UNIVERSIDADE FEDERAL DO CEARÁ

CAMPUS QUIXADÁ

CURSO DE GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO

KASSIANE LOPES FAÇANHA

MONITORAMENTO DE DISTANCIAMENTO SOCIAL UTILIZANDO APRENDIZADO DE MÁQUINA E VISÃO COMPUTACIONAL

QUIXADÁ

2021

KASSIANE LOPES FAÇANHA

MONITORAMENTO DE DISTANCIAMENTO SOCIAL UTILIZANDO APRENDIZADO DE MÁQUINA E VISÃO COMPUTACIONAL

Trabalho de Conclusão de Curso apresentado ao Curso de Graduação em Ciência da Computação do Campus Quixadá da Universidade Federal do Ceará, como requisito parcial à obtenção do grau de bacharel em Ciência da Computação.

Orientador: Prof. Dr. Paulo Victor Barbosa de Sousa

QUIXADÁ

2021

Dados Internacionais de Catalogação na Publicação
Universidade Federal do Ceará
Biblioteca Universitária
Gerada automaticamente pelo módulo Catalog, mediante os dados fornecidos pelo(a) autor(a)

F123m Façanha, Kassiane Lopes.
Monitoramento de distanciamento social utilizando Aprendizado de Máquina e Visão Computacional / Kassiane Lopes Façanha. – 2022.
42 f. : il. color.

Trabalho de Conclusão de Curso (graduação) – Universidade Federal do Ceará, Campus de Quixadá, Curso de Ciência da Computação, Quixadá, 2022.
Orientação: Prof. Dr. Paulo Victor Barbosa de Sousa.

1. Cluster (Sistema de computador)-Detecção. 2. Visão Computacional. 3. Aprendizado do computador. I. Título.

CDD 004

KASSIANE LOPES FAÇANHA

MONITORAMENTO DE DISTANCIAMENTO SOCIAL UTILIZANDO APRENDIZADO DE MÁQUINA E VISÃO COMPUTACIONAL

Trabalho de Conclusão de Curso apresentado ao Curso de Graduação em Ciência da Computação do Campus Quixadá da Universidade Federal do Ceará, como requisito parcial à obtenção do grau de bacharel em Ciência da Computação.

Aprovada em: ___/___/___.

BANCA EXAMINADORA

Prof. Dr. Paulo Victor Barbosa de Sousa (Orientador)
Universidade Federal do Ceará (UFC)

Prof. Dr. Paulo Armando Cavalcante Aguilar
Universidade do Federal do Ceará (UFC)

Prof. Dr. Regis Pires Magalhães
Universidade do Federal do Ceará (UFC)

Prof. Dr. Wladimir Araujo Tavares
Universidade do Federal do Ceará (UFC)

Dedico este trabalho primeiramente a Deus, por ser essencial na minha vida, à minha família, por sua capacidade de acreditar em mim e investir em mim, e aos meus avôs, Valtemi e Manuel, “In Memorian”, por me ensinarem os valores que jamais esquecerei.

AGRADECIMENTOS

Agradeço primeiro à Deus, por ter me mantido na trilha certa durante este projeto de pesquisa com saúde e forças para chegar até o final.

Aos meus pais, Cléia e Marcos, pelos sacrifícios que fizeram para que eu pudesse obter uma boa educação, pelo amor, ensinamentos e compreensão em toda minha caminhada, pelo apoio e incentivo a cada decisão que tomei e que serviram de alicerce para as minhas realizações. Obrigada por me mostrarem o caminho do conhecimento, por acreditarem nos meus sonhos e por serem meus maiores exemplos de fé.

À minha irmã, Karine, pelos incentivos, conselhos e cuidados que sempre teve comigo e não desistir de me ajudar. Agradeço à minha irmã, Kauane, por nunca duvidar de onde eu poderia chegar e por sempre estar ao meu lado quando precisei. Obrigada pela amizade e atenção dedicadas sempre que precisei. Com vocês compartilhei meus melhores momentos.

Sou grata à minha família pelo apoio que sempre me deram durante toda a minha vida, em especial ao meu tio, Jandir, por me introduzir o mundo da informática que hoje sou apaixonada e por todo incentivo e acolhimento. Agradeço à minha vó, Neném, por ser um exemplo de mulher, pelo amor e sabedoria que me inspira e por me presentear com meu primeiro notebook, que proporcionou meus primeiros passos para chegar onde estou. Agradeço á minha vó, Maria, por tanto amor, carinho, acolhimento e cuidado que sempre teve comigo. Agradeço à minha vó de coração, Lurdes, pelo cuidado e amor que tem por mim e minha família.

Ao professor, Paulo Victor Barbosa de Sousa, por toda orientação, pela paciência e pelas valiosas contribuições dadas durante todo o processo. Seus conhecimentos fizeram grande diferença no resultado final deste trabalho.

Aos professores, Paulo Armando Cavalcante Aguilar, Regis Pires Magalhães e Wladimir Araujo Tavares, pela disponibilidade em participar da banca deste trabalho e pelas excelentes colaborações e sugestões.

Agradeço ao PACCE, por ser o meu ponto de apoio durante a graduação, e aos amigos de bolsa, que me mostraram que trabalhar em equipe pode ser a melhor coisa do mundo. Sou grata pelas amizades que essa bolsa me proporcionou e que vou levar para vida toda, em especial, à Marisa, Késsia, Patrícia, Gabriela e Lorena, que compartilharam seu amor comigo e que passaram a ser minha família de coração. Obrigada pelos melhores momentos durante toda graduação. Agradeço aos professores David Sena e Valdemir Queirós, por todo incentivo, confiança, amizade e orientação que me foram dadas.

Agradeço aos meus amigos de faculdade, por me mostrarem que juntos podemos mais, em especial, agradeço à galera da área de convivência, Gabriel, Roberta, Felix, Walleson, Wesley, Rian e Karine, por me estimularem todos os dias a vencer os obstáculos, pelos dias e noites estudando juntos, pelas risadas, choros, fofocas, festas, brincadeiras, caronas e planos que fizemos juntos.

Agradeço ao Pastor Naldo e família, que me acolheram em sua casa como filha e me proporcionaram um ponto de apoio familiar sempre que precisei. Agradeço à minha melhor amiga, Dálete, por ser minha confidente de todas as horas, pelas fofocas, risadas, filmes, séries, organizações de festas, fotos e momentos especiais que sempre vou guardar. Obrigada por me aceitar como eu sou e ser sempre a pessoa que me coloca no caminho quando não vejo saída. Agradeço à Dona Ivanilda e família, que me recebeu em sua casa no momento crucial no início da minha graduação. Agradeço à minha amiga, Daiane, por ter sido a ponte para que pudesse vir a estudar em Quixadá e ser uma inspiração para mim.

Sou grata à toda igreja Bela Vista em Quixadá, Boa Viagem e Sobral, por me apoiarem com suas orações e incentivos, em especial, aos conjuntos de jovens que me proporcionaram momentos de descontração e refúgio nos momentos de angústia. Agradeço aos meus líderes de jovens e irmãos em Cristo, Argentino, Claudilene, Jocélio, Mayrla e Rael.

À escola profissionalizante de Boa Viagem e à todos os funcionários, por marcarem minha história e contribuírem para minha formação profissional e pessoal. Aos professores da área técnica que me introduziram ao mercado da área e as tecnologias que hoje tenho prazer em pesquisá-las. Agradeço aos professores, que compartilharam comigo seus conhecimentos e sabedorias de vida. Sou grata aos amigos de turma que foram a ponte para que eu chegasse até aqui, em especial minha outra melhor amiga, Érika Layanne, que sempre foi meu refúgio e alegria durante todo meu ensino médio e que amo de todo meu coração. Obrigada por me fazerem acreditar que eu podia ser bem mais do que pensava.

Ao campus UFC Quixadá e a toda sua direção, quero deixar uma palavra de gratidão por ter me recebido de braços abertos e por me proporcionarem as condições para um aprendizado muito rico. Sempre terei orgulho de carregar o emblema dessa universidade. Agradeço à todas as pessoas que fazem parte desse campus e que proporcionaram, com seu trabalho, um ambiente caloroso, cuidado e amigo durante toda minha graduação.

Agradeço a todos os professores por me proporcionar o conhecimento não apenas racional, mas a manifestação do caráter e afetividade da educação no processo de formação

profissional, por tanto que se dedicaram a mim, não somente por terem me ensinado, mas por terem me feito aprender a aprender.

E, por fim, agradeço todas as pessoas que, de alguma forma, foram essenciais para que alcançasse este objetivo com o qual sempre sonhei.

"Que os nossos esforços desafiem as impossibilidades. Lembrai-vos de que as grandes proezas da história foram conquistadas do que parecia impossível."

(Chales Chaplin)

RESUMO

Considerando a história das doenças virais humanas, desde a pandemia de gripe até a atual pandemia de COVID-19, estamos sempre em busca de formas cada vez mais inovadoras de combater essas doenças. Pesquisadores de diferentes áreas estão pesquisando incansavelmente métodos eficazes para ajudar as pessoas a se protegerem. Alguns estudos importantes têm sido propostos na área de computação, como boné inteligente, que pode ajudar na proteção por meio da detecção de altas temperaturas e avisos de quebra de distância social. Também temos detectores de distância social que utilizam câmeras de segurança e redes neurais profundas e análises de riscos de infecção através desse monitoramento. Neste trabalho é proposto o desenvolvimento de uma ferramenta com técnicas de aprendizado de máquina e visão computacional para detecção de aglomeração em espaços públicos. Foi utilizado o YOLOv3 (You Only Look Once) para detectar pessoas e calcular a distância euclidiana entre pares e verificar se há violação de distância mínima de 1,5m, conforme estabelecido pela Organização Mundial de Saúde (OMS). Por conta de ser uma conversão simples, essa a distância mínima não chega a ser a exata em relação a distância real mas chega próximo. Os testes foram realizados utilizando um vídeo pré-gravado e um vídeo em tempo real no local por onde as pessoas costumam caminhar. Foi feita uma análise qualitativa e foi calculada a acurácia de cada vídeo em determinados momentos. A acurácia média foi de 94,66% para vídeos pré-gravados, 90,28% para vídeos gravados em tempo real e 92,47% para vídeos gravados em tempo real, sem uso de GPU (Unidade de Processamento Gráfico) e alto FPS (Frames por segundos), sendo considerado muito eficiente, em questão de velocidade do vídeo e da detecção, principalmente em comparação com outros trabalhos em que as GPUs ajudam. Como resultado, foi desenvolvido uma ferramenta capaz de ajudar no combate, não só ao COVID-19, mas a várias doenças de contágio similar.

Palavras-chave: Cluster (Sistema de computador)-Detecção. Visão Computacional. Aprendizado do computador.

