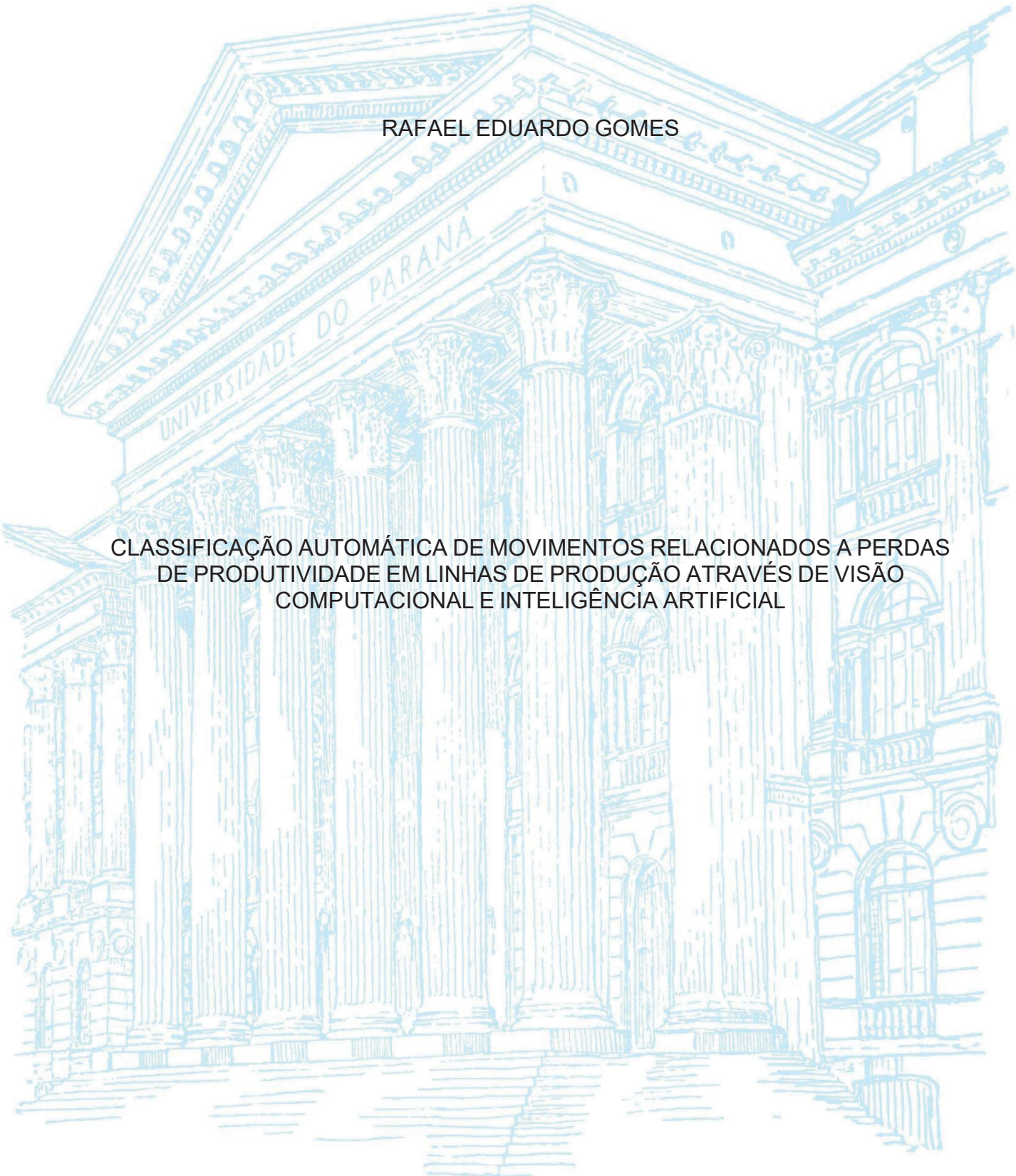
UNIVERSIDADE FEDERAL DO PARANÁ

RAFAEL EDUARDO GOMES



CLASSIFICAÇÃO AUTOMÁTICA DE MOVIMENTOS RELACIONADOS A PERDAS DE PRODUTIVIDADE EM LINHAS DE PRODUÇÃO ATRAVÉS DE VISÃO COMPUTACIONAL E INTELIGÊNCIA ARTIFICIAL

CURITIBA
2022

RAFAEL EDUARDO GOMES

CLASSIFICAÇÃO AUTOMÁTICA DE MOVIMENTOS RELACIONADOS A PERDAS DE PRODUTIVIDADE EM LINHAS DE PRODUÇÃO ATRAVÉS DE VISÃO COMPUTACIONAL E INTELIGÊNCIA ARTIFICIAL

Trabalho de Conclusão de Curso apresentado ao curso de Pós-Graduação em Inteligência Artificial Aplicada, Setor de Educação Profissional e Tecnológica, Universidade Federal do Paraná, como requisito parcial à obtenção do título de Especialista em Inteligência Artificial.

Orientador: Prof. Dr. Razer Anthom Nizer Rojas Montaño

CURITIBA
2022

TERMO DE APROVAÇÃO

Os membros da Banca Examinadora designada pelo Colegiado do Programa de Pós-Graduação INTELIGÊNCIA ARTIFICIAL APLICADA da Universidade Federal do Paraná foram convocados para realizar a arguição da Monografia de Especialização de **RAFAEL EDUARDO GOMES** intitulada: **CLASSIFICACAO AUTOMATICA DE MOVIMENTOS RELACIONADOS A PERDAS DE PRODUTIVIDADE EM LINHAS DE PRODUCAO ATRAVES DE VISAO COMPUTACIONAL E INTELIGENCIA ARTIFICIAL**, que após terem inquirido o aluno e realizada a avaliação do trabalho, são de parecer pela sua APROVAÇÃO no rito de defesa.
A outorga do título de especialista está sujeita à homologação pelo colegiado, ao atendimento de todas as indicações e correções solicitadas pela banca e ao pleno atendimento das demandas regimentais do Programa de Pós-Graduação.

Curitiba, 24 de Novembro de 2022.

RAZER ANTHOM NIZER ROJAS MONTAÑO
Presidente da Banca Examinadora

LUCAS FERRARI DE OLIVEIRA
Avaliador Interno (UNIVERSIDADE FEDERAL DO PARANÁ)

Classificação automática de movimentos relacionados a perdas de produtividade em linhas de produção através de visão computacional e inteligência artificial

Rafael Eduardo Gomes
Especialização em Inteligência Artificial Aplicada
Universidade Federal do Paraná (UFPR)
Curitiba, Brasil
rafaelgomes1@ufpr.br

Razer Anthom Nizer Rojas Montaño
Especialização em Inteligência Artificial Aplicada
Universidade Federal do Paraná (UFPR)
Curitiba, Brasil
razer@ufpr.br

Resumo - A produtividade da mão de obra é um pilar fundamental na gestão industrial. O processo de automatização e digitalização da indústria - Indústria 3.0 e 4.0 - trouxeram avanços em relação ao controle das operações baseados em dados provenientes das máquinas e comandos numéricos, entretanto a obtenção de dados referentes ao trabalho humano segue dependendo de observação e aquisição manual. O propósito deste trabalho foi a comparação da acuracidade de um modelo baseado em poses e movimentos com aquisição de dados por câmeras genéricas e análise com inteligência artificial. O objetivo, identificação de perdas com atividades que não agregam valor (NVAA) em relação à análises realizadas por especialistas com uma assertividade de 90%, foi validado em aplicações industriais reais de diversos segmentos.

Palavras-chave: Lean Manufacturing, Eficiência, Estimativa de Pose, Visão Computacional, Inteligência Artificial

Abstract - Labor productivity is a fundamental pillar in industrial management. The process of automation and digitization of industry - Industry 3.0 and 4.0 - brought advances in relation to the control of operations based on data from machines and numerical commands, however obtaining data related to human work continues to depend on observation and manual acquisition. The purpose of this work was to compare the accuracy of a model based on poses and movements with data acquisition by generic cameras and analysis with artificial intelligence. The objective, identification of losses with activities that do not add value (NVAA) in relation to the analyzes carried out by specialists with an assertiveness of 90%, was validated in real industrial applications of several segments.

Keywords: Lean Manufacturing, Efficiency, Pose Estimation, Computer Vision, Artificial Intelligence

I. INTRODUÇÃO

A busca por eficiência é uma constante nas atividades industriais [13]. Porém, ao contrário das máquinas que são capazes de produzir dados sobre sua produtividade, a mão de obra em um processo industrial demanda de análises criteriosas por engenheiros e técnicos para que seja possível medir, evidenciar e calcular oportunidades de melhoria.

A literatura apresenta 7 (sete) tipos de desperdícios, identificados pela empresa Toyota, que são classificados como qualquer atividade realizada que absorve recursos e não cria nenhum valor (NVAA). Assim, o desperdício pode ser entendido como um custo, pois ele não agrega valor ao produto, tornando-se um gasto desnecessário na elaboração do produto ou serviço [3].

Uma linha de produção enxuta propõe-se a realizar sua produção buscando o conceito de zero desperdícios, com a máxima utilização dos recursos disponíveis. Para isto acontecer é necessário estar continuamente em busca de eliminar os desperdícios e aperfeiçoar o processo constantemente. Contudo, antes é primordial saber identificar os desperdícios da cadeia com eficiência e agilidade.

Para permitir uma varredura mais ágil a respeito da aplicação desta mão de obra em um processo industrial, a proposta deste trabalho é aplicar métodos computacionais de visão computacional para extração de dados, mineração de dados e utilização de inteligência artificial para classificação de dados de acordo com parâmetros pré-determinados.

Como resultado, a proposta é a obtenção de um extrato rápido a respeito da parcela do tempo de cada operador que está sendo empregada na agregação de valor ao produto, e assim direcionar de forma mais assertiva os trabalhos de melhoria da equipe técnica.

A. Justificativa

O processo de análise do trabalho na indústria de transformação baseia-se no acompanhamento das atividades por um analista, segmentando o tempo avaliado

em micro atividades e/ou micro movimentos e classificando-os de acordo com as diretrizes adotadas pela companhia, entre condições de agregação de valor ou não agregação de valor.

Este trabalho demanda uma análise individual e detalhada, seja em borda de linha, seja através de gravações. A morosidade deste processo impede que ele seja feito na frequência, amplitude ou profundidade necessária, inclusive pela dificuldade apresentar-se diretamente proporcional à quantidade de ciclos analisados.

Apresentar um modelo automático para segmentação deste tempo e que forneça dados em tempo real para diagnósticos de perdas representa um ganho substancial no tempo e qualidade para tomada de decisões e avaliação de processos produtivos.

B. Objetivos

O objetivo deste trabalho é o desenvolvimento e validação de um modelo capaz de identificar e mensurar o tempo e duração de movimentos e condições pré-determinadas em relação ao corpo humano. Através de bibliotecas e quadroworks que identificam e decodificam em uma imagem pontos referentes às juntas, selecionar os dados e características necessárias para a classificação da pessoa em relação a condições pré treinadas.

Objetivo geral

Desenvolvimento de modelo de classificação de posições e movimentos relacionados ao corpo humano a ser aplicado em ambientes de produção industrial, capaz de identificar condições classificadas como perdas (*Non Value Added Activities* - NVAA) ou agregadoras de valor (*Value Added Activities* - VA) com uma assertividade de no minimo 90% em relação às referências existentes na documentação da Engenharia Industrial dos processos avaliados.

