

# Plano de Iniciação Científica

Graduando *Mateus Matzkin Gurza*, orientado por *Roberto Marcondes César Júnior*

## 1. Introdução

O desenvolvimento de tecnologias baseadas em visão computacional tem mostrado um crescimento significativo nos últimos anos, especialmente no contexto de aplicações urbanas. As cidades, como epicentros da vida moderna, estão cada vez mais dependentes da inovação tecnológica para lidar com desafios de mobilidade, acessibilidade e segurança. Nesse cenário, a inteligência artificial (IA) se tornou uma ferramenta importante para pensar o futuro da eficiência dos sistemas urbanos.

O presente projeto de iniciação científica se insere na temática da utilização de visão computacional para a análise das cidades e suas particularidades. O objetivo principal é aplicar e avaliar os modelos de Convolutional Neural Networks (CNN's)<sup>1</sup>, usados em problemas de detecção e tracking<sup>2</sup>, em circunstâncias realistas, ampliando o domínio do uso da IA na exploração do meio urbano. Por meio dessa abordagem prática, espera-se obter dados e indicadores relevantes sobre os cenários de estudo e, no processo, entender formas de aprimorar a maneira como empregamos algoritmos de detecção de automóveis, pedestres e outras entidades que se movimentam na esfera sócio-urbana.

Este estudo pretende colaborar com o ecossistema de pesquisa centrado na viabilização de cidades mais inteligentes, acessíveis, sustentáveis, seguras e com maior mobilidade, através do uso de Inteligência Artificial. Especificamente, o projeto está inscrito na produção acadêmica do grupo Creativision da Universidade de São Paulo. O núcleo é responsável por pesquisas avançadas nas áreas de visão computacional e modelos multimodais na USP. Considerando o exposto, o trabalho, aqui proposto, também terá como foco a interação com outras iniciativas no campo.

## 2. Revisão Bibliográfica

O intuito desta revisão bibliográfica é expor as linhas de atuação e iniciativas do mundo acadêmico no que tange a implementação de métodos de IA, com foco em visão computacional, nas cidades. Deseja-se delimitar o estado da arte dessa ferramenta e levantar casos de uso relevantes nesse contexto. Com isso feito, será possível, posteriormente, situar o trabalho proposto neste plano de pesquisa, entendendo onde ele se insere e a que fins contribuirá.

Para situar o leitor acerca dos diferenciais do uso

de IA para o estudo do meio sócio-urbano, abordar-se-á, de antemão, um rápido exemplo. Imagine que se quer saber qual caminho conectando o ponto A ao ponto B é mais acessível para idosos. Uma forma antiquada de realizar isso seria fazer uma pessoa percorrer e avaliar todos esses percursos para depois comparar cada caminho. Esse método obsoleto tem como desvantagens: a interferência da subjetividade, dada a inclusão do avaliador humano e sua interpretação do que seria uma obstrução hostil a um idoso; o grande consumo de tempo, dada a necessidade de calcular parâmetros que permitam comparar os trajetos; e a ausência de uma padronização, dada a dificuldade de analisar igualmente todos os percursos. Outra maneira, mais moderna, de abordar esse problema seria utilizar visão computacional em vídeos desses trajetos para identificar obstáculos neles e comparar, de maneira automática, parâmetros obtidos de forma padronizada sobre todos os caminhos percorridos; indicando, por fim, o percurso mais acessível.

Assim, a IA aumenta a eficiência com a qual podemos analisar e identificar indicadores importantes sobre as cidades. Esse ganho em poder de análise se traduz em um potencial de entender e, portanto, impactar diversos aspectos do urbanismo através do design e controle urbano. Entre eles, a acessibilidade é um fator sob o qual a utilização da inteligência artificial pode ter grande efeito, servindo para guiar a construção de "human-centered smart cities"<sup>3</sup>, conforme explicado em "The future of Urban Accessibility: The Role of AI"[2] (*FROEHLICH* et al., 2024). A capacidade da IA de gerar conhecimento para guiar decisões urbanísticas e logísticas, possibilitará a transformação do ambiente da cidade em um espaço de utilidade pública mais seguro e acessível. Para que se alcance esse objetivo, é necessário, então, através de machine learning, capturar indicadores importantes sobre as interações entre os elementos constitutivos do espaço estudado, por exemplo: calçadas, pedestres, carros, ruas e outras infraestruturas e agentes; para que assim seja possível usar os parâmetros encontrados para facilitar a tomada de decisões estratégicas sobre esse ambiente.