ABSTRACT

Considering the history of human viral diseases, from the flu pandemic to the current COVID-19 pandemic, we are always looking for more and more innovative ways to combat these diseases. Researchers from different fields are relentlessly researching effective methods to help people protect themselves. Some important studies have been proposed in the area of computing, such as a smart cap, which can help in protection by detecting high temperatures and warnings of breaking social distance. We also have social distance detectors that use security cameras and deep neural networks and analysis of infection risks through this monitoring. This work proposes the development of a tool with machine learning and computer vision techniques to detect agglomeration in public spaces. YOLOv3 (You Only Look Once) was used to detect people and calculate the Euclidean distance between pairs and verify if there is a minimum distance violation of 1.5m, as established by OMS. Because it is a simple conversion, this minimum distance is not exactly the same in relation to the real distance, but it comes close. The tests were carried out using pre-recorded video and real-time video at a public place where people usually walk. A qualitative analysis was performed and the accuracy of each video was calculated at certain times. The average accuracy was 94.66% for pre-recorded videos, 90.28% for videos recorded in real-time and 92.47% for videos recorded in real-time, without using GPU (Graphics Processing Unit) and high FPS (Frames per Second), which is considered very efficient, in terms of video speed and detection, especially compared to other jobs where GPUs help. As a result, we have a tool capable of helping to fight not only COVID-19, but also several diseases WITH similar contagion.

Keywords: Cluster (Computer System)-Detection. Computer vision. Computer learning.

LISTA DE FIGURAS

LISTA DE TABELAS

LISTA DE ABREVIATURAS E SIGLAS

OMS	Organização Mundial de Saúde
YOLO	*You Only Look Once*
IA	Inteligência Artificial
CV	Visão Computacional
OpenCV	*Open Source Computer Vision Library*
MS COCO	*Microsoft Common Objects in Context*
CCTV	Câmera de vigilância fechada
TL	*Transfer Learning*
IPM	*Inverse Perspective Mapping*
NMS	*Non-maximum Suppression*
DM	Distância Mínima
CDE	Cálculo de Distância Euclidiana

SUMÁRIO

1 INTRODUÇÃO

As doenças virais têm sido uma preocupação para humanidade desde os tempos antigos. Como exemplo, a pandemia de influenza de 1918 onde quase 50 milhões de pessoas morreram. Já a gripe asiática em 1957, custou cerca de 1,1 milhão de vidas (COSTA; MERCHAN-HAMANN, 2016). Atualmente, há o surto de COVID-19 e iniciado em dezembro de 2019, afetando quase todos os países do mundo.

Com surto inicial na China, a detecção do vírus SARS-CoV-2 responsável pela pandemia de COVID-19, em fins de 2019, a população mundial tem lutado contra uma grande ameaça de nível global. Um grande número de casos ao redor do mundo surgiu de maneira muito rápida, levando a OMS a decretar uma Emergência de Saúde Pública de Importância Internacional em 30 de janeiro de 2020 (WHO, 2020b) e uma pandemia no dia 11 de março de 2020 (WHO, 2020a).

A transmissão de COVID-19 se dá através da tosse, espirros, distribuição de ar ao falar, objetos contaminados, ou por via fecal-oral (ONG *et al.*, 2020), (WANG *et al.*, 2020), podendo ser bem mais rápida devido as pessoas assintomáticas, pré-sintomáticas ou com leves sintomas que podem transmitir o vírus sem saber (KIMBALL *et al.*, 2020). Algumas opções de prevenção têm sido usadas, principalmente medidas de saúde pública não farmacológicas, usadas frequentemente no controle de epidemias, especialmente quando não há vacinas e medicamentos de prevenção. Entre as medidas indicadas para o combate do vírus estão o isolamento, a quarentena e o distanciamento social (WILDER-SMITH; FREEDMAN, 2020).

Na presente situação de pandemia, todos os países estão lutando contra a COVID-19 e procurando medidas efetivas e práticas para solucionar esse problema. Aeroportos e locais públicos eram alvos de soluções tecnológicas para controlar o contágio, entre eles, aparelhos térmicos para aferir a temperatura das pessoas e câmeras e sistemas inteligentes para controlar a distância segura.

O distanciamento social, também conhecido como "distanciamento físico", significa manter um espaço seguro entre você e as outras pessoas. O objetivo é reduzir a transmissão, diminuir o tamanho do pico da epidemia e estender os casos por mais tempo para reduzir a pressão sobre o sistema de saúde. Com isso, pesquisadores de todas as áreas trabalham para conhecer mais sobre o vírus e criar novos artefatos de combate e prevenção. Assim surgiram projetos de aprendizagem de máquina e visão computacional como o DeepSOCIAL, que é um sistema de monitoramento à distância social e avaliação de risco de infecção na pandemia

COVID-19 (REZAEI; AZARMI, 2020), e o Monitoramento de violação de distanciamento social usando *You Only Look Once* (YOLO) para detecção de pessoas (PUNN *et al.*, 2020).

Neste trabalho, propomos o desenvolvimento de uma ferramenta com técnicas da aprendizagem de máquina e visão computacional para detecção de aglomeração em espaços públicos. Como resultado, temos uma ferramenta capaz de ajudar no combate, não só ao COVID-19, mas a várias doenças de contágio similar. A longo prazo pode-se ter várias finalidades que vão desde a análise de vitrines e lojas, onde pode-se identificar onde as pessoas gostam de olhar/entrar e saber a hora certa de mudar algo para chamar mais a atenção dos clientes, até mesmo em termostatos para controle de temperatura ambiente.

1.1 Objetivos

1.1.1 Objetivo Geral

Neste trabalho, propomos o desenvolvimento de uma ferramenta com técnicas da aprendizagem de máquina e visão computacional para detecção de aglomeração em espaços públicos, ajudando na prevenção de doenças virais. Também contribui-se com a comunidade de desenvolvedores e pesquisadores através da implementação de um *software* de código aberto.

1.1.2 Objetivos Específicos

1. Desenvolver método de detecção utilizando algoritmo de Aprendizado de Máquina e Visão computacional
2. Desenvolver método para cálculo de distanciamento social
3. Realizar testes com vídeos pré-gavados e gravados em tempo real em locais públicos

1.2 Estrutura do Trabalho

O restante deste trabalho está organizado da seguinte maneira: no capítulo 2 são apresentados os principais conceitos para o desenvolvimento deste trabalho, que são Inteligência Artificial, Aprendizado de Máquina, Visão Computacional; no capítulo 3 realiza-se a apresentação e discussão dos trabalhos relacionados; no capítulo 4 a metodologia a ser empregada para o desenvolvimento deste trabalho é explicada; no capítulo 5 são apresentados os resultados deste trabalho; e, por fim, no capítulo 6 é apresentado a conclusão.

2 FUNDAMENTAÇÃO TEÓRICA

Nesta seção, serão apresentados alguns dos conceitos principais necessários para o entendimento e desenvolvimento do projeto proposto neste trabalho e a relação com o projeto.

2.1 Inteligência Artificial

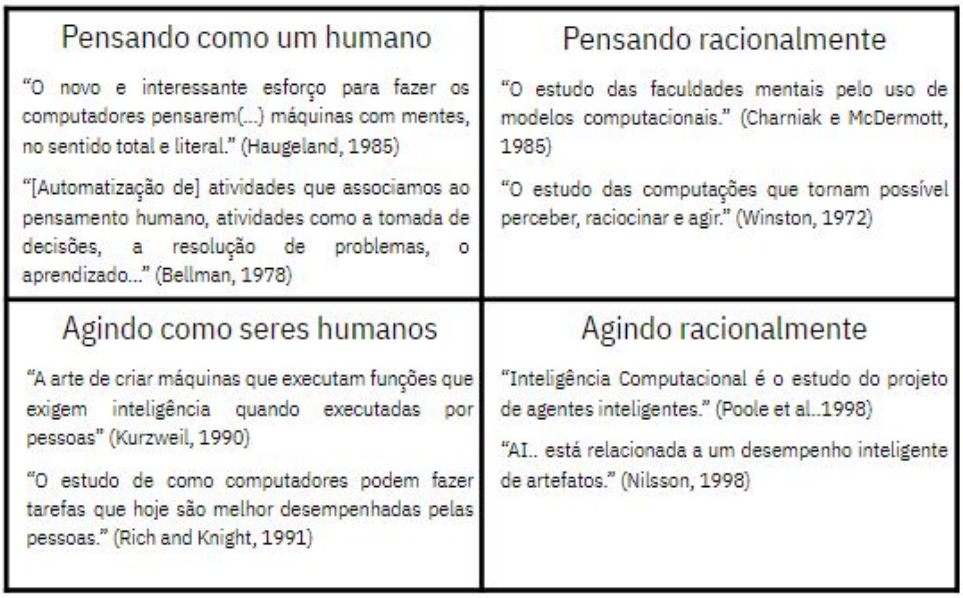
A Inteligência Artificial (IA), é um enorme campo computacional que traz grandes questionamentos, começando com sua definição. O conceito de IA é bastante discutido e não se tem uma definição fixa (HAENLEIN KAPLAN *et al.*, 2019). Já existem sistemas e robôs inteligentes. Por isso a definição de IA é bastante modificada. Os pesquisadores Haenlein Kaplan *et al.* (2019) definiram a inteligência artificial como "a capacidade de um sistema de interpretar corretamente os dados externos, aprender com esses dados e usar esse aprendizado para atingir metas e tarefas específicas por meio de adaptação flexível".

Além de Haenlein Kaplan *et al.* (2019), Russell e Norvig (2002) também buscaram propor uma definição para IA. Nisso eles propuseram uma divisão da inteligência artificial em quatro categorias. A primeira categoria se concentra em criar máquinas que pensam como humanos, logo procura-se reproduzir o processo, a manifestação e o resultado do pensamento humano na máquina. A segunda é dedicada a criar máquinas como humanos, e busca se concentrar na ação. A terceira está focada no desenvolvimento de máquinas para o pensamento racional, logo as decisões a serem feitas por essas máquinas devem ser pensadas logicamente. Finalmente, a quarta a se concentra no desenvolvimento de máquinas que se comportam de maneira lógica e são projetados para sempre fazer a coisa certa ou se comportar da maneira certa, no sentido lógico (RUSSELL; NORVIG, 2002). A Figura 1 mostra as divisões de IA segundo Russell e Norving.

Alan Turing (1950), propôs um teste para fornecer uma definição aceitável de inteligência, conhecido até hoje como teste de Turing. Esse teste, em outras palavras, é feito para saber se um computador pode se passar por um humano. O computador é interrogado por uma pessoa, com perguntas por escrito, e só se passa no teste se as pessoas não conseguirem distinguir se as respostas escritas vêm de um computador ou de um humano. As capacidades que o computador precisa ter são:

- **processamento de linguagem natural** para permitir que ele se comunique com sucesso em um idioma natural;

Figura 1 – Divisões de IA de Russell e Norvig



Fonte: (RUSSELL; NORVIG, 2002)

- **representação de conhecimento** para armazenar o que sabe ou ouve;
- **raciocínio automatizado** para usar as informações armazenadas com a finalidade de responder a pergunta e tirar novas conclusões;
- **aprendizado de máquina** para se adaptar a novas circunstâncias e para detectar e extrapolar padrões.