Objetivos específicos

- Avaliação dos parâmetros de influência (em relação à biblioteca e quadrowork para identificação do corpo humano), considerando a própria câmera, resolução, distância das pessoas, quantidade de pessoas na cena e os tipos de posições encontradas;
- Mineração de dados para tratamento e seleção de dados relevantes, avaliando necessidade de normalização ou inserção de cálculos específicos para incrementar a qualidade da informação a ser carregada nos modelos neurais;
- Determinação e classificação das condições a serem monitoradas, definindo o objeto de análise de classificação baseado na teoria da Engenharia Industrial;
- Desenvolvimento de redes neurais para classificação das condições buscando a melhor acuracidade possível através da exploração dos modelos de redes, quantidades de camadas, funções de ativação e demais parâmetros;
- Teste e validação do modelo baseado em simulação de uma operação real, realizando a filmagem de uma operação e comparando o resultado entregue pelo modelo com a análise manual de um engenheiro industrial em relação às poses, movimentos e tempos.

II. REVISÃO DE LITERATURA

O capítulo buscou identificar na produção científica os principais conceitos para uma correta compreensão sobre o tema da pesquisa e analisar como a visão computacional pode ser utilizada como ferramenta para detecção de MUDA (termo utilizado para expressar perdas relacionadas ao desperdício da aplicação de recursos) em postos de montagem. Para isso, primeiramente, foi definido o conceito do *Lean Manufacturing*, seguido pela fundamentação teórica sobre MUDA, caracterizando os 7 desperdícios, conforme Taiichi Ohno.

Por fim, será apresentada as tecnologias de visão computacional, redes neurais e *deep learning*, contextualizando suas definições, caracterizações, interfaces e interações, em relação à arquitetura de hardware e software. A pergunta a ser respondida pela revisão de literatura é relativa à efetividade da aplicação da visão computacional para detecção automática de movimentos relacionados a perdas de produtividade em postos de montagem.

A. Lean manufacturing

Indústria e Lean Manufacturing

Conceito da Manufatura Enxuta, ou Lean Manufacturing, foi criado pela empresa automotiva japonesa Toyota nos anos 1960, focado na ideia de melhoria da qualidade, redução de custos e aumento de flexibilidade [3]. O modelo apresentou um resultado tão importante que se tornou referência para as empresas ao redor do planeta, sendo adotado em larga escala nos mais diversos segmentos.

A implementação do Lean Manufacturing é desejada pelas companhias para a manutenção ou ganho de competitividade através de melhorias de produtividade e aprimoramento da qualidade [1]. A competição entre as indústrias, ambientes econômicos e escassez de matéria prima são desafios que tornam essencial a implementação de sistemas de produção cada vez mais eficientes.

Os objetivos do modelo criado pela Toyota, como dito, giram em torno da Qualidade, Custo e Entrega, sendo representados como o telhado da casa do sistema, conforme Figura 1. Segundo Womack [4] estes objetivos são suportados pelos dois pilares principais: a) *Just-in-Time*, que preconiza produzir apenas o que for preciso, na quantidade demandada e no momento em que for demandado, sem excessos e sem estoque; b) *Jidoka*, que por sua vez significa “automatização com um toque humano” preconizando que em todos os processos, manuais ou automáticos, as anomalias sejam rapidamente detectadas e sanadas, não permitindo que defeitos sejam passados à frente no ciclo produtivo.

Como base para este conceito, o modelo preza pela estabilidade, sendo esta uma condição básica para a manutenção do sistema. Esta estabilização do processo é suportada pelos conceitos do Heijunka - um sistema de nivelamento da produção, evitando sobrecargas ou ociosidade - o Trabalho Padronizado que consiste em definir as tarefas necessárias para a execução de uma

atividade de produção e dividi-las entre os operadores buscando a garantia de que todos a farão da mesma maneira e Kaizen, o conceito chave da melhoria contínua que propõe uma incessante busca por oportunidades/perdas no processo e suas respectivas soluções.

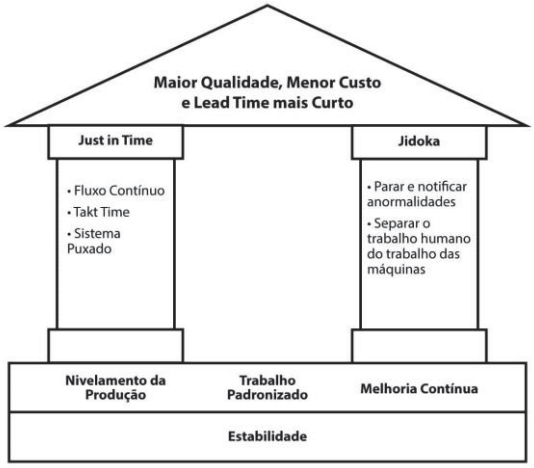


Figura 1 – Diagrama da "casa" do Sistema Toyota de Produção
FONTE: Womack (1990)

Além dos conceitos, diversas ferramentas foram desenvolvidas visando maximizar a utilização dos recursos, reduzir tempo ciclo, lead time e estoques [3]. Seguindo esta linha o Lean Manufacturing visa combater o que foi elencado como os sete desperdícios [4]: Transporte, Estoque, Movimento, Espera, Superprodução, Superprocessamento e Defeitos, buscando maneiras de eliminá-los, recordando que desperdício é qualquer atividade que absorve recursos, mas não cria valor [4].

Análise do Trabalho, MURI, MURA e MUDA

Dentre as ferramentas e métodos preconizados na implementação do sistema Lean, têm-se especial atenção as métricas relacionadas à gestão da mão de obra e os desperdícios embutidos nas atividades dos operadores [3].

Uma das bases do Lean, o "Trabalho Padronizado", é uma receita de atividades estabelecida durante o desenvolvimento do produto e processo. Ela representa todos os movimentos e ações que devem ser realizados para a transformação dos produtos, especificando o local, momento e ferramentas, buscando sempre o menor tempo possível e a menor probabilidade de geração de defeitos.

Em relação aos tipos de desperdícios [3] que também estão relacionados à mão de obra e embutidos ou não no trabalho padronizado, destacam-se os conceitos de MURI, MURA e MUDA (conhecidos como os 3 M's). Conforme a Figura 2, pode-se compreender os termos como: MURI sendo a sobrecarga dos operadores, exigindo que operem em ritmo mais intenso, acelerado, empregando mais força ou esforço, por um período maior de tempo de o que podem suportar, incluindo-se neste ponto os riscos ergonômicos.

MURA é falta de regularidade [13] ou ritmo em uma operação, resultando em um trabalho irregular com os operadores trabalhando com picos de intensidade e depois momentos de espera. Por fim, MUDA são as atividades que consomem recursos (tempo) sem criar valor para o cliente. Estas atividades podem ser inerentes ao ciclo de trabalho, como múltiplas movimentações, carregar peças, posicionar peças e até mesmo o caminhar ao longo da estação de trabalho, como também desperdícios de tempo em ações não contempladas no ciclo, como retrabalhos ou tempos ociosos de espera.

Figura 2 – Perdas MURI, MURA e MUDA
FONTE: Lean.org (2008)

Shingo [13] listou estas atividades contempladas por MUDA como NVAA (*Non Value Added Activities*), ou seja, atividades que não agregam valor ao produto final, sendo irrelevantes para o cliente, porém, consumidoras de recursos das empresas. Ao contrário, as atividades que agregam valor, que estão ligadas diretamente à transformação física dos materiais, são chamadas de VA (*Value Added Actitivies*).

B. VISÃO COMPUTACIONAL

O recente progresso das tecnologias científicas está produzindo novas técnicas que direcionam o mundo para era da inteligência artificial. Com a capacidade de perceber e compreender o mundo visual, com mais acurácia do que o olho humano, as novas técnicas de inteligência artificial, visão computacional e aprendizagem de máquinas, são mais eficientes em uma série de tarefas específicas. Essa competência é possível, principalmente, por sistemas que incluem algoritmos e suas implementações de hardware, permitindo ao usuário ensinar a máquina a entender o mundo físico a partir da visão. As entradas da visão computacional, em específico, buscam habilitar o sistema a ver, identificar e compreender o mundo visual de forma autônoma, simulando da mesma forma que a visão humana faz [5].

Em relação à visão computacional, segundo Gonzales *et. al*. [6], uma imagem pode ser definida como uma função bidimensional onde x e y são coordenadas em um plano espacial e a amplitude da função representa a intensidade dos níveis de cinza deste ponto determinado. Quando os valores de x, y e intensidade são finitos e discretos, podemos chamar esta imagem de imagem digital.

No espectro do processamento computacional destas imagens consideram-se 3 tipos de processos: Baixo; Médio e Alto Nível [5]. Processos de baixo nível envolvem operações básicas como pré-processamento para redução de ruído, melhoramento de contraste e nitidez. No nível médio, diferente do baixo nível onde as saídas são imagens, o processamento resulta em tarefas como a segmentação e extração de características, descrição e redução (particionamento de uma imagem em regiões ou objetos

específicos) em formas de atributos, como contornos e bordas. Processamento de alto nível engloba tarefas relacionadas à identificação e reconhecimento de objetos, sendo parte desta etapa análise de imagens e funções cognitivas da visão computacional [6].

Ainda conforme Gonzales *et. al*. [6], esta divisão pode ser vista e aprofundada observando a Figura 3, considerando seu início na aquisição da imagem. A base de conhecimento da visão computacional e tratamento de imagens digitais é circundada por diversas tarefas, descritas a seguir.

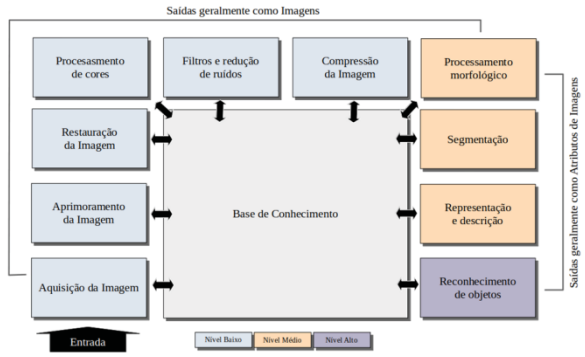


Figura 3 - Passos fundamentais de um processamento digital
Fonte: Adaptado de Gonzales et. al. (2008)

O aprimoramento da imagem é o processo de manipulação da imagem buscando uma melhoria do aspecto desta em relação a sua apresentação original.

A restauração da imagem é a área do processamento que além de melhorar aspectos da imagem utiliza técnicas matemáticas e probabilísticas em partes degradadas.

O processamento de cores, sendo estas importantes descritores que simplificam a análise da imagem e extração de características. Esta área do processamento trabalha com a manipulação dos modelos de cores (RGB, CMYK, HSI) assim como a interação destas para escalas de cinza.

Os filtros e redução de ruídos são tarefa para executar limpeza de ruídos com manutenção da nitidez das bordas da imagem.

A compressão tem a função de transformar a reduzir o espaço necessário para o armazenamento das imagens.

O processamento morfológico lida com ferramentas para extrair componentes da imagem que são úteis na representação e descrição da forma.

A segmentação particiona uma imagem em suas partes ou objetos constituintes. Em geral, quanto mais precisa a segmentação, maior a probabilidade de o processo de reconhecimento obter sucesso.

A representação e descrição trabalha para representar este agregado resultante de pixels segmentados em uma forma adequada para processamento posterior. Basicamente, representar uma região envolve duas escolhas: representar a região em termos de suas características externas, seu limite, ou representá-la em termos de suas características internas, os pixels que compõem a região.

