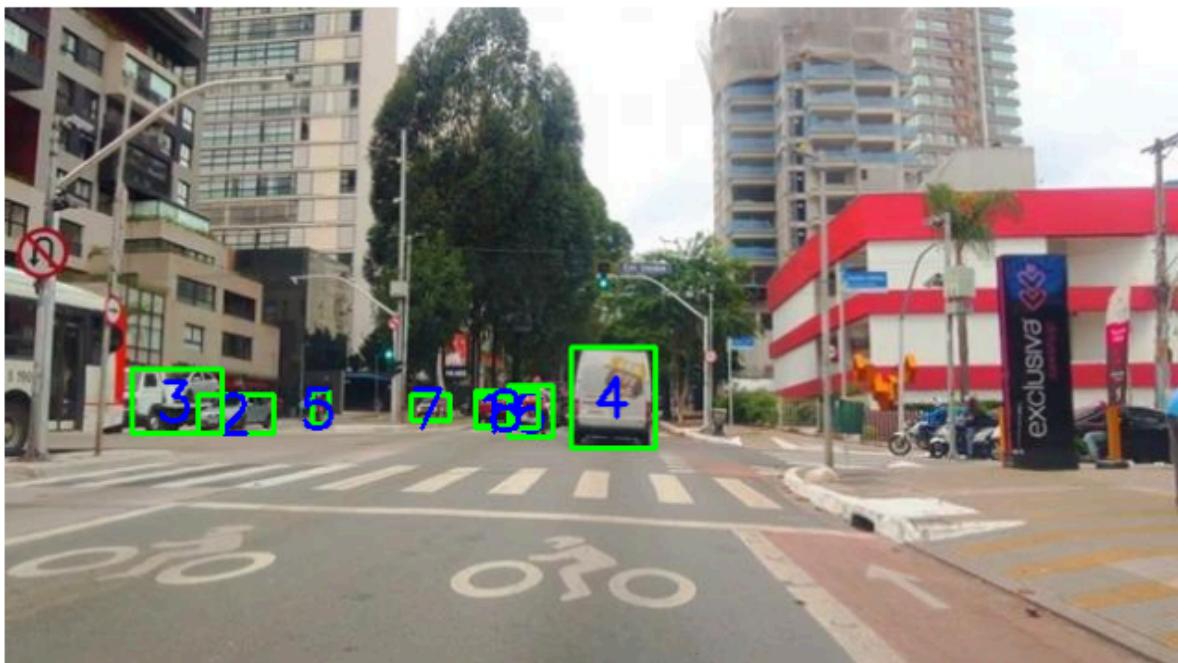


Figura 9: Predições do modelo retreinado na segunda imagem de análise

2.2.3: Análise de Casos Específicos - Frame 3

Passado para o terceiro caso, notamos uma extensão do desafio anterior. Como podemos observar, o modelo retreinado falhou em identificar o semáforo, que antes era identificado pelo modelo base. Tal fato indica novamente que a quantidade de épocas em que o modelo foi retreinado não foi suficiente para que o mesmo aprendesse toda a distribuição do conjunto de dados, dando ainda mais suporte para o que foi discutido na análise quantitativa.