Turing acreditava que a a simulação física era desnecessária para inteligência, por isso não é necessário a interação física entre o interrogador e o computador, mas o teste de Turing Total que inclui o sinal de vídeo, para testar as habilidades de percepção. Para o computador passar nesse teste, além das capacidades anteriores, é necessário:

- **visão computacional** para perceber objetos e
- **robótica** para manipular objetos e movimentar-se.

O teste de Turing é altamente relevante e permanece valorizado por mais de 60 anos. Pode-se afirmar que a maior parte de IA está disposta entre essas seis capacidades.

Para nosso protótipo ser inteligente para fazer todas a funções necessárias precisamos de algumas dessa habilidades como aprendizado de máquina e visão computacional.

2.2 Visão computacional

A Visão Computacional (CV), é um campo de estudo que trabalha no desenvolvimento de técnicas para ajudar o computador a "ver" e compreender o conteúdo das imagens

digitais, onde o principal objetivo trazer as habilidades da visão humana para a computação, ou seja, desenvolver sistemas que podem realizar tarefas que o sistema visual humano pode realizar, isso tudo de forma autônoma. Podemos citar também a forma de pensar na visão computacional que é a que procura criar um modelo computacional do sistema visual humano. A visão computacional é, portanto, uma extração automatizada de informações de imagens que podem incluir desde modelos 3D, posição da câmera, detecção e reconhecimento de objetos até agrupamento e busca de conteúdo de imagem. Nossa visão é capaz de captar informações como cor, formas e muitos outros dados que fazem com que a nosso cérebro raciocine e apresente uma conclusão sobre essas informações recebidas. Da mesma forma a visão computacional vem para trabalhar esses aspectos da visão. Estudar a visão biológica requer a compreensão da percepção realizada pelos os olhos e a interpretação da percepção realizada pelo cérebro. Além disso, o mundo visual tem uma enorme complexidade inerente, o que torna o problema mais desafiador. Muito progresso foi feito neste campo, mas ainda há um longo caminho a percorrer. Os pesquisadores da área procuram se inspirar nessa ideias para projetar seus sistemas e como resposta podemos estudar ainda mais como funciona o sistema visual humano (SZELISKI, 2010).

2.2.1 Biblioteca OpenCV

Para que possamos desenvolver a nossa ferramenta vamos utilizar uma biblioteca de visão computacional e aprendizado de máquina que é a *Open Source Computer Vision Library* (OpenCV). Ao todo são mais de 500 algoritmos e cerca de 10 vezes mais funções que compõem ou suportam esses algoritmos e foi construída para fornecer uma infraestrutura comum para aplicativos de visão computacional e para acelerar o uso da percepção da máquina em produtos comerciais. Ela é equipada com mais de 2.500 algoritmos otimizados, que incluem um conjunto abrangente de algoritmos de visão computacional e aprendizado de máquina clássicos e de última geração capazes de fazer o reconhecimento de objetos e cenas inteiras. As interfaces implementadas nela são em C++, Python, Java e MATLAB [1] e suporta os sistemas operacionais Windows, Linus, Android e Mac OS (KAEHLER; BRADSKI, 2016).

A biblioteca OpenCV tem uma ampla gama de aplicações, como unir imagens de rua a vídeo de vigilância em Israel para detecção de intrusos, equipamentos de vigilância de minas na China, suporte à navegação de robôs e coleta de objetos em Willow Garage, detecção de acidentes de afogamento em piscinas na Europa, exibindo arte interativa na Espanha e em

[1] MATLAB (Matrix Laboratory) é um *software* interativo que se destina a cálculos numéricos e gráficos científicos.

Nova York, verificação de buracos nas estradas da Turquia, inspecione rótulos de produtos em fábricas ao redor do mundo e detecção rápida da face humana no Japão.[2]

A biblioteca é amplamente utilizada em empresas, grupos de pesquisa e por órgãos governamentais. Logo podemos usar essa biblioteca para o desenvolvimento da nossa ferramenta utilizando as funções e algoritmos já disponibilizados pela biblioteca.

2.3 Aprendizado de Máquina

O aprendizado de máquina é o ramo da inteligência artificial que vem trazer o estudo de como criar algoritmos para melhorar desempenho do computador, por meio do qual usam-se dados com base nos *feedbacks* (MITCHELL *et al.*, 1997). Fazendo uma analogia com o ser humano, para que o individuo aprenda algo ele passa por um processo de treinamento da mente, e quando pequenos, somos apresentados a letras, depois sílabas, depois palavras, e depois as frases com a junção das anteriores. Depois desse processo aprendemos a formular frases sem precisar aprender novamente as letras do começo. Esse é um método que é escolhido para aprender começando primeiramente com as letras e seguir gradativamente até a formação de frases. Da mesma forma é o aprendizado de máquina, para que o sistema tenha sucesso é preciso criar um processo onde se tem que escolher o tipo de treinamento, definir o que o sistema vai aprender e como será representado e o algoritmo que será usado no treinamento.

A aprendizagem envolve trabalhar com hipóteses e combinações de resultados anteriores. Dessa forma existindo diversas hipóteses é preciso procurar a que mais se adéqua para a resposta da função, considerando os *feedbacks* do treinamento. Logo pode-se usar a matemática e suas funções, árvores de decisão, regras simbólicas, redes neurais e entre outros (MITCHELL *et al.*, 1997).

Dentro do aprendizado de máquina existem algumas subdivisões, podemos citar as duas mais conhecidas que é o aprendizado supervisionado e não supervisionado. A aprendizagem supervisionada é uma dos métodos utilizados nesse ramo. O nome bem sugestivo nos traz a ideia de um "supervisor", logo temos uma espécie de manual que ajuda o sistema a designar os rótulos para os dados de treinamento, isso para ajudar na classificação. Os algoritmos são responsáveis pela criação de base de dados de treinamento, que posteriormente podem servir para classificar dados ainda não rotulados. Importante ressaltar que a principal diferencia de outros métodos é que utiliza dos dados de treinamento para definir e rotular novos dados(RUSSELL; NORVIG,

[2] Informações obtidas no *site*: http://opencv.org.

2002). Um exemplo análogo que podemos citar é a montagem de um carrinho de bebê que vem com um manual onde diz qual peça se encaixa na outra peça e é consultado para saber quais as próximas montagens a serem feitas.

Fazendo uma relação, se o aprendizado supervisionado depende de dados de treinamento, o aprendizado não supervisionado não depende desses dados. Ele está focado no aprendizado na entrada de dados e sem que nenhum dado seja usado para treinamento. Busca também compreender os padrões e relacionamentos de dados com métodos de análise e junção de dados principais. Estes são frequentemente usados como pioneiros exploratórios para métodos de aprendizagem guiada (BARLOW, 1989).

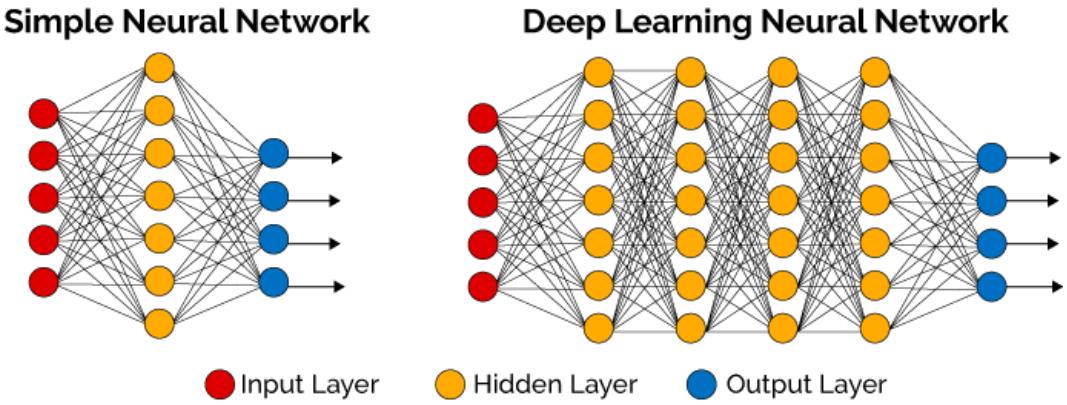
O aprendizado de máquina será de extrema importância para nosso projeto, pois ele fará a parte principal, para que nosso método faça a detecção de pessoas. Exitem várias técnicas e algoritmos utilizados na aprendizagem e vamos apresentar alguns mais adiante.

2.3.1 Deep Learning

Deep Learning, é um subcampo do aprendizado de máquina que se preocupa com algoritmos inspirados na estrutura e função do cérebro chamadas redes neurais artificiais. A rede neural precisa aprender a todo o tempo a resolver tarefas de forma mais qualificada ou mesmo a utilizar vários métodos para proporcionar um melhor resultado. Ao obter novas informações no sistema, aprende como agir de acordo com uma nova situação. O aprendizado se torna mais profundo quando as tarefas que você resolve se tornam mais difíceis. A rede neural profunda representa o tipo de aprendizado de máquina quando o sistema usa muitas camadas de nós para derivar funções de alto nível das informações de entrada. Uma Rede Neural Profunda é benéfica quando você precisa substituir o trabalho humano por um trabalho autônomo, sem comprometer sua eficiência. Na Figura 2 temos uma representação de como é a entrada, processamento e saída de uma rede neural simples e de uma rede neural profunda (GOODFELLOW *et al.*, 2016).

Algumas empresas já utilizam essas técnicas para ajudar a sociedade como a SENSTIME, que uma empresa chinesa criadora de um sistema de reconhecimento automático de rosto para identificar criminosos, que usa câmeras em tempo real para localizar um infrator na multidão. Um outro exemplo é a PONY-AI, uma empresa americana que desenvolveram um sistema para carros de IA que pode funcionar sem motorista. Ele reconhece pessoas, sinais de trânsito e outras marcações como árvores e outros objetos importantes. Todas essas empresas utilizaram algoritmos de detecção de objetos baseados em Deep Learnig como YOLO, R-CNN:

Figura 2 – Rede Neural Simples e Rede Neural Profunda (Deep Learning)



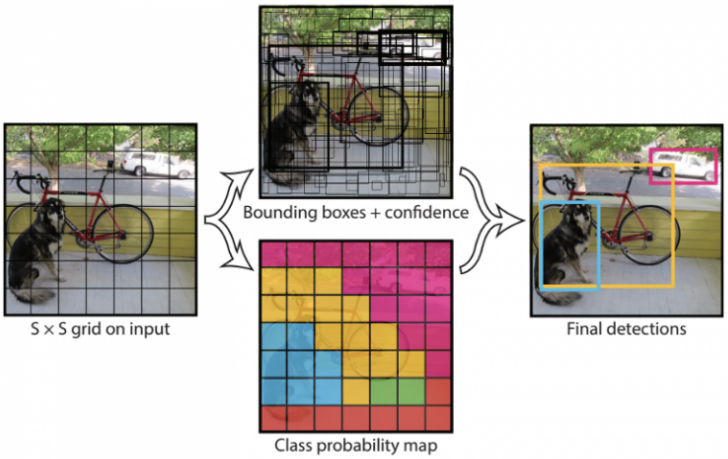
Fonte: (CASTRO; CASTRO, 2001)

Regions with CNN feature, Fast-R-CNN: *Fast Region-based Convolutional Networks for object detection*. Assim como os sistemas dessas empresas também vamos usar dessas técnicas para fazer nosso sistema. [3]

2.3.2 YOLO

YOLO (*You Only Look Once*) significa, você só olha uma vez, ele é um sistema de detecção de objetos em tempo real. Ele aplica uma única rede neural à imagem completa e divide a imagem em regiões e prevê caixas delimitadoras e probabilidades para cada região assim como está representada pela Figura 3.