O reconhecimento de objetos é o processo que atribui um rótulo a um objeto com base em seus descritores, por exemplo, o reconhecimento de uma xícara, rotulada como tal.

Os componentes básicos de um sistema de visão computacional e tratamento digital de imagens são demonstrados na Figura 4.

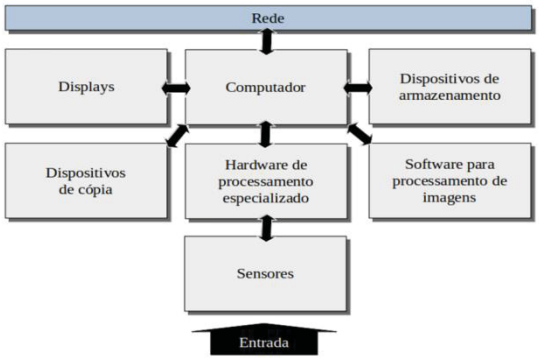


Figura 4 - Componentes de um sistema de processamento de imagens
Fonte: Adaptado de Gonzales et. al. (2008)

O primeiro componente, o Sensor, tem a função de adquirir/detectar as imagens e dois elementos são necessários para as imagens digitais. O primeiro é um dispositivo físico sensível à energia irradiada pelo objeto. O segundo, digitalizador, é um dispositivo para converter a saída do dispositivo de detecção física em forma digital. Este digitalizador representa o hardware de processamento de imagem especializado, porém este componente executa outras operações primitivas, por exemplo, executando operações aritméticas e lógicas em paralelo em imagens inteiras para fazer a média das imagens tão rapidamente quanto elas são digitalizadas, para fins de redução de ruído.

O computador em um sistema de processamento de imagem é um computador de uso geral e pode variar de um PC simples a um supercomputador. Em aplicativos dedicados, às vezes, computadores personalizados são usados para atingir um nível necessário de desempenho.

O software para processamento de imagens consiste em módulos especializados que realizam tarefas específicas. Pacotes de software mais sofisticados permitem a integração desses módulos e comandos de software de uso geral a partir de pelo menos uma linguagem de computador.

Capacidade de armazenamento em massa é fundamental em aplicativos de processamento de imagem. Dados e processamento digital de imagens geralmente demandam uma grande capacidade de armazenamento, que podem ser enquadrados em três categorias principais: Armazenamento de curto prazo para uso durante o processamento, armazenamento on-line para recuperação relativamente rápida e armazenamento de arquivamento, caracterizado por acesso infrequente.

Os visores de imagem em uso hoje são principalmente monitores de TV em cores, porém também existem diversas outras interfaces, por exemplo, óculos de realidade aumentada ou tablets. Os monitores são acionados pelas saídas de placas de exibição de imagens e gráficos que são parte integrante do sistema do computador.

Dispositivos de cópia impressa para gravar imagens incluem impressoras a laser, câmeras de filme, dispositivos sensíveis ao calor, unidades de jato de tinta e unidades digitais, como discos ópticos e CD-ROM.

Por fim, a rede é uma componente integradora do sistema computacional, provendo a interligação de todos estes sistemas.

C. *Estimativa de pose*

Determinar a localização espacial das articulações do corpo de uma pessoa a partir de uma determinada imagem é a função dos modelos de estimativa de pose [7].

Segundo Papandreou [9], o processo de estimativa de pose divide-se em duas etapas principais: a) localização das articulações e/ou pontos-chave do corpo humano e; ii) agrupamento destas articulações em configurações válidas de pose humana. Na primeira etapa, o foco principal é encontrar a localização de cada ponto-chave dos seres humanos, como, por exemplo, ombros, braços, mãos, joelhos e tornozelos. A segunda etapa é caracterizada pelo agrupamento destes pontos em uma pose humana válida, transformando estes pontos em um "esqueleto".

A identificação humana tem como base diferentes bases de conjuntos de imagens, como COCO e MPII. Existem duas abordagens distintas em relação à avaliação das imagens para identificação da(s) pose(s) [7], a abordagem Descendente (de cima para baixo) e a Ascendente (de baixo para cima).

Na abordagem Descendente, o processamento é realizado partindo de uma versão de baixa resolução para alta resolução, com a identificação do corpo humano primeiramente através de uma *bounding box* e a partir desta etapa, seguindo com a determinação da pose, conforme Figura 5.

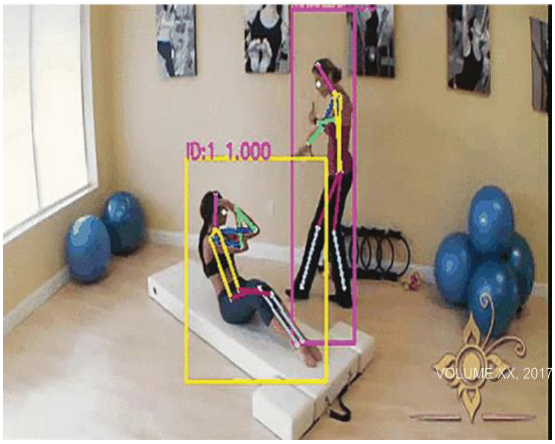


Figura 5 - Processamento com abordagem descendente
Fonte: Munea et.al (2020)

No modelo Ascendente, ao inverso do exposto anteriormente, o processamento inicia-se com versões em alta resolução, migrando para baixa resolução durante o processo. Ele começa localizando entidades semânticas sem identidade e, em seguida, agrupando-as em uma instância de pessoa. Seu resultado é exposto diretamente na identificação dos pontos e ligações, conforme figura 6.

Figura 6 - Processamento com abordagem ascendente
Fonte: Munea et.al (2020)

Outro ponto fundamental destacado por Munea [7] são os componentes principais de um modelo de estimativa de pose:

- Arquitetura *Backbone*: São arquiteturas de rede deep learning utilizadas na identificação do corpo humano e segmentação em pontos, utilizando camadas de convolução, como a AlexNet, VGG e ResNet;
- Funções de Perda: Como parte do aprendizado de máquina, os modelos de estimativa de pose aprendem por funções de perda. As funções de perda avaliam a qualidade com que um algoritmo específico modela o conjunto de dados fornecido;
- Base de Dados: Bases de imagens são fundamentais para o treinamento dos modelos. Entre estas bases, destacam-se FLIC, LSP, MPII Human Pose e COCO. Exemplificando o MS-COCO ou COCO (Common Objects in Context) é um produto da Microsoft (MS) que consiste em uma coleção de um conjunto de dados muito grande com anotação de detecção de objeto, detecção de ponto-chave, segmentação de material, segmentação panorâmica e legenda de imagem, totalizando 200.000 imagens com 250.000 pessoas contidas, cada uma delas identificadas e rotuladas com 17 pontos chave;
- Métricas de avaliação comuns: Os diferentes modelos de estimativa de pose utilizam-se de avaliadores para comparar e contrastar seu desempenho. Dentre as métricas elencadas têm-se:

  i) Porcentagem de partes corretas (PCP), medindo a quantidade de membros do corpo encontradas;

  ii) Porcentagem de Juntas Detectadas (PDJ), definindo uma articulação detectada corretamente se a distância entre a localização da articulação prevista e a localização da articulação verdadeira está dentro de uma certa fração do diâmetro do tronco;

  iii) Porcentagem de pontos-chave corretos (PCK), que também mede a distância entre a localização prevista da junta e a verdadeira localização da junta. A métrica de avaliação PCK mede a precisão da localização das articulações do corpo;

  iv) A área sob a curva (AUC) medindo os diferentes limites de intervalo de PCK e v) Object Keypoint

Similarity (OKS) fornecendo uma medida de como um ponto-chave predito se aproxima da verdade básica.

Um dos modelos de estimativa de pose disponíveis e baseado no *quadrowork* TensorFlow, em colaboração com o Google Creative Lab, é o PoseNet. Este modelo pode ser usado para estimar uma única pose ou várias poses. Para exemplificar o conceito da estima de pose conceituado, Oved [8] expõe como o PoseNet executa os passos: a) imagem RGB de entrada é alimentada por uma rede neural convolucional e b) algoritmo de decodificação de pose única ou multi-pose é usado para decodificar poses com suas pontuações de confiança e posições de pontos-chaves com suas pontuações de confiança.

Nesta etapa de decodificação, tem-se a imagem da pose humana, onde o PoseNet retornará um objeto que contém uma lista de pontos-chaves e uma pontuação de confiança em nível de instância para cada pessoa detectada, conforme Figura 7.

Figura 7 - Detecção de pessoas e níveis de confiança
FONTE: Posenet - TensorFlow (2018)

Os Pontos-Chaves são caracterizados por uma parte da pose de uma pessoa que foi estimada, como nariz, orelha direita, joelho esquerdo, pé direito etc. O PoseNet detecta atualmente 17 pontos-chave, conforme Figura 8.

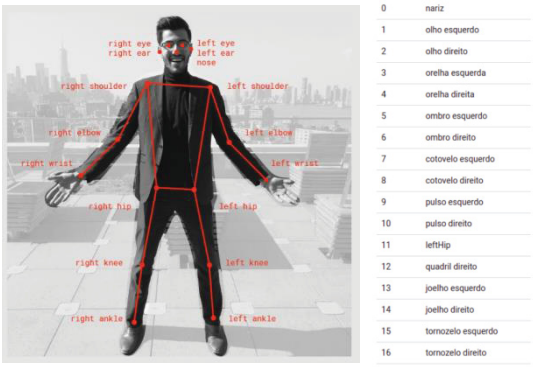


Figura 8 - Pontos chaves de detecção – Posenet
FONTE: Posenet - TensorFlow (2018)

Cada Ponto-Chave possui um valor de confiança de quanto uma localização estimada do ponto-chave é precisa, além de uma posição, que são as coordenadas 2D (X e Y) na imagem de entrada original onde um ponto-chave foi detectado.

D. Redes neurais e deep learning

Deep Learning é um sub-campo do Aprendizado de Máquinas e consequentemente, da Inteligência Artificial [12]. Pode-se posicioná-la, em uma escala, um degrau acima das Redes Neurais conforme a Figura 9. De forma básica, o que diferencia a *Deep Learning* de uma simples rede é a sua profundidade, a quantidade de camadas, trazendo como característica a capacidade de aprendizado.

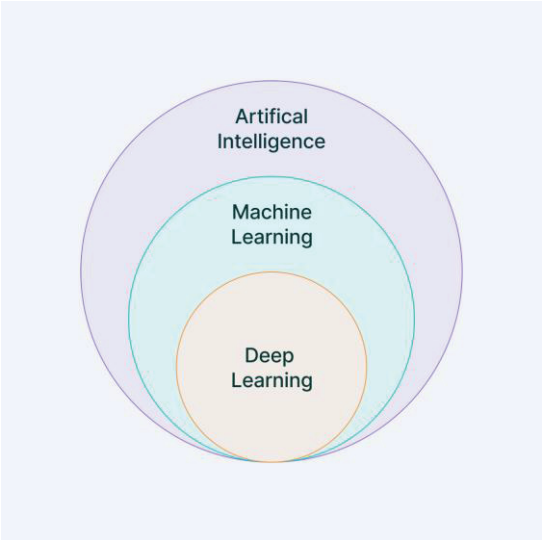


Figura 9 - Representação da hierarquia dos campos da I.A.
FONTE: Autor (2022)

A história do *Deep Learning* teve seu início ainda na década de 1940, quando uma rede computacional inspirada em redes neurais biológicas foi criada por Walter Pitts e Warren McCulloch. Neurônios biológicos consistem em diversas entidades, entretanto, pode-se comparar as redes artificiais e humanas conforme:

- Sinapse: Recepção dos sinais de entrada
- Dendritos: Definição/atribuição dos pesos
- Corpo Celular: Somas e integração
- Axônio: Transporte do sinal
- Terminal axônico: Resultado de saída

Na Figura 10 demonstra-se uma arquitetura básica de como as redes neurais profundas são, na verdade, redes neurais artificiais com diversas camadas. Cada camada é responsável por extrair alguma informação, representada por resultados e pesos e enviá-la para as próximas camadas. A camada de entrada (*Input*) se encarrega de coletar os dados, as camadas ocultas se encarregam de guardar os pesos atribuídos a elas e a camada de saída (*Output*) produz os resultados.

