

UNIVERSIDADE TECNOLÓGICA FEDERAL DO PARANÁ

YURI NAGY DE OLIVEIRA

**DETECÇÃO DE MATERIAIS ILÍCITOS EM IMAGENS DE RAIO-X DE BAGAGENS
EM AEROPORTOS**

PONTA GROSSA

2022

YURI NAGY DE OLIVEIRA

**DETECÇÃO DE MATERIAIS ILÍCITOS EM IMAGENS DE RAIOS-X DE BAGAGENS
EM AEROPORTOS**

Detection of illegal materials in x-ray images of baggage at airports

Trabalho de conclusão de curso de graduação apresentada como requisito para obtenção do título de Bacharel em Ciência da Computação da Universidade Tecnológica Federal do Paraná (UTFPR).
Orientador(a): Prof(a). Dr(a). Simone Bello Kaminski Aires.

PONTA GROSSA

2022



[4.0 Internacional](https://creativecommons.org/licenses/by/4.0/)

Esta licença permite compartilhamento, remixe, adaptação e criação a partir do trabalho, mesmo para fins comerciais, desde que sejam atribuídos créditos ao(s) autor(es). Conteúdos elaborados por terceiros, citados e referenciados nesta obra não são cobertos pela licença.

YURI NAGY DE OLIVEIRA

**DETECÇÃO DE MATERIAIS ILÍCITOS EM IMAGENS DE RAIOS-X DE BAGAGENS
EM AEROPORTOS**

Trabalho de conclusão de curso de graduação
apresentada como requisito para obtenção do título de
Bacharel em Ciência da Computação da Universidade
Tecnológica Federal do Paraná (UTFPR).

Data de aprovação: 17 de novembro de 2022

Simone Bello Kaminski Aires
Doutorado
Universidade Tecnológica Federal do Paraná - Campus Ponta Grossa

Ionildo José Sanches
Doutorado
Universidade Tecnológica Federal do Paraná - Campus Curitiba

Erikson Freitas de Moraes
Doutorado
Universidade Tecnológica Federal do Paraná - Campus Ponta Grossa

**PONTA GROSSA
2022**

Dedico este trabalho à minha família, pelos
momentos de ausência.

AGRADECIMENTOS

Meu primeiro agradecimento vai a Deus por me dar força e foco para realização desta metodologia, e todo discernimento em conjunto ao foco que precisei para chegar até esse momento da graduação. Agradeço à minha orientadora Simone Bello Kaminski Aires por todo apoio e dedicação durante o desenvolvimento da pesquisa. A minha mãe que sempre esteve comigo em todos os momentos da graduação, me dando todo apoio necessário para que eu nunca perdesse o foco. Aos amigos da universidade que estiveram ao meu lado em grupos de estudos durante os momentos de distração.

Quero também parabenizar os professores e professoras que exerceram excelência em seu trabalho, podendo agregar com minha evolução tanto profissional quanto acadêmica.

Eu denomino meu campo de Gestão do Conhecimento, mas você não pode gerenciar conhecimento. Ninguém pode. O que você pode fazer, o que a empresa pode fazer é gerenciar o ambiente que otimize o conhecimento.
(DAVENPORT; PRUSAK, 2012).

RESUMO

O transporte aéreo cresce de uma maneira significativa no Brasil e o fluxo de usuários está em constante ampliação. Diante disso, garantir a segurança dos usuários deve acompanhar este crescimento. Dentre as medidas de segurança essenciais nos aeroportos, tem-se a inspeção de bagagem, essa medida acontece durante o fluxo de embarque e possui o intuito de buscar objetos que possam conter algum risco durante o voo dos usuários. Essa tarefa acontece por meio de um raio-X onde as bagagens são submetidas, gerando uma imagem em tempo real, esta imagem é utilizada para análise dos operadores humanos, auxiliando na detecção de objetos que possam oferecer perigo dentro da bagagem. Essa verificação necessita de muita atenção e treinamento dos operadores, para que não ocorram enganos durante a análise dos objetos apresentados na imagem. Fatores como: aptidão, habilidade, motivação, entre outros, podem afetar diretamente a análise, qualquer falha, pode deixar os voos expostos a ameaças. Durante o procedimento de expedição, alguns aeroportos contam com o auxílio de sistemas integrados que realizam a simples classificação de objetos. Neste trabalho, foi proposta uma rede neural convolucional capaz de classificar os objetos contidos nas malas e até mesmo apontar os possíveis objetos que apresentam perigo para o voo, com um fator de confiança que possa auxiliar de maneira mais significativa os operadores humanos. Ao decorrer da metodologia discute um modelo de rede neural convolucional, baseada na arquitetura da YOLO que visa a classificação de objetos em tempo real. Foi realizado uma busca para encontrar alguns bancos de imagens de bagagens em raio-X. A partir da definição dos bancos de imagens foram aplicadas técnicas de processamento de imagem, para melhor evidenciar as características das imagens além, de deixar as imagens padronizadas seguindo os padrões da rede utilizada. Notou-se a necessidade da aplicação de *data argumentation* método que realizou a expansão dos bancos de imagem. Foram propostos dois métodos de classificação: binária e não binária, onde objetos como: facas, armas, carregadores portáteis, entre outros, como perigosos. Cada método realizou a rotulação das imagens de forma distinta, o método binário foi rotulado utilizando uma área da imagem em que pode ou não conter um objeto perigoso, já o modelo não binário foi realizado a rotulação especificamente nos objetos que possam trazer perigo. A partir da finalização da rotulação foi realizado o treinamento da YOLO, rede neural convolucional utilizada na metodologia. Os dois métodos utilizados na classificação dos *datasets* tiveram grandes resultados durante a classificação dos objetos contidos nas bagagens, alcançando uma precisão de 96% para a classificação não binária e 67% de precisão no modelo binário.

Palavras-chave: YOLO; raio-X; inspeção em aeroportos; detecção de objetos; imagens de raio-X de bagagem.

ABSTRACT

Air transport is growing significantly in Brazil and the flow of users is constantly expanding. Given this, ensure the safety of users who must accompany this growth. Among the essential security measures at airports is the airworthiness of baggage. This measure takes place during the boarding flow and intends to look for objects that may contain some risk during the users' flight. This task takes place through an X-ray in which the luggage is kept, generating a real-time image, used by human operators to carry out the analysis, and thus detect whether the luggage contains any object that could pose danger. However, this verification needs a lot of attention and training of the operators, so that they do not cause mistakes during the analysis of the objects presented in the image. Factors such as: skill, motivation, among others, can directly affect the analysis, any failure can expose flights to threats. During the dispatch procedure, some airports rely on the aid of integrated systems that carry out simple object classification. We propose a convolutional neural network capable of classifying the objects contained in the bags and even saying the possible objects that present a danger to the flight, with a confidence factor that can help human operators in a more significant way. During the methodology, it discusses a convolutional neural network model, based on the YOLO architecture, which aims to classify objects in real time. A search was carried out to find some X-ray baggage image banks. From the definition of the image banks, image processing techniques were applied, to better highlight the characteristics of the images, in addition to leaving the standardized images following the standards of the network used. It was noted the need to apply a data argumentation method that carried out the expansion of image banks. Two classification methods were proposed: binary and non-binary, where objects such as: knives, weapons, portable chargers, among others, as dangerous. Each method carried out the labeling of the images in a different way, the binary method was labeled using an area of the image that may or may not contain a dangerous object, while the non-binary model carried out the labeling specifically on objects that may be dangerous. From the finalization of the labeling, the training of the YOLO convolutional neural network used in the methodology was carried out. The two methods used in classifying the datasets had great results during the classification of objects contained in luggage, achieving an accuracy of 96% for the non-binary classification and 67% accuracy for the binary model.

Keywords: YOLO; X-ray; airport inspection; baggage detection; baggage x-ray images.

LISTA DE ILUSTRAÇÕES

Figura 1 - Bagagens com arma branca e arma de fogo	15
Figura 2 - Conjunto de imagens de inspeção de bagagens em aeroportos.....	16
Figura 3 - Métodos de realce de contraste.....	18
Figura 4 - Algoritmo de reconstrução 3D(<i>Space Carving</i>).....	19
Figura 5 - Aplicação da metodologia.....	20
Figura 6 - Exemplo de aplicação de Escala de Cinza.....	23
Figura 7 - Exemplo de aplicação do Flip Horizontal.....	23
Figura 8 - Exemplo de aplicação do Flip Vertical	24
Figura 9 - Exemplo de aplicação da Rotação.....	24
Figura 10 - Exemplo de aplicação do método de Isolar Objetos.....	24
Figura 11 - Exemplo de aplicação da função Cisalhamento.....	25
Figura 12 - Exemplo dataset Compass-Xp.....	26
Figura 13 - Exemplo banco de imagem SIXRay	27
Figura 14 - Exemplo dataset PIDray	28
Figura 15 - Exemplo dataset CLCXray.....	28
Figura 16 - Exemplo do fluxo de uma Rede Neural Convolutiva	29
Figura 17 - Exemplo de aplicação de filtro na camada de convolução	30
Figura 18 - Exemplo de saída de aplicação de filtro	31
Figura 19 - Exemplo de execução da camada totalmente conectada.....	31
Figura 20 - Exemplo de execução da YOLO.....	32
Figura 21 - Arquitetura das camadas da YOLO	33
Figura 22 - Exemplo de classificação da YOLO.....	34
Figura 23 - Exemplo de aplicação da metodologia	39
Figura 24 - Imagens de entrada modelo binário	45
Figura 25 - Comparação da precisão de treinamento binário	46
Figura 26 - Imagens de entrada modelo não binário.....	47
Figura 27 - Comparação da precisão do treinamento não binário.....	48

LISTA DE TABELAS

Tabela 1 - Comparativo de modelos operacionais da YOLOv5.....	35
Tabela 2 - Geração do novo banco de classe binária.	41
Tabela 3 - Políticas de divisão de treinamento e validação da rede.	42
Tabela 4 - Classes utilizadas no modelo não binário.....	42
Tabela 5 - Resultados do treinamento binário.....	46
Tabela 6 - Resultados do treinamento não binário.	49
Tabela 7 - Comparação dos métodos.....	50
Tabela 8 - Comparação das metodologias.....	51

LISTA DE ABREVIATURAS E SIGLAS

2D	Bidimensional
3D	Tridimensional
AP	<i>Average Precision</i> (Precisão Média)
CNN	<i>Convolutional Neural Networks</i> (Redes Neurais Convulacionais)
FN	<i>False Negative</i> (Falso Negativo)
FP	<i>False Positive</i> (Falso Positivo)
FPS	<i>Frame Per Second</i> (Quadro por Segundo)
HOG	<i>Histogram of Oriented Gradients</i> (Histograma de gradientes orientados)
HSV	<i>Hue Saturation Value</i> (Valor de Saturação da Tonalidade)
mAP	<i>Mean Average Precision</i> (Média da Precisão Média entre as classes)
SURF	Speeded Up Features (Características de Aceleração)
SVM	<i>Support Vector Machines</i> (Máquina de vetores de suporte)
TN	<i>True Negative</i> (Verdadeiro Negativo)
TP	<i>True Positive</i> (Verdadeiro Positivo)
YOLO	<i>You Only Look Once</i> (Você só olha uma vez)

LISTA DE SÍMBOLOS

<i>l</i>	<i>Large</i> (Grande)
<i>m</i>	<i>Medium</i> (Médio)
<i>n</i>	Nano
<i>s</i>	<i>Small</i> (Pequeno)
<i>x</i>	<i>XLarge</i> (Extra Grande)

SUMÁRIO

1	INTRODUÇÃO	14
1.1	Problema e justificativa	17
1.2	Objetivos	17
1.2.1	Objetivo geral	17
1.2.2	Objetivos específicos	17
2	TRABALHOS RELACIONADOS	18
3	FUNDAMENTAÇÃO TEÓRICA	22
3.1	Processamento de imagem	22
3.1.1	Escala de cinza	23
3.1.2	Flip	23
3.1.3	Rotação	24
3.1.4	Recorte	24
3.1.5	Redimensionar	25
3.1.6	Cisalhamento	25
3.2	Banco de imagens	25
3.2.1	Compass - XP	26
3.2.2	SIXray	26
3.2.3	PIDray	27
3.2.4	CLCXray	28
3.3	Rede neural convolucional	29
3.4	YOLO	32
3.5	Critérios de avaliação	35
3.5.1	Precisão	36
3.5.2	Taxa de verdadeiro positivo	36
3.5.3	Taxa de falso positivo	36
3.5.4	Média de precisão	37
3.5.5	Recall	37
3.6	Ambientes de desenvolvimento	37
4	DESENVOLVIMENTO	39
4.1	Experimento 1 - modelo binário	40
4.2	Experimento 2 - modelo não binário	42
4.3	Rede neural utilizada nos experimentos	43
5	RESULTADOS	44

5.1	Modelo de rede binário	44
5.2	Modelo de rede não binário	47
5.3	Análise sobre a precisão de outras metodologias	50
6	CONCLUSÃO	52
6.1	Trabalho Futuros	53
6.1.1	Itens de aprimoramento.....	53
	REFERÊNCIAS.....	54

1 INTRODUÇÃO

Tendo em vista que o transporte aéreo cresce de uma maneira significativa no Brasil e sabe-se que o fluxo de usuários está aumentando nos aeroportos (ANAC, 2019), assim a responsabilidade para garantir a segurança de seus usuários também cresce. Os locais públicos devem conter efetivas medidas de segurança porque a exposição a ameaças é diária (ANAC, 2019). A forma em que são tomadas as medidas preventivas compreende tarefas distintas, tais como: centrais de segurança, guardas, inspeção de bagagens e controle de acesso.