Figura 3 – You Only Look Once: YOLO)

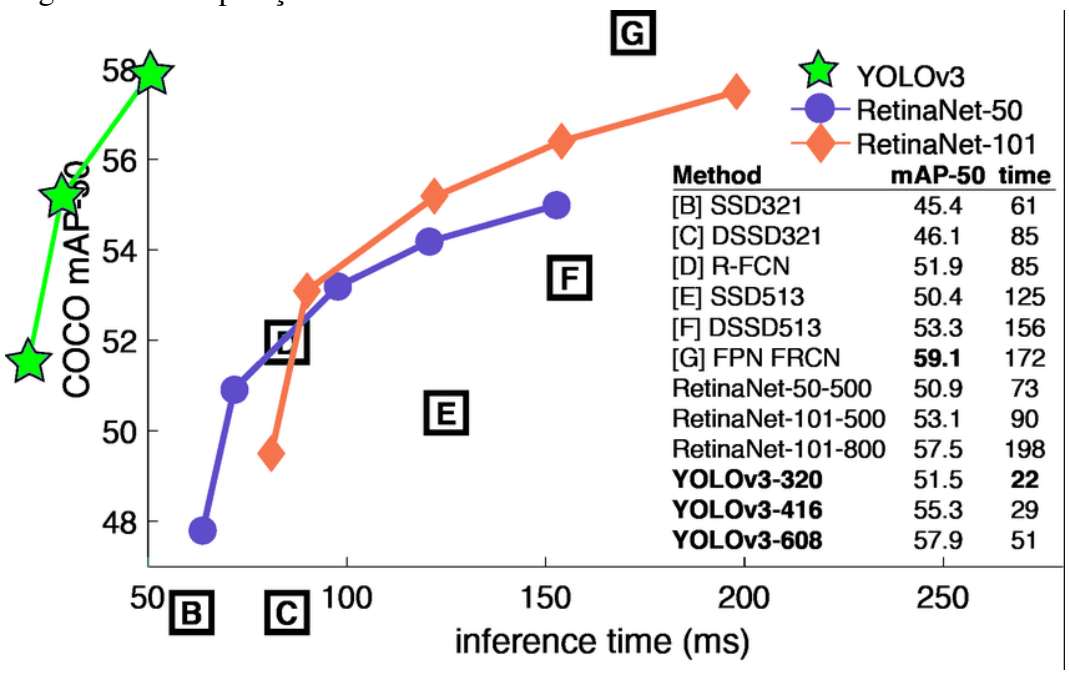


Fonte: (CASTRO; CASTRO, 2001)

[3] R-CNN é um método para detectar e localizar objetos. Ele visa "encontrar"objetos de interesse em uma imagem e desenhar caixas delimitadoras ao redor deles, enquanto classifica suas classes. O nome R-CNN significa Regiões com recursos CNN (Convolutional Neural Networks), o mesmo autor do método de algoritmo R-CNN resolveu algumas deficiências e construiu um algoritmo de detecção de objetos mais rápido, chamado Fast R-CNN.

Ele pode alcançar o dobro da precisão média de outros sistemas de tempo real. Porém a limitação do algoritmo YOLO é que ele pode ter dificuldade em detectar pequenos objetos, perdendo precisão, sendo esse problema melhorado em sua versão mais atual. Isso se deve às restrições espaciais do algoritmo. Com o YOLO, também é possível executar com a entrada de uma *webcam*. As previsões são feitas por meio de uma única avaliação da rede, diferentemente de sistemas como o R-CNN, que exigem milhares de avaliações em uma única imagem. Isso o torna muito rápido, mais de 1000x mais rápido que o Regions-CNN e 100x mais rápido que o Fast Regions-CNN. No mAP medido em 0,5 IoU, YOLOv3 está no mesmo nível que a Perda Focal, mas cerca de 4 vezes mais rápido. Além disso, você pode alternar facilmente entre velocidade e precisão simplesmente alterando o tamanho do modelo, sem treinamento adicional, assim como podemos observar no gráfico da Figura 4. É possível analisar que o time do YOLO, indicado como a estrela verde, é executado em menos tempo do que os outros, indicados como B, C, D, E, F, G, linha lilás e beje, também é possível notar pela legenda do gráfico no canto direito.

Figura 4 – Comparação YOLO com outros detectores



Fonte: (REDMON; FARHADI, 2018)

2.3.3 MS COCO

O conjunto de dados, *Microsoft Common Objects in Context* (MS COCO), é um conjunto de dados de detecção, segmentação, detecção de pontos-chave e legendas de objetos em grande escala. O conjunto de dados consiste em imagens de 328K. O COCO significa Objetos

Comuns em Contexto porque o objetivo de criar conjuntos de dados de imagem é avançar no reconhecimento de imagem. O conjunto de dados COCO contém conjuntos de dados de visão desafiadores e de alta qualidade para visão computacional, especialmente redes neurais de última geração (CHEN *et al.*, 2015).

Compreender uma cena visual é um objetivo primário da visão computacional; incluindo identificar quais objetos estão presentes, localizar objetos em 2D e 3D, determinar propriedades de objetos e caracterizar relacionamentos entre objetos. Portanto, algoritmos para detecção e classificação de objetos podem ser treinados usando o conjunto de dados. Engenheiros de aprendizado de máquina e visão computacional geralmente usam o conjunto de dados COCO para vários projetos de visão computacional (VISO.AI, 2021).

Iremos utilizar o YOLO para detecção de pessoas e ter o MS COCO como o conjundo de dados de treinamento dessa detecção.

3 TRABALHOS RELACIONADOS

Nesta seção, alguns trabalhos relacionados com o projeto proposto neste trabalho são apresentados.

3.1 DeepSOCIAL: Social Distancing Monitoring and Infection Risk Assessment in COVID-19 Pandemic

Os autores propuseram o DeepSOCIAL (REZAEI; AZARMI, 2020), um modelo de detector humano com base em uma rede neural profunda para detectar e rastrear pessoas estáticas e dinâmicas em locais públicos para monitorar as métricas de distanciamento social nesse período de pandemia.

Figura 5 – Detecção de pessoas, rastreamento e avaliação de risco no Oxford Town Center, usando uma câmera CCTV pública. (a) Monitoramento do distanciamento social; (b) Risco de infecção acumulado (zonas vermelhas) devido a violações do distanciamento social.

(a) (b)

Fonte: (REZAEI; AZARMI, 2020)

Os modelos de detecção de objetos geralmente são construídos a partir de um conjunto de modelos de redes neurais que consistem em uma espinha dorsal (backbone), pescoço (neck) e cabeça (head) que juntas executam tarefas de detecção e classificação. As principais previsões são caixas delimitadoras para os objetos identificados e suas associadas classes (BOCHKOVSKIY *et al.*, 2020).

A rede escolhida pelos autores é o YOLO, que consiste principalmente em três componentes principais.

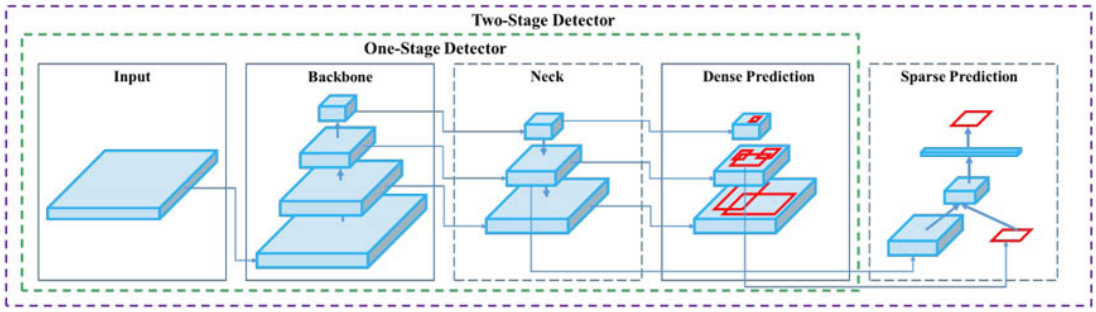
- Backbone - Rede neural convolucional que agrega e forma recursos de imagem em granularidade refinada de diferentes imagens, depois de análises foi escolhido usar o modelo CSPDarkNet53.
- Pescoço : Uma série de camadas de rede que misturam e combinam recursos de imagem e

passam os recursos de imagem para a camada de previsão, nesse caso escolhido a SPP/PAN e SAM.

– Cabeça : Preveja recursos de imagem, gere caixas delimitadoras e preveja categorias, sendo escolhido o YOLO.

A rede da espinha dorsal extrai os features da imagem usada como entrada, enquanto a rede da cabeça é treinada na tarefa supervisionada para prever as bordas da caixa delimitadora e classificar seu conteúdo. A adição da rede do pescoço permite que a rede da cabeça processe os features a partir de etapas intermediárias da rede da espinha dorsal. Todo o pipeline processa as imagens apenas uma vez, daí o nome *You Only Look Once* (YOLO). Podemos visualizar essa arquitetura na Figura 6.

Figura 6 – O diagrama da arquitetura



Fonte: (BOCHKOVSKIY *et al.*, 2020)

Foi aplicado também a função de perda completa e um aumento nos dados de mosaico nos conjuntos de dados do MS COCO e na imagem aberta do Google de vários pontos de vista para enriquecer a fase de treinamento, o que levou a um detector humano eficiente e preciso, que use qualquer tipo de Câmera de vigilância fechada (CCTV), sendo aplicável em vários tipos de ambiente. Para avaliar o modelo apresentado foi utilizado o conjunto de dados do centro da cidade de Oxford. O sistema foi capaz de lidar com vários desafios, incluindo mudanças de iluminação, sombras e visibilidade parcial e, em comparação com as três tecnologias, o sistema obtém ótimos resultados em termos de acurácia (99,8%) e velocidade (24,1 FPS). O sistema funciona em tempo real (REZAEI; AZARMI, 2020).

Foi decidido seguir adaptando o mapeamento de perspectiva inversa, que é método para visualização de imagem por uma câmera através dos ângulos, e algoritmo de rastreamento SORT para estimar a distância interpessoal e rastrear as trajetórias de movimento das pessoas, avaliação de risco e análise de risco, assim podendo colaborar com a prevenção e ajudar o pessoal da saúde e o governo. Ele pode ser integrado e aplicado a todos os tipos de câmeras de vigilância

CCTV com qualquer resolução *VGA Full HD*, e tem desempenho em tempo real. Ele fornece um algoritmo de classificação humana, onde independentemente do ângulo e da posição da câmera, os resultados do estudo, segundo os autores, podem ser aplicados diretamente a um grupo mais amplo de pesquisadores, não apenas nas áreas de computação, inteligência artificial e saúde, mas também em outras aplicações industriais. Podemos analisar a detecção de pessoas, rastreamento e avaliação de risco no Oxford Town Center, usando uma câmera CCTV pública desenvolvido pelo trabalho na Figura 5. Geralmente, qualquer aplicativo que requeira detecção humana se tornará o foco das atenções (REZAEI; AZARMI, 2020).