A partir deste evento, a *Deep Learning* vem evoluindo de forma constante, com apenas 2 grandes eventos em seu desenvolvimento. Henry Kelley recebeu créditos em 1960 pelo desenvolvimento da métrica de um modelo contínuo de *Back Propagation*, basicamente um modelo que utilizar perdas de saída para recalcular pesos para o propósito de

treinamento. Entretanto, o conceito ainda era ineficiente e não foi útil até 1985.

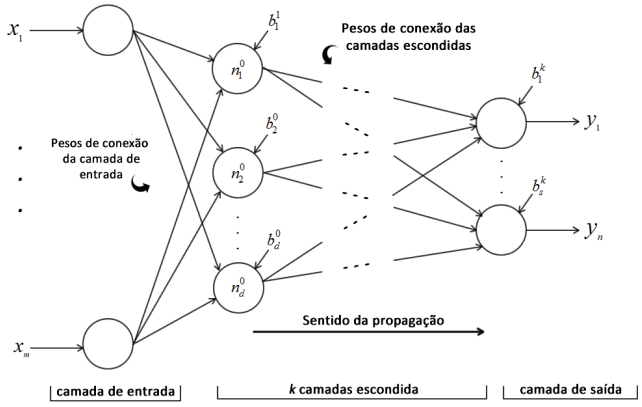


Figura 10 - Representação básica de uma rede neural
FONTE: Autor (2022)

Posteriormente, um grande passo foi dado em 1979 com o desenvolvimento das primeiras Redes Neurais Convolucionais (CNN) por Kunihiko Fukushima, batizada de Neocognitron. Este modelo era composto de múltiplas camadas de pooling e convoluções, um sistema que consiste na aplicação de filtros repetidamente para extrair em características de imagens em diversos níveis de profundidade. Estas redes são capazes de aprender a reconhecer padrões visuais.

A partir de 1985 o uso de *Back Propagation* voltou a tornar-se interessante, principalmente em 1989 quando Yann LeCun, combinando-a com redes convolucionais, demonstrou um caso prático de reconhecimento de dígitos manuscritos no centro Bell Labs com grande sucesso.

A década de 1990 apesar de um período conturbado para Inteligência Artificial devido a criação de grandes expectativas, foi o período onde outros dois importantes atores foram desenvolvidos [11], a *Support Vector Machine* (SVM) no mapeamento e reconhecimento de dados similares e a *Long Short-term Memory* (LSTM) para redes neurais recorrentes. A partir deste momento, os grandes avanços foram observados de forma ostensiva no ganho de potencial de processamento dos computadores, com um incremento de praticamente 1000 vezes na velocidade em 10 anos, inclusive com o ganho de protagonismo das placas gráficas (GPU) para processamento de dados.

Em 2012, Google Brain demonstrou os resultados de um experimento chamado “The Cat Experiment”, fundado por Andrew Ng, onde modelos de Deep Learning em mais de 1.000 computadores e 10 milhões de imagens retiradas aleatoriamente do YouTube, usaram aprendizado não supervisionado (quando não são fornecidas as “respostas” para o treinamento de redes) para gerar em sua última camada uma resposta fortemente positiva para reconhecer imagens de gatos.

Para auxiliar no desenvolvimento de aplicações baseadas em Deep Learning existem diversas bibliotecas estruturadas, como Theano para expressões matemáticas e matrizes, Keras - TensorFlow com vasta amplitude de modelos de redes neurais, assim como outras bibliotecas como Darknet , Torch e Caffe com algoritmos voltados a *Machine Learning* e *Deep Learning* baseadas tanto em C quanto Python e com suporte para rodar tanto em CPU quanto GPU [11].

Segundo Bolhasani [10], entre as principais aplicações do Deep Learning, pode-se elencar:

- Visão Computacional cuja demanda de processamento de entrada (pixels) é surpreendentemente grande;
- Reconhecimento de Fala, onde há a necessidade de compreensão e tradução de sinal acústico em palavras;
- Processamento de Linguagem Natural (NLP), com redes capazes de traduzir, ler, completar e até mesmo escrever textos de forma autônoma.
- Além de diversas outras aplicações onde grandeza de escala e aprendizado não supervisionado são características demandadas.

III. MATERIAL E MÉTODOS

O objetivo do método proposto é a identificação e segregação do tempo onde há agregação de valor em uma atividade manual dentro de uma linha ou célula de produção. Esta identificação deve ocorrer de forma autônoma, independente do tipo de produto ou processo, com o mínimo possível de intervenção (setup) e ser capaz de replicar o resultado da medição realizada pelo Engenheiro de Produção com métodos alternativos empregados atualmente, por exemplo, através de fichas de cronoanálise ou análise posterior por vídeo.

A. Hardware e software

Para o desenvolvimento do algoritmo e redes neurais, foi utilizado um computador equipado com processador Ryzen 1700X de 3.400Ghz e 8 núcleos e uma GPU Nvidia GTX1060 com memória de 6GB.

Os vídeos de entrada foram gravados a partir de uma câmera GoPro Hero4, com resolução de entrada de 1080p (1920 x 1080), 60 FPS e ângulo de abertura de 120 graus.

Estas gravações foram processadas e tiveram a resolução alterada para 848x476 e 5FPS para a entrada no modelo de Estimativa de Pose PoseNet. O modelo utilizou a arquitetura Resnet50 para o processamento das redes de classificação das poses humanas.

Todo o sistema foi modelado na linguagem Python utilizando a interface PyCharm e rodando em sistema operacional Ubuntu, versão 20.4.

B. Arquitetura básica do modelo

O modelo seguiu uma arquitetura conforme o diagrama da Figura 12. A base é fundamentada na especificação e posicionamento da câmera, no algoritmo de extração de pose, na definição das situações e condições de perda (NVAA e VA). A partir da obtenção de imagens via câmera, os pontos-chaves são extraídos com o PoseNet e processados de acordo com as demandas (coordenadas de posições, ângulos e distâncias) conforme exemplo na Figura 11.

Os dados extraídos para um arquivo CSV são processados novamente para regularização e divididos para a criação da base de treino e teste. Esta base é rotulada para um treinamento supervisionado, com valores baseados nas tabelas I e II. O treinamento ocorre junto ao desenvolvimento das próprias redes neurais para cada condição.

As redes treinadas são destinadas ao próximo passo, que é a classificação de cada pose em cada momento para extrair tanto a condição para aquele momento quanto o período e duração de cada uma.

O resultado do modelo é o arquivo CSV com a identificação da situação do operador a cada segundo, dividida entre VA e NVAA.

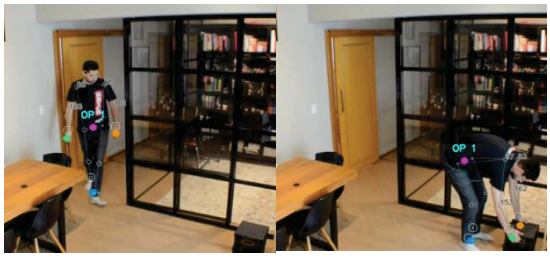


Figura 11 - Exemplo da saída do modelo de estimativa de pose
FONTE: Autor (2022)

C. Determinação das condições alvo

De acordo com estudos de trabalho em engenharia industrial, decidiu-se no trabalho segmentar cada pose em 9 condições:

1) 5 Posições de braço direito;
2) 5 Posições de braço esquerdo;
3) 2 posições de tronco;
4) 2 posições de pernas;
5) 2 condições de movimento de braço direito;
6) 2 condições de movimento de braço esquerdo
7) 2 condições de movimento de tronco;
8) 2 condições de movimento de pernas;
9) 1 condição de deslocamento;

Estas condições são classificadas de acordo com as tabelas I e II.

Com base nas condições apresentadas, situações especificas devem ser interpretadas, conforme a Tabela III. Estas situações, juntamente com o posicionamento do operador na sua estação de trabalho, resultam no estabelecimento da condição de agregação de valor ou não em uma atividade industrial.

A Tabela III, para a aplicação final, é de grande relevância, pois a partir dela, por exemplo, poderá ser diferenciado o momento em que um operador esteja apenas caminhando de um momento em que ele esteja carregando algo, visto que teremos um cruzamento dos dados das pernas (movimentando em deslocamento) e dos braços, que podem estar em balanço para baixo, como no caminhar, ou com cotovelos angulados e mãos à frente, como quando se segura algo.

TABELA I
CONDIÇÕES DE POSIÇÃO

Membro	Condição de posição	Descrição
Braço Direito	0	Braço relaxado para baixo
	1	Antebraço dobrado com cotovelo rente ao corpo
	2	Antebraço dobrado com cotovelo afastado do corpo
	3	Braços esticados à frente
	4	Mãos acima da cabeça com cotovelo abaixo da linha do ombro
	5	Mãos e cotovelos acima da cabeça
Braço Esquerdo	0	Braço relaxado para baixo
	1	Antebraço dobrado com cotovelo rente ao corpo
	2	Antebraço dobrado com cotovelo afastado do corpo
	3	Braços esticados à frente
	4	Mãos acima da cabeça com cotovelo abaixo da linha do ombro
	5	Mãos e cotovelos acima da cabeça
Tronco	0	Tronco ereto
	1	Tronco curvado
Pernas	0	Pernas esticadas
	1	Pernas dobradas (abaixado)

FONTE: Autor (2022)

TABELA II
CONDIÇÕES DE MOVIMENTO

Membro	Condição de movimento	Descrição
Braço Direito	0	Braço estático
	1	Braço em movimento
Braço Esquerdo	0	Braço estático
	1	Braço em movimento
Tronco	0	Tronco estático
	1	Tronco em movimento
Pernas	0	Pernas estáticas
	1	Pernas em movimento
Deslocamento	0	Pessoa parada
	1	Pessoa caminhando

FONTE: Autor (2022)

TABELA III
SITUAÇÕES DO OPERADOR

Situação	
Walking	Caminhando com braços livres
Carrying	Caminhando carregando
Crouching	Agaixando
Crouch Arm Mov Up	Agaixado com braços em movimento sobre cabeça
Crouch Arm Mov Down	Agaixado com braços em movimento abaixo cabeça
Crouch Arm Mov Front	Agaixado com braços em movimento esticados a frente
Crouch Arm stop Up	Agaixado com braços estáticos sobre cabeça
Crouch Arm stop Down	Agaixado com braços estáticos abaixo cabeça
Crouch Arm stop Front	Agaixado com braços estáticos esticados a frente
Bending	Curvando tronco
Bend Arm Mov	Tronco curvado com braços em movimento esticados
Bend Arm Stop	Tronco curvado com braços estáticos esticados
Stand Arm Mov Up	Em pé com braços em movimento sobre cabeça
Stand Arm Mov Down	Em pé com braços em movimento abaixo cabeça
Stand Arm Mov Front	Em pé com braços em movimento esticados a frente
Stand Arm stop Up	Em pé com braços estáticos sobre cabeça
Stand Arm stop Down	Em pé com braços estáticos abaixo cabeça
Stand Arm stop Front	Em pé com braços estáticos esticados a frente