Modelo 1



Figura 10: Predições do modelo base na terceira imagem de análise

Modelo 2

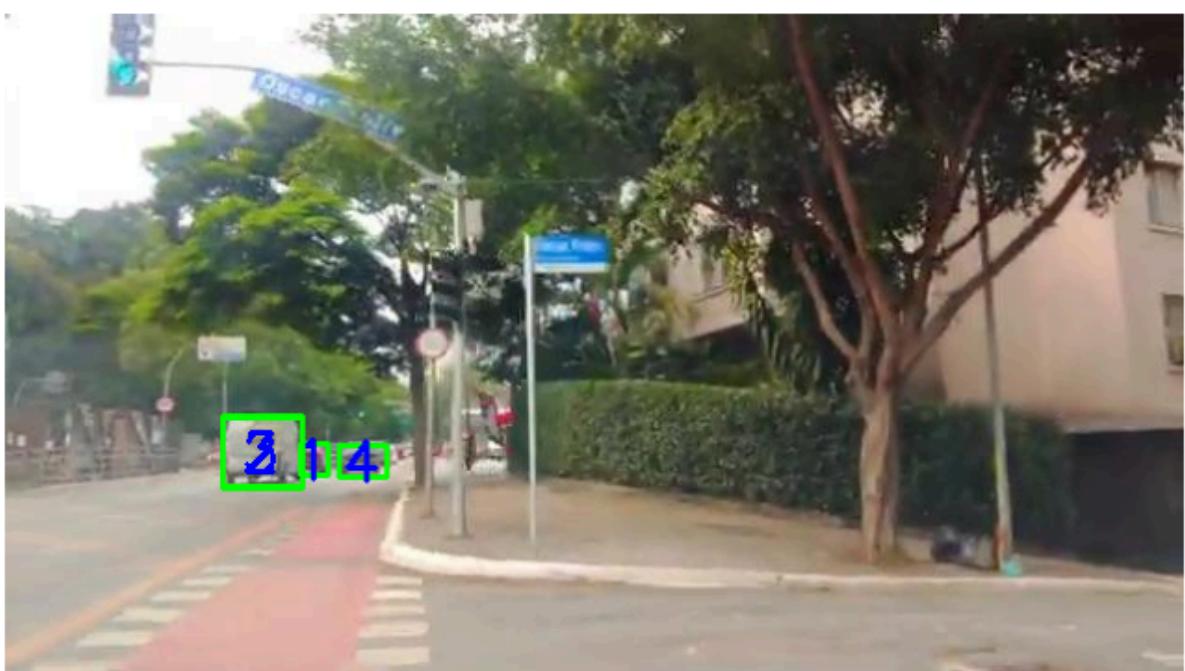


Figura 11: Predições do modelo retreinado na terceira imagem de análise

2.2.4: Análise de Casos Específicos - Frame 4

Como último caso de análise, decidimos trazer essa última imagem pois ela ilustra uma tendência clara observada.

Como podemos ver, o primeiro modelo falha novamente em identificar todos os semáforos presentes nas imagens. O segundo modelo também apresenta dificuldades claras na identificação da maioria dos objetos de análise nesse caso. No entanto, podemos ver que os semáforos do canto superior direito foram identificados pelo modelo retreinado. Isso é relevante pois nossa GAN gerou casos parecidos com certa frequência (com os semáforos de lado). Isso corrobora com a hipótese de que as imagens geradas estão sim adicionando conhecimento ao modelo, e que precisamos apenas aumentar o volume de épocas e de imagens geradas para treinamento para obtermos resultados mais satisfatórios.