A partir da leitura do trabalho de Rezaei e Azarmi (2020), foi decidido utilizar também a ideia de uma rede neural profunda para fazer a classificação no nosso projeto. A ideia de utilizar o YOLO seria de grande importância para o projeto.

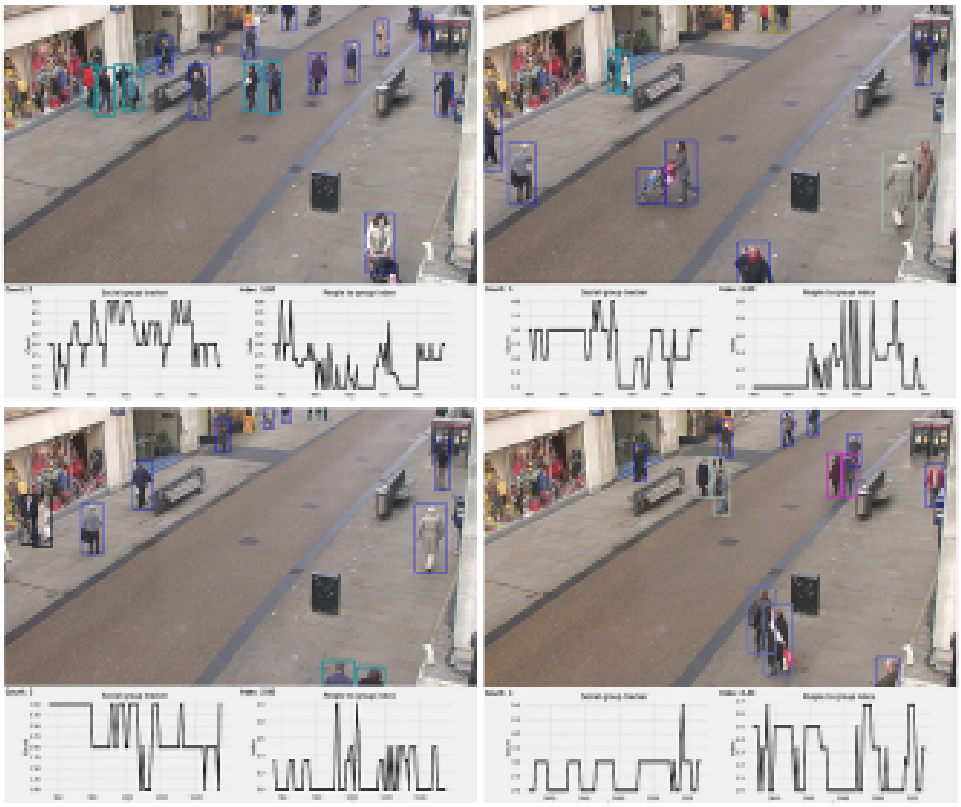
3.2 Monitoring COVID-19 social distancing with person detection and tracking via fine-tuned YOLOv3 and Deepsort techniques

Visto o impacto da atual pandemia do COVID-19, os autores desse trabalho propõem uma estrutura baseada em aprendizagem profunda que pode executar automaticamente a tarefa de usar vídeo de vigilância para monitorar distância social. A estrutura proposta usa o modelo de detecção de objeto YOLOv3 para separar as pessoas do fundo e usa o método Deepsort para rastrear as pessoas identificadas com a ajuda de caixas delimitadoras e IDs atribuídos. Os resultados do modelo YOLOv3 são ainda comparados com outros modelos populares, por exemplo, o *Fast R-CNN*, traduzindo, é uma rede neural convolucional rápida baseada em região, e detector de disparo único (SSD) em termos de acurácia média (mAP), quadros por segundo (FPS) e o valor da perda definido pela classificação e localização do objeto.

Em seguida, a norma L2 vetorizada em pares é calculada com base no espaço de recurso tridimensional obtido usando as coordenadas do centroide e as dimensões da caixa delimitadora. É recomendada pelos autores a utilização do termo índice de violação para quantificar a não utilização de protocolos de distanciamento social. Pode ser visto a partir da análise experimental que YOLOv3 com o esquema de rastreamento Deepsort mostra os melhores resultados enquanto equilibra as pontuações de mAP e FPS para monitorar a distância social em tempo real (PUNN *et al.*, 2020). Também é possível ver a saída das amostras pela ferramenta desenvolvida na Figura 7.

Chamou-nos a atenção as comparações feitas com outros algoritmos de detecção,

Figura 7 – Saída da amostra da estrutura proposta para monitorar o distanciamento social em imagens de vigilância de Oxford Town Centro.



Fonte: (PUNN *et al.*, 2020)

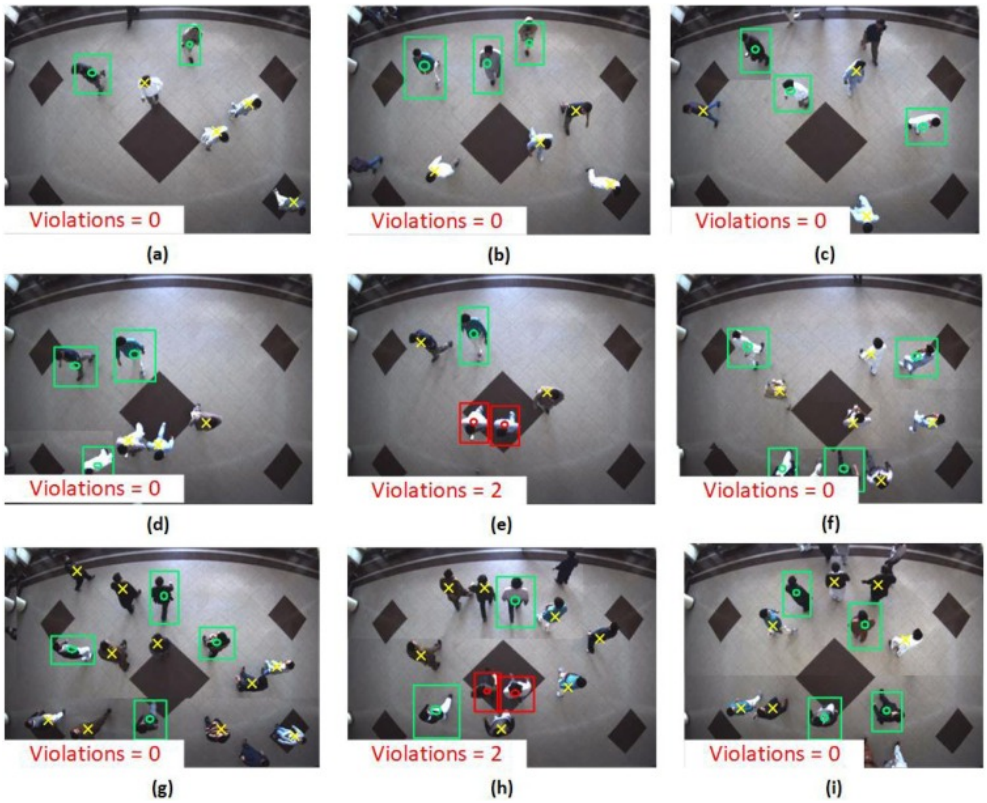
confirmando ainda mais a ideia de usar o YOLO para nossa detecção.

3.3 A deep learning-based social distance monitoring framework for COVID-19

Esse trabalho fornece uma plataforma de aprendizado profundo para rastreamento social remoto usando uma perspectiva de sobrecarga (AHMED *et al.*, 2021). A estrutura usa o paradigma de reconhecimento de objeto YOLOv3 para reconhecer humanos em sequências de vídeo. Um método de *Transfer Learning* (TL) também foi implementado para melhorar a acurácia do modelo. Desta forma, o algoritmo de detecção usa um algoritmo pré-treinado que é conectado a uma camada adicional treinada com um conjunto de dados humanos sobrecarregado (AHMED *et al.*, 2021).

O modelo de detecção usa as informações da caixa delimitadora detectada para identificar a pessoa. Usando a distância euclidiana, determina a distância entre o par de pessoas detectadas e o centro de massa da caixa delimitadora. Para estimar a violação da distância social entre as pessoas, é usado uma aproximação da distância física ao pixel e defini-se um limite. Estabeleça um limite de violação para avaliar se o valor da distância viola o limite inferior de

Figura 8 – Monitoramento de distância social de uma visão aérea usando um modelo de detecção pré-treinado. Nos quadros de amostra, as pessoas em retângulos verdes são aquelas que mantêm o distanciamento social. As pessoas que violam o limite da distância social são mostradas em retângulos em vermelho. A cruz amarela positiva rotulada manualmente mostra detecções de falha.



Fonte: (AHMED *et al.*, 2021)

distância social. Além disso, algoritmos de rastreamento são usados para detectar indivíduos na sequência de vídeo, a fim de também rastrear pessoas que violam o limite de distância social. Além disso é usado a aprendizagem por transferência, que é uma variante da aprendizagem supervisionada, podemos usá-la quando nos deparamos com tarefas com um número limitado de exemplos rotulados. Outro motivo para usá-lo é que, se a escassez de dados não for um problema, você pode aproveitar a aprendizagem por transferência quando quiser evitar a grande quantidade de recursos necessários para treinar um modelo que consome muitos dados (AHMED *et al.*, 2021).

Os resultados mostram que a estrutura desenvolvida distingue com sucesso indivíduos que caminham muito perto e que violam a distância social, além disso, o método de aprendizagem por transferência melhora a eficiência geral do modelo. Os modelos de detecção sem e com aprendizagem de transferência alcançam taxas de acurácia de 92% e 98%, respectivamente. A acurácia de rastreamento do modelo é de 95% (AHMED *et al.*, 2021).

Assim como os autores, devido aos trabalhos utilizarem mais o YOLOv3 foi decidido começar o projeto utilizando o ele. Com os resultados de acurácia foi decidido também utilizar o cálculo de distância euclidiana.

3.4 Análise Comparativa

O autor do DeepSocial (REZAEI; AZARMI, 2020), propôs um modelo de detector humano baseado em rede neural profunda, denominado DeepSOCIAL, que é usado para detectar e rastrear pessoas estáticas e dinâmicas em locais públicos para monitorar a distância social na era COVID-19 e além. Após análise e pesquisa, foi escolhido o CSPDarkNet53 como backbone, SPP / PAN e SAM como pescoço e YOLOv4 como pescoço. Além disso, a função de perda IoU completa e o aprimoramento de dados do Mosaic são aplicados aos conjuntos de dados MS COCO e Google Open Image de vários ângulos, enriquecendo a fase de treinamento. Para entrada de dados é utilizado qualquer tipo de câmera de vigilância.