FONTE: Autor (2022)

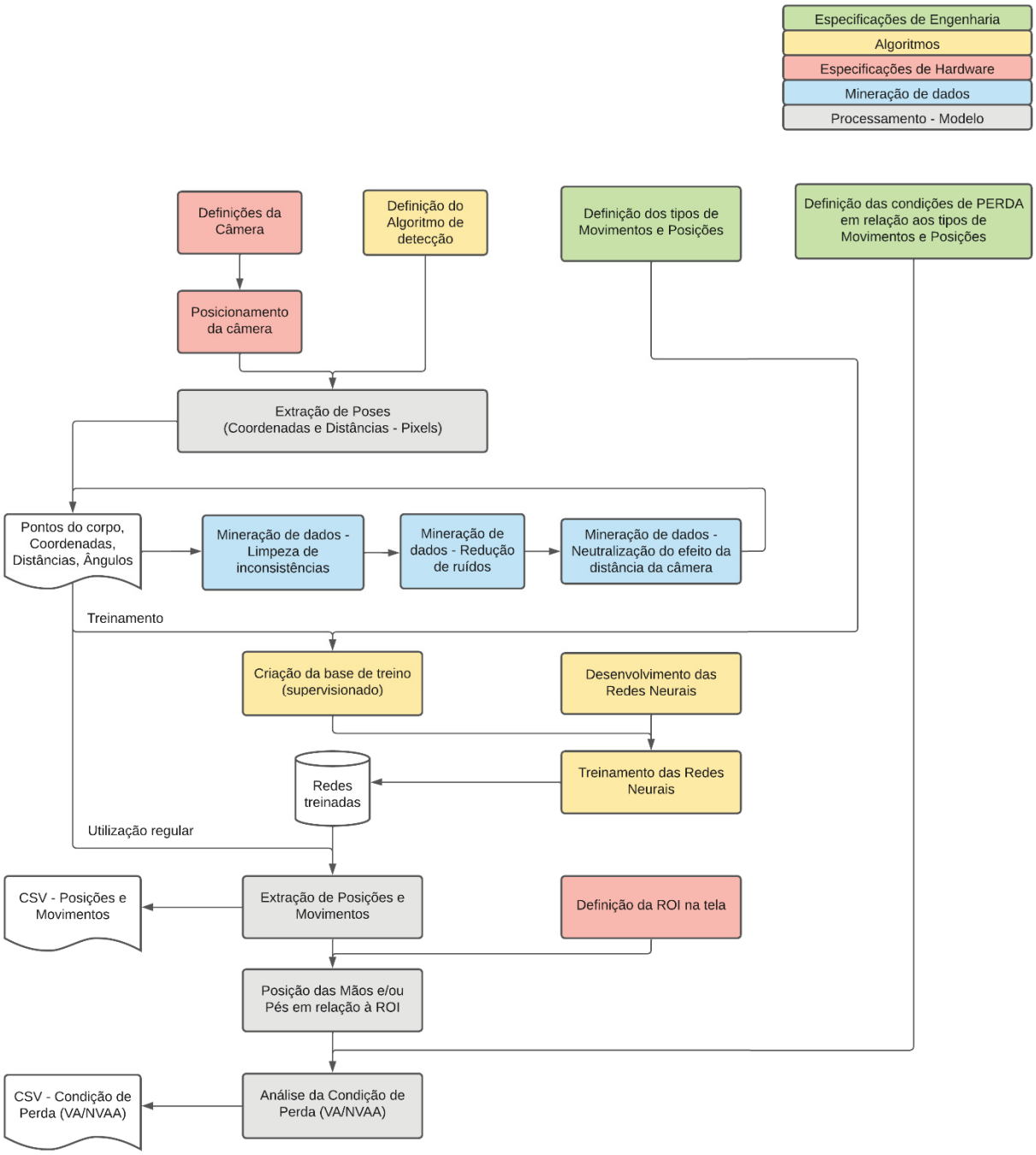


Figura 12 - Arquitetura básica do modelo
FONTE: Autor (2022)

Estas situações, aliadas ao posicionamento espacial do operador na estação de trabalho, identificada através de Regiões de Interesse na tela, definem se a atividade identificada em cada quadro enquadra-se como VA (Agregação de valor) ou NVAA (Não agregação de valor).

D. Definição do modelo de validação

Considerando o objetivo de avaliar o modelo na capacidade e acuracidade da classificação das poses em relação às situações elencadas, o resultado é o percentual de acertos da duração destas situações em comparação a uma análise realizada manualmente.

Como a aplicação visa situações diversas de coleta das imagens, serão realizadas filmagens em ângulos distintos, devendo o modelo ser capaz de identificar as condições em todas estas situações.

Um somatório do tempo de cada condição, para cada situação (ângulo e local de filmagem) será calculado para esta validação e espera-se acerto de no mínimo 90% na presença e duração destes eventos - VA ou NVAA.

Esta validação é resultante dos três fatores principais: A detecção da figura humana pelo PoseNet, a capacidade das redes neurais em identificar corretamente as poses e movimentos treinados e o correto desdobramento destas situações em relação ao posicionamento espacial do operador.

E. Extração dos dados iniciais

Convencionou-se a segmentação dos vídeos em 5FPS, ou seja, um quadro a cada 0.2 segundos. Este tempo representa a menor fração de tempo necessária para evidenciar movimentos humanos em um processo produtivo, sendo que qualquer valor inferior a este seria irrelevante, e valores superiores a 0.2 segundos poderiam fazer com que alguns movimentos rápidos não fossem evidenciados.

Após a gravação e identificação dos pontos do corpo humano, os dados foram extraídos de cada quadro de acordo com a Tabela IV. Devido às condições adversas das filmagens, seja por sobreposição de objetos com a figura humana, pouco contraste e baixa acuracidade, em geral, do modelo utilizado, fez-se necessária uma suavização dos dados, adotando em cada quadro o valor da mediana entre o identificado imediatamente anterior à este quadro, o valor do quadro atual e o valor imediatamento posterior à este quadro (mediana entre 3 valores) a fim de neutralizar momentos de perda de referência. O resultado deste tratamento pode ser evidenciado através do exemplo dos valores da distância Pulso ao Ombro esquerdo (Ple_OmbroY) em um vídeo de referência na Figura 13.

F. Definição das redes neurais

Para cada condição a ser identificada, de acordo com os itens do capítulo 3.3, foi desenvolvida uma rede neural específica. As características e especificações gerais destas redes foram as seguintes:

Posição dos braços (6 níveis de resposta)

- Modelo com 2 camadas densas
- Dropout entre a primeira e segunda camada
- Ativação *relu*
- Camada de saída tamanho 6 e ativação *softmax*
- Otimizador *Adam* com taxa de aprendizagem 0.0005
- 200 épocas

Movimentos (braços ou pernas) e posição das pernas e tronco

- Modelo com 2 camadas densas
- Dropout entre a primeira e segunda camada
- Ativação *relu* e *sigmoid*
- Camada de saída tamanho 1 e ativação *sigmoid*
- Otimizador *Adam* com taxa de aprendizagem 0.0001
- 500 épocas

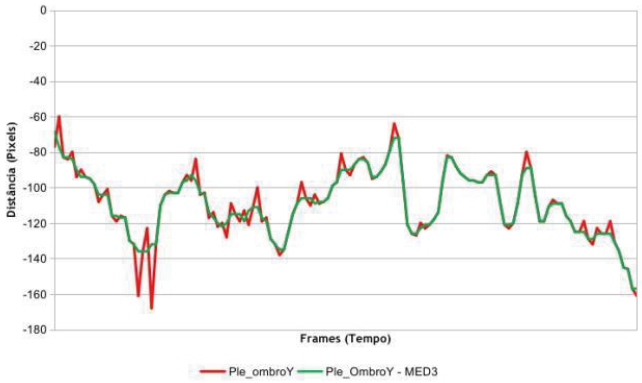


Figura 13 – Gráfico resultante da eliminação de ruidos
FONTE: Autor (2022)

G. Treinamento das redes neurais

Caracterizado como um modelo capaz de identificar as poses e movimentos específicos em diversas condiçõe de aplicação, foram segregados fragmentos de 9 vídeos distintos, contemplando ângulos, iluminação e distâncias variadas entre eles, conforme exemplos de quadros na Figura 14.

Para estes fragmentos, foram identificadas manualmente as condições de saída para cada quadro, sendo as funções de movimento avaliadas de forma contínua e as condições de posição em quadros aleatórios, totalizando uma base de 5300 linhas.

Considerando que a extração dos dados ocorre quadro a quadro e uma das condições a ser desecoberta é referente à presença ou não de movimento, fez-se necessário transformar a tabela de entrada.

A transformação dos pontos extraídos do Posenet, demonstrados na Tabela IV deu-se de 3 formas:

- Dados referenciais diretos, cujos valores são baseados em diferenças das distâncias entre pontos como, por exemplo, o dado “altura” cuja dimensão tende a referenciar a distância entre pés à cabeça, ou “Ole_ContY” considerando a distância entre o ombro

e quadril, avaliando alterações referentes a flexões de coluna;

- Dados brutos, onde o valor da matriz é o mesmo extraído pelo algoritmo, sem alterações. Como exemplo deste tipo de dado, cita-se o “pley”, valor do pulso esquerdo no eixo Y;
- Dados angulares, utilizando 3 pontos específicos para estabelecer ângulos aparentes nas articulações dos braços e pernas como, por exemplo, o “AngBrLe” representando o ângulo formado entre o quadril, ombro e cotovelo cuja função é avaliar se o braço está junto ao corpo ou levantado.

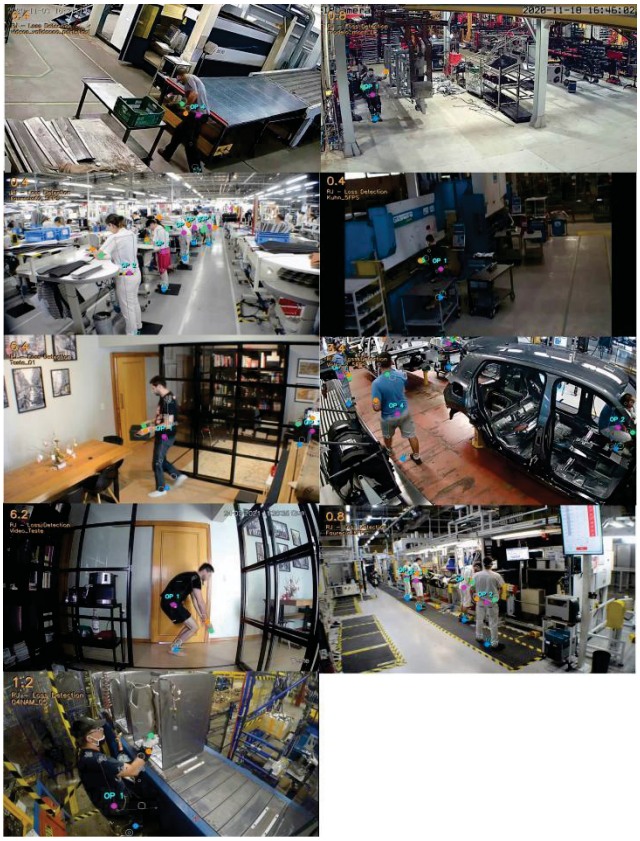


Figura 14 - Videos de coleta de dados para treinamento das redes
FONTE: Autor (2022)

As redes referentes à estas condições foram abastecidas a partir de funções que calculam a distância euclidiana entre o ponto do quadro atual com o do quadro anterior, observando um índice de variação e deslocamento em pixels para o sistema cartesiano da imagem e em graus para os cálculos baseados em ângulo.