Modelo 1

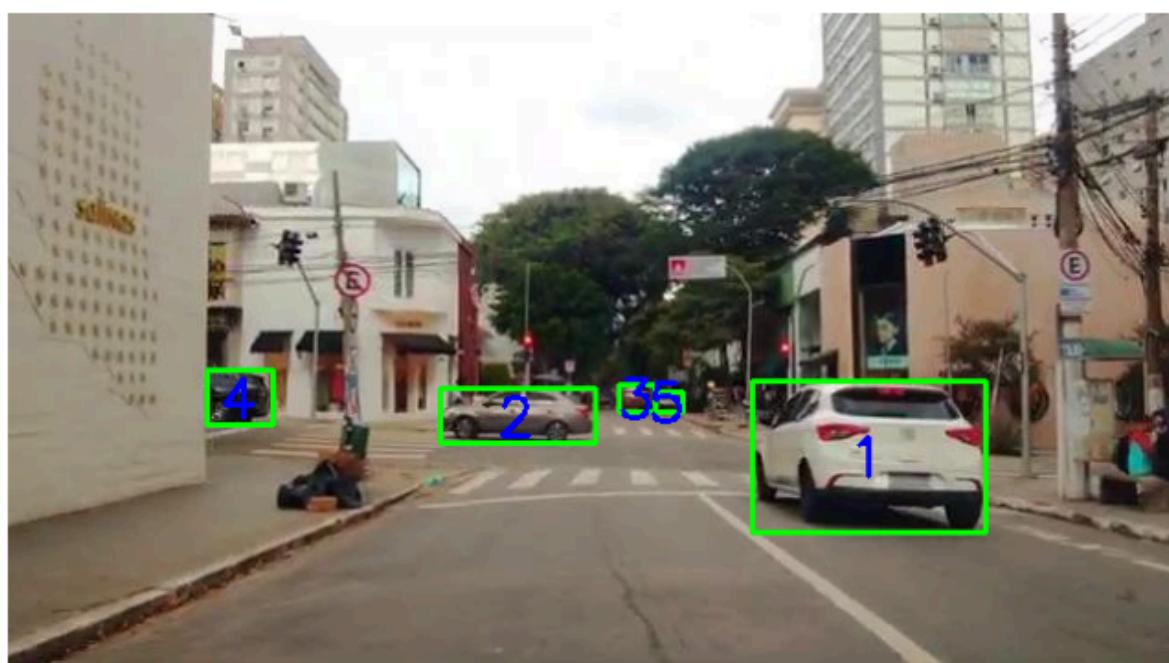


Figura 12: Predições do modelo base na quarta imagem de análise

Modelo 2

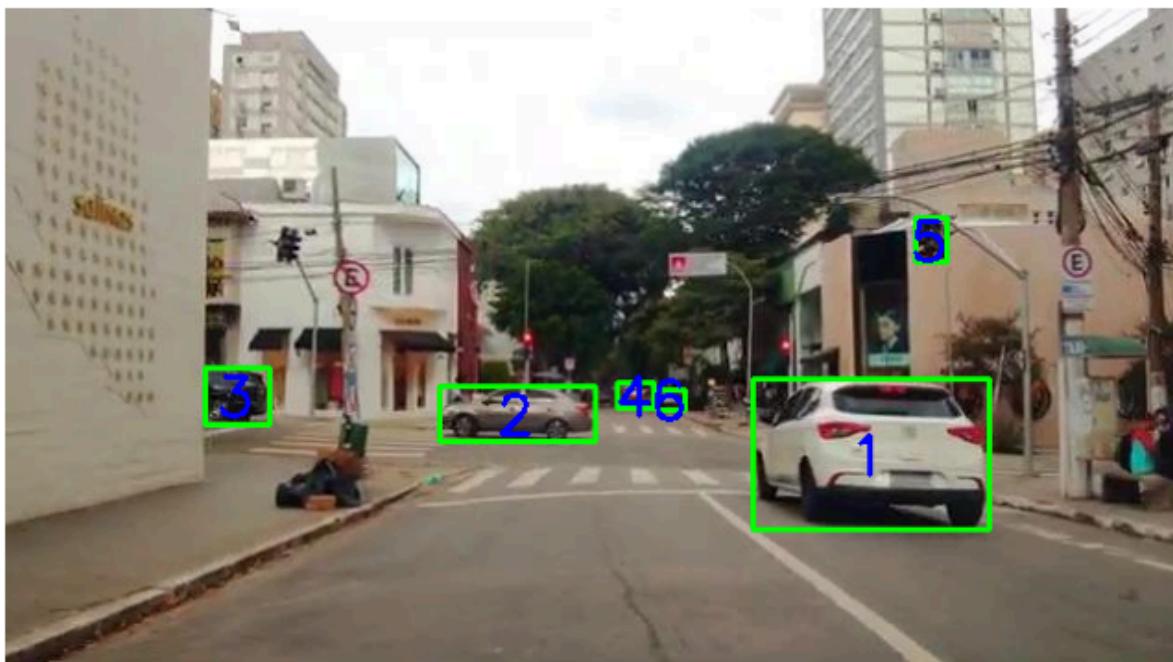


Figura 13: Predições do modelo retreinado na quarta imagem de análise

LIMITAÇÕES

Este trabalho enfrenta várias limitações que impactam diretamente sua abrangência e os resultados obtidos. Uma das principais limitações é o poder computacional disponível. A capacidade de processamento restrita limita a complexidade dos algoritmos utilizados tanto para a geração de imagens quanto para a detecção de objetos. Isso implica na necessidade de simplificar os modelos, reduzir o tamanho dos conjuntos de dados. Tais restrições impactaram a qualidade das imagens geradas e podem ter reduzido a capacidade de generalização dos detectores de

objetos, potencialmente diminuindo a eficácia do modelo em cenários complexos do mundo real.

Adicionalmente, a quantidade e a diversidade de dados disponíveis para treinamento também representam uma limitação significativa. Os dados foram coletados e rotulados manualmente a partir de frames de vídeo, um processo demorado e que pode resultar em um conjunto limitado em termos de variedade e volume. Isso pode não refletir adequadamente a diversidade de cenários encontrados em aplicações práticas, comprometendo a generalização do modelo para novas situações não vistas durante o treinamento.

A representatividade dos dados sintéticos gerados pelas GANs em relação aos dados reais é outra área de preocupação. Embora as GANs sejam capazes de produzir imagens visualmente similares às reais, discrepâncias e artefatos ainda são presentes e podem distorcer a percepção do modelo treinado. Essas discrepâncias são especialmente problemáticas em aplicações críticas como segurança e vigilância, onde a precisão na representação e detecção de objetos é crucial.