No trabalho de Punn *et al.* (2020), um sistema baseado em aprendizado profundo em tempo real foi desenvolvido para automatizar o processo de monitoramento da distância social por meio de métodos de detecção e rastreamento de objetos, onde todos são identificados em tempo real com a ajuda de caixas delimitadoras. Essas caixas geradas ajudam a identificar as pessoas que atendem aos atributos de proximidade calculados com a ajuda do método de pares de vetores. Confirme o número de violações calculando o número de grupos formados e essas violações são calculadas como a proporção entre o número de pessoas e o número de grupos. Extensos experimentos foram conduzidos usando modelos de detecção de objetos populares e mais avançados, como: RCNN, SSD e YOLOv3, onde YOLOv3 demonstra desempenho eficiente com pontuações FPS e mAP balanceadas. Uma vez que esta abordagem é altamente sensível à localização espacial da câmera, a mesma abordagem pode ser ajustada para melhor ajuste com o campo de visão correspondente.

Já no trabalho apresentado por Ahmed *et al.* (2021), desenvolveu-se uma estrutura de monitoramento de distância social baseada em aprendizagem profunda usando o paradigma do YOLOv3 um modelo pré-treinado para detecção humana, juntamente com a aprendizagem por transferência. O TL é adotado para melhorar o desempenho do modelo pré-treinado, já que a aparência, visibilidade, escala, tamanho, forma e pose de uma pessoa variam significativamente de uma visão aérea. O modelo é treinado em um conjunto de dados de sobrecarga e a camada recém-treinada é anexada ao modelo existente. O modelo de detecção fornece informações de

caixa delimitadora, contendo informações de coordenadas do centroide. Usando a distância euclidiana, as distâncias centróides de pares entre as caixas delimitadoras detectadas são medidas. Para verificar violações de distância social entre pessoas, uma aproximação da distância física ao pixel é usada e um limite é definido. Um limite de violação é usado para verificar se o valor da distância viola a distância social mínima definida ou não. Além disso, um algoritmo de rastreamento de centróide é usado para rastrear pessoas na cena.

Os resultados do trabalho de Rezaei e Azarmi (2020) foi avaliado para o conjunto de dados Oxford Town Center, incluindo 7530 frames, e aproximadamente 150.000 detecção de pessoas e estimativa de distância. O sistema foi capaz de realizar em uma variedade de desafios, incluindo oclusão, variações de iluminação, sombras e visibilidade parcial, ou seja, ofereceu um algoritmo de classificação humana independente do ângulo e da posição da câmera, e provou ser um grande desenvolvimento em termos de acurácia (99,8%) e velocidade (24,1 fps) em comparação com três técnicas de última geração. O sistema funcionava em tempo real usando uma plataforma GPU básica ou um Plataforma de CPU multi-core/multi-thread de 10ª geração ou superior. Adaptamos um *Inverse Perspective Mapping* (IPM) geométrico e algoritmo de rastreamento SORT para nosso aplicativo para estimar as distâncias entre pessoas, e para rastrear as trajetórias móveis das pessoas, avaliação de risco de infecção e análise para o benefício de autoridades de saúde e governos. O resultado desta pesquisa é diretamente aplicável para uma comunidade mais ampla de pesquisadores, não apenas nos setores de visão computacional, IA e saúde, mas também em outras aplicações industriais incluindo detecção de pedestres para sistemas de assistência ao motorista, veículos autônomos, detecções de comportamento e anomalia em público e multidão, sistemas de vigilância de segurança, reconhecimento de ação em esportes, e principalmente em lugares públicos; e, geralmente, quaisquer aplicativos em que a detecção humana caia no centro das intenções.

Já os resultados de cada modelo obtido ao final da fase de treinamento, do trabalho de Punn *et al.* (2020), com o valor do tempo de treinamento, número de iterações, mAP e perda total, é mostrado nos resultados onde observou que o modelo RCNN mais rápido obteve perda mínima com o mAP máximo, porém, possui o FPS mais baixo, o que o torna não adequado para aplicações em tempo real. Além disso, em comparação com SSD, YOLOv3 obteve melhores resultados com mAP, tempo de treinamento e pontuação de FPS balanceados. O modelo YOLOv3 treinado é então utilizado para monitorar o distanciamento social no vídeo de vigilância.

Por fim, os resultados experimentais do trabalho de Ahmed *et al.* (2021), indicaram

que a estrutura identifica com eficiência as pessoas que andam muito perto e violam o distanciamento social; além disso, a metodologia de TL aumenta a eficiência e a acurácia geral do modelo de detecção. Para um modelo pré-treinado sem aprendizagem por transferência, o modelo atinge uma acurácia de detecção de 92% e 95% com aprendizagem por transferência. A acurácia de rastreamento do modelo é de 95%. O trabalho pode ser aprimorado no futuro para diferentes ambientes internos e externos. Diferentes algoritmos de detecção e rastreamento podem ser usados para ajudar a rastrear a pessoa ou pessoas que estão violando ou violando o limite de distanciamento social.

Foi proposto uma junção dos métodos para classificação e detecção de pessoas e monitoramento de distanciamento social. Sendo assim utilizar o algoritmo YOLOv3 para detecção de pessoas, que por sua vez pré-treinado com o conjunto de dados COCO. Para monitoramento vamos utilizar de uma função que usa a detecção de pessoas para verificar as distâncias, e como principal conceito vamos utilizar o cálculo de distância euclidiana. A ideia é que seja possível usar com qualquer tipo de câmera e em qualquer local. Neste trabalho, considerando trabalhos que utilizam GPU e resultam em FPS alto, alcançamos uma precisão média superior a 90% sem uso de GPU e FPS baixo. Assim, podemos ver que este método fornece resultados eficientes mesmo que você não tenha uma GPU para fornecer o processamento mais rápido, mesmo dentro de seus limites.

Quadro 1 – Quadro de comparação dos trabalhos

	Rezaei e Azarmi (2020)	Punn *et al.* (2020)	Ahmed *et al.* (2021)	*Este Trabalho*
Versão YOLO	YOLOv4	YOLOv3	YOLOv3	YOLOv3
Método	IPM	Deepsort	CDE e TL	CDE
Conjunto de dados	MS COCO e OPI	PASCAL-VOC and MS-COCO	IMSciences	MS-COCO
GPU	Nvidia RTX 2080 GPU	Nvidia GTX 1060 GPU	-	Sem GPU
FPS	24,1 FPS	23 FPS	-	2.0 FPS
Resultados numéricos	99,8% de acurácia	-	95% de acurácia	92,47%

Fonte: Autora.

4 PROCEDIMENTOS METODOLÓGICOS

Neste capítulo são apresentados os passos necessários para a elaboração deste trabalho. A seguir, na Figura 9, é mostrada uma visão geral da metodologia adotada nesta obra.

Figura 9 – Procedimentos Metodológicos

Fonte: Autora

4.1 Detecção de pessoas usando YOLO

A estrutura YOLO (*You Only Look Once*) lida com a detecção de objetos de uma maneira diferente. Ela obtém a imagem inteira em uma única instância e prevê as coordenadas da caixa delimitadora e as probabilidades de classe para essas caixas. A maior vantagem de usar YOLO é sua velocidade considerada rápida e pode processar 45 quadros por segundo. YOLO também entende representação generalizada de objetos. Ele é um dos melhores algoritmos para detecção de objetos e mostrou um desempenho comparativamente semelhante aos algoritmos R-CNN para a detecção de objetos(REDMON; FARHADI, 2018). Neste trabalho, utilizamos o YOLOv3 (FARHADI; REDMON, 2018) para a detecção de pessoas no quadro geral, criando uma função apenas para detecção de pessoas. A partir do resultado da detecção, apenas a classe pessoas foi usada e outras classes de objeto que podem ser detectadas com YOLO são ignoradas nesta ferramenta. Ao detectar a pessoa é posta uma caixa delimitadora que melhor se ajusta a cada pessoa detectada e é desenhada na imagem, e esses dados de pessoas detectados serão usados para a medição da distância euclidiana.

A função de detecção realiza um pré-processamento do quadro que por sua vez precisou da construção de um BLOB, que significa *Binary Large Object* e é um tipo de dado que pode armazenar dados binários e usado para armazenar arquivos de mídia como imagens, vídeo e arquivos de áudio, e assim poderíamos realizar a detecção de objetos com YOLO e OpenCV. A cada detecção é retornada a identificação e a confiança (probabilidade) da pessoa detectada,

assim é verificada se a detecção atual é uma pessoa e a confiança mínima é atendida ou excedida. Tendo esses resultados, é calculado as coordenadas da caixa delimitadora e, em seguida, é derivado o centro (isto é, centróide) da caixa delimitadora. Em seguida, é atualizado cada uma de nossas listas e aplicamos uma *Non-maximum Suppression* (NMS). A biblioteca OpenCV dispõe desse método que tem por objetivo, suprimir caixas delimitadoras fracas e sobrepostas das detecções. Assumindo que o resultado do NMS produz pelo menos uma detecção, fazemos um *loop* sobre eles, extraímos as coordenadas da caixa delimitadora e atualizamos nossa lista de resultados que consiste em: (1) Confiança na detecção de cada pessoa (2) Caixa delimitadora de cada pessoa (3) Centróide de cada pessoa. Por fim, é retornado os resultados para a função de chamada.

4.2 Cálculo de distância em pares

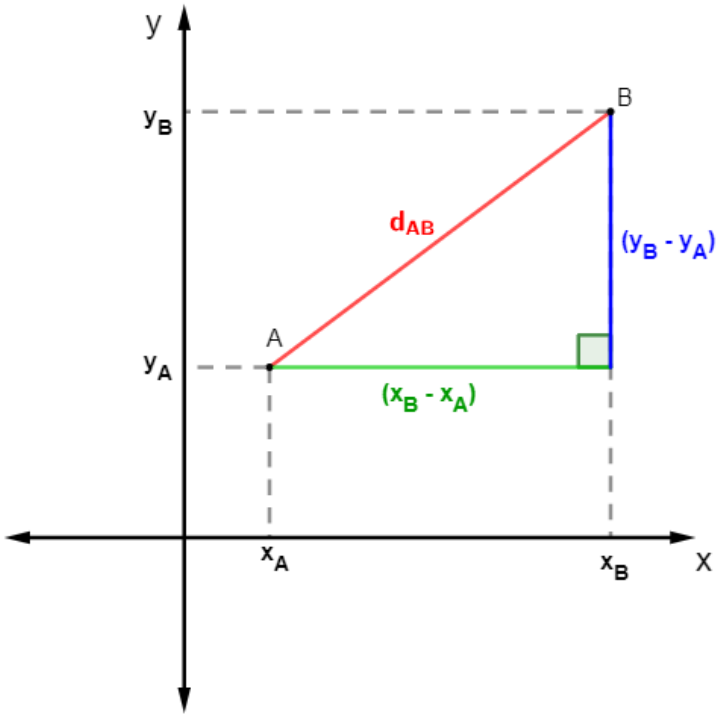
Quando todas as pessoas são detectadas, o próximo passo foi medir as distâncias entre cada indivíduo. É definido primeiramente a Distância Mínima (DM) entre pares de pessoas, este trabalho utiliza a medida de 1,5m conforme as medidas de segurança estabelecidas pela OMS. Importante ressaltar que essa distancia em metros é convertida em pixels e por isso a distância não é exatamente a distância real, pois depende também do ângulo da câmera mas chega a ser o mais próximo possível da real. As dimensões de um vídeo de entrada para teste podem ser muito grandes, então é redimensionado cada quadro enquanto é mantido a proporção de 700px, ou seja, é redimensionado o vídeo de entrada em 700px. Com os resultados da função de detecção de pessoas que é feita com o YOLO, sendo configurada com a cor verde em todas as caixas delimitadoras de detecção de pessoas e se inicia as verificações de distância entre elas. Para medir todas as distâncias possíveis entre duas pessoas, é usado as listas de todas as pessoas detectadas com YOLOv3 e é medido em pares. Supondo que pelo menos duas pessoas foram detectadas no quadro, é encaminhado para o Cálculo de Distância Euclidiana (CDE) (DANIELSSON, 1980) entre todos os pares de centróides.