O estudo das redes de transporte, por exemplo, é um campo onde iniciativas direcionadas para a captação desses indicadores vem ocorrendo por meio de ferramentas baseadas em visão computacional, tal como os métodos de video-based "Automated Passenger Counting" (APCs) (*PRONELLO; GARZÓN*, 2020). Entre as infraestruturas que compõe o meio urbano, as redes de

<sup>1</sup>Redes neurais convolucionais profundas

<sup>2</sup>Tracking seria o rastreamento do objeto em movimento em um vídeo

<sup>3</sup>Tradução Livre: cidades inteligentes centradas nas pessoas

transporte podem ser facilmente associadas às temáticas de: sustentabilidade, pela emissão de gases de estufa; desigualdade socio-espacial, pela capacidade - ou sua ausência - de conectar pontos distantes; e segurança, pelas interações entre veículos e entre estes e pedestres. Visto isso, a aplicação de IA para analisar essas redes e obter parâmetros que permitam pensar em formas de melhorá-las é crucial para a consolidação de "Intelligent Transport Systems (ITS)"<sup>4</sup>, uma das formas como, na atualidade, pode-se, segundo Pronello e Garzón (2023, p. 1) "contribute to a safer, more efficient, and environmentally friendly transport network."<sup>5</sup> e assim, "make different aspects of urban mobility smarter..."<sup>6</sup>.

Além das redes de transporte, a infraestrutura para pedestres tem uma posição central na forma como as pessoas interagem com as cidades. Entender os problemas de acessibilidade e segurança associados às condições físicas e distribuição espacial dessa infraestrutura é essencial para guiar a maneira como ela será repensada, em prol da resolução desses problemas, no futuro. Nesse cenário, métodos de mapeamento por imagens aéreas para a produção de dados geoespaciais sobre as calçadas, por exemplo, é uma das aplicações de visão computacional destinada à construção do entendimento mais claro desta face do meio urbano; entendimento, esse, que busca corroborar com o avanço de "pedestrian-oriented city designs" e a melhora do ambiente para quem se desloca à pé na cidade. Em "Towards global-scale crowd + AI techniques to map and assess sidewalks for people with disabilities" (HOSSEINI et al., 2023) pode-se enxergar esse esforço, por meio da aplicação de IA, direcionado ao mapeamento e avaliação da infraestrutura em foco.

Entende-se que o papel da Inteligência Artificial neste contexto é o de produzir eficientemente parâmetros e indicadores que tornem mais tangível a análise do espaço urbano. Facilitar a visualização e interpretação dessas informações capturadas pelos algoritmos acerca do ambiente é um fator que incentiva a criação de novas tecnologias. Exemplos dessa influência podem ser vistos na busca de melhorar o mapeamento de calçadas, capturando as conexões da rede que interliga elas, como no projeto "Mapping the walk" (HOSSEINI et al., 2023), e até na elaboração de novos frameworks de visualização, como no Urban Toolkit para visual analytics, que tem o objetivo de facilitar a integração e visualização de dados urbanos distintos (MIRANDA et al., 2024). Ambos projetos, entre muitos outros no campo, tem como intuito a consolidação de formas claras de utilizar a informação obtida por IA para a tomada de decisões de design urbano e logística, facilitando a descoberta de insights claros e açãoáveis.