Mitigação e Melhorias Propostas: Para mitigar essas limitações, algumas estratégias podem ser consideradas. O aumento do poder computacional, seja através de hardware mais avançado ou pelo uso de recursos de computação em nuvem, pode permitir a utilização de modelos mais complexos e o processamento de conjuntos de dados maiores. Além disso, métodos automatizados e mais eficientes de coleta de dados podem ser explorados para enriquecer os conjuntos de dados com uma maior diversidade de imagens. Finalmente, aperfeiçoamentos contínuos nos algoritmos de geração de GANs devem ser priorizados para melhorar

a fidelidade e reduzir os artefatos nas imagens geradas, aproximando-as ainda mais da realidade.

É crucial que futuras investigações enderecem estas limitações para aprimorar a aplicabilidade e eficácia dos modelos de detecção de objetos baseados em aprendizado profundo, garantindo que eles possam operar de forma confiável em uma ampla gama de condições práticas.

CONCLUSÃO

Neste estudo, investigamos a aplicação de Redes Adversariais Generativas (GANs) para aumento de dados (data augmentation) e seu impacto no desempenho do modelo YOLO V8n para detecção de semáforos em cenários de trânsito. O objetivo principal foi avaliar se a utilização de imagens sintéticas geradas por GANs poderia melhorar a precisão do modelo YOLO V8n.

Os resultados obtidos foram promissores. Observamos que o modelo retreinado com as novas imagens geradas apresentou melhorias em alguns aspectos. Verificamos que o modelo retreinado foi capaz de identificar alguns semáforos que o modelo YOLO V8n original não conseguia detectar. Embora o contrário também tenha ocorrido, onde o novo modelo não identificou alguns semáforos que o modelo original identificava, o fato de o modelo ter sido capaz de descobrir novos padrões é bastante interessante. Isso indica que a abordagem de usar GANs para gerar imagens adicionais é viável e pode contribuir para a melhoria dos modelos de detecção de semáforos.

No entanto, o estudo também revelou algumas limitações que apontam direções claras para trabalhos futuros. Aumentar o número de imagens geradas pela GAN e prolongar o tempo de treinamento do YOLO V8n são passos essenciais para potencializar ainda mais os resultados. Estes ajustes são necessários para explorar plenamente o potencial desta abordagem.