É possível calcular a distância entre dois pontos e, representados em um espaço euclidiano n-dimensional, ou seja, a distância pode ser calculada para n pontos. Para um vídeo com duas dimensões, no plano euclidiano, deixe ponto p tem coordenadas cartesianas (x0,x1) e deixe ponto q tem coordenadas (y0,y1). Então a distância euclidiana entre x e y é dado por:

$$d = \sqrt{(y_1 - x_1)^2 + (y_2 - x_2)^2} \tag{4.1}$$

Seu cálculo é baseado no teorema de pitágoras, por pode ser chamado de distância de Pitágoras, tendo o segmento de reta de x para y como sua hipotenusa. As duas fórmulas quadradas dentro da raiz quadrada fornecem as áreas dos quadrados nos lados horizontal e vertical, e a raiz quadrada externa converte a área do quadrado na hipotenusa no comprimento da hipotenusa. A distância entre dois objetos que não são pontos, geralmente é definida como a menor distância entre pares de pontos dos dois objetos. As fórmulas são conhecidas por calcular distâncias entre diferentes tipos de objetos, como a distância de um ponto a uma linha. Temos uma representação visual desse cálculo na Figura 10.

Figura 10 – Distância entre dois pontos



Fonte: (EDUCAÇÃO, 2017)

4.3 Verificação de distância

Depois de medir as distâncias de pessoas entre todas as outras pessoas presentes no quadro de câmera fornecido, o dispositivo verifica se a distância é de pelo menos 1,5 metro para classificá-los como não violadores. Se houver violação da distância, a caixa delimitadora que é desenhada na detecção de pessoas passa a ser vermelha e o número de violações gerais aumentam. Esse número de violações gerais é mostrado em tempo real. O trabalho foi executado utilizando a linguagem de programação *Python*.

4.4 Testes

Para testar a ferramenta é feito testes com entrada de vídeo pré-gravado, de um conjunto de dados do centro da cidade de Oxford disponibilizada pelo Google Imagem, e vídeo gravado em tempo real no loteamento Santa Clotílde na cidade de Quixadá, local onde as pessoas costumam caminhar. Tanto no pré-gravado quanto no vídeo gravado em tempo real a câmera que captura as imagens está localizada em um local alto onde há mais pessoas caminhando. Importante ressaltar que a câmera utilizada no vídeo gravado é uma CCTV e a utilizada no vídeo em tempo real é uma *webcam* 8000k Microfone Embutido 6 Ledcom, com 4.0 megapixels para vídeos (2301x1726), vídeo de 176x144 a 640x480 (30FPS) e vídeo de 800x600 a 3200x2400 (15FPS).

5 RESULTADOS

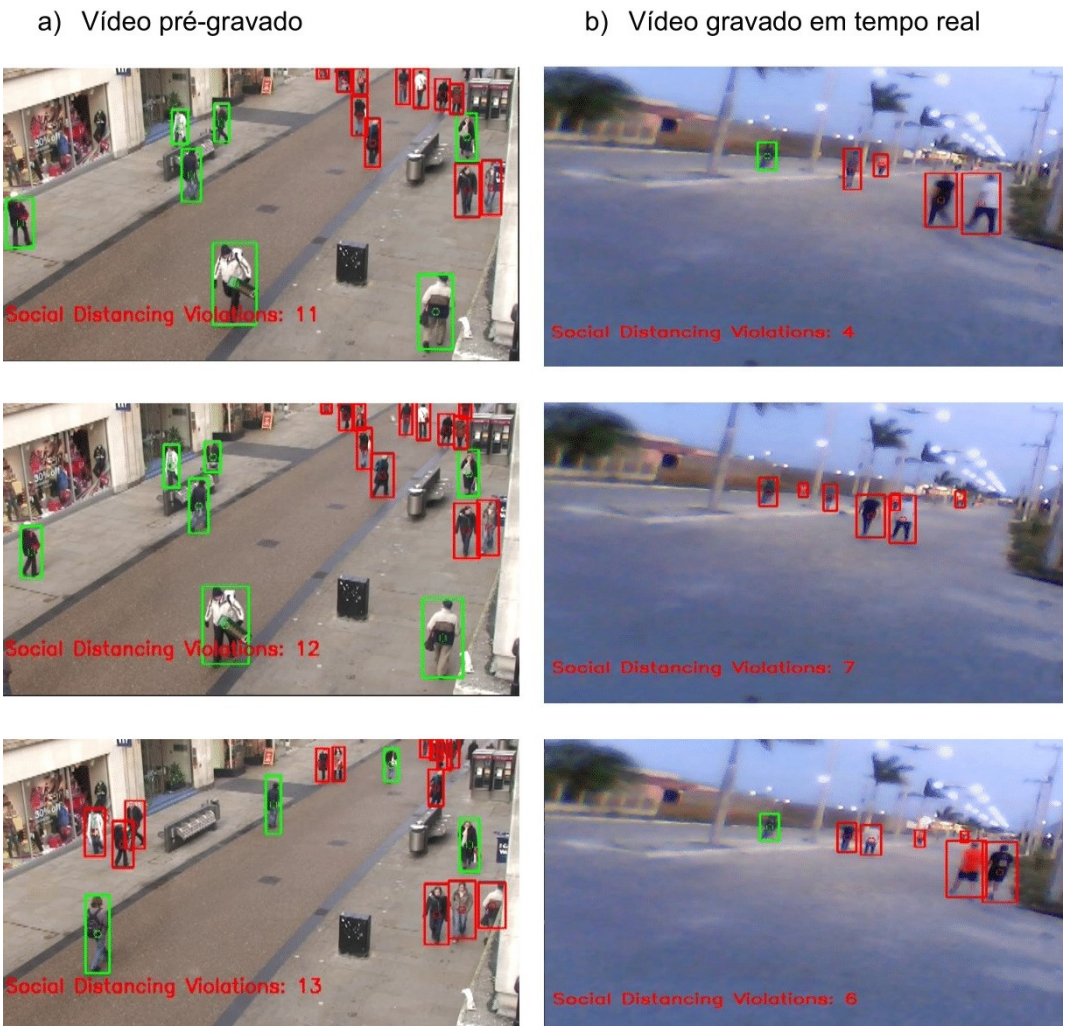
Este capítulo apresenta os resultados obtidos na execução procedimentos metodológicos.

A Figura 11 a) mostra pedestres caminhando em vias públicas. Neste trabalho, o quadro do vídeo é fixado em um ângulo específico em relação à rua assim a visão em perspectiva do quadro de vídeo é de cima para baixo para estimar as medições de distância com mais precisão. Conseguimos notar que é possível detectar corretamente todas as pessoas no vídeo e ignorar os objetos que estão no quadro. Mostra as violações encontradas que variaram entre 11 e 13 violações durante o vídeo de 52 segundos de acordo com a caminhada das pessoas. Embora os números parecem corretos, é possível notar que alguma pessoas já estão andando em pares então não seria uma violação. Importante ressaltar que não foi utilizada GPU, rodou com 2.0 FPS, onde para um vídeo gravado é um razoável FPS, mas para tempo real isso é muito baixo. Com uma GPU o FPS aumenta e quanto maior FPS mais fluído e velocidade real é o vídeo e quanto menor menos fluído e velocidade real do vídeo é diminuída. Então com um FPS alto o vídeo em tempo real sairia no tempo mais real possível, no caso, da nossa ferramenta não afetou a detecção.

A Figura 11 b) mostra pessoas caminhando em uma via pública da cidade de Quixadá, a câmera também foi colocada em uma ângulo específico onde se consegue uma visão de cima para baixo pegando em perspectiva maior número de pessoas caminhando. Foi possível notar o número total de violações que variaram entre 4 e 8 violações durante o vídeo de 30 segundos de acordo com a caminhada das pessoas. Importante ressaltar que para fazer o teste em tempo real foi preciso calibrar a distância mínima manualmente por contra da quantidade de pixel que é necessária para 1,5m em uma *webcam*. Também não foi utilizado GPU, rodou com a mesma quantidade de FPS, sendo 2.0. A Tabela 1 apresenta os dados de forma mais clara em comparação dos resultados dos dois vídeos.

É feita uma análise quantitativa e é calculada a acurácia das detecções pela ferramenta desenvolvida. O cálculo da acurácia se deu pela divisão de violações detectadas pela ferramenta e as violações reais observadas nos vídeos multiplicadas por 100. Na Tabela 2, é possível analisar que a acurácia do vídeo pré-gravado ficou acima de 90% e chegou a 100% em parte do vídeo. Já no vídeo gravado em tempo real, embora com uma baixa qualidade, ficou acima de 88% e chegou também a 100% em parte do vídeo. Importante ressaltar que apesar da qualidade da câmera ser baixa, a YOLO continua tendo uma alta capacidade de detecção de pessoas. Na

Figura 11 – Detecção de pessoas pela ferramenta desenvolvida



Fonte: Autora

Tabela 1 – Tabela de resultado dos testes

Vídeo pré-gravado	Vídeo gravado em tempo real
Sem GPU	Sem GPU
2.0 FPS	2.0 FPS
11 e 13 violações	4 e 8 violações
4 segundos em tempo real	2 minutos em tempo real
52 segundos depois da detecção	30 segundos depois da detecção
sem calibração de DM	com calibração de DM
câmera de segurança pública	Webcam USB

Fonte: Autora

Figura 12, temos o gráfico das médias das acurácias em cada vídeo e uma média geral chegando a mais de 92%.