Uma tarefa de segurança muito importante durante um fluxo de embarque é a inspeção de bagagens e passageiros. Eventualmente é feita a análise das imagens de raio-X dos objetos com o intuito de buscar dentro de bagagens materiais ilícitos ou que apresentem um certo grau de letalidade, ou seja, materiais que de alguma forma apresentem ameaça durante o voo. Materiais esses como: bombas, facas, líquidos inflamáveis, lâminas de barbear e até mesmo correntes, entre outros. Tal tarefa é realizada por operadores humanos, e estes são auxiliados por sistema computadorizado durante o processo (ANAC, 2017).

Para Magalhães (2020) o fato de a verificação de bagagens ser majoritariamente feita por operadores humanos pode ocorrer falhas, pois há necessidade de um período de treinamento rígido para que os operadores dominem o perfil de riscos contidos durante aquele voo. Alguns fatores influenciam na tomada de decisões durante o processo de verificação da bagagem, tais como: aptidão, treinamento, habilidade, atenção, motivação. Estes fatos influenciam durante o processo de análise das imagens. Qualquer tipo de falha pode deixar os voos expostos a ameaças.

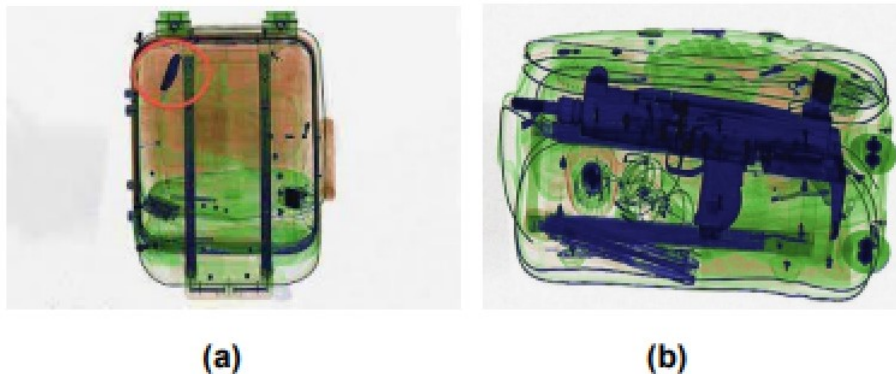
A atividade de inspeção de bagagens pode ser auxiliada através de sistema que realiza a classificação de objetos, sendo está uma das subáreas do processamento de imagens. A tarefa de classificação de objetos baseia-se em padrões, classes de padrões, descritores quantitativos e qualitativos (GONZALEZ; WOODS, 2010). O reconhecimento de padrões através do aprendizado de máquina é utilizado em diversas áreas, até mesmo na medicina (OLIVEIRA, 2018).

As imagens de raio-X exigem extrema atenção durante a análise humana, não podendo haver nenhum deslize, lapso ou equívoco durante a análise. Segundo Arcúrio (2014), os equívocos cometidos pelos operadores durante a análise são os

mais difíceis de serem detectados, isso retrata a necessidade de um agente que não tenha equívocos e não ocorra falhas durante a análise. O manuseio do equipamento de raio-X também requer uma extrema distinção dos objetos, então se aplica filtros que auxiliam durante o processo. Arcúrio, Fortes e Armborst (2014) apresentam um erro relacionado a limitação da racionalidade humana ou de um modelo incompleto ou inexato do espaço do problema. Um exemplo de erro humano é relacionado à expressão "Olhos que não veem, coração que não sente", neste caso o operador coloca-se no lugar do passageiro e simplifica o processo de inspeção de segurança.

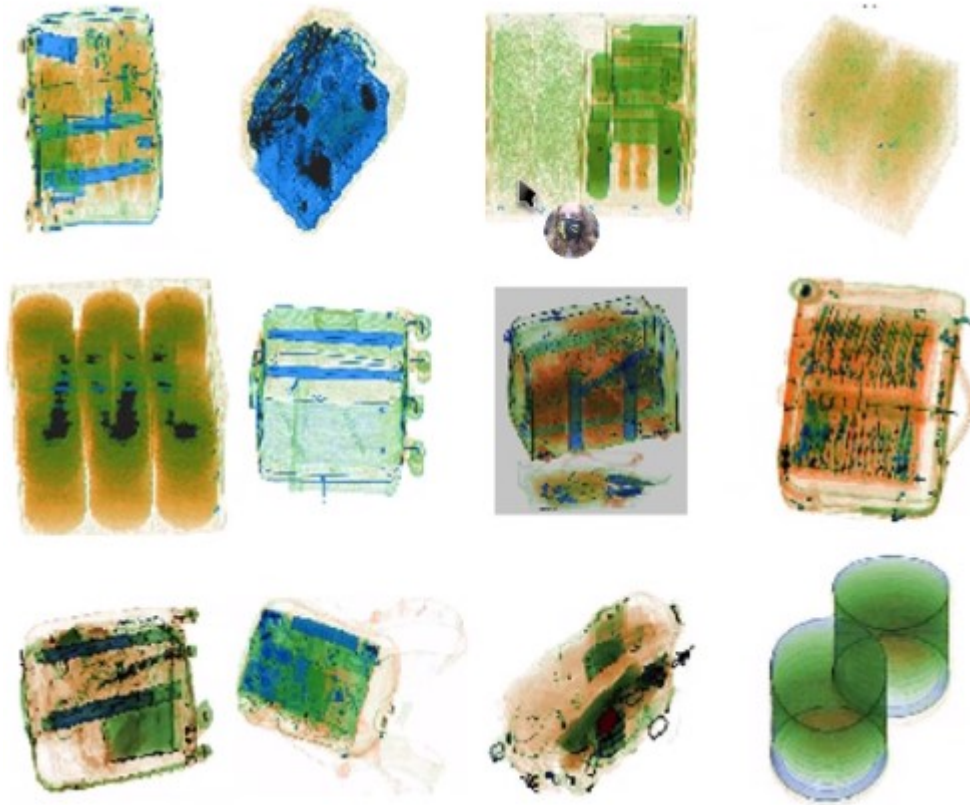
A Figura 1 apresenta imagens de raio-X de bagagens. Observa-se que os objetos não são de fácil identificação aos olhos humanos, não apenas pela qualidade da imagem, mas também porque os objetos podem estar sendo obstruídos por outros objetos. Deve-se considerar possível que um conjunto de objetos que não trazem risco ao voo formem a estrutura de outro objeto pois as cores apresentadas pelas imagens são relativamente limitadas. A Figura 1 (a) apresenta uma mala contendo uma possível arma branca (faca) e a Figura 1 (b) apresenta uma mala com uma arma de fogo (AIRWAY, 2016). A Figura 2 apresenta um conjunto de imagens geradas durante a inspeção de bagagens em aeroportos.

Figura 1 - Bagagens com arma branca e arma de fogo



Fonte: AirWay (2016)

Figura 2 - Conjunto de imagens de inspeção de bagagens em aeroportos



Fonte: Autoria própria (2022)

Ao analisar o conjunto de exemplos de bagagens contidos na Figura 2, percebe-se o quão difícil é para os agentes de inspeção encontrarem objetos que tragam risco durante o voo. Pois as imagens não possuem um conteúdo trivial para a análise, não são nítidas, os objetos aparecem distorcidos e até mesmo podem ficar sobrepostos a outros trazendo dúvida durante a inspeção.

Estudos na área de classificação de objetos através de imagens de raio-X apresentam desafios similares aos apresentados durante a inspeção de bagagens, durante o momento da definição de quais objetos podem trazer algum risco ou não. Mas por sua vez, a análise feita por agentes não humanos (computadores) tende a ter uma chance de falhas menores que as feitas por humanos. O objetivo deste trabalho é propor uma rede neural convolucional, capaz de detectar e classificar objetos contidos em bagagens que possam trazer algum risco ao ser humano como: lâminas, armas, bombas, bastão de choque, algemas, martelos, chaves utilitárias, canivetes, *spray*, tesouras, objetos de vidro, pilhas, isqueiro e bala.

1.1 Problema e justificativa

A identificação de objetos que apresentam alguma ameaça em aeroportos é uma das atividades essenciais. Sabe-se que essa tarefa hoje em dia é feita através de agentes humanos, podendo conter falhas durante a análise, sendo este um processo delicado e realizado num curto prazo de tempo (RAMOS, 2016). As empresas aéreas buscam eficiência durante o período de embarque. O embarque mais rápido e seguro traz uma melhor experiência para o usuário (LOPES, 2018). Este trabalho se propõe a contribuir para a área de classificação de objetos em bagagens através de imagens de raio-X. Como base da pesquisa será treinado modelos de redes convolucionais buscando a classificação dos objetos contidos nas bagagens tais como os trabalhos desenvolvidos por Miao et al. (2019), Akcay et al. (2016) e Kundegorski et al. (2018).

1.2 Objetivos

A pesquisa organizou os objetivos entre objetivo geral e objetivos específicos.

1.2.1 *Objetivo geral*

O objetivo geral deste trabalho é desenvolver uma metodologia capaz de reconhecer objetos ilícitos ou letais em imagens de raio-X de bagagens em aeroportos.

1.2.2 *Objetivos específicos*

- Analisar as bases de dados de imagens de raio-X de bagagens disponíveis.
- Compreender e avaliar Redes Neurais Convolucionais.
- Buscar, compreender e aplicar algoritmos de pré-processamento de imagens.
- Criação de um novo *dataset* rotulado com apenas duas categorias de dados: *safe* e *danger*.
- Realizar treinamento de uma rede neural convolucional.
- Analisar os resultados obtidos.

2 TRABALHOS RELACIONADOS

A busca de processos que possam agilizar e trazer maior segurança durante o embarque em aeroportos são objetos de diversas pesquisas, cada uma propõe diferentes metodologias para o reconhecimento de materiais ilícitos em malas e bagagens. Algumas pesquisas relacionam-se com o trabalho proposto utilizando imagens de raio-X para detecção de objetos ilícitos.

O trabalho desenvolvido por Sajid Khan e Yin Wang (2012) propõe a melhoria das imagens de raio-X, realçando o contraste e trazendo uma redução de ruídos. Esse procedimento visa diminuir a taxa de alarmes falsos durante a inspeção realizada pelos agentes. A implementação da metodologia é baseada em *High Energy* e *Low Energy*, que são a intensidade dos raio-X emitidos sobre os objetos gerando dois tipos de imagem *High* e *Low (Energy)*. Através das duas imagens os pesquisadores pretendem distinguir os tipos de objetos por seu número atômico, pois grande parte dos itens explosivos são compostos por matérias orgânicas compostas de carbono, hidrogênio, nitrogênio e oxigênio. As drogas por sua vez são ricas em hidrogênio e carbono. A abordagem consiste em três estágios principais: 1) fusão (combinação das duas imagens), 2) redução de ruído e 3) aprimoramento de imagem. Os autores utilizam a técnica de *Denoising* e o histograma para redução de ruído.

A pesquisa relata que a técnica de pré-processamento pode auxiliar na extração de características e que tem bons resultados na redução nos falsos positivos. A Figura 3 mostra a saída das imagens com (a) *high* e (b) *low Energy*. Onde o método (b) destaca mais as informações sobre o objeto silício (placa mãe de um computador).

Figura 3 - Métodos de realce de contraste

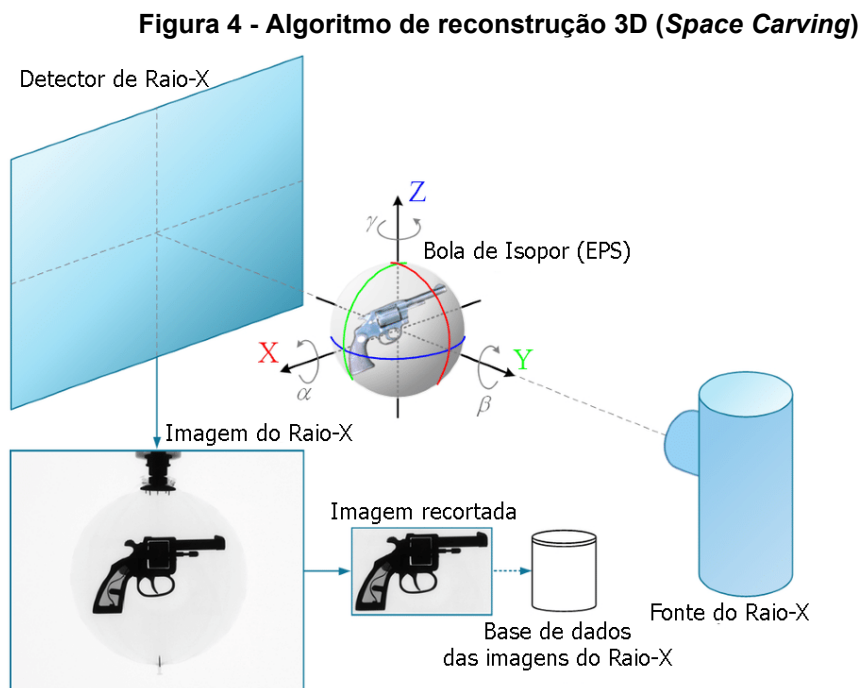


Fonte: Khan e Wang (2012)

A pesquisa realizada por Vladimir Riffo, Ivan Godoy e Domingo Mery (2019) propôs a detecção de objetos que ofereçam algum tipo de ameaça para os

passageiros durante o voo e embarque. O procedimento de reconhecimento é baseado na reconstrução dos objetos em imagens 3D a partir de um conjunto de imagens 2D de espectro único. Para gerar as imagens 3D, cada imagem do conjunto 2D representa uma parte do contorno do objeto, o nome dado a essa técnica é *Space Carving*, pode-se observar o funcionamento do algoritmo na Figura 4.