Considerando a importância da produção sistemática de indicadores precisos através dessas novas tec-

nologias, obter dados que descrevam com exatidão as particularidades de cada circunstância explorada no ambiente urbano é uma necessidade, visto que estes são cruciais para a construção de modelos mais exatos das experiências na cidade. Em outras palavras, a evolução da coleta de dados é essencial porque a qualidade deles e a disponibilidade de fontes multimodais são imperativos para a eficiência de algoritmos de aprendizado de máquina. No entanto, essa evolução não é uma tarefa fácil; o ambiente das cidades é repleto de obstáculos e condições adversas à aplicação de IA. Podemos ressaltar a poluição visual, sobreposição de objetos, perspectivas e ângulos de visão pouco convencionais, entre outros desafios. Projetos como o Streetaware (PIADYK et al., 2023), procuram, nesse contexto, disponibilizar dados de alta qualidade para a possibilitar a extração de insights mais claros sobre o espaço urbano e assim elevar o patamar da IA aplicada a este contexto. O Streetaware, especificamente, realiza essa tarefa para o cenário de interseções urbanas movimentadas, como em Brooklyn, Nova York. O projeto apresenta dados multimodais de alta resolução — incluindo áudio, vídeo e LiDAR — em três interseções, totalizando cerca de 8 horas de dados. Esses dados foram capturados por sensores sincronizados, oferecendo múltiplas perspectivas e modalidades de informação, possibilitando uma análise mais detalhada, como a descoberta de objetos oclusos, associação de eventos sonoros a suas representações visuais, rastreamento de objetos ao longo do tempo e medição da velocidade de pedestres. Esse tipo de recursos, como explicado, é muito valioso para estudos de interações urbano-ambientais e a aplicação de aprendizado de máquina nesse contexto.

Ademais, ainda sobre a captura de dados, coletá-los de forma realista, eficiente e personalizada para abordar os problemas que os métodos de IA se propuserem a mapear, é uma tarefa imprescindível para o avanço dos estudos do meio urbano. Projetos como o SideSeeing (DAMACENO et al., 2024), que apresenta um framework para coleta, processamento e análise de dados de uma perspectiva "egocêntrica" - baseada na perspectiva do pedestre e no ambiente como qual interage diretamente - representam o movimento de especialização e personalização dos dados para que apresentem mais precisamente a realidade das circunstâncias a serem analisadas através de visão computacional e modelos multimodais; No caso do SideSeeing, essa personalização busca aproximar a informação o máximo possível da vivência do humano inserido nos múltiplos cenários das cidades, entendendo as características dos percursos que tomam para ir para o hospital, por exemplo, capturando as nuances da perspectiva do indivíduo sujeito ao trajeto e, assim, possibilitando um análise mais detalhada das condições de acessibilidade de rotas importantes para a saúde pública e segurança.

Em suma, pesquisadores de diferentes áreas, impulsionados pela necessidade de repensar as cidades de forma mais inteligente para que estas atendam eficien-

<sup>4</sup>Tradução Livre: Sistemas de Transporte Inteligentes

<sup>5</sup>Tradução Livre: contribuir para uma rede de transporte mais segura, eficiente e ambientalmente amigável.

<sup>6</sup>Tradução Livre: tornar diferentes aspectos da mobilidade urbana mais inteligentes...

temente a população, vem se dedicando a projetos que utilizam AI para viabilizar isso. Nesse cenário, a visão computacional emerge como uma ferramenta fundamental. Diversos estudos destacam o uso dela em soluções urbanas, com ênfase na análise de redes de transporte, infraestrutura para pedestres e mapeamento de espaços públicos. Pesquisas recentes abordam desde o monitoramento de tráfego até a avaliação da acessibilidade em calçadas, revelando como a tecnologia permite a análise da cidade em prol de melhorar a experiência cotidiana para indivíduos em diversas condições de vida.

Entretanto, a aplicação de visão computacional em ambientes urbanos apresenta desafios únicos. Podem ser citados: poluição visual, ângulos de captação de dados não convencionais, dados que não refletem a realidade das circunstâncias sendo analisadas, sobreposição de objetos e condições climáticas adversas, como chuva ou baixa luminosidade. Projetos recentes, como o Streetaware e o SideSeeing, ilustram o esforço em superar essas dificuldades, ressaltando a importância da adaptação das tecnologias para as especificidades do ambiente urbano, garantindo a captura realista e eficiente dos dados necessários para gerar insights acionáveis.