Neste trabalho, conseguimos uma ótima média de acurácia com o resultado acima de 90%, sem GPU e com o FPS baixo, levando em consideração os trabalhos que utilizam GPU e resultaram em um FPS alto. Logo, conseguimos ver que o método proporcionou um resultado

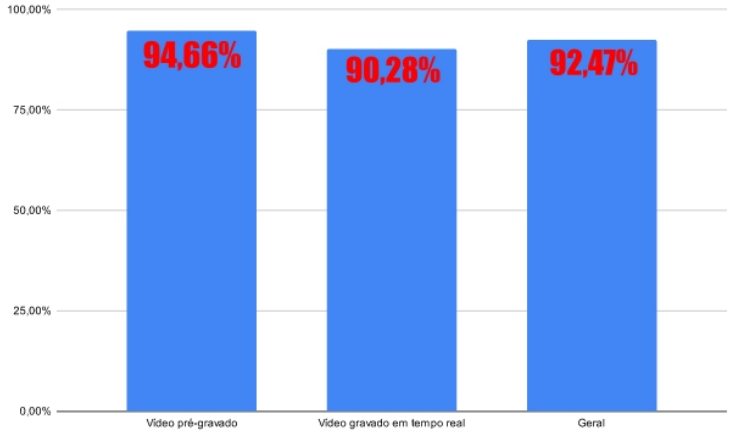
Tabela 2 – Acurácia

Vídeo de Entrada	Violações Detectadas	Violações Reais	Acurácia
Vídeo pré-gravado	11	12	92%
Vídeo pré-gravado	12	13	92%
Vídeo pré-gravado	13	13	100%
Vídeo gravado em tempo real	4	5	80%
Vídeo gravado em tempo real	7	8	88%
Vídeo gravado em tempo real	5	5	100%

Fonte: Autora

eficiente, mesmo quando não se tem uma GPU para proporcionar o mais rápido processamento, mesmo dentro de suas limitações.

Figura 12 – Média das acurácias



Fonte: Autora

Finalizando, conseguimos uma ferramenta capaz de fazer um monitoramento adequado de distanciamento social e identificar as violações da distancia mínima com a média de acurácia acima de 90%. Os pontos de melhoria seriam a calibração de câmera ser automática e usar GPU para melhorar o processamento.

Pensando também a longo prazo, essa ferramenta teria diversos casos de uso para as mais diversas pessoas. Na utilização em ruas, seria possível o gestor da cidade obter dados das ruas mais movimentadas. Ao utilizar no *shopping*, pode-se analisar quais vitrines e lojas as pessoas mais gostam de olhar/entrar e saber o momento certo de mudar algo para chamar mais atenção dos clientes. Na utilização do termostato inteligente IoT, dispositivo destinado a manter constante a temperatura de um determinado sistema, através de regulação automática, é possível analisar a quantidade de pessoas de um ambiente, e assim calcular uma temperatura ideal, trazendo benefícios para diversos tipos de ambientes.

6 CONCLUSÃO

Foi desenvolvida uma ferramenta de detecção de distanciamento social, onde foi possível medir uma distância segura entre pessoas em espaços público. Os métodos de aprendizado de máquina e técnicas de visão computacional foram utilizados nesse trabalho. Inicialmente, uma rede de detecção de objetos de código aberto baseada no algoritmo YOLOv3 foi usada para detectar o pessoas no quadro do vídeo, onde é desenhada uma caixa delimitadora com uma cor verde. Tendo isso, através do cálculo de distância euclidiana CDE tem-se a distância entre as pessoas e logo após, é feita uma verificação para as distância maiores que a distância mínima pré-estabelecida. Essa a distância mínima não chega a ser a exata em relação a distância real mas chega próximo. Ao violar essa distância mínima é acrescentado ao número total de violações no quadro e as pessoas que estão violando tem sua caixa delimitadora em vermelho. Após testes, podemos analisar a ferramenta em ação onde conseguimos obter resultados esperados e tanto em um vídeo pré-gravado quando em um vídeo em tempo real, onde ela foi capaz de identificar as violações. Para comprovar sua eficiência, realizamos uma análise qualitativa e calculamos a acurácia de cada vídeo em determinados momentos. A acurácia média foi de 94,66% para vídeos pré-gravados, 90,28% para vídeos gravados em tempo real e 92,47% é a média geral, sem uso de GPU e alto FPS. Isso é considerado muito eficiente, em questão de velocidade do vídeo e da detecção, principalmente em comparação com outros trabalhos em que as GPUs ajudam.

A análise deste trabalho mostra que o combate contra as pandemias teria maior efetividade quando apoiada por tecnologias. Buscando a garantia do princípio da integralidade nas ações de prevenção como, quarentena e o isolamento social, impostas pelas autoridades internacionais, nacionais e locais, e considerando o benefício que este sistema traria para nossa população, impactando principalmente no combate da atual pandemia mundial, torna-se fundamental o uso das técnicas aqui apresentadas, proporcionando meios de combate a outros possíveis problemas na saúde pública.

6.1 Trabalhos futuros

Para trabalhos futuros pretende-se adicionar a calibragem de câmera automaticamente, aumentar o nível de precisão de distância e usar GPU para melhorar o processamento.

O código fonte deste trabalho pode ser reutilizado em futuros trabalhos e se encontra no repositório https://github.com/kassianefacanha/distanciamento_social disponível no GitHub.

REFERÊNCIAS

AHMED, I.; AHMAD, M.; RODRIGUES, J. J.; JEON, G.; DIN, S. : A deep learning-based social distance monitoring framework for covid-19. **Elsevier**, [*S. l.*], v. 65, p. 102571, 2021.

BARLOW, H. B. Unsupervised learning. **Neural computation**, MIT Press, [*S. l.*], v. 1, n. 3, p. 295–311, 1989.

BOCHKOVSKIY, A.; WANG, C.; LIAO, H. M. **YOLOv4:**: Optimal speed and accuracy of object detection. [*S.l*], 2020. Disponível em: https://arxiv.org/abs/2004.10934. Acesso em: 23 nov. 2021.

CASTRO, F. D.; CASTRO, M. D. **Redes neurais artificiais**. [*S. l.*]: Unicamp, 2001.

CHEN, X.; FANG, H.; LIN, T.-Y.; VEDANTAM, R.; GUPTA, S.; DOLLÁR, P.; ZITNICK, C. L. Microsoft coco captions: Data collection and evaluation server. **arXiv preprint arXiv:1504.00325**, [*S. l.*], 2015.

COSTA, L. M. C. da; MERCHAN-HAMANN, E. Pandemias de influenza e a estrutura sanitária brasileira: breve histórico e caracterização dos cenários. **Revista Pan-Amazônica de Saúde**, [*S. l.*], v. 7, n. 1, p. 15–15, 2016.

DANIELSSON, P.-E. Euclidean distance mapping. **Computer Graphics and image processing**, Elsevier, [*S. l.*], v. 14, n. 3, p. 227–248, 1980.

EDUCAÇÃO, U.-M. **Distância entre dois pontos**. [*S. l.*], 2017. Disponível em: https://mundoeducacao.uol.com.br/matematica/distancia-entre-dois-pontos.htm. Acesso em: 12 dez. 2021.

FARHADI, A.; REDMON, J. Yolov3: An incremental improvement. **Springer**, [*S. l.*], p. 1804–2767, 2018.

GOODFELLOW, I.; BENGIO, Y.; COURVILLE, A. **Deep learning**. MIT press, [*S. l.*], 2016.

HAENLEIN KAPLAN, M.; ANDREAS; TAN, C.-W.; ZHANG, P. Artificial intelligence (ai) and management analytics. **Journal of Management Analytics**, Taylor & Francis, [*S. l.*], v. 6, n. 4, p. 341–343, 2019.

KAEHLER, A.; BRADSKI, G. **Learning OpenCV 3:**: computer vision in c++ with the opencv library. [*S. l.*]: "O'Reilly Media, Inc.", 2016.

KIMBALL, A.; HATFIELD, K. M.; ARONS, M.; JAMES, A.; TAYLOR, J.; SPICER, K.; BARDOSSY, A. C.; OAKLEY, L. P.; TANWAR, S.; CHISTY, Z. *et al.* Asymptomatic and presymptomatic sars-cov-2 infections in residents of a long-term care skilled nursing facility—king county, washington, march 2020. **Morbidity and Mortality Weekly Report**, Centers for Disease Control and Prevention, v. 69, n. 13, p. 377, 2020.

MITCHELL, T. M. *et al.* **Machine learning**. New York: McGraw-hill, 1997.

ONG, S. W. X.; TAN, Y. K.; CHIA, P. Y.; LEE, T. H.; NG, O. T.; WONG, M. S. Y.; MARIMUTHU, K. Air, surface environmental, and personal protective equipment contamination by severe acute respiratory syndrome coronavirus 2 (sars-cov-2) from a symptomatic patient. **Jama**, American Medical Association, v. 323, n. 16, p. 1610–1612, 2020.

PUNN, N. S.; SONBHADRA, S. K.; AGARWAL, S. **Monitoring COVID-19 social distancing with person detection and tracking via fine-tuned YOLO v3 and Deepsort techniques**. 2020. Disponível em: https://arxiv.org/abs/2005.01385. Acesso em: 02 nov. 2021.

REDMON, J.; FARHADI, A. Yolov3: An incremental improvement. **arXiv**, [*S. l.*], 2018.

REZAEI, M.; AZARMI, M. **DeepSOCIAL:**: Social distancing monitoring and infection risk assessment in covid-19 pandemic. Applied Sciences, 2020. v. 10. ISSN 2076-3417. Disponível em: https://www.mdpi.com/2076-3417/10/21/7514. Acesso em: 02 nov. 2021.

RUSSELL, S.; NORVIG, P. **Artificial intelligence:**: a modern approach. [*S. l. s. n.*], 2002.

SZELISKI, R. **Computer vision:**: algorithms and applications. Springer Science & Business Media, [*S. l.*], 2010.

VISO.AI. **What is the COCO Dataset? What you need to know in 2021**. [*S.l*], 2021. Disponível em: https://viso.ai/computer-vision/coco-dataset/. Acesso em: 24 nov. 2021.

WANG, W.; XU, Y.; GAO, R.; LU, R.; HAN, K.; WU, G.; TAN, W. *Detection of SARS-CoV-2 in different types of clinical specimens*. **Jama**, American Medical Association, [*S. l.*], v. 323, n. 18, p. 1843–1844, 2020.

WHO, W. H. O. **WHO Director-General's opening remarks at the media briefing on COVID-19-11 March 2020 Geneva: WHO**. [*S. n.*], 2020. Disponível em: https://www.who.int/dg/speeches/detail/who-director-general-s-opening-remarks-at-the-media-briefing-on-covid-19---11-march-2020). Acesso em: 30 nov. 2020.

WHO, W. H. O. **WHO Director-General's statement on IHR Emergency Committee on Novel Coronavirus (2019-nCoV) Geneva: WHO**. [*S. n.*], 2020. Disponível em: https://www.who.int/news-room/detail/23-01-2020-statement-on-the-meeting-of-the-international-health-regulations-(2005)-emergency-committee-regarding-the-outbreak-of-novel-coronavirus-(2019-ncov). Acesso em: 30 nov. 2020.

WILDER-SMITH, A.; FREEDMAN, D. O. Isolation, quarantine, social distancing and community containment: pivotal role for old-style public health measures in the novel coronavirus (2019-ncov) outbreak. **Journal of travel medicine**, Oxford University Press, v. 27, n. 2, p. taaa020, 2020.