

# Protótipo De Reconhecimento De Expressões Faciais Utilizando Inteligência Artificial

GABRIEL ORLANDIN GOMES<sup>1</sup>  
JOÃO LUÍS TAVARES DA SILVA<sup>2</sup>

Data de submissão: 20/02/2021. Data de publicação: 31/05/2021.

## RESUMO

O objetivo deste artigo foi realizar uma análise de trabalhos relacionados às emoções visuais e sua aplicabilidade no varejo, fornecendo uma visão geral das tecnologias baseadas em visão computacional e os recursos necessários para aplicação da tecnologia. A ideia partiu do desenvolvimento de um protótipo utilizando um dispositivo com uma câmera integrada para efetuar uma análise das imagens com dependência parcial de uma rede de internet. A fase de testes demonstrou que a Jetson Nano seria o dispositivo ideal para este protótipo e que o Yolo era a rede neural com melhor performance para esta aplicação em tempo real. O Yolo, um algoritmo de detecção de objetos baseado em redes neurais convolucionais, foi utilizado para detecção de faces juntamente com um algoritmo de análise de emoções, gênero e idade. Os resultados obtidos nos experimentos alcançaram uma acurácia de 90% na inferência de gênero e idade, enquanto a detecção de emoções faciais obteve uma acurácia geral de aproximadamente 70%, e demonstraram o emprego destas tecnologias no desenvolvimento de um protótipo que entrega informações analíticas e não intrusivas a respeito dos consumidores das redes de varejo físicas. Um protótipo de armazenamento das informações coletado pela câmera durante o dia e um painel gráfico das análises de cada cliente foi desenvolvido como suporte na análise de satisfação do cliente.

**Palavras-chave:** Sentimentos. Expressões Faciais. Aprendizagem Profunda. Jetson Nano.

## ABSTRACT

The objective of this work is to carry out an analysis of related works that were developed in the last years for the analysis of visual emotions and their applicability in retail, providing an overview of technologies based on computer vision and the necessary resources for applying

---

<sup>1</sup> Graduado em Engenharia da Computação (Uniftec).

<sup>2</sup> Professor do Centro Universitário Uniftec – Unidade de Caxias do Sul, Graduação em Ciência da Computação com Doutorado em Mathematique et Informatique pela Université Joseph Fourier (França) e Mestrado em Processamento de Língua Natural pela Pontifícia Universidade Católica do Rio Grande do Sul. Pós-doutorado em Tecnologias Computacionais para a Educação na Universidade Federal do Rio Grande do Sul.

the technology. The idea came from the development of a prototype using a device with an integrated camera to perform an analysis of the images with partial dependence on an internet network. The testing phase demonstrated that the Jetson Nano would be the ideal device for this prototype and that Yolo was the neural network with the best performance for this application in real time. Yolo, an object detection algorithm based on convolutional neural networks, was used for face detection along with an algorithm for analyzing emotions, gender and age. The results obtained in the experiments reached an accuracy of 90% in the inference of gender and age, while the detection of facial emotions obtained an overall accuracy of approximately 70%, and demonstrated the use of these technologies in the development of a prototype that delivers analytical and non-intrusive information about consumers in physical retail chains. A prototype for storing the information collected by the camera during the day and a graphic panel of the analysis of each customer was developed to support the analysis of customer satisfaction.

**Keywords:** Feelings. Facial expressions. Deep Learning. Jetson Nano.

## 1 INTRODUÇÃO

No âmbito geral, todas as lojas visam a satisfação do cliente. A maioria das empresas do varejo utilizam métodos convencionais de pesquisa de satisfação como, por exemplo, totens de avaliação. De acordo com Bouzakraoui, Sadiq, Emneya (2017), esses métodos são ultrapassados e carecem de confiabilidade e meios para distinguir entre uma resposta racional e emocional. A inovação se torna necessária para ter mais assertividade e se destacar no mercado através dos *insights*.

Através dos profissionais de marketing da empresa, o uso de dados se tornou muito relevante para oferecer maior valor e relevância para seus produtos. Ao analisar o comportamento dos clientes, muitos sinais relacionados à percepção de produto e valor podem ser extraídos.

Por meio das técnicas de Visão Computacional é possível interpretar dados como, por exemplo, a emoção humana através da captura de algumas imagens obtidas por uma câmera e utilizando-as como parâmetro para uma tomada de decisão. Estas técnicas são utilizadas no reconhecimento visual de objetos através do aprendizado de máquina, chamado de Aprendizado Profundo ou *Deep Learning*.