Nesse contexto, o presente projeto se propõe a aprofundar o estudo sobre a aplicação de modelos de visão computacional no meio urbano, explorando em particular a tarefa de detecção para tracking de entidades em movimento nesse espaço. A seguir, apresenta-se o escopo da pesquisa, onde será delineado como o projeto busca contribuir para o avanço da inteligência artificial como ferramenta de análise das cidades.

### 3. Escopo do trabalho

Este projeto de iniciação científica, inserido no ecossistema das produções acadêmicas do núcleo Creativision da Universidade de São Paulo (USP), tem como objetivo aplicar os modelos de visão computacional, reconhecidos como estado da arte, para a detecção de entidades no meio sócio-urbano. Em especial, o foco será a análise de dados provenientes de vídeos para a identificação de corpos em movimento nas cidades. O intuito dessa iniciativa será estudar os conceitos e fundamentos de redes neurais convolucionais de detecção de objetos e avaliar o desempenho e aplicabilidade desses modelos em condições marcadas pelas particularidades do meio urbano. A pesquisa utilizará algoritmos de ponta atuais, como a versão mais recente do YOLO (You Only Look Once), EfficientDet ou SSD (Single Shot MultiBox Detector), para a detecção de automóveis e pedestres em tempo real.

A análise será inicialmente realizada para casos base simplificados, ou seja, sem movimento de câmera e utilizando apenas vídeos dentre as modalidades de informação capturadas. Os dados utilizados serão aqueles captados pelo grupo Creativision, os quais também estão sendo utilizados em outros projetos. O método de obtenção é baseado em 3 câmeras fixas dispostas estratégicamente em diferentes perspectivas de um mesmo

ponto; isso será feito para múltiplos locais movimentados do campus Butantã da USP. Este setup inicial permitirá entender os efeitos diretos - sem interferência de fatores como movimento da câmera ou informações não visuais - das condições do ângulo de visão na tarefa de detecção. Concomitantemente, esse caso base introduzirá as dificuldades intrínsecas do cenário da cidade em circunstâncias de baixa luminosidade, chuva, poluição visual, entre outros. Segue uma imagem na qual é possível identificar pontos de oclusão (barras metálicas do ponto de ônibus), ângulos de perspectiva não convencionais (visualização da parte traseira do caminhão), ambientes poluídos e objetos pequenos (carro e pessoa à distância no fundo da imagem).



Figure 1: Imagem representativa de captura realizada, pelo Creativision, nas condições explicadas.

O objetivo da aplicação inicial dos modelos será a quantificação e identificação de agentes no ambiente durante suas interações com ele. Ou seja, a pesquisa buscará calcular indicadores como fluxo de pessoas e veículos, com o intuito de gerar dados relevantes sobre a dinâmica urbana. Nesse processo, serão estabelecidos frameworks claros de avaliação e métricas que possibilitem a comparação entre os desempenhos em diferentes cenários, permitindo uma análise abrangente e crítica.

Como dito anteriormente, este projeto se insere em um contexto de colaboração e interdisciplinaridade, interagindo com outras pesquisas em andamento no Creativision. Essa sinergia permitirá a troca de informações e técnicas entre projetos que também exploram a coleta e interpretação de dados no ambiente urbano, incluindo o mapeamento das características das infraestruturas urbanas. Com isso em mente, existem perspectivas de expansão dos casos abordados nesta pesquisa para abranger os frameworks e ferramentas em desenvolvimento no Laboratório, como por exemplo o SideSeeing. Assim, dada a interlocução com outros esforços na área de IA, associados à rede integrada do Creativision e acadêmicos parceiros, espera-se que este trabalho contribua para o avanço da aplicação de visão computacional em ambientes urbanos, ajudando a identificar limitações e potencialidades dos algoritmos existentes e, também, buscando fornecer insights para o desenvolvimento de abordagens mais robustas e adaptáveis ao contexto das cidades.