Outro ponto relevante é que todas as redes foram também alimentadas com alguns dados chave para que o modelo fosse capaz de identificar situações que, em um primeiro momento, poderia distorcer o resultado.

Por tratar-se de um modelo baseado em uma imagem 2D, um passo de uma pessoa que esteja a vários metros de distância da câmera representaria, em pixels, uma distância irrelevante em relação uma pessoa que estivesse próxima à câmera. Para que o modelo pudesse receber uma percepção de profundidade, o campo “Femur” foi utilizado como denominador para as medições de distância referentes ao deslocamento, portanto uma figura humana que estivesse em uma escala pequena na tela, demandaria uma variação de poucos pixels para ser classificada como “caminhando”, pois estaria referenciada por uma distância pequena entre o joelho e quadril (a origem do dado “Femur”).

Outro campo importante para a análise foi o “Orient”, resultado da diferença no Eixo X entre o ponto do ombro direito com o esquerdo. Este campo demonstra a orientação da figura humana em relação à tela, se está de frente, de costas de ou lado. Este parâmetro é relevante para a avaliação dos ângulos, já que, por exemplo, estando de frente, não é possível calcular o ângulo de dobra do joelho.

Após as funções de transformação dos dados, prosseguiu-se com a alimentação das redes e treinamento, estando a medição da acurácia das redes disposta na Tabela V. Os resultados demonstraram uma capacidade de detecção dentro o esperado para a aplicação, com um resultado bastante positivo para os índices de condição de pernas e tronco, porém uma precisão bem mais reduzida na condição dos braços, conforme a Matriz de Confusão da Figura 15a e 15b.

		Predição					
		0	1	2	3	4	5
Referência	0	490	80	20	10	0	0
	1	110	2540	40	60	0	0
	2	0	60	660	10	5	0
	3	0	80	0	560	15	20
	4	0	10	20	0	100	0
	5	0	0	0	0	5	160

15a - Matriz da condição de posição

		Predição	
		0	1
Ref.	0	720	630
	1	60	3640

15b - Matriz da condição de movimento

Figura 15 - Matrizes de confusão dos braços
FONTE: Autor (2022)

H. Aplicação do modelo em condições controladas

Com os modelos gerados, foi realizada uma simulação para analisar cada etapa do algoritmo de identificação.

Um video em ambiente neutro foi gravado, contemplando a ação de caminhar até uma caixa, agaixar, pegar a caixa, retornar ao ponto inicial carregando a caixa, caminhar caminhar novamente até a pilha de caixas, curvar-se sem dobrar as pernas e pegar uma segunda caixa, levando-a ao ponto inicial. As linhas de movimentação referentes à este video e atividade podem ser vistas na Figura 16.

TABELA IV
DADOS EXTRAÍDOS DO PROCESSAMENTO INICIAL

Nome	Descrição
altura	Diferença no eixo Y entre o ponto do nariz e tornozelo (em pixels)
pcx	Valor médio entre quadril esquerdo e quadril direito, no eixo X
pcy	Valor médio entre quadril esquerdo e quadril direito, no eixo Y
X	Valor médio entre o tornozelo esquerdo e direito, no eixo X
Y	Valor médio entre o tornozelo esquerdo e direito, no eixo Y
HandLDx	Extensão do ponto do pulso direito, baseado no tamanho do umero, no eixo X
HandLDy	Extensão do ponto do pulso direito, baseado no tamanho do umero, no eixo Y
HandLEx	Extensão do ponto do pulso esquerdo, baseado no tamanho do umero, no eixo X
HandLEy	Extensão do ponto do pulso esquerdo, baseado no tamanho do umero, no eixo Y
tlex	Valor do tornozelo esquedo no eixo X
tley	Valor do tornozelo esquedo no eixo Y
tldx	Valor do tornozelo direito no eixo X
tldy	Valor do tornozelo direito no eixo Y
clex	Valor do cotovelo esquedo no eixo X
cley	Valor do cotovelo esquedo no eixo Y
cldx	Valor do cotovelo direito no eixo X
cldy	Valor do cotovelo direito no eixo Y
plex	Valor do pulso esquedo no eixo X
pley	Valor do pulso esquedo no eixo Y
pldx	Valor do pulso direito no eixo X
pldy	Valor do pulso direito no eixo Y
jlex	Valor do joelho esquedo no eixo X
jley	Valor do joelho esquedo no eixo Y
jldx	Valor do joelho direito no eixo X
jldy	Valor do joelho direito no eixo Y
olex	Valor do ombro esquedo no eixo X
oley	Valor do ombro esquedo no eixo Y
oldx	Valor do ombro direito no eixo X
oldy	Valor do ombro direito no eixo Y
Cin_baseY	Diferença no eixo Y entre o valor médio dos quadris direito e esquedo ao ponto médio do tornozelo esquedo e direito
Ole_CintY	Diferença entre o valor no eixo Y do quadril esquedo ao ponto ombro esquerdo
Old_CintY	Diferença entre o valor no eixo Y do quadril direito ao ponto ombro direito
Ple_ombroY	Diferença entre o valor no eixo Y do pulso esquedo ao ponto ombro esquerdo
Pld_ombroY	Diferença entre o valor no eixo Y do pulso direito ao ponto ombro direito
Cle_ombroY	Diferença entre o valor no eixo Y do cotovelo esquedo ao ponto ombro esquerdo
Cld_ombroY	Diferença entre o valor no eixo Y do cotovelo direito ao ponto ombro direito
PulOmLe	Distancia euclidiana entre o pulso e ombro esquedos (em pixels)
PulOmLd	Distancia euclidiana entre o pulso e ombro direitos (em pixels)
PulCtLe	Distancia euclidiana entre o pulso e cotovelo esquedos (em pixels)
PulCtLd	Distancia euclidiana entre o pulso e cotovelo direitos (em pixels)
JoToLe	Distancia euclidiana entre o joelho e tornozelo esquedos (em pixels)
JoToLd	Distancia euclidiana entre o joelho e tornozelo direitos (em pixels)
AngBrLe	Angulo formado pelos pontos do quadril, ombro e cotovelo esquerdos
AngBrLd	Angulo formado pelos pontos do quadril, ombro e cotovelo direitos
AngCtLe	Angulo formado pelos pontos do ombro, cotovelo e pulso esquerdos
AngCtLd	Angulo formado pelos pontos do ombro, cotovelo e pulso direitos
AngJoLe	Angulo formado pelos pontos do quadril, joelho e tornozelo esquerdos
AngJoLd	Angulo formado pelos pontos do quadril, joelho e tornozelo direitos
AngCol	Angulo formado pelos pontos centrais dos joelhos, quadril e ombros
Femurle	Distancia euclidiana entre o joelho e quadril esquedos (em pixels)
Femurld	Distancia euclidiana entre o joelho e quadril direitos (em pixels)
Femur	Valor médio entre Femurle e Femurld (em pixels)
Umerole	Distancia euclidiana entre o cotovelo e pulso esquedos (em pixels)
Umerold	Distancia euclidiana entre o cotovelo e pulso direitos (em pixels)
Umero	Valor médio entre Umerole e Umerold (em pixels)
Orient	Valor da diferença no eixo X entre o ponto do ombro direito e esquerdo (em pixels)

FONTE: Autor (2022)

TABELA V
ACURÁCIA DAS REDES

Condição			Output	Acc
Posição	Perna	Em pé	1	97%
		Agaixado	1	
	Tronco	Ereto	1	97%
		Curvado	0	
	Braços	Esticado para baixo	0	89%
		Dobrado para baixo	1	
		Dobrado a frente	2	
		Esticado a frente	3	
		Dobrado acima cabeça	4	
		Esticado acima cabeça	5	
Movimentos	Perna	Agachando	1	84%
		Estático	0	
	Tronco	Curvando	1	85%
		Estático	0	
	Braços	Movendo	1	86%
		Estático	0	
Deslocamento	CG	Movendo	1	88%
		Estático	0	

FONTE: Autor (2022)

Figura 16 - Caminho percorrido no video de validação
FONTE: Autor (2022)

Como resultado, obteve-se na escala de tempo (quadros) a classificação de cada uma das redes neurais, de acordo com o gráfico da Figura 17 e posteriormente sua tradução em movimentos no gráfico da Figura 18. O modelo conseguiu, por exemplo, evidenciar e diferenciar um agaixamento de uma curvatura de coluna.

No momento entre 3,8 e 5,4 segundos, a linha vermelha, representando a posição da coluna, alterou seu valor de 1 para 0, ou seja, o algoritmo evidenciou uma flexão de coluna, assim como no momento entre 9,8 e 11,4 segundos. Entretanto, no primeiro caso, a linha azul, que representa a flexão das pernas, não sofreu a mesma alteração, mantendo-se em 1 entre 3,8 e 5,4 segundos. Esta condição indica que apenas a coluna foi flexionada, mas não as pernas. No segundo evento de flexão de coluna, observou-se também o valor da flexão de pernas com valor 0, ou seja, o modelo identificou também uma flexão das pernas.

Neste caso, o que pode ser evidenciado através destes eventos na Figura 17 é que em um primeiro momento a pessoa identificada pegou um objeto no chão curvando apenas a coluna e em um segundo momento, dobrando as pernas. Esta situação é explicitada no gráfico da Figura 18, após a tradução do conjunto de situações.

Pode-se observar no ponto do segundo 4 a situação "Bending" e no momento do segundo 11 a situação "Crouching", representando os movimentos citados anteriormente.

I. Relacionamento das situações identificadas com as perdas industriais

Para a aplicação prática, o último estágio de desenvolvimento foi estabelecer a relação entre estas poses e situações com a agregação ou não de valor ao produto.

Para tanto, foi necessário criar uma forma de identificar em tela a região onde o operador realizar interface física com o produto através do desenho de Regiões de Interesse (ROI). Estas ROI definem onde, em determinadas situações, as mãos ou os pés dos operadores devem estar posicionados para que o tempo seja classificado como VA ou NVAA.

A relação entre ROI e Situação pode ser vista na Tabela VI. Os movimentos realizados pelo operador, obtidos através do processamento das redes, são comparados com a posição espacial das mãos e/ou dos pés, sendo divididos entre dentro ou fora da ROI. Algumas condições, como o "caminhar", são automaticamente classificados como NVAA, pois são movimentos definidos como perda em qualquer que seja a situação da operação [13].

Entretanto, alguns movimentos, como o operador com as pernas paradas, porém movimentando apenas os braços, pode estar condicionada a uma atividade de agregação de valor ou não. O fator determinante para esta classificação é o seu posicionamento em relação à ROI. Estando suas mãos dentro da ROI definida como zona de agregação, o resultado será VA, caso contrário, NVAA.