Foram usadas sequências que consistem em 90 imagens de raios-X para reconstrução de uma imagem, autores relatam sucesso com acurácia de 97% no reconhecimento de armas de fogo.



Fonte: Adaptado de Riffo, Godoy e Mery (2019)

A metodologia de Samet Akcay e Toby Breckon (2020) propõe uma análise dos algoritmos clássicos de aprendizado de máquinas e os convencionais, focados na classificação de objetos, detecção, segmentação e detecção de anomalias. O estudo foi dividido em 2 etapas: a primeira fase discute superficialmente as técnicas clássicas de aprendizado de máquina; e a etapa final é focada nos algoritmos modernos de aprendizado profundo. Os autores apontam que a combinação do FAST-SURF com o treinamento de uma SVM em um grande conjunto de dados, obteve-se bons resultados com precisão de 94%.

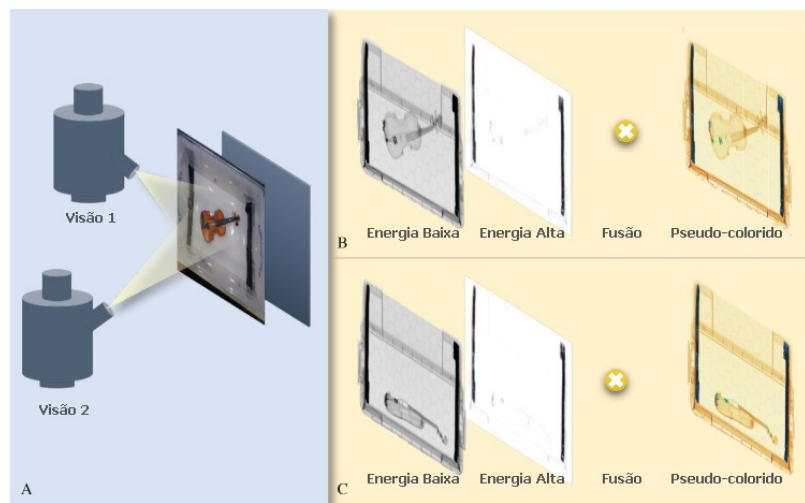
Para detecção de objetos relatam uma abordagem que utiliza um classificador SVM linear em conjunto de HOG. O autor relata outra metodologia que utiliza o conceito de *multi-view*, as imagens resultantes são enviadas para um classificador k-

NN. Sendo um conjunto de 120 imagens para o treinamento obteve resultados muito significativos alcançando 95,7% precisão.

Na etapa inicial foi realizado a fusão das imagens com *high energy* e *low energy*. Para o aprimoramento das imagens foi utilizado uma rede neural de multicamadas para dizer qual é a melhor técnica de aprimoramento com base na entrada. Foi utilizado TIP (*Threat Image Projection*) para criar um treinamento sintético das imagens que contém ameaças, além de ser possível utilizar essa técnica para o algoritmo de detecção de objetos que obteve resultados significativos que renderam 93% de precisão.

É apresentado no artigo algumas implementações utilizando Rede Neural Convolutacional com resultados muito promissores durante a etapa de classificação, detecção e segmentação. O artigo apresenta alguns bancos de imagens de raio-X utilizados para o treinamento, sendo eles: Durham Baggage Patch/Full Image Dataset com 15.449 imagens, GDXray com 19.407 imagens, SIXray contendo 1.509.231 imagens, entre outros bancos menores. A Figura 5 apresenta superficialmente como a metodologia se comporta.

Figura 5 - Aplicação da metodologia



Fonte: Adaptado de Akcay e Breckon (2020)

A pesquisa realizada por Boying Wang *et al.* (2021) focou na criação de um novo banco de imagens de raio-X, utilizou outros seis bancos de imagens como base. A criação desse novo banco de imagens foi denominado de PIDRay contentando 47.677 imagens. Além da criação do banco de imagem os pesquisadores desenvolveram uma rede neural convolutacional capaz de classificar 12 classes de objetos denominada de *selective dense attention network* (SDANet), que é nada mais

de um aprimoramento da ResNet que foi utilizada como a base da nova rede desenvolvida.

Depois da execução da etapa de convolução da rede, a matriz com o conjunto de características é enviada para uma camada de totalmente conectada que acontece o refinamento monitorado, a rede obteve uma média de 61,6% de precisão.

A proposta apresentada por Jianjun Wu e Shaowen Liao (2022) utiliza um modelo de rede neural convolucional que utiliza grades de detecção chamada de YOLO (*You Only Look Once*).

Além de utilização do modelo da YOLO, foi implementado uma detecção utilizando uma etapa de cascata cruzada, que traz uma etapa de *pooling* com diferentes tamanhos de filtros. Depois da aplicação de cada filtro é realizado uma concatenação entre os filtros. O resultado final é enviado para uma camada CBL que possui uma etapa de convolução, normalização e regularização do dado.

O método de cascata cruzada teve comparações com três modelo de redes e obteve o melhor índice de precisão. A metodologia do autor obteve 88% com a detecção executando a 392 FPS, enquanto o segundo melhor algoritmo comparado executou a 50 FPS e 87% de precisão.

Este capítulo apresentou alguns trabalhos relacionados que utilizam imagens de raio-X para identificação de objetos ilícitos ou objetos que apresentem ameaça para seus próprios portadores e até mesmo pessoas ao redor.

Os trabalhos que apresentaram técnicas de aprendizado de máquina com resultados expressivos durante a classificação são os de: Samet Akcay e Toby Breckon (2020), Jianjun Wu e Shaowen Liao (2022) e Boying Wang *et al.* (2021). A metodologia de Wang Kan (2012) apresenta técnicas de como ajustar imagens e realçar suas características e padronizando-as.

3 FUNDAMENTAÇÃO TEÓRICA

Neste capítulo serão apresentadas técnicas utilizadas no desenvolvimento deste trabalho, tais como: algoritmos de processamento de imagens, banco de imagem, modelos de redes neurais e as medidas de avaliação de desempenho.

3.1 Processamento de imagem

Segundo Rafael C. Gonzalez e Richard E. Woods (2000) o interesse em métodos de processamento de imagens digitais decorre de duas áreas principais de aplicação: melhoria de informação visual para a interpretação humana e o processamento de dados de cenas para percepção automática através de máquinas.

Reconhecimento automático de caracteres, visão computacional, inspeção de produtos, reconhecimento militar, processamento automático de impressões digitais, análise de resultados de raios X e análise sanguínea, são alguns problemas típicos que utilizam do processamento de imagem para propor soluções no momento de percepção de dados utilizando máquinas (GONZALEZ; WOODS, 2000).

Considerando as pesquisas apresentadas na seção 2 o processamento de imagem pode ser uma etapa essencial para preparar *datasets* que serão consumidos por *deep learnings*, não sendo uma etapa obrigatória, porém pode ser muito importante.

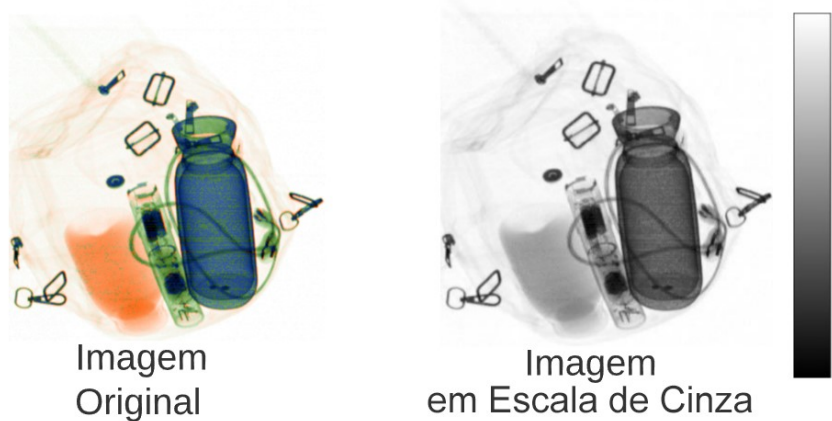
Como na maioria das vezes os *dataset* são criados por comunidades e as imagens inseridas, nem sempre obedecem a um padrão, podendo ter cores, formatos e até mesmo tamanhos diferentes, nesses casos o pré-processamento torna-se uma etapa essencial para aprimorar a qualidade dos dados. Como apresentado por Henrique M. A. Gomes (2019), que aponta o quão importante foi o processamento de imagens para criação do seu banco de imagens, por conta das limitações do aparelho utilizado para gerar as imagens.

Um exemplo de aplicação do processamento de imagem é durante o processo de *data argumentation* que aplica um conjunto de técnicas para expansão de um banco de imagens. A seguir vamos explorar algumas técnicas de processamento de imagem.

3.1.1 Escala de cinza

Sendo um novo tipo de representação de uma imagem a escala de cinza (*grayscale*), tem o propósito de realçar características em uma imagem. Pois o mesmo converte a imagem de entrada em uma saída de escala de cinza com valores que iniciam de 0 a 255 (0 representa o ponto mais escuro e 255 o ponto mais claro) (GONZALEZ; WOOD, 2000). A Figura 6 exemplifica a nova representação da imagem de entrada.

Figura 6 - Exemplo de aplicação de Escala de Cinza



Fonte: Autoria própria (2022)

3.1.2 Flip

Flip ou *Flipping* é uma técnica que realiza uma inversão na imagem original, onde essa inversão pode ser na horizontal ou na vertical como apresentado nas Figuras 7 e 8 respectivamente (CLARO *et al.*, 2020).

Figura 7 - Exemplo de aplicação do Flip Horizontal



Fonte: Adaptado de Claro *et al.* (2020)

Figura 8 - Exemplo de aplicação do Flip Vertical

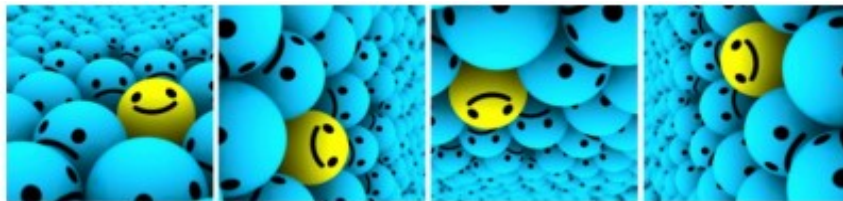


Fonte: Adaptado de Claro *et al.* (2020)

3.1.3 Rotação

Na aplicação da técnica de rotação é possível realizar o movimento de “girar” uma imagem nos sentidos horário e anti-horário. Em ambos sentidos pode-se utilizar os ângulos múltiplos de 90° para o movimento de rotação, como ilustrado na Figura 9 (CLARO *et al.*, 2020).

Figura 9 - Exemplo de aplicação da Rotação



Fonte: Adaptado de Claro *et al.* (2020)

3.1.4 Recorte

O procedimento de recorte (isolar objetos) aplica um corte na imagem em regiões delimitadas pelo usuário, sendo essas delimitações *labels* que representam um objeto, como apresentado na figura 10, onde aplica-se um corte para evidenciar o objeto na cena, removendo as informações irrelevantes ao redor dele.

Figura 10 - Exemplo de aplicação do método de Isolar Objetos



Fonte: Autoria própria (2022).

3.1.5 Redimensionar

Redimensionar é o procedimento de ajustar as dimensões de uma imagem, altura ou largura ou ambas. Podendo preservar a proporção durante o redimensionamento ou não se preocupar com a proporção deixando a imagem com detalhes menos evidentes (ALEXANDROVA; TATLOCK; CAKMAK, 2015).

3.1.6 Cisalhamento

A aplicação de cisalhamento (*shear*) é similar a função rotação. O diferencial da execução é que se utiliza de um eixo fixado, e o mesmo é esticado em um determinado ângulo, dando um aspecto que a imagem aconteceu a rotação da imagem (CLARO *et al.*, 2020). A Figura 11 demonstra o funcionamento do método de cisalhamento.

Figura 11 - Exemplo de aplicação da função Cisalhamento



Fonte: Adaptado de Claro *et al.* (2020)

3.2 Banco de imagens

Para realizar o treinamento de uma rede neural é necessário possuir um banco de imagens. Essa base de dados precisa ser classificada, para que a rede consiga classificar os objetos contidos, que são as imagens de entrada (CLARO *et al.*, 2020).