Embora nosso estudo tenha focado na detecção de semáforos, a metodologia pode ser aplicada a outros tipos de objetos de trânsito. É provável que resultados semelhantes possam ser alcançados, melhorando a precisão dos modelos de detecção para uma variedade de objetos.

Concluímos que a utilização de GANs para aumento de dados é uma técnica promissora. O uso de perturbações adversariais em data augmentation, conforme desenvolvido por Behpour et al. (2019), mostrou ser uma estratégia eficaz, levando a um modelo mais promissor capaz de performar bem em condições de teste adversas, refletindo a importância desta abordagem para o desenvolvimento futuro de sistemas de detecção de objetos mais eficientes. Antoniou et al. (2017) argumentam que "as GANs são essenciais para a geração de dados robustos para treinamento de algoritmos de visão computacional". Dewi et al. (2021) complementam essa visão ao mostrar especificamente como diferentes GANs podem ser empregadas para gerar dados de treinamento que melhoraram a detecção de sinais de trânsito, proporcionando um aumento quantificável na precisão dos modelos de detecção. Com a devida exploração e alocação de recursos adequados, esta metodologia tem o potencial de se tornar uma ferramenta valiosa na área de detecção de objetos de trânsito, oferecendo melhorias significativas em precisão e desempenho.

Além disso, futuras pesquisas podem considerar a aplicação desta abordagem em outras versões do YOLO, como o YOLO V8 em suas diferentes variantes, bem como versões mais atualizadas, como o YOLO V9 e futuras evoluções do modelo. Explorar estas versões pode proporcionar insights adicionais e potencialmente melhorar ainda mais a eficácia na detecção de objetos de trânsito

REFERÊNCIAS

RADFORD, Alec; METZ, Luke. Unsupervised representation learning with deep convolutional generative adversarial networks. 7 jan. 2016. Disponível em: <https://arxiv.org/pdf/1511.06434.pdf>. Acesso em: 8 jun. 2024.

GOODFELLOW, Ian J. et al. Generative Adversarial Nets. In: *Advances in Neural Information Processing Systems*, 2014. Disponível em: https://proceedings.neurips.cc/paper_files/paper/2014/file/5ca3e9b122f61f8f06494c97b1afccf3-Paper.pdf. Acesso em: 8 jun. 2024.

DEWI, Christine et al. Yolo V4 for advanced Traffic Sign Recognition with synthetic training data generated by various GAN. 2021. Disponível em: https://web.archive.org/web/20220522005156id_/_https://ieeexplore.ieee.org/ielx7/6287639/6514899/09471877.pdf. Acesso em: 8 jun. 2024.

ZHU, J.-Y. et al. Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks. In: Proceedings of the IEEE International Conference on Computer Vision (ICCV), 2017. Disponível em: https://openaccess.thecvf.com/content_ICCV_2017/papers/Zhu_Unpaired_Image-To-Image_Translation_ICCV_2017_paper.pdf. Acesso em: 16 jun 2024.

SHRIVASTAVA, A. et al. Learning from Simulated and Unsupervised Images through Adversarial Training. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017. Disponível em: <https://arxiv.org/pdf/1612.07828.pdf>. Acesso em: 16 jun 2024.

ANTONIOU, A.; STORKEY, A.; EDWARDS, H. Data Augmentation Generative Adversarial Networks. arXiv preprint arXiv:1711.04340, 2017. Disponível em: <https://arxiv.org/pdf/1711.04340.pdf>. Acesso em: 16 jun 2024.

BEHPOUR, Sima; KITANI, Kris M.; ZIEBART, Brian D. ADA: Adversarial Data Augmentation for Object Detection. [S.l.]: Univ. of Illinois at Chicago; Carnegie Mellon University, 2019. Disponível em: <https://par.nsf.gov/servlets/purl/10098118>. Acesso em: 16 jun 2024.