IV. APRESENTAÇÃO DOS RESULTADOS

Foram selecionadas algumas aplicações industriais para a validação, realizando através da empresa de consultoria industrial RFJ Consultoria Empresarial LTDA provas de conceito nos seguintes segmentos:

a) Área de fabricação de uma industria metalúrgica de grande porte em São José dos Pinhais-PR;

b) Estação de linha de montagem de máquinas agrícolas em multinacional o setor em São José dos Pinhais-PR;

c) Estação de linha de montagem de painéis automotivos de uma multinacional do setor em São José dos Pinhais-PR;

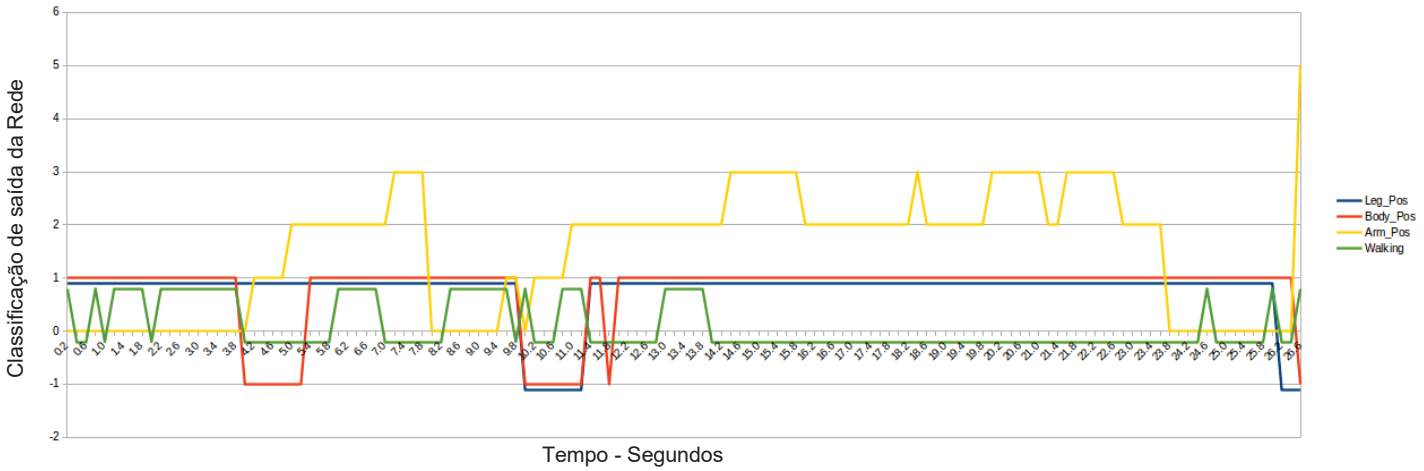


Figura 17 – Gráfico da classificação da posição dos braços, pernas, tronco e deslocamento ao longo do tempo do vídeo (quadros - no eixo X)
FONTE: Autor (2022)

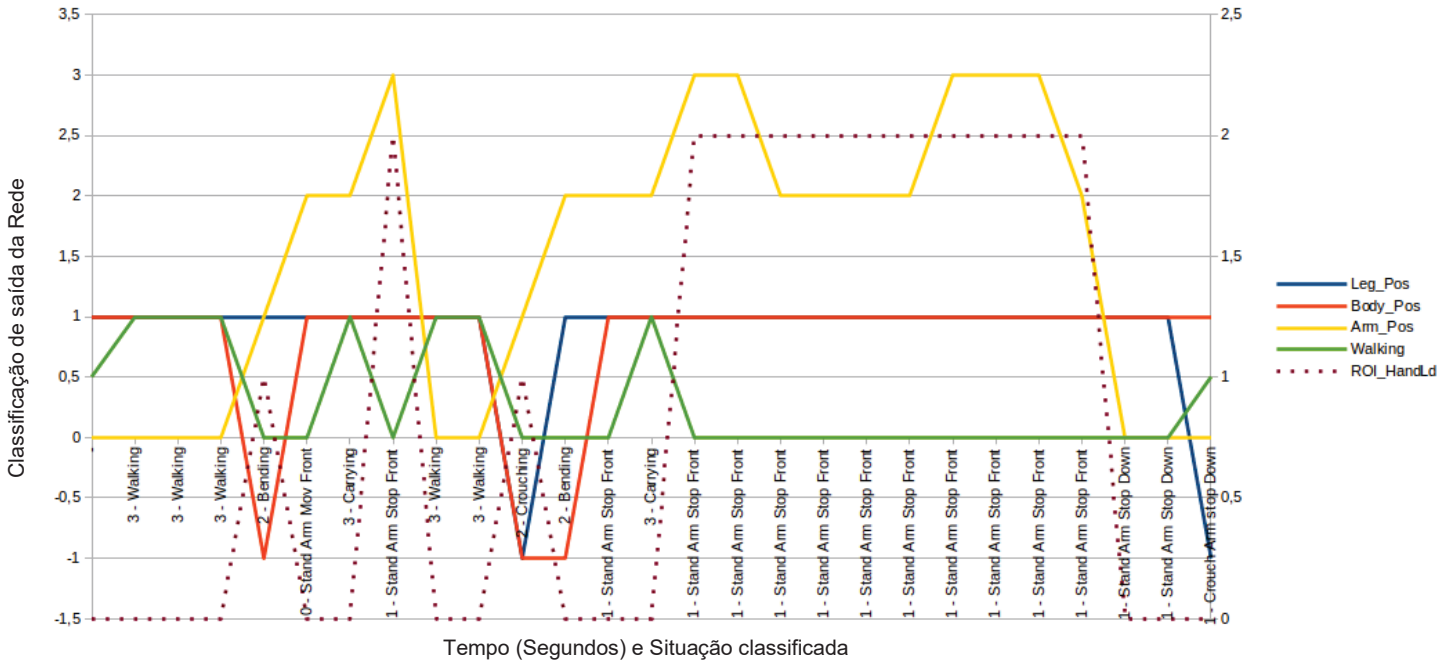


Figura 18 – Gráfico da classificação da situação da figura humana ao longo do tempo do vídeo (situação majoritária a cada segundo - no eixo X)
FONTE: Autor (2022)

d) Estação de montagem de motores de uma multinacional fabricante de caminhões em Curitiba-PR;

e) Estação de montagem de caminhões de uma multinacional do setor em Curitiba-PR;

f) Estação de montagem de lavadoras de roupa de uma multinacional do setor em Joinville-SC;

g) Estação de montagem de refrigeradores de uma multinacional do setor em Curitiba-PR;

h) Estação de montagem de secadoras de roupa de uma multinacional do setor em Marion-Ohio-EUA.

Em todos estes casos, foram selecionadas junto à área de Engenharia Industrial uma atividade, ou uma pequena sequência de atividades de montagem realizadas sempre pelo mesmo operador. O ângulo e posicionamento das câmeras em cada uma das aplicações pode ser vista na Figura 19

Com base nesta seleção, foi realizado o posicionamento da câmera para que não houvessem sombras (pontos sem que fosse possível avistar o operador) e onde fosse possível observar as mãos o maior tempo possível.

Além desta seleção, foi identificada também onde seria a região, na tela, onde o operador faz efetivamente a interface com o produto que está sendo produzido.

As gravações registraram diversos ciclos das mesmas operações, para que fosse possível uma análise mais ampla e menos sujeita ter o resultado afetado por eventos especiais.

Após a gravação e processamento dos dados, os resultados obtidos e comparados com os padrões existentes nas Engenharias Industriais das respectivas empresas estão expressos na Tabela VII. Estes padrões foram fornecidos pelas respectivas empresas.

Os resultados expressos também foram segmentados para estudos futuros de acordo com o tamanho da estação de montagem, onde é possível evidenciar as diferenças entre atividades realizadas em uma pequena área ou até mesmo sentado de operações onde o operador caminha ao redor do produto. Outro ponto registrado foi o tempo ciclo, ou seja, o tempo de duração das atividades entre um produto e o próximo.

Estes índices demonstram que em estações maiores e com tempos ciclos maiores, o modelo identificou uma parcela de perdas NVAA ligeiramente maiores do que o especificado pela documentação técnica da engenharia da empresa, mas ainda dentro da margem de aceitação.

Por exemplo, no segundo item da Tabela VII, o percentual do tempo calculado manualmente pela empresa classificado como VA foi 27,3%, ou seja, do tempo total da operação, em apenas 27% do tempo o operador está realizando atividades que agregam valor ao produto. Já o modelo proposto evidenciou em apenas 24,8% do tempo observado estas condições, uma diferença de 2,5% sobre o especificado.

Considerando o objetivo de, através de forma autônoma, chegar na margem de 90% de conformidade em relação ao padrão medido manualmente pelos engenheiros das empresas, de forma geral pode-se aceitar que o modelo é capaz de realizar uma medição superficial do nível de perdas de mão de obra, apresentando um erro médio de apenas 0,7%, variando entre -2.5% e +3.6% entre as aplicações, e com desvio padrão de 0,5% a 4,8% entre as tomadas, entretanto, variações entre as atividades industriais são inerentes ao trabalho com operadores.

V. CONSIDERAÇÕES FINAIS

Observar atentamente a produção é uma das práticas difundidas pelo Sistema Toyota de Produção, já que permite a identificação de oportunidades de melhoria no chão de fábrica. O sistema aqui proposto, baseado em visão computacional para identificação das anomalias de um processo produtivo, é uma ferramenta impulsionadora destas práticas. Como é baseado em bibliotecas de reconhecimento de elementos em uma imagem captada por câmeras de filmagem, se mostra rápida e isenta de vícios de um observador humano. Trazem também uma precisão de marcação do tempo dos eventos desejados uma vez que o processamento destas se dão em quadros por segundo.

Em relação ao reconhecimento e identificaçãos do MUDA, é válido salientar que próprio conceito de classificação das atividades em VA e NVA varia entre as diversas empresas, tornando a medição um pouco subjetiva, entretanto, a forma com que é definida a Região de Interesse pode ser ajustada para refinar o dado.

Tratando-se de um modelo em que mesmo a linha de produção e o produto sendo processado são desconhecidos e mesmo assim foi possível, com uma câmera de especificações básicas de mercado, atingir um nível de conformidade como o que foi obtido é de grande potencial.

Variações em relação ao nível e proporção do tempo de agregação de valor (VA) em processos industriais também são observados entre analistas distintos mensurando uma mesma operação. É algo inerente à capacidade de observação de cada um. Este fato corrobora a virtude deste modelo, mesmo com as diferenças entre o padrão da empresa.

Embora o resultado seja superficial, não indicando detalhadamente atividade por atividade, pois não existem gatilhos para que seja possível identificar os eventos onde o operador finaliza uma operação e inicia outra, seria possível realizar uma varredura em toda a linha de montagem de forma bastante rápida e validar o trabalho prévio de engenharia assim como ser uma base para priorização de novos projetos.

Ainda existem oportunidades de melhoria no modelo. A própria qualidade do modelo de detecção de poses PoseNet não é referência em acuracidade e outros modelos com processadores mais eficientes podem oferecer resultados ainda mais confiáveis.

Da mesma forma, alimentar o treinamento das redes com mais dados também auxiliaria um desempenho mais estável.

Dentre os pontos fortes, importante salientar a forma contínua de coleta, sendo possível a análise de várias horas de atividade de forma imediata.