Em um primeiro momento, é necessário conseguir um conjunto de imagens de raio-X de bagagens pois todos os dias há inúmeras fiscalizações em aeroportos, estações de trens entre outras plataformas onde usuários necessitam consumir algum recurso que necessita de alta segurança.

Porém a tarefa exaustiva está na classificação dos objetos da imagem, pois até mesmo para olhos humanos esses objetos nem sempre estão bem-dispostos nas imagens podendo causar confusão durante a classificação.

A seguir será apresentado e detalhada as características encontradas sobre algumas das bases de imagens pesquisadas para a realização dos experimentos.

3.2.1 *Compass - XP*

As imagens desta base são todas criadas em ambientes controlados, contendo 1.928 imagens divididas em 366 classes de objetos. O criador da base de dados (LEWIS; MATTHEW; JERONE, 2019) disponibilizou alguns formatos da mesma imagem: 1) Técnica de baixa energia (*low energy*); 2) Técnica de alta energia (*high energy*); 3) Técnica de densidade, onde a representação dos objetos é a partir da inferência do mesmo entre os dois canais de cor. Para esses 3 formatos as imagens são de 8-bit em escala de cinza. A Figura 12 apresenta um modelo das imagens apresentadas no conjunto de dados.

Figura 12 - Exemplo dataset Compass-XP

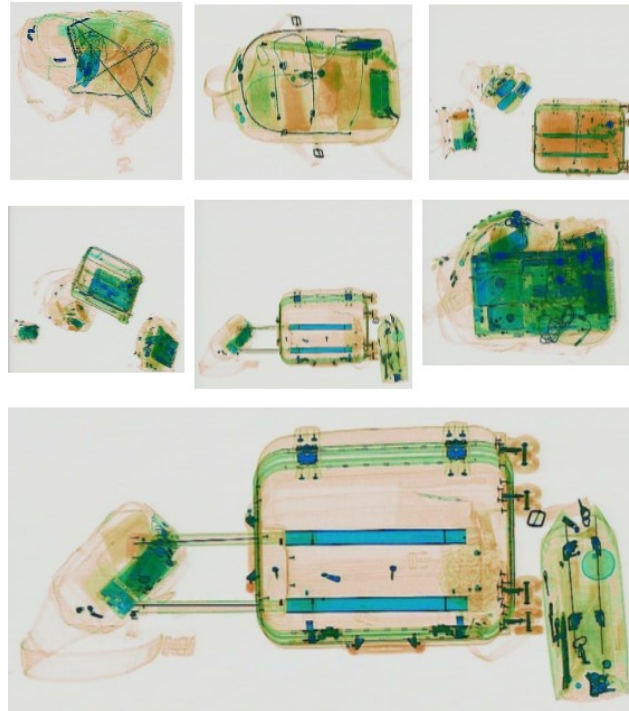


Fonte: Adaptado de Lewis, Matthew e Jerone (2019)

3.2.2 *SIXray*

Durante a pesquisa realizada, este foi o maior banco de imagens encontrado com imagens reais de bagagens, contendo 10.509.231 de imagens. Porém não são disponibilizadas as imagens com objetos proibidos, a quantidade de imagens desse banco é elevada pois os autores aplicaram o procedimento de *data argumentation* (MIAO *et al.*, 2019). A base de imagens conta com 6 classes de objetos proibidos, mas na versão disponível para download não foi possível o acesso a esses dados. A Figura 13 apresenta o modelo de imagens contido no banco.

Figura 13 - Exemplo banco de imagem SIXRay



Fonte: Adaptado de Miao et al. (2019)

3.2.3 PIDray

Para a construção da base de imagens PIDRay, foram utilizados 6 bancos de imagens. Considera-se que esta base foi a que obteve uma atenção especial em virtude da qualidade dos rótulos e a forma em que foram divididas as imagens nos conjuntos. A metodologia de Yanlu Wei et al. (2020) dividiu os objetos em faixas como: (1) fácil (*easy*), (2) difícil (*hard*), (3) escondido (*hidden*).

A Figura 14 apresenta alguns exemplos de cada faixa. Contendo 12 classes de objetos e tendo num total de 47.677 imagens.

- 1) Assim como apresentado na Figura 14, a faixa fácil apresenta os dados mais evidentes na imagem, a maioria dos objetos perigosos estão sobre outros objetos, assim o objeto fica em evidência sendo facilmente identificado. A imagem dessa faixa contém um nível de detecção mais fácil que as demais faixas.
- 2) Os objetos da segunda faixa (difícil) tem algumas obstruções na sua área de ocupação. Em alguns casos objetos não tem uma disposição na cena dando impressão que o mesmo está quebrado.

- 3) A faixa escondida apresenta alguns objetos que estão próximos de objetos com alguma alta representatividade de algum tipo de metal, assim esses objetos podem esconder parte ou o todo do objeto a ser detectado.

Figura 14 - Exemplo dataset PIDray



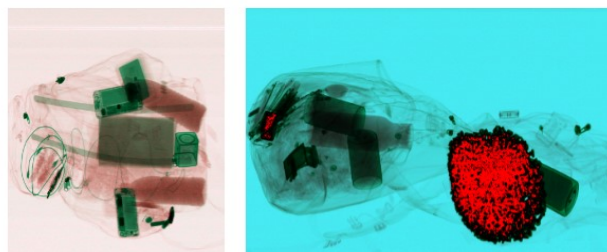
Fonte: Adaptado de Wei *et al.* (2020)

3.2.4 CLCXray

Esta base de dados contém imagens reais de bagagens em raio-X. O banco apresenta 9.565 imagens, porém todas são de objetos proibidos, dividido em 12 classes. Observa-se neste *dataset* que as imagens contidas não possuem um padrão de cores, algumas imagens estão num padrão de HSV (*hue saturation value*) e a maior parte são originais sem aplicação de qualquer filtro, como apresentado na Figura 15.

Figura 15 apresenta a falta de padrão do banco de imagens, pois foi distribuído o filtro HSV de forma aleatória entre as imagens.

Figura 15 - Exemplo dataset CLCXray



Fonte: Adaptado de Bhowmik e Neel (2022)

3.3 Rede neural convolucional

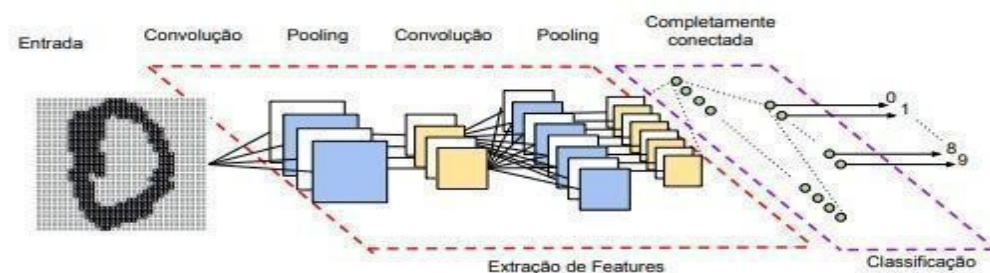
A aprendizagem profunda não é uma metodologia tão recente, porém existem diversas pesquisas promissoras no âmbito de aprendizado de máquinas, sendo utilizadas em diversas áreas de aplicações, e sabe-se que o uso de aprendizado de máquina para classificação de imagens cresce radicalmente (MARTINS, 2018). A utilização de *deep learning* para extração de características se dá pela alta eficiência durante as camadas de convolução e *pooling*, utilizando de abstrações convencionadas por seus módulos (FEITOZA; SILVA; CALUMBY, 2019).

Rede neural convolucional é uma metodologia que une visão computacional e inteligência artificial, sendo caracterizado por uma das técnicas de aprendizado profundo. As CNNs tornaram-se relevantes no meio de classificação de imagens pela efetividade durante a classificação de objetos (MESQUITA *et al.*, 2022).

Atualmente existem inúmeras pesquisas para aprimorar os métodos de execução de uma rede neural convolucional, pois sua implementação é muito complexa, além de utilizar um alto custo computacional durante o treinamento da rede e grandes bases de dados de entrada da rede.

Algumas das pesquisas tem o intuito de promover maior eficiência da rede através dos avanços computacionais. Podendo utilizar do poder computacional para diminuir o tempo de execução do treinamento da rede e até mesmo a estrutura de dados utilizada no momento em que a rede entra em execução (FANK; SCHREINER; DUARTE, 2021). A Figura 16 exemplifica as etapas de treinamento da uma Rede Neural Convolucional onde é abstraída cada etapa sendo: entrada, convolução, *pooling* e classificação.

Figura 16 - Exemplo do fluxo de uma Rede Neural Convolucional



Fonte: Vargas, Carvalho e Vasconcelos (2018)

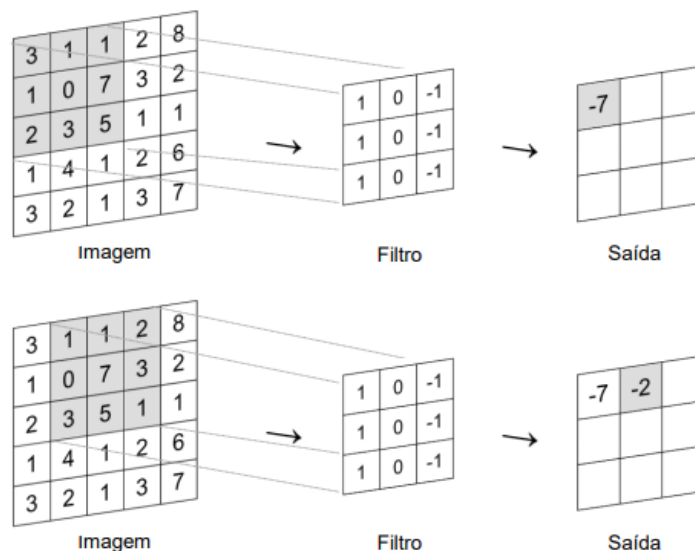
Uma Rede Neural Convolutiva é constituída por uma estrutura de grade, composta em k camadas, onde cada camada é responsável por abstrair mais informações sobre a imagem de entrada (GOODFELLOW; BENGIO; COURVILLE, 2016).

A etapa de convolução possui um fluxo de execução trivial que inicialmente é comum a aplicação de filtros em todas as posições da imagem até o momento em que os filtros estejam todos sobrepostos à imagem. Os filtros são responsáveis pela extração de formas, detecção de bordas, supressão de ruídos entre outras operações.

Uma camada de convolução é composta por n neurônios, cada camada é responsável por aplicar um filtro específico sobre o dado de entrada. Ela está diretamente ligada a um conjunto de pixels da camada anterior. Cada neurônio tem uma conexão com um conjunto específico de pixels, são atribuídos pesos a cada uma das conexões, e a saída das ligações é enviada para a próxima camada (VARGAS; CARVALHO; VASCONCELOS, 2018).

A Figura 17 exemplifica a aplicação de um filtro dentro da camada de convolução, utilizando um filtro de 3x3 em uma imagem de entrada de 5x5.

Figura 17 - Exemplo de aplicação de filtro na camada de convolução

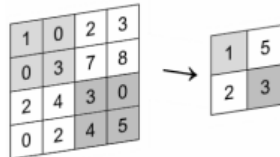


Fonte: Bessa (2021)

A camada de *pooling* é intercalada com a camada de convolução. Na operação de *pooling* é realizado a redução das dimensões do mapa de características gerado pela etapa de convolução. O resultado gerado é uma compactação dos dados para ser enviada para a próxima camada de convolução, operação que influencia no

tempo de processamento (BESSA, 2021). Camada em que ajusta a saída da rede em um resumo estatístico de saídas, tendo a influência de todas as camadas anteriores (GOODFELLOW; BENGIO; COURVILLE, 2016). Figura 18 exemplifica a operação de compactação dos dados de saída na camada de convolução.

Figura 18 - Exemplo de saída de aplicação de filtro

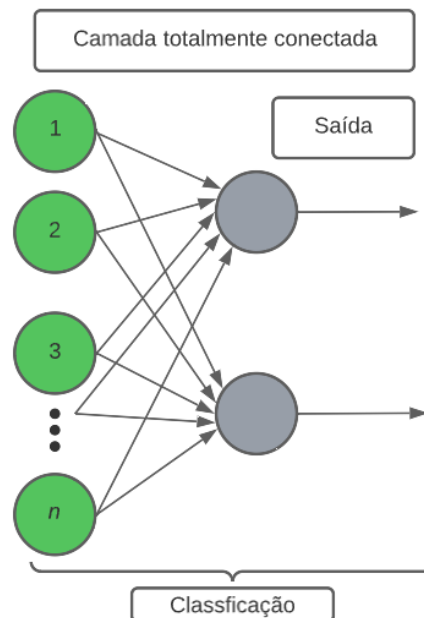


Fonte: Bessa (2021)

Camada totalmente conectada tem o propósito de classificação dos dados, geralmente ela é implementada como sendo a última etapa de execução de uma rede neural. Todas as redes devem possuir pelo menos uma etapa de classificação.

Existem redes como a YOLO que possui mais de uma camada totalmente conectada (REDMON; DIVVALA; GIRSHICK; FARHADI, 2016). Figura 19 exemplifica a execução da etapa totalmente conectada.

Figura 19 - Exemplo de execução da camada totalmente conectada



Fonte: Adaptado de Camolesi (2020)

Na próxima seção vamos explorar um modelo de rede neural convolucional, que utiliza um método de grade para a extração de características e possui nas etapas finais duas camadas de classificação.