Atualmente, nas redes de varejo, as empresas aplicam pesquisas de satisfação pelos métodos convencionais, porém esses métodos são ultrapassados. O grande potencial da informação visual faz com que a visão computacional e sua aplicação seja um grande desafio para ser explorado de forma detalhada.

Compreender os níveis de satisfação do cliente é muito importante para que as organizações possam administrar com sabedoria a imagem positiva da empresa e dos níveis aceitáveis de qualidade de seus produtos e serviços e, conseqüente, a manutenção e engajamento de seus clientes.

Objetiva-se desenvolver um protótipo para avaliar a satisfação dos clientes por meio da coleta e análise de dados de câmeras de vídeo em que serão analisadas as emoções dos clientes em relação a um produto com base em expressões faciais. O foco deste estudo será na previsão da emoção do cliente em relação a um produto específico.

## **2 FUNDAMENTAÇÃO TEÓRICA**

O mercado global de reconhecimento facial deverá crescer de US \$ 3,2 bilhões em 2019 para US \$ 7,0 bilhões até 2024, a uma taxa de crescimento de 16,6% entre 2019 e 2024. Os principais fatores que impulsionam o mercado incluem o aumento dos avanços tecnológicos em todo o mundo. Portanto, espera-se que o segmento de reconhecimento facial obtenha uma alta taxa de adoção em todo o mundo. O monitoramento específico da atração de um produto pelo consumidor durante sua publicidade é importante para melhorar as campanhas publicitárias. Obviamente o uso de instrumentos invasivos, como por exemplo, Eletrocardiograma e Eletroencefalograma, sensores biométricos, podem afetar o comportamento dos sujeitos e, em particular, adulterar a espontaneidade e, conseqüentemente, as emoções experimentadas por eles (CECCACCI et al., 2018). Dessa forma, as técnicas não invasivas são alternativas mais tangíveis para monitorar a satisfação dos consumidores.

O comportamento do consumidor em uma loja é uma resposta de natureza cognitiva e emocional a estímulos desse ambiente. Dentro das linguagens não verbais, as expressões faciais têm um peso de 55%, ou seja, a expressão facial contribui com 55% da impressão geral do indivíduo (MEHRABIAN; RUSSELL, 1974).

O conjunto das emoções foi considerado como universal por apresentarem as mesmas características em indivíduos de cultura, raça, gênero e idade diferentes. O Facial Action Coding System (FACS) é um modelo criado por Paul Ekman e Wallace Friesen que categoriza as aparências faciais originadas de contrações musculares em Unidades de Ação (Action Units - AUs), que, combinadas ou não, representam todas as expressões faciais possíveis (EKMAN; FRIESEN, 1976).

Os seres humanos percebem a estrutura tridimensional do mundo com uma aparente facilidade. No entanto, modelar o mundo visual em toda a sua rica complexidade para o mundo computacional não é tarefa fácil. A visão computacional é um campo que engloba diversas áreas do conhecimento como geometria, matemática e processamento de sinais com o intuito de extrair informações a partir de imagens ou vídeos para recuperar a forma tridimensional e a aparência dos objetos nas imagens (SZELISKI, 2010).

O processo de análise de uma imagem por sistemas de visão computacional é efetuado em etapas, sendo que o resultado de uma etapa é utilizado como entrada para a etapa seguinte. As etapas são: aquisição da imagem, pré-processamento, segmentação, extração de características, reconhecimento e interpretação (GONZALEZ; WOODS, 2010).

Segundo Oliveira e Jaques (2008), para a identificação das expressões faciais e posterior classificação, inicialmente é necessário encontrar a face humana em uma imagem. Após obtenção da região onde se encontra um rosto, é necessário localizar as características faciais relevantes numa expressão facial, como boca e olhos, comumente efetuado pelo classificador de *haar-like features*. O processo seguinte consiste em determinar pontos na face que serão utilizados para realizar cálculos de deformação geométrica das características faciais que permitem identificar as AUs (Figura 1). Portanto, uma emoção é definida com a combinação de várias AU's.

A *Jetson Nano* apresenta uma performance computacional de 472 GFLOPS, possui um processador quad-core ARM A57 de 1,43 GHz e uma GPU de 128 cores NVIDIA Maxwell com 4 GB de memória RAM LPDDR4, USB 3.0 e opera com baixo consumo. É uma placa pequena, desenvolvida para aplicação de Inteligência Artificial e ideal para embarcá-la em experimentos, produtos ou protótipos.

Figura 1 - Descrição das AU's.

Upper Face Action Units					