TABELA VI
RELAÇÃO DE SITUAÇÕES COM AGREGAÇÃO DE VALOR

Situação		Classificação	
		Mãos na ROI	Mãos fora da ROI
Walking	Caminhando com braços livres	NVAA	NVAA
Carrying	Caminhando carregando	NVAA	NVAA
Crouching	Agaixando	NVAA	NVAA
Crouch Arm Mov Up	Agaixado com braços em movimento sobre cabeça	VA	NVAA
Crouch Arm Mov Down	Agaixado com braços em movimento abaixo cabeça	VA	NVAA
Crouch Arm Mov Front	Agaixado com braços em movimento esticados a frente	VA	NVAA
Crouch Arm stop Up	Agaixado com braços estáticos sobre cabeça	NVAA	NVAA
Crouch Arm stop Down	Agaixado com braços estáticos abaixo cabeça	NVAA	NVAA
Crouch Arm stop Front	Agaixado com braços estáticos esticados a frente	NVAA	NVAA
Bending	Curvando tronco	NVAA	NVAA
Bend Arm Mov	Tronco curvado com braços em movimento esticados	VA	NVAA
Bend Arm Stop	Tronco curvado com braços estáticos esticados	VA	NVAA
Stand Arm Mov Up	Em pé com braços em movimento sobre cabeça	VA	NVAA
Stand Arm Mov Down	Em pé com braços em movimento abaixo cabeça	VA	NVAA
Stand Arm Mov Front	Em pé com braços em movimento esticados a frente	VA	NVAA
Stand Arm stop Up	Em pé com braços estáticos sobre cabeça	NVAA	NVAA
Stand Arm stop Down	Em pé com braços estáticos abaixo cabeça	NVAA	NVAA
Stand Arm stop Front	Em pé com braços estáticos esticados a frente	NVAA	NVAA

FONTE: Autor (2022)

TABELA VII
RESULTADOS DA APLICAÇÃO DO MODELO EM CONDIÇÕES REAIS

Segmento	**Produto**	**Tamanho estação de trabalho (metros)**	**Tempo ciclo (minutos)**	**% VA Especificado pela Engenharia**	**Quantidade de Observações**	**% VA Identificado pelo modelo**		**Variação Padrão vs Modelo (%)**	
						Média	**Desv. Pad.**	**Média**	**Desv. Pad.**
Metalurgia	Chapas de aço	5	122	7,5%	15	6,9%	0,5%	-0,6%	0,5%
Máquinas Agrícolas	Pulverizador agrícola	7	275	27,3%	6	24,8%	1,7%	-2,5%	1,7%
Automotiva	Painel automotivo	3	0,58	56,8%	18	60,4%	4,2%	3,6%	4,2%
Caminhões	Motor	3	45	45,9%	22	48,1%	3,3%	2,2%	3,3%
Caminhões	Caminhão	12	240	33,2%	16	30,8%	2,1%	-2,4%	2,1%
Eletrodomésticos	Lavadora de roupas	2	0,42	47,1%	35	49,1%	3,4%	2,1%	3,4%
Eletrodomésticos	Refrigerador	4	0,75	42,6%	42	44,1%	3,1%	1,5%	3,1%
Eletrodomésticos	Secadora de roupas	2	0,18	67,4%	32	69,2%	4,8%	1,9%	4,8%
VARIAÇÃO MÉDIA (EM %)								**0,7%**	**3,7%**

Tempo Ciclo = Tempo de operação padrão para produção de 01 unidade
% VA = % do tempo ciclo com Atividades que Agregam valor ao produto

FONTE: Autor (2022)

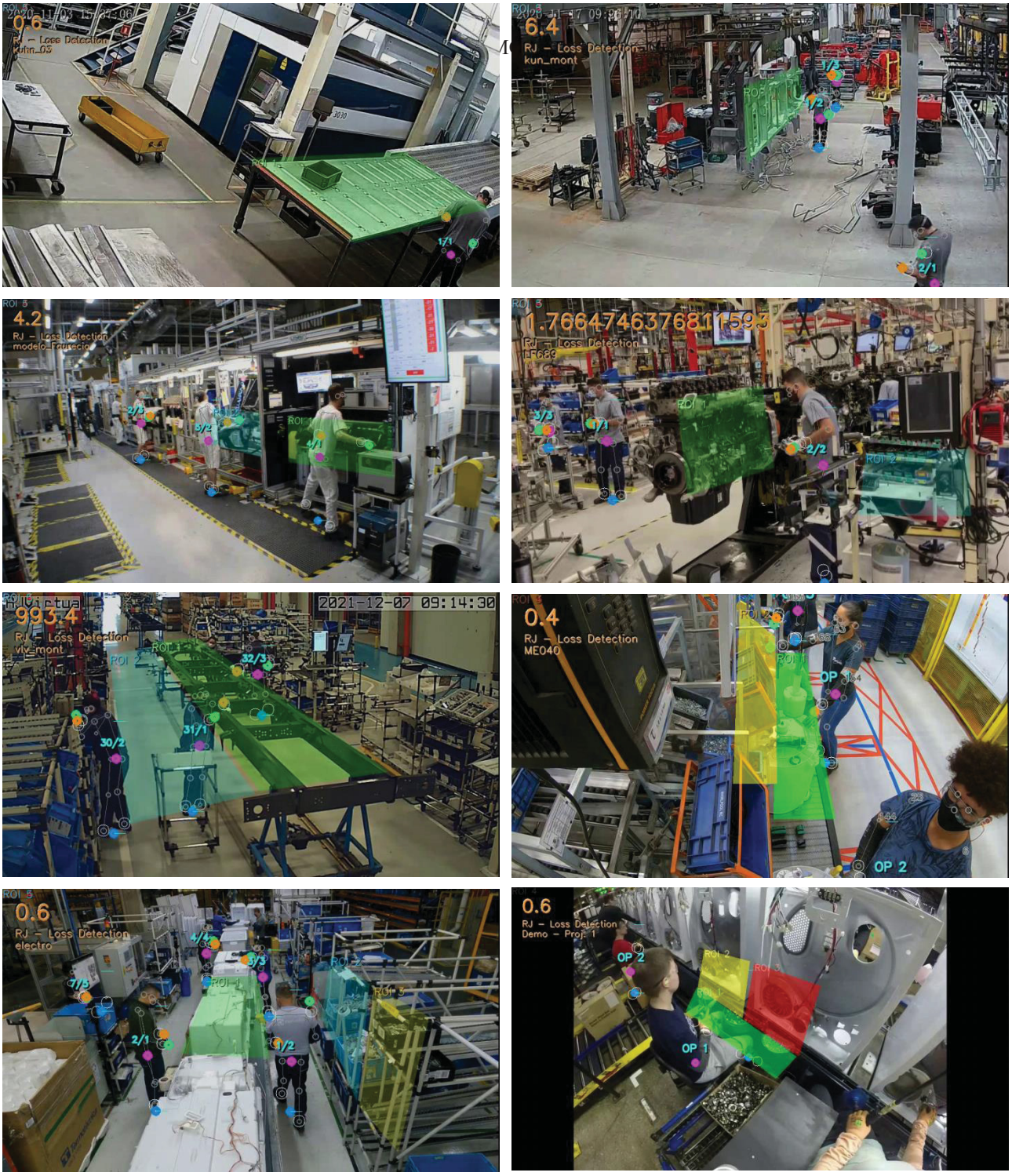


Figura 19 - Imagens da esquerda superior para direita inferior: 1) Área de fabricação, máquina de corte a laser 2) Montagem de tubulação hidráulica de máquina agrícula 3) Montagem de painel automotivo 4) Montagem de motor 5) Montagem de caminhão 6) Montagem de lavadora de roupas 7) Montagem de refrigerador 8) Montagem de secadora

FONTE: Autor (2022)

Como uma restrição observada na aplicação prática, a necessidade da câmera estar posicionada de forma fixa, para que as ROI marcadas na tela não sejam deslocadas de suas referências físicas, demonstrou-se um entrave para a flexibilidade do modelo, assim como a instalação em linhas de montagem tracionadas precisa de um estudo profundo de deslocamentos para o correto posicionamento.

A. Recomendações para trabalhos futuros

Alguns pontos relevantes para os próximos estudos podem ser elencados:

a) Implementar um modelo multi-câmera, mapeando a atividade em 3 dimensões: No trabalho desenvolvido, houveram casos de sobreposição de operadores e também dificuldade em estabelecer uma região de interesse devido a não haver uma análise mais aprofundada da profundidade do operador em relação à câmera. Dois pontos que estejam nas mesmas coordenadas X e Y da tela podem estar em distâncias muito diferentes na realidade.

b) Comparar o modelo com algoritmos já existentes de identificação poses de movimentos, principalmente eventos de "caminhar" e relacionados ao deslocamento humano e detecção de movimento. Apesar do modelo proposto neste trabalho visa identificar condições específicas relacionadas às perdas industriais, a base comparativa com estes algoritmos pode prover uma avaliação mais controlada a respeito da base do sistema, focada na capacidade de detecção inicial dos movimentos.

c) Utilizar métricas de reconhecimento de objetos para substituir a definição das regiões de interesse (ROI) na tela, acelerando o processo de setup para a aplicação real. O reconhecimento de objetos específicos e relacionados à agregação de valor, no caso o produto em processo, pode ser utilizado na avaliação do posicionamento do operador em relação à estação de trabalho, sendo a referência para a identificação das situações de perda ou agregação de valor.

VI. REFERÊNCIAS

[1] PALANGE, A.; DHATRAK, P. Materials Today: Proceedings: Lean manufacturing a vital tool to enhance productivity in manufacturing, 2021.

[2] VUKADINOVIC, S.; MACUZIC, I.; DJAPAN, M.; MILOSEVIC, M. Safety Science: Early management of human factors in lean industrial systems, 2019.

[3] WOMACK, J.P.; JONES, D.T. The Machine That Changed the World: The Story of Lean, Simon and Schuster, New York (1990)

[4] WOMACK, J.P.; JONES, D.T. Lean Thinking: Banish Waste and Create Wealth in Your Corporation, Simon and Schuster, New York (1996)

[5] FENG, X.; JIANG, Y.; YANG, X.; DU, M.; LI, X. Computer vision algorithms and hardware implementations: a survey. Integration, [S.L.], v. 69, p. 309-320, nov. 2019. Elsevier BV.

[6] GONZALEZ, R. C.; RICHARD E. W. Digital image processing. 3rd ed. Prentice Hall, 2008.

[7] MUNEA, T. L.; JEMBRE, Y. Z.; WELDEGEBRIEL, H. T; CHEN, l.; "The Progress of Human Pose Estimation: A Survey and Taxonomy of Models Applied in 2D Human Pose Estimation," in IEEE Access, vol. 8, pp. 133330-133348, 2020, doi: 10.1109/ACCESS.2020.3010248.

[8] BLOG THE TENSORFLOW. Real-time Human Pose Estimation, Dan Oved, 2018, Disponível em: http://blog.tensorflow.org/2018/05/real-time-human-pose-estimation-in.html

[9] PAPANDREOU, G.; ZHU, T.; CHEN, L.; GIDARIS, S.; TOMPSON, J.; MURPHY, K. PersonLab: Person Pose Estimation and Instance
ArXiv, 22 Mar 2018

[10] BOLHASANI, H; MOHSENI, M; RAHMANI, A. Deep learning applications for IoT in health care: a systematic review. Informatics In Medicine Unlocked, [S.L.], p. 100550, mar. 2021. Elsevier BV.

[11] GOODFELLOW, I.; BENGIO, Y.; COURVILLE, A. Deep Learning. MIT Press, 2016.

[12] FOOTE, K. A Brief History of Deep Learning., 2017.

[13] SHINGO, S. O Sistema Toyota de Produção – Do ponto de vista da engenharia de produção. Ed. Bookman: Porto Alegre, 1996.