3.4 YOLO

É um modelo de detecção de objetos de única passagem como seu próprio nome sugere “Você só olha uma vez”. A metodologia utiliza uma rede convolucional para seu extrator de características. É relativamente simples sua execução tendo 3 etapas principais: (1) extração dos mapas de características, (2) execução única da rede neural convolucional e (3) limitação das detecções através da confiança apresentada pelo modelo. Com apenas uma execução da rede, ela é capaz de gerar várias caixas de predição com suas probabilidades (REDMON; DIVVALA; GIRSHICK; FARHADI, 2016). Figura 21 apresenta o modelo de execução da YOLO.

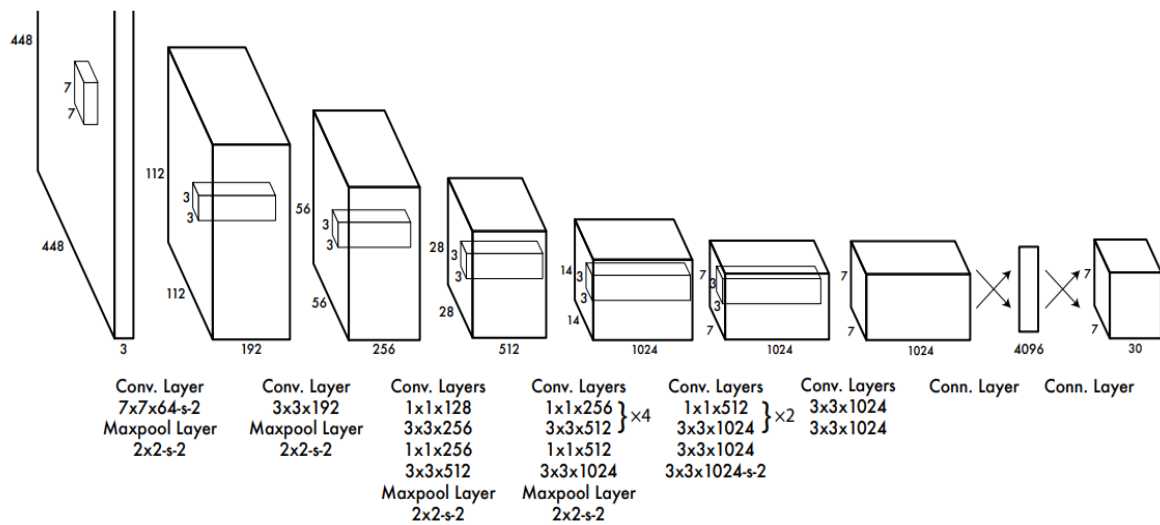
Figura 20 - Exemplo de execução da YOLO



Fonte: Redmon, Divvala, Girshick e Farhadi (2016)

O modelo da YOLO tem como base a GoogleNet, que foi inspirada na rede neural convolucional clássica. A arquitetura da rede é dada por 24 camadas convolucionais, seguidas por 2 camadas totalmente conectadas (REDMON; DIVVALA; GIRSHICK; FARHADI, 2016). Figura 22 apresenta todas as camadas contidas na rede.

Figura 21 - Arquitetura das camadas da YOLO



Fonte: Redmon, Divvala, Girshick e Farhadi (2016)

A YOLO utiliza um método unificado de execução. Durante a detecção a rede, usa para prever as classes contidas na imagem de entrada *bounding box* (caixa delimitadora).

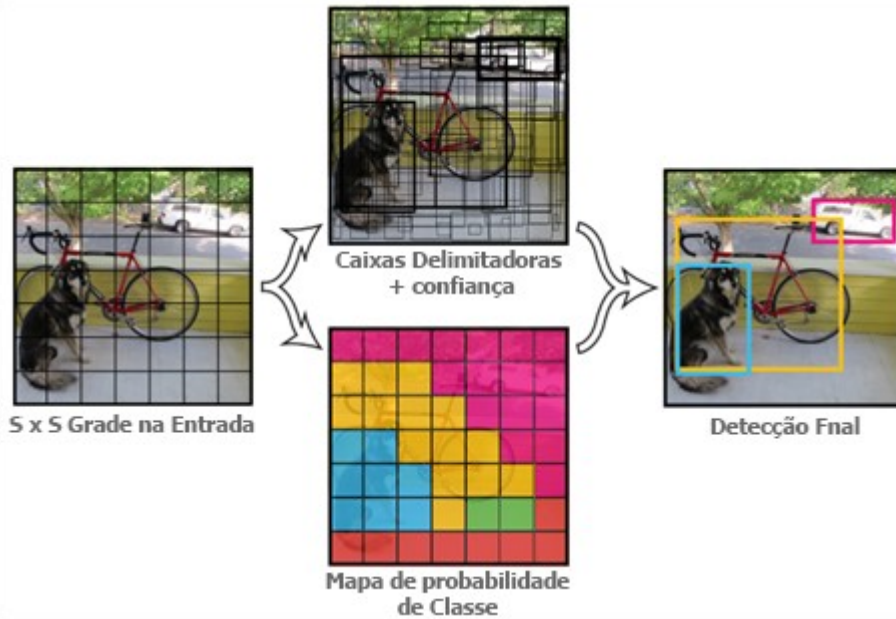
Todas as previsões acontecem simultaneamente, pois é utilizado uma única imagem durante toda a execução dos passos de classificação da rede. Pois são geradas caixas delimitadoras a partir da previsão de cada célula do *grid*. Além de que a YOLO permite o treinamento de ponta a ponta em tempo real, conseguindo manter a maior média do classificador.

O sistema divide a imagem de entrada em uma matriz, apresentado na Tabela 1 como tamanho, que é o responsável por gerar as células de previsão. Cada bloco recebe uma previsão realizada pela YOLO, a partir disso é realizado um procedimento de regressão entre os blocos comparando suas probabilidades. Se o centro do objeto cair em uma das células, ele é responsável por detectar qual é o objeto.

Cada célula tem uma previsão B e um valor de confiança, através de cada célula da caixa delimitadora é realizado o procedimento de regressão para determinar o valor probabilístico da caixa. Cada caixa delimitadora consiste em 5 previsões: x, y, w, h, e o valor de confiança.

Sendo (x, y) as coordenadas que representam o centro da caixa em relação aos limites da matriz. Para w e h, está diretamente relacionado ao valor total da imagem onde w recebe o valor de largura da imagem original e h a altura (REDMON; DIVVALA; GIRSHICK; FARHADI, 2016). Figura 23 apresenta a aplicação dos grids e a caixa delimitadora, durante a classificação de objetos.

Figura 22 - Exemplo de classificação da YOLO



Fonte: Adaptado de Redmon, Divvala, Girshick e Farhadi (2016)

A YOLO atualmente está na sua versão de número 5, contendo 10 modelos pré-treinados, com tamanhos da densidade de camadas de convolução diferentes.

A Tabela 1 apresenta os dados informativos de execução dos modelos, sendo o tamanho dos *grids* gerados na imagem de entrada, tamanho de *flops* utilizados no treinamento que está relacionado a capacidade de processamento, a quantidade parâmetros utilizados, tempo de treinamento e sua média de precisão sobre a média geral (mAP).

A YOLO possui 5 tipos bases de modelos sendo eles: *n*, *s*, *m*, *l* e *x*, que determinam a quantidade de neurônios entre as conexões de cada camada de convolução. A diferença entre os modelos está entre a densidade da camada de convolução e a quantidade de células geradas no grid de classificação (ULTRALYRICS, 2020).

Tabela 1 - Comparativo de modelos operacionais da YOLOv5

Modelo	Tamanho (pixels)	mAP 0.5:0.95	mAP 0.5	Velocidade CPU b1 (ms)	Parâmetros (m)	Flops @640 (b)
YOLOv5n	640	28.0	45.7	45	1.9	4.5
YOLOv5s	640	37.4	56.8	98	7.2	16.5
YOLOv5m	640	45.4	64.1	224	21.2	49.0
YOLOv5l	640	49.0	67.3	430	46.5	109.1
YOLOv5x	640	50.7	68.9	766	86.7	205.7
YOLOv5n6	1280	36.0	54.4	153	3.2	4.6
YOLOv5s6	1280	44.8	63.7	385	12.6	16.8
YOLOv5m6	1280	51.3	69.3	887	35.7	50.0
YOLOv5l6	1280	53.7	71.3	1784	76.8	111.4
YOLOv5x6	1280	55.0	72.7	3136	140.7	209.8
+ TTA	1536	55.8	72.7	-	-	-

Fonte: Adaptado de Ultralytics (2020)

As classes de modelos: *n*, *s*, *m*, *l*, *x*, não possuem mais ou menos camadas. Digamos que é uma rede maior, e que seu caminho é mais denso até chegar ao classificador, onde as etapas de execuções executam multiplicadores de escala que é dado pela largura e profundidade, através destes multiplicadores pode-se refinar seus mapas de características. Por isso os modelos 5n possuem uma precisão menor que qualquer modelo 5x. Os modelos com prefixo 6 possuem uma maior taxa de células geradas, através disso cada parte de um objeto pode estar contido em apenas uma célula diminuindo a chance de múltiplas classes em uma célula.

A plataforma Roboflow de Sonya Alexandrova, Zachary Tatlock e Maya Cakmak (2015) disponibilizou um modelo da YOLO que eles personalizam. Por se tratar de ser uma empresa privada não foi disponibilizado dados sobre como eles construíram seu modelo da YOLO.

3.5 Critérios de avaliação

Durante o treinamento de uma rede neural convolucional é realizada uma análise sobre os dados gerados pelo treinamento, para poder monitorar o comportamento da rede durante o treinamento.

Nesta seção será explorado algumas técnicas que podem ser utilizadas como métodos estatísticos para medir o desempenho de uma rede e se uma rede está sendo capaz de classificar algum objeto ou não. Em casos específicos, utilizando as saídas de cada época do treinamento para que seja possível apontar o percentual de confiança da rede.

Para o cálculo das métricas foram utilizadas as expressões apresentadas por (AKCAY; BRECKON, 2020) que utiliza como variáveis centrais os seguintes dados: verdadeiros positivos (TP), falso positivo (FP), verdadeiros negativos (TN) e falsos negativos (FN). Através dessas variáveis pode ser realizado o cálculo de precisão, taxa de verdadeiro positivo, taxa de falso positivo e média de precisão.

3.5.1 *Precisão*

Sendo referente ao número de amostras com resultados corretamente previstos sobre o número total de previsões. A fórmula da precisão é dada pela representação a seguir.

$$\text{Fórmula de Precisão} \\ \frac{TP + TN}{(TP + TN) + (FP + FN)}$$

3.5.2 *Taxa de verdadeiro positivo*

A taxa de verdadeiros positivos é realizada pela proporção de amostras positivas com previsões corretas, podendo ser representada matematicamente a seguir.

$$\text{Fórmula da taxa de verdadeiro positivo} \\ \frac{TP}{(TP + FN)}$$

3.5.3 *Taxa de falso positivo*

A taxa de falsos positivos é similar ao cálculo de verdadeiros positivos, utilizando a quantidade de falsos positivos previstos, sobre a adição de total de amostras negativas com a quantidade de previsões falsos positivos, a equação pode ser dada pela representação a seguir.

Fórmula da Taxa de Falso Positivo

$$\frac{FP}{(FP + TN)}$$

3.5.4 Média de precisão

O método de média de precisão (mAP) é definido como a média de cada precisão de todas as classes contidas no treinamento. Utiliza-se do somatório das médias de cada categoria (AP) para realizar o produto com o resultado de (1/N) onde N é o número de classes contido no treinamento. A fórmula está representada a seguir.

Fórmula de mAP

$$\left\{ \begin{array}{l} AP = \int_0^1 p(r) dr \\ mAP = \frac{1}{n} \sum_{i=1}^N AP_i \end{array} \right.$$

Além disso, esse método também utiliza interpolação, a partir da intersecção sobre a união, para o cálculo do (AP), assim chamado de mAP[0.5:0.95]. Pode-se utilizar apenas um ponto de interpolação, método denominado como mAP[0.5] (COCO, 2022).

3.5.5 Recall

A equação do *Recall* é definida pelo somatório dos (TP) sobre o somatório dos (TP) mais o somatório dos (FN), para os somatórios é considerado todas as classes contidas no sistema.

Fórmula do Recall.

$$\frac{\Sigma(tp)}{(\Sigma(tp) + \Sigma(fn))}$$

3.6 Ambientes de desenvolvimento

Os experimentos foram realizados no Google Colab e RoboFlow. Cada modelo utilizado já possui o ajuste predefinido para uma detecção geral de objetos que é o foco da rede YOLO, todos os modelos da rede foram retirados na página

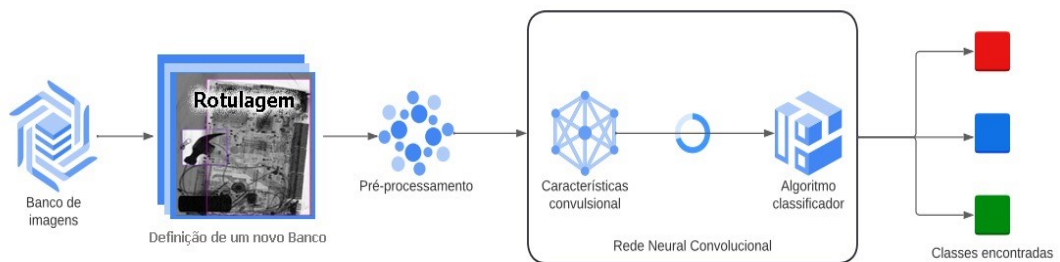
oficial do GitHub (ULTRALYRICS, 2020). O Roboflow não disponibilizou informações sobre o ambiente de treinamento, já no ambiente do Colab foi disponibilizado uma GPU Tesla T4 16GB.