## 4. Método

A detecção de objetos é um problema amplamente estudado na área de Visão Computacional, que consiste em identificar e localizar objetos de interesse em imagens e vídeos. Tipicamente, as ocorrências desses objetos são indicadas por caixas delimitadoras que os envolvem. Quando múltiplas classes de objetos são consideradas, as caixas também recebem rótulos de classe, identificando o tipo de objeto detectado.

Modelos de redes neurais profundas, como o YOLO (You Only Look Once), realizam a detecção de objetos de forma eficiente, em um único processo de ponta-a-ponta. Para treinar esses modelos, o problema de otimização é tratado como uma combinação de duas tarefas: classificação, que lida com a identificação da classe do objeto, e regressão, que determina a localização e o tamanho da caixa delimitadora. Uma vez devidamente treinado, o modelo pode ser utilizado para detectar objetos de mesmas classes em novas imagens ou quadros de vídeo, gerando uma lista de objetos detectados, cada um associado a uma caixa delimitadora e a uma pontuação que indica a probabilidade das classes.

Os principais desafios enfrentados na detecção de objetos incluem a variação de escala (objetos muito pequenos ou muito grandes), oclusão parcial dos objetos, variação na aparência (como cor, textura, formato ou diferentes ângulos de visão), além de questões como a qualidade das imagens capturadas. Apesar dessas dificuldades, as redes neurais profundas têm demonstrado alto desempenho na detecção, principalmente de objetos comuns como veículos, pedestres e alguns tipos de animais, mesmo em cenários urbanos complexos.

Para compreender o funcionamento de redes como o YOLO, é essencial ter um entendimento sólido sobre os fundamentos de redes neurais, incluindo conceitos de propagação direta (forward propagation) e reversa (back propagation), função de perda (loss function) e o uso do algoritmo de descida do gradiente para otimização (gradient descent). Além disso, o conhecimento sobre redes neurais convolucionais (CNNs) é crucial, pois estas formam a base dos principais modelos de detecção de objetos.

No contexto da avaliação de desempenho de algoritmos de detecção, para utilizar de forma crítica os modelos propostos é importante, também, entender métricas como mean Average Precision (mAP), cuja utilização é comum para a avaliação da performance de detectores, e a qual é calculada com base na Intersection over Union (IoU). A IoU mede a sobreposição entre a caixa delimitadora predita e a caixa delimitadora real, sendo um indicador importante da precisão do modelo na delimitação e localização de objetos.

Neste projeto, serão estudados esses conceitos fundamentais de redes neurais e detecção de objetos, com foco na implementação e teste de redes YOLO ou de arquiteturas similares para a detecção de veículos e pedestres em sequências de vídeo capturadas por câmeras,

inicialmente, estáticas. O objetivo prático final do modelo será classificar e contar o número de entidades em movimento em um intervalo de tempo determinado.

## 5. Cronograma

- **Estudo dos fundamentos e conceitos básicos – 3 meses**

O aluno dedicará este período ao estudo de redes neurais, redes neurais convolucionais (CNNs) e outros conceitos fundamentais de aprendizado de máquina, como funções de perda, otimização por descida do gradiente e métricas de avaliação de desempenho.

- **Estudo do problema de detecção de objetos e modelos de redes neurais para detecção de objetos – 2 meses**

Nesse estágio, o foco será o estudo aprofundado do problema de detecção de objetos em imagens e vídeos, com ênfase nas arquiteturas da família YOLO e outros modelos relevantes. Também serão abordados métodos para adaptação desses modelos ao contexto urbano.

- **Implementação e testes de modelos YOLO em conjuntos de dados padrão de detecção de objetos – 2 meses**

O aluno irá implementar e testar os modelos YOLO em datasets padrão de detecção de objetos, com adaptações para processar sequências de vídeo. Isso permitirá familiarização com as ferramentas e algoritmos e verificação do desempenho em diferentes cenários.

- **Aplicação do modelo aos dados de vídeo coletados no projeto temático – 3 meses**

Neste estágio, os modelos treinados serão aplicados aos dados de vídeo coletados em ambientes urbanos. Isso incluirá a ajuda na anotação dos quadros dos vídeos para gerar dados precisos que possam ser usados nas análises subsequentes.