4 DESENVOLVIMENTO

Neste capítulo será apresentada a metodologia proposta para o desenvolvimento deste trabalho. Sendo está baseada em um fluxo de 4 etapas: 1) definição e rotulação de um banco de imagens 2) pré-processamento no banco de imagens, 3) treinamento e validação da Rede Neural Convolutacional e a 4) classificação da imagem de entrada.

A Figura 23 apresenta as etapas da metodologia proposta.

Figura 23 - Exemplo de aplicação da metodologia



Fonte: Autoria própria (2022)

Pode-se observar na Figura 23 as etapas utilizadas para a aplicação da metodologia. Etapa inicial utilizada no desenvolvimento da metodologia, é a seleção dos bancos de imagens e aplicação da nova rotulação, que é essencial para os modelos de experimentos criados explanados na seção 4.1 e 4.2.

A metodologia também utilizou um conjunto de imagens, já rotuladas que contém uma mistura entre as imagens. Também foi criado um novo banco de imagens que foram rotulados pelo autor, as imagens utilizadas foram de dois bancos de imagens citados na seção 3.2.

As imagens utilizadas foram produzidas em ambientes controlados e até mesmo imagens reais geradas durante a fiscalização. Todos os bancos apresentados nessa metodologia foram agrupados em uma lista no GitHub (BHOWMIK, 2022).

Dada uma imagem de entrada, esta passa por uma etapa de pré-processamento, no qual as imagens serão padronizadas e aperfeiçoadas. A seguir, a imagem é enviada para uma rede neural convolutacional para que possam ser extraídas e evidenciadas as informações contidas na imagem de entrada, gerando uma matriz de características que é utilizada na etapa de classificação, assim a rede retorna as classes encontradas na imagem.

Na etapa de classificação dos objetos foi realizada observando dois aspectos. O clássico que foi notado na maioria das metodologias que buscam a detecção de objetos ilícitos em bagagens, elas apresentam a detecção através de cada objeto (classe) como: facas, armas, bombas, entre outros. E o não trivial que busca a classificação utilizando uma área delimitada que pode conter um objeto que possa trazer risco.

O primeiro experimento (chamado de modelo binário) utiliza a perspectiva de saber se a mala poderia ser classificada como *safe* ou *danger*, se limitando apenas às duas classes, sem a preocupação de classificar diretamente qual é o objeto perigoso contido na bagagem. Basta saber que a bagagem é perigosa. O diferencial desse experimento, está diretamente ligado na rotulação do banco de imagem. Pois as classes não são delimitadas diretamente por um objeto e sim toda uma área em que o objeto pode estar contido.

O segundo experimento (chamado de modelo não binário) foi realizado conforme as demais metodologias apresentadas na literatura. Os objetos perigosos representam uma classe e devem ser identificados na mala. Utilizamos esse método para que possamos saber como a YOLO se comporta durante a detecção de cada classe de objeto com as demais redes.

4.1 Experimento 1 - modelo binário

O primeiro modelo treinado foi utilizado apenas duas classes: *safe* e *danger*. Foram utilizados para o treinamento da rede do experimento em questão apenas o modelo *l* da YOLO e um modelo disponibilizado pelo Roboflow de (ALEXANDROVA; TATLOCK; CAKMAK, 2015).

Foi realizada a criação de um novo banco de imagem que foi chamada de COMPIray, que se baseou apenas em uma mistura entre os bancos de imagens.

Foram utilizadas 1.100 imagens aleatórias do COMPASS-XP e 776 do PIDray que foram removidos da faixa *easy*. Durante a seção 3.2 foi detalhado as características principais de cada banco de imagens. A Tabela 2 demonstra como ficou a estruturação desse novo banco de imagens.

Tabela 2 - Geração do novo banco de classe binária

Banco de Imagens	Total de Imagens	Imagens Utilizadas	Classes
COMPASS-XP	1.928	1.100	366
PIDray	47.677	776	12
COMPIray	1.876	-	2
COMPIray (Expandido)	3.752	-	2

Fonte: Autoria própria (2022)

A mistura das imagens dos bancos de imagens utilizados formou um novo banco de imagens. Foram criados rótulos em cada imagem onde cada rotulado sendo denominado como: *safe* ou *danger*. Foram rotuladas manualmente 1.876 imagens e gerando um total de 3.592 rótulos, sendo: 2.219 *danger* e 1.373 *safe*.

Além do procedimento de rotulação, foi aplicado *data argumentation* sobre o novo banco de imagens, pois inicialmente ele ficou com um número pequeno de amostras para ser utilizado no treinamento da rede. A Tabela 2 demonstra os resultados obtidos depois da aplicação da argumentação de dados. Para realizar toda a rotulação e preparar o banco foi utilizado o Roboflow de (ALEXANDROVA; TATLOCK; CAKMAK, 2015) ferramenta indicada pela própria comunidade da YOLO (ULTRALYRICS, 2020).

Foi baseado nas técnicas apresentadas na seção 3.1 para aplicação de *data argumentation*. Foram algumas técnicas e métodos de processamento de imagem aplicadas, a exemplo de: *flip* tanto na horizontal como vertical, rotação utilizando ângulo de 45° no sentido horário e anti-horário (foram aplicados de maneira aleatória entre as imagens) e também *shear* sendo 20° na horizontal e 30° na vertical.

Foi necessário a padronização das imagens, já que as imagens contidas nos bancos disponíveis não possuem um tamanho padrão e as cores entre cada banco de imagens eram muito distintas. Então foi necessário aplicar escala de cinza e redimensionamento das imagens para a resolução de (560x448), o tamanho foi definido com base na média das imagens contidas no COMPIray.

Para realizar os experimentos descritos na seção 5 foi utilizado a técnica de 70 por 30 para divisão entre imagens de treinamento da rede e teste da rede. O próprio Roboflow indica a utilização desse método para divisão de um banco de imagens quando se trata de uma rede YOLO a qual foi utilizada nesta metodologia

(ALEXANDROVA; TATLOCK; CAKMAK, 2015). A Tabela 3 apresenta como ficou a divisão entre as imagens do novo banco.

Tabela 3 - Políticas de divisão de treinamento e validação da rede

Operação	Imagens	Porcentagem
Treinamento	2.626	70%
Validação	750	20%
Testes	376	10%

Fonte: Autoria própria (2022)

4.2 Experimento 2 - modelo não binário

No modelo não binário as classes utilizadas são apresentadas na Tabela 2. Todas as imagens utilizadas nos experimentos são do banco de imagens PIDray.

Foram selecionadas do banco de imagens 29.429 imagens, que contém 12 classes e 39.683 rótulos onde cada rótulo representa uma classe apresentada na Tabela 4. Todos os rótulos já estavam preparados pelos autores que disponibilizaram a base de imagens (WEI *et al.*, 2020).

Tabela 4 - Classes utilizadas no modelo não binário

Classe	Número de Exemplos
Carregador portátil	5.155
Tesouras	4.353
Chave inglesa	4.350
Alicate	4.236
Isqueiro	4.139
Martelo	3.546
Faca	3.290
Spray	2.989
Arma	2.179
Algemas	2.096
Bala	1.837
Cassetete	1.513
Total:	29.429

Fonte: Adaptado de Alexandrova, Tatlock e Cakmak (2015)

Utilizando um limiar de pelo menos 2.200 imagens para definir as imagens com maior índice de representatividade. Pode-se observar na Tabela 4 que a maioria das classes possuem quantidades significativas de exemplos para utilizar no treinamento, exceto as classes: Armas, Algemas, Bala e Cassetetes.

Conforme apresentado na seção 3.4, os modelos utilizados no treinamento foram YOLOv5: *m*, *l*, *x* e o default (disponibilizado pelo Roboflow). A partir do banco de imagens original foi necessário aplicar *data argumentation*, sendo as técnicas de pré-processamento de imagem utilizadas: *flip* horizontal e vertical, rotação de 45° e *shear* de 45° na horizontal e 45° na vertical. Através desses métodos foi gerado um novo banco de imagens com um total de 70.619 imagens.

Também foram aplicados os procedimentos de escala de cinza e redimensionamento para o treinamento, todas as imagens foram ajustadas para a resolução de (440x448). A resolução aplicada foi a média entre o tamanho das imagens contidas no banco de imagens utilizado.

4.3 Rede neural utilizada nos experimentos

Os modelos da rede utilizados nos testes serão YOLOv5: *m*, *l*, *l6* e *x*, conforme apresentado na Tabela 1. A diferença em cada modelo está no tamanho das camadas de convolução. Apenas o modelo *l6* trabalha com um *grid* sobre a imagem maior que os demais modelos, conseguindo chegar a trabalhar com o dobro das células de classificação sobre a imagem, método esse que foi detalhado na seção 3.4.

A plataforma Roboflow disponibilizou um ambiente de treinamento com um modelo da YOLOv5 que foi personalizado por seus pesquisadores. Os ajustes realizados neste modelo não foram divulgados, porém este modelo será utilizado como parâmetro de comparação durante a etapa de treinamento, para evidenciar a eficiência do modelo para os modelos de classificação apresentados.

5 RESULTADOS

Nesta seção serão apresentados os resultados dos experimentos realizados. Serão apresentados dois métodos de testes utilizando banco com imagens similares, porém, em cada banco as imagens foram rotuladas de maneiras diferentes.

5.1 Modelo de rede binário

Os rótulos inseridos na rede, são os objetos classificados no *dataset* como *safe* e *danger*. Sendo eles cada objeto que pode ou não apresentar risco, contido numa imagem de bagagem.

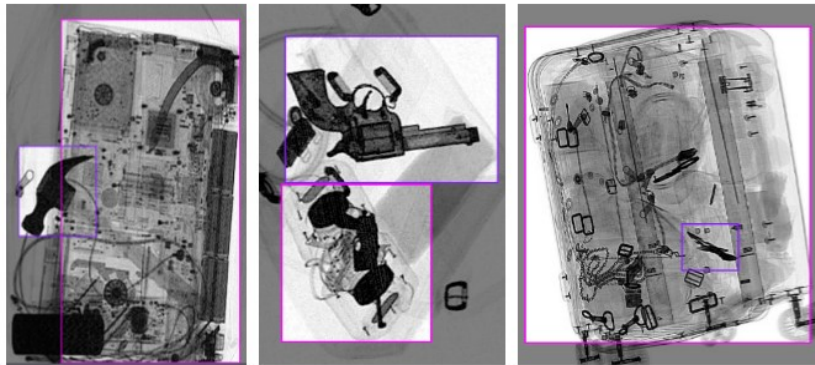
Como a rotulação binária foi mais abrangente sobre a área do objeto das classes *safe* ou *danger*, os delimitadores (rótulos) podem conter mais de um objeto pois na maioria dos casos a rotulação foi realizada não só sobre as bordas do objeto, mas sim pela área em que o objeto está disposto na cena.

Em alguns casos durante a rotulação, ocorre a sobreposição de objetos, mais de um tipo de objeto acaba recebendo a mesma classificação por estar contido na mesma área em que o objeto principal está ocupando. Pois não ocorre a distinção de objetos e sim é levado em consideração a classe *safe* ou *danger* da bagagem. A Figura 24 apresenta um exemplo de imagens de entrada, rotuladas utilizando o Roboflow.

Para os experimentos foram utilizados os modelos da YOLOv5: *m* e *l*, conforme apresentado na Tabela 1. Sendo que a diferença nos dois modelos da rede YOLOv5 está na profundidade de suas camadas de convolução. A Tabela 5 apresenta os experimentos realizados em dois modelos de arquiteturas diferentes e o modelo do Roboflow.

Nas áreas delimitadas apresentadas na Figura 24, verifica-se que as bordas em rosa apresentam objetos da classe *safe* e em roxo pertencem à classe *danger*.

Figura 24 - Imagens de entrada modelo binário



Fonte: Autoria própria (2022)

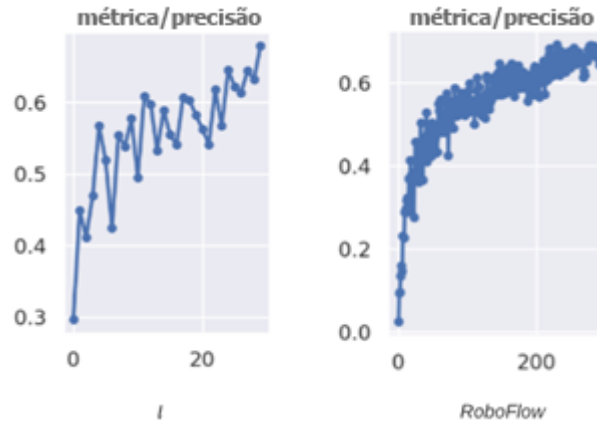
Nos modelos disponibilizados pela comunidade da YOLO apresentados na seção 3.4, foram utilizadas 10 e 30 épocas como testes iniciais, alguns modelos da YOLO haviam registrados valores de precisão já significativos, assim optou-se utilizar 10 e 30 épocas, já o Roboflow utilizou 300 épocas sendo esse número indicado pela comunidade (ULTRALYRICS, 2020). O número de épocas é baseado na análise do gráfico que apresenta a precisão no treinamento, conforme Figura 25.

A análise de um conjunto de testes sobre treinamentos com épocas muito distintas, foi o método utilizado para detectar o que estava ocorrendo com a rede que não estava apresentando melhoria na precisão durante o treinamento, então através desse parâmetro foi detectado uma oscilação. Durante o treinamento utilizando um número diferente de épocas notasse que a oscilação sempre acontece.

Foram utilizadas as métricas apresentadas na seção 3.5 para realizar a análise de comportamento da rede durante o treinamento. Não foi realizado um valor fixo de épocas para cada modelo e através do gráfico de precisão que foi utilizado para explorar o ocorrido. E através das épocas diferentes foi possível observar a oscilação da precisão na Figura 25, que está sendo apresentada na comparação dos modelos: / e Roboflow.

A figura 25 apresenta os dados de saída durante o treinamento do modelo *large* e Roboflow. No eixo x está disponibilizado as épocas de treinamento e eixo y está disponibilizado cada precisão referente à época x.

Figura 25 - Comparação da precisão de treinamento binário



Fonte: Autoria própria (2022)

A Tabela 5 apresenta os resultados gerais considerando as duas classes: *danger* e *safe*.

Tabela 5 - Resultados do treinamento binário

Modelo	Épocas	Métricas de saída do treinamento		
		Precisão	Recall	mAP 0.5
<i>m</i>	10	0,378	0,323	0,268
<i>m</i>	30	0,535	0,568	0,520
<i>l</i>	10	0,618	0,568	0,577
<i>l</i>	30	0,679	0,620	0,595
<i>ROBOFLOW</i>	300	0,662	0,574	0,580

Fonte: Autoria própria (2022)

Nota-se através da Tabela 5 que o modelo utilizado no teste que possui o menor número de conexões entre as camadas, conforme detalhado na seção 3.4. Necessita de um número maior de épocas para que métricas obtenham resultados significativos.

O modelo que possui uma rede mais densa, com poucas épocas de execução, já consegue aprender a classificar objetos com valores significativos. Fica evidente a diferença gerada pela densidade da rede, mostrando uma eficiência maior durante cada época de execução para fazer a predicação de cada classe.

O modelo que apresentou melhores resultados foi o YOLOv5l (*large*) que utilizou um número menor de épocas e obteve resultado de precisão 67% em comparação ao Roboflow que utiliza 10x mais épocas e obteve uma precisão de 66%.

Foi observado através da Tabela 5, que para essa base de imagens o número de épocas necessárias para o treinamento da rede é de 30 épocas.

5.2 Modelo de rede não binário

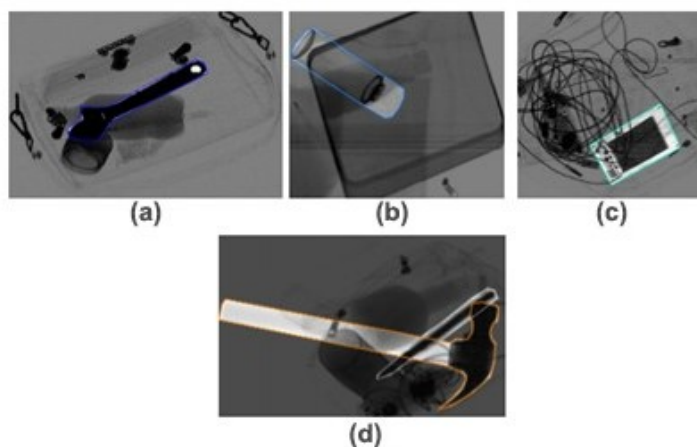
Pode-se observar na Figura 26 alguns dos rótulos utilizados no treinamento e a forma em que foram delimitados. O banco de imagens utilizado PIDray delimita exatamente a área em volta do objeto (bordas).

Os experimentos no modelo não binário, utilizou 4 modelos da YOLOv5: *m*, *l*, *x* e *l6*, todos os modelos utilizaram o mesmo banco de imagem, para que seja possível realizar uma comparação do comportamento de cada modelo da rede durante o treinamento. Pois como apresentado na própria documentação da rede e demonstrado na seção 3.4 a variação de tamanho das camadas totalmente conectadas, como a diferença dos parâmetros e conexões para cada extensão da YOLO (ULTRALYRICS, 2020).

Os experimentos foram realizados com o objetivo de verificar se um modelo mais denso da YOLO poderia comportar-se igual ou melhor a um modelo menos denso para treinamentos considerando número de épocas similares.

Na Figura 26 é possível verificar alguns rótulos inseridos nas imagens da base. A Figura 26 (a) está delimitando uma chave inglesa; em Figura 26 (b) é apresentado o rótulo em um *spray*; na Figura 26 (c) um carregador portátil; e na Figura 26 (d) um alicate e um martelo.

Figura 26 - Imagens de entrada modelo não binário



Fonte: A autoria própria (2022)

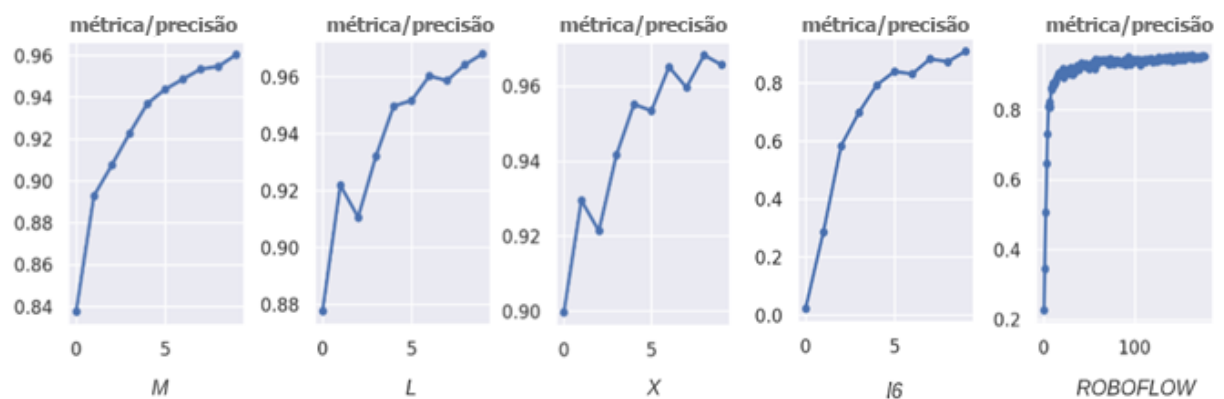
Foram realizados experimentos nos modelos mais densos da YOLO (conforme a Tabela 1), sendo eles: x e $l6$, para verificar os resultados obtidos considerando nos treinamentos um número de épocas igual às utilizadas nas arquiteturas menos conectadas. A Tabela 6 apresenta os melhores resultados de precisão encontrados durante o treinamento de cada modelo.

Através das métricas apresentadas na seção 3.5 foi gerado as previsões estatísticas de performance do modelo. A Figura 27 apresenta o gráfico de precisão de cada modelo treinado.

Observa-se que todos possuem um “salto” nas primeiras épocas e todos demonstram uma linha de aprendizado crescente similar entre eles. Porém, apenas o modelo m , é o único que demonstra um contínuo aprendizado, como pode ser observado na Figura 27 a curva de precisão está sendo crescente.

Os modelos l , x e $l6$ apresentam uma boa precisão, porém pode-se analisar na Figura 27 a curva de precisão que contém algumas variações durante o treinamento. O gráfico do Roboflow demonstra um grande salto em sua linha de aprendizado entre as 15 primeiras épocas de treinamento. Porém as épocas sucessivas apresentam uma linha constante no aprendizado, isso pode demonstrar que para o problema em questão, a rede não necessita de 300 épocas para realizar o treinamento conforme indicado pelo Roboflow e pela própria comunidade da YOLO.

Figura 27 - Comparação da precisão do treinamento não binário



Fonte: Autoria própria (2022)

O modelo treinado que obteve maiores índices de precisão 96,8% com 10 épocas foi o l (*large*), além de ter os maiores índices em todas as outras categorias. Porém todos os modelos treinados mantiveram uma taxa base de 90% de precisão com 10 épocas.

Através métricas utilizadas para poder avaliar a rede, ficou evidente que o Recall e mAP do *l6* retratado na Tabela 6, demonstra que não obteve um bom resultado comparado aos demais modelos da YOLO. Acredita-se que o modelo necessita de maior tempo de treinamento de forma a abstrair melhor o conjunto de características das imagens.

Observando o modelo Roboflow, este utilizou vinte vezes épocas a mais que as demais redes utilizadas nos experimentos, teve um resultado médio como apresentado na Tabela 6. Observa-se que o treinamento foi mais “pesado” e não obteve taxas tão significativos comparados aos modelos originais da YOLO.

Tabela 6 - Resultados do treinamento não binário

Modelo	Épocas	Métricas de saída do treinamento			
		Precisão	Recall	mAP 0.5	mAP 0.5:0.95
<i>m</i>	10	0.960	0.920	0.962	0.717
<i>l</i>	10	0.968	0.932	0.967	0.746
<i>x</i>	10	0.965	0.938	0.970	0.749
<i>l6</i>	10	0.910	0.645	0.730	0.490
<i>ROBOFLOW</i>	200	0.950	0.883	0.930	0.650

Fonte: Autoria própria (2022)

O modelo binário obteve resultados inferiores ao modelo não-binário, deve-se considerar que o modelo binário possui um banco de imagens menor e com um conjunto de rótulos menos preciso, pois apresenta dados sobre diversos objetos misturados em uma única classe e a delimitação de cada classe está sobre um quadrado que contém mais de um objeto.

Assim, cada classe contém apenas um tipo de objetos. Porém cada objeto contido na classe possui algumas características diferentes, a exemplo a classe arma possui: revolver, metralhadora.

Durante o momento da extração de características necessita de um conjunto maior de amostras de determinada classe. Para que seja possível a rede aprender um pouco sobre todos os possíveis objetos que pode conter em uma única classe. Além das possíveis novas formas que a delimitação pode gerar.

A Tabela 7 apresenta a comparação dos melhores resultados dos modelos binário e não binário utilizando os critérios de avaliação apresentados na seção 3.5.

O modelo binário apresentou melhores resultados com 30 épocas e o não binário com apenas 10 épocas. Fica claro que o método não binário teve melhores resultados, isso acontece pois foi dividido em classes de objetos, sendo um modelo mais linear no momento da extração de características.

Tabela 7 - Comparação dos métodos

Método	Modelo	Épocas	Métricas de saída do treinamento			
			Precisão	Recall	mAP 0.5	mAP 0.5:0.95
Binário	/	30	0,679	0,620	0,595	0,279
Não Binário	/	10	0.968	0.932	0.967	0.746

Fonte: Autoria própria (2022)

5.3 Análise sobre a precisão de outras metodologias

Uma breve análise sobre outras metodologias que buscam a classificação de objetos como realizado no experimento não-binário.

Considerando as pesquisas de Boying Wang *et al.* (2021) que utilizou a rede ResNet101 e abordou um treinamento como N classes, a rede obteve uma precisão de 62% na detecção geral de objetos.

Outra metodologia que aborda um treinamento similar foi a de (MIAO *et al.*, 2019) utilizando como core da aplicação uma ResNet34 o melhor resultado obtido em precisão é de 77,2%.

Uma das pesquisas mais recentes encontrada de (LIU *et al.*, 2022) que aborda experimentos utilizando a SIXray que foi um dos bancos de imagens utilizado durante os experimentos dessa metodologia. O autor utilizou uma rede proposta por eles que é chamada de *dual multi-instance attention* que possui uma ResNet para a etapa de aprendizado, e o resultado de precisão apresentado foi de 84%.

A Tabela 8 apresenta uma comparação sobre a análise das precisões obtidas por cada metodologia.

Tabela 8 - Comparação das metodologias

Autor	Método	Métricas de saída do treinamento
		Precisão
Autoria Própria. (2022)	YOLOv5	96%
WANG, B. <i>et al.</i> (2021)	SDANet	62%
MIAO, C. <i>et al.</i> (2019)	ResNet34	77,2%
AKCAY, S. <i>et al.</i> (2018)	YOLOv2	88,5%
KUNDEGORSKI, M. E. <i>et al.</i> (2016)	GoogLeNet	98,4%
WEI, Y. <i>et al.</i> (2020)	FCOS+DOAM	82,4%
LIU, S. <i>et al.</i> (2022)	DMAResNet	84%
ZHOU, C. <i>et al.</i> (2021)	YOLOv4'	88,1%

Fonte: Autoria própria (2022)

A rede apresentada por Mikolaj E. Kundegorski *et al.* (2016) apresenta o resultado um pouco superior, porém foi utilizada a técnica de transferência de aprendizado (*transfer learning*) durante o treinamento que ao invés de ir atualizando os pesos conforme o treinamento, os pesos são congelados a partir do estado original, a rede através dos pesos congelados começa a se ajustar para as novas características encontradas no banco utilizado no treinamento.