- **Aplicação do modelo aos dados de vídeo coletados no projeto temático – 3 meses**

Neste estágio, os modelos treinados serão aplicados aos dados de vídeo coletados em ambientes urbanos. Isso incluirá a ajuda na anotação dos quadros dos vídeos para gerar dados precisos que possam ser usados nas análises subsequentes.

- **Aplicação do modelo aos dados de vídeo coletados no projeto temático – 3 meses**

Neste estágio, os modelos treinados serão aplicados aos dados de vídeo coletados em ambientes urbanos. Isso incluirá a ajuda na anotação dos quadros dos vídeos para gerar dados precisos que possam ser usados nas análises subsequentes.

- **Avaliação dos resultados – 1 mês**

Os resultados da aplicação dos modelos serão avaliados, utilizando métricas de desempenho como mean Average Precision (mAP) e Intersection over Union (IoU). Será feita uma análise crítica do desempenho em cenários reais e identificação de possíveis melhorias.

- **Redação do relatório científico – 1 mês**

Ao final do ciclo de trabalho, será redigido um relatório científico documentando as metodologias, resultados e conclusões obtidas durante o projeto.

- **Possível renovação do projeto (Meses 13 a 24)**

Dependendo dos resultados e do progresso alcançado nos primeiros 12 meses, o projeto poderá ser renovado, visando aprofundar as análises ou explorar novas abordagens de detecção e análise de dados urbanos.

## 6. Justificativa e Motivação

Considerando os pontos levantados durante a revisão bibliográfica, esta seção tem como objetivo expor os motivos por trás da iniciativa de pesquisa aqui proposta, justificando-a diante do cenário atual do campo de estudos da aplicação de IA em questões de urbanismo. Para tanto, é importante entender que, atualmente, a aplicação de visão computacional, somada a outros métodos de inteligência artificial, é uma ferramenta cujas características permitem uma abordagem de analítica importante para aprofundar o conhecimento das dinâmicas do meio urbano. As características citadas envolvem: a produção padronizada de resultados sobre os dados capturados; a automação do processamento dessas informações obtidas e, também, a consolidação de análises livres de subjetividade humana.

Como explicado anteriormente, ao longo da seção 2, as dinâmicas urbanas sob estudo referem-se ao diálogo entre diferentes entidades e infraestruturas, ou seja, as interações, por exemplo, entre pedestres, automóveis, calçadas e redes de transporte. Estas definem experiências intrínsecas à cidade, às quais todos os indivíduos estão submetidos de formas distintas, dado que a acessibilidade e a vivência nesse ambiente estão intimamente ligadas à realidade pessoal, socioeconômica e espacial de cada um. A forma como as pessoas de diferentes faixas etárias, condições físicas e financeiras estão sujeitas ao meio urbano são os pontos-chave de atenção na busca pela construção de cidades mais acessíveis, seguras, sustentáveis e igualitárias. Portanto, a capacidade de medir e analisar essas condições e produzir informações confiáveis acerca delas é crucial para a tomada de decisões no que tange à engenharia e ao design do meio sociourbano. Nesse contexto, a inteligência artificial possibilita a produção sistemática de conhecimento sobre a relação entre os indivíduos e as infraestruturas, e, como a busca por melhorar as interações entre essas entidades molda as

cidades, cada vez mais a inteligência artificial ajudará a repensar e reconfigurar o espaço urbano.

As vantagens da utilização de IA como método de análise, se comparada à alternativa mais antiquada, indicada no exemplo da seção 2, são o seguinte conjunto de capacidades notáveis: processar grandes volumes de dados muito mais rápido que o humano; obter análises em tempo real; identificar padrões e nuances complexas que humanos deixam passar; automatizar fluxos de informação por meio da conexão das tecnologias às redes e computação em nuvem; e, por fim, a produção de métricas não subjetivas, as quais podem representar numericamente, com uma nota por exemplo, ou categoricamente, com uma classificação, as condições do fenômeno urbano analisado, tornando mais tangível a análise do cenário que o envolve. Tarefa pouco replicável sem a utilização de aprendizado de máquina. A IA então, é uma ferramenta importante para capturar as dinâmicas dos elementos urbanos, e produzir insights cruciais para a tomada de decisões estratégicas nas cidades.