6 CONCLUSÃO

Este trabalho apresenta um modelo de rede neural convolucional com cinco (5) variações de modelos que possuem configurações sobre suas camadas de convolução diferentes. Todos sendo capazes de detectar e classificar objetos contidos em imagens de raio-X de bagagens que possam trazer algum risco ao ser humano como: lâminas, armas, bombas, bastão de choque, algemas, martelos, chaves utilitárias, canivetes, *spray*, tesouras, objetos de vidro, pilhas, isqueiro e bala.

O problema de entrada, conforme apresentado durante o desenvolvimento da metodologia, baseia-se nas falhas ocorridas pelos operadores humanos durante a inspeção de bagagens em aeroportos, um processo maçante que exige uma grande atenção. Esta pesquisa tem o intuito de auxiliar durante a inspeção de bagagens trazendo maior agilidade e confiança ao processo, assim podendo melhorar a experiência e segurança dos usuários durante o embarque.

A metodologia proposta realiza experimentos utilizando dois métodos de classificação de bagagens, experimento binário e experimento não-binário. Com base nos resultados dos experimentos, os modelos demonstraram uma boa prospecção através dos gráficos apresentados, e suas detecções, apesar do modelo binário não alcançar um valor de segurança acima dos 90%, alcançou taxa de precisão igual a 67%.

O modelo não binário apresentou bons resultados, alcançando taxas entre 91% e 96% de precisão. Observa-se que a rede YOLO obteve sucesso na detecção de objetos, mesmo casos em que os objetos que podem estar oclusos em bagagens que dificulta a extração de características e classificação. O modelo não binário de rede apresentado, que foi implementado utilizando a Yolov5l obteve o melhor resultado entre as redes utilizadas nos experimentos, conseguindo uma taxa de 96% de precisão. Os modelos implementados que obtiveram piores resultados durante o treinamento binário foram: m com apenas 37% e Roboflow com 66%.

Conclui-se que a YOLO é uma rede que permite uma boa classificação de objetos contidos em bagagens, com taxas de 96% para o modelo não binário. Com isso pode-se verificar que é possível utilizar tal modelo para auxiliar no processo de detecção de objetos ilícitos em bagagens. O modelo apresentado pode ser utilizado em aplicações de segurança, inicialmente como de maneira supervisionada por operadores humanos.

6.1 Trabalho Futuros

Essa pesquisa serve de estímulo para que próximos pesquisadores utilizem da mesma técnica, porém com um conjunto de dados maior, podendo realizar uma expansão do *dataset* de entrada, para que possa verificar se haverá o aumento do valor de precisão dos modelos apresentados para classe binária, e não se limitando aos 67% atingidos com os experimentos realizados.

Outra possível implementação seria aplicar transferência de aprendizado com os modelos já treinados para objetos ilícitos da YOLO, para verificar avanços que a rede pode possuir durante o treinamento.

6.1.1 *Itens de aprimoramento*

- Aprimoramento da rotulação.
- Adicionar exemplos do CLCXray no COMPIray.
- Realizar um novo treinamento no modelo binário com o COMPIray expandido.
- Adicionar mais faixas do banco de imagens PIDray no COMPIray.
- Expansão dos objetos com menor representatividade no PIDray.
- Realizar mais experimentos com um número maior de épocas durante o treinamento.

REFERÊNCIAS

- AirWay. **Como o raio-x dos aeroportos enxergam por dentro das malas.** Disponível em: <<https://www.airway.com.br/como-o-raio-x-dos-aeroportos-enxergam-por-dentro-das-malas/>> Acesso em: 11 de julho de 2021.
- AKCAY, S. et al. Using Deep Convolutional Neural Network Architectures for Object Classification and Detection Within X-Ray Baggage Security Image. In: Conferência IEEE/CVF sobre Transações em Matéria Forense e Segurança da Informação, 2018. **Anais [...]** p. 2203-2215.
- AKCAY, Samet; BRECKON, T.. **Towards automatic threat detection:** A survey of advances of deep learning within X-ray security imaging. arXiv preprint arXiv:2001.01293, 2020.
- ALEXANDROVA, S.; TATLOCK, Z.; CAKMAK, M.. RoboFlow: A flow-based visual programming language for mobile manipulation tasks. In: IEEE/ICRA Conferência Internacional sobre Robótica e Automação. 2015. **Anais [...]** p. 5537-5544.
- ANAC, **Inspeção de segurança em aeroportos.** 2019. Disponível em: <<https://www.anac.gov.br/noticias/2019/inspecao-de-seguranca-em-aeroportos>> Acesso em: 10 de julho de 2021.
- ANAC, **Painel de Indicadores do Transporte Aéreo.** 2018. Disponível em: <<https://www.gov.br/anac/pt-br/assuntos/dados-e-estatisticas/mercado-do-transporte-aereo/painel-de-indicadores-do-transporte-aereo/painel-de-indicadores-do-transporte-aereo-2019>> Acesso em: 10 de julho de 2021.
- ANAC, **Procedimentos de Inspeção no Transporte Aéreo.** 2017. Disponível em: <<https://www.anac.gov.br/noticias/2017/procedimentos-de-inspecao-no-transporte-aereo>> Acesso em: 10 de julho de 2021.
- ARCÚRIO, M. S. F., FORTES, J. A. A. S., ARMBORST, T.. Fatores e erros humanos na inspeção de segurança da aviação civil brasileira. **Revista Psicologia: Organizações e Trabalho**, Florianópolis, v.16, n. 3, p. 259 – 273. Disponível em: <<http://pepsic.bvsalud.org/pdf/rpot/v16n3/v16n3a05.pdf>> Acesso em: 11 de julho de 2021.
- ARCÚRIO, M. S. F.. **Impacto dos fatores e erros humanos no processo de inspeção de segurança da aviação civil brasileira.** 2014. xiii, 318 f., il. Dissertação (Mestrado em Transportes) - Universidade de Brasília, Brasília, 2014.
- BESSA, O. G. W.. **Redes neurais convolucionais e ampliação de dados para detecção de antracnose em folhas de feijoeiro.** 2021. Dissertação (Mestrado em Computação Aplicada) - Universidade Estadual de Ponta Grossa, Ponta Grossa, 2021.
- BHOWMIK, NEEL. **Brief history of X-ray security imaging in Computer Vision.** 2021. Disponível em: <<https://github.com/NeelBhowmik/xray/blob/main/README.md>> Acesso em 07 de agosto de 2022.

CAMOLESI, M. A.. **Treinamento de autômatos celulares para reconhecimento de padrões em imagens de textura**. Universidade Estadual de Campinas, Campinas 2020.

CLARO, M.; et al. **Utilização de Técnicas de Data Augmentation em Imagens: Teoria e Prática**. Sociedade Brasileira de Computação, 2020.

COCO, Dataset. **Metrics**. Disponível em: <<https://cocodataset.org/#detection-eval>> Acesso em: 25 de setembro de 2022.

DE MESQUITA, I. M.; et al. Análise comparativa entre redes neurais convolucionais para classificação de pacientes com Covid-19. **Brazilian Journal of Health Review**, v. 5, n. 2, p. 5165-5180, 2022.

FANK, E. A.; SCHREINER, G. A.; DUARTE, D.. Estudo comparativo de plataformas de Deep Learning: Apache Singa, Graphlab e H2O. In: XVI Escola Regional de Banco de Dados. SBC, 2021 **Anais** [...] p. 11-20.

FEITOZA, M. C.; DA SILVA, W. B.; CALUMBY, R. T.. Exploring deep features and transfer learning for plant species recognition. In: XV Simpósio Brasileiro de Sistemas de Informação. 2019 **Anais** [...] p. 1-8.

GOMES, H. M. A.; et al. **Aplicação das técnicas de processamento digital de imagens na caracterização de carvão ativado da biomassa do coco babaçu**. 2019. Dissertação (Mestrado em Engenharia de Materiais) - Instituto Federal de Educação, Ciência e Tecnologia do Piauí, Teresina, 2019

GONZALEZ, R. C; WOODS, R, E. **Processamento Digital de imagens**: 3. ed. São Paulo: Editora Pearson, 2010.

GOODFELLOW, I.; BENGIO, Y.; COURVILLE, A.. **Deep learning**. Cambridge: MIT press, 2016. Disponível em: <http://www.deeplearningbook.org>. Acesso em: 03 out. 2019.

Griffin, L. D.; Caldwell, M.; Andrews, J.T.; **Computational Security Science Group**, UCL 2019.

KHAN, S. U.; WANG, Y. C.. An image enhancement technique of X-Ray carry-on luggage for detection of contraband/illicit Object (s). **International Journal of Computer Science Issues (IJCSI)**, v.9, n.5, p. 205-211, 2012.

KUNDEGORSKI, M. E. et. al. Transfer learning using convolutional neural networks for object classification within X-ray baggage security imagery. In: Conferência IEEE/CVF sobre Processamento de Imagem (ICIP), 2016. **Anais** [...] p. 1057-1061.

LIU, Shuoyan et al. DMA-Net: Dual multi-instance attention network for X-ray image classification. **IET Image Processing**, v. 16, n. 13, p. 3518-3528, 2022.

LOPES, G. F. V.. **Lean para a redução do tempo de entrega de bagagem no aeroporto Humberto Delgado**, 2018. Disponível em: <<http://hdl.handle.net/10362/96485>> Acesso em: 11 de julho de 2021.

MAGALHÃES, V. C.. **A Receita Federal e aplicação da análise de riscos no controle das cargas aéreas no comércio exterior**: enfoque nos aeroportos. 2020.

- MARTINS, V. E. **Deep Learning Para Classificação De Imagens**. Versão Preliminar de Trabalho de Conclusão de Curso, Universidade Estadual de Londrina, 2018.
- MARVIN, L. M.; SEYMOUR, A.; BOTTOU, L. **Perceptrons**: An Introduction to Computational Geometry. MIT press, 2017.
- MIAO, C. et al. SIXray: A Large-scale Security Inspection X-ray Benchmark for Prohibited Item Discovery in Overlapping Images. In: Conferência IEEE/CVF sobre Visão e Reconhecimento de Padrões, 2019. **Anais [...]** p. 2119-2128.
- OLIVEIRA, M. **Reconhecimento de padrões em biomecânica de corrida usando aprendizado de máquina**. 2018. 65 f. Monografia (Graduação em Engenharia da Computação) - Instituto de Ciências Exatas e Aplicadas, Universidade Federal de Ouro Preto, João Monlevade, 2018.
- QUEIROZ, C. S. 1. Radiações Ionizantes – Exposição dos Trabalhadores do Setor de Inspeção de Bagagens com Raios X em Aeroporto. In: VIII Congresso Vertentes e Desafios da Segurança, 2018 **Anais [...]** p. 119 - 125.
- RAMOS, W. L. **Eficiência operacional nos processos de embarque e desembarque de passageiros e restituição de bagagens em aeroportos brasileiros**. 2016. Dissertação (Mestrado) - Programa de Pós-Graduação em Administração - Centro Universitário Alves Faria, Goiânia, 2016.
- Redmon, J.; Divvala, S.; Girshick, R.; Farhadi, A.. You Only Look Once: Unified, Real-Time Object Detection. In: Conferência IEEE/CVPR sobre Visão Informática e Reconhecimento de Padrões, 2016. **Anais [...]** p. 779-788.
- RIFFO, V.; GODOY, I.; MERY, D.. Handgun detection in single-spectrum multiple X-ray views based on 3d object recognition. **Journal of Nondestructive Evaluation**, v. 38, n. 3, p. 1-11, 2019.
- SCHWANINGER, A. Why do airport security screeners sometimes fail in covert tests? In: 43ª Conferência Internacional IEEE Carnahan sobre Tecnologia de Segurança, 2009. **Anais [...]** p. 41 - 45.
- ULTRALYTICS. **YOLOV5**. Disponível em: <<https://github.com/ultralytics/yolov5>> Acesso em 08 de julho de 2022.
- VARGAS, A. C. G; CARVALHO, A. M. P.; VASCONCELOS, C. N. **Um estudo sobre Redes Neurais Convolucionais e sua aplicação em detecção de pedestres**. Instituto de Computação, Universidade Federal Fluminense, 2018.
- WANG, B. et al. Towards real-world prohibited item detection: A large-scale x-ray benchmark. In: Conferência Internacional IEEE/CVF sobre Visão Informática, 2021. **Anais [...]** p. 5412-5421.
- WEI, Y.; et al. Occluded prohibited items detection: An x-ray security inspection benchmark and de-occlusion attention module. In: 28ª Conferência Internacional ACM sobre Multimídia. 2020. **Anais [...]** p. 138-146.

WU, J.; LIAO, S.. Intelligent Detection of Dangerous Goods in Security Inspection Based on Cascade Cross Stage YOLOv3 Model. **Tehnički vjesnik**, v. 29, n. 3, p. 888-895, 2022.

ZHOU, C.; et al. X-ray Security Inspection Image Detection Algorithm Based on Improved YOLOv4. In: IEEE/ECICE 3^a Conferência da Eurásia sobre IOT, Comunicação e Engenharia. 2021. **Anais** [...] p. 546-550.