Frente à necessidade de entender os pontos chaves das interações humanas com as cidades, para torná-las mais acessíveis, seguras, sustentáveis e igualitárias, e considerando as capacidades de análise por meio de IA, entende-se que esta se tornará uma ferramenta indispensável para pensar o futuro do urbanismo. Em função disso, a compreensão dos desafios associados à aplicação dos métodos de Inteligência Artificial, no contexto estudado, ganha cada vez mais relevância. Ao entendê-los, torna-se possível identificar e guiar a forma como as iniciativas na área podem lidar com eles e, consequentemente, contribuir para o aprimoramento da utilização de IA como método de análise urbana.

Assim, a pesquisa se justifica por contribuir para a ampliação do uso sistêmico da inteligência artificial como método de promover avanços na engenharia do meio urbano. Isso porque essa se beneficia de entender como aplicar e avaliar os modelos - atualmente representantes do estado da arte da visão computacional - em tarefas referentes à exploração de métricas relevantes dos cenários das cidades. A pesquisa pretende facilitar esse entendimento, buscando otimizar tais modelos para a detecção de entidades em movimento e, sobretudo, consolidar o conhecimento acerca de sua utilização nesse contexto.

## Referências

- [1] R. J. P. Damaceno et al. *SideSeeing: A multimodal dataset and collection of tools for sidewalk assessment*. 2024. arXiv: 2407.06464 [cs.CV]. URL: <https://arxiv.org/abs/2407.06464>.
- [2] Jon E Froehlich et al. “The Future of Urban Accessibility: The Role of AI”. In: (2024).
- [3] Maryam Hosseini et al. “Mapping the walk: A scalable computer vision approach for generating sidewalk network datasets from aerial imagery”. In: *Computers, Environment and Urban*

*Systems* 101 (2023), p. 101950. ISSN: 0198-9715.  
DOI: <https://doi.org/10.j.compenvurbsys.2023.101950>. URL: <https://www.sciencedirect.com/science/article/pii/S0198971523000133>.

- [4] Maryam Hosseini et al. “Towards global-scale crowd+ AI techniques to map and assess sidewalks for people with disabilities”. In: *arXiv preprint arXiv:2206.13677* (2022).
- [5] Deepak Kumar Jain et al. “Robust multi-modal pedestrian detection using deep convolutional neural network with ensemble learning model”. In: *Expert Systems with Applications* 249 (2024), p. 123527.
- [6] Yangke Li and Xinman Zhang. “Multi-modal deep learning networks for RGB-D pavement waste detection and recognition”. In: *Waste Management* 177 (2024), pp. 125–134.
- [7] Fabio Miranda et al. “The Urban Toolkit: A Grammar-Based Framework for Urban Visual Analytics”. In: *IEEE Transactions on Visualization and Computer Graphics* 30.1 (2024), pp. 1402–1412. DOI: [10.1109/TVCG.2023.3326598](https://doi.org/10.1109/TVCG.2023.3326598).
- [8] Yurii Piadyk et al. “Streetaware: A high-resolution synchronized multimodal urban scene dataset”. In: *Sensors* 23.7 (2023), p. 3710.
- [9] Caio Pieroni et al. “Big data for big issues: Revealing travel patterns of low-income population based on smart card data mining in a global south unequal city”. In: *Journal of Transport Geography* 96 (2021), p. 103203.
- [10] Cristina Pronello and Ximena Rocio Garzón Ruiz. “Evaluating the Performance of Video-Based Automated Passenger Counting Systems in Real-World Conditions: A Comparative Study”. In: *Sensors* 23.18 (2023). ISSN: 1424-8220. DOI: [10.3390/s23187719](https://doi.org/10.3390/s23187719). URL: <https://www.mdpi.com/1424-8220/23/18/7719>.
- [11] Claudio T Silva et al. “Integrated Analytics and Visualization for Multi-modality Transportation Data”. In: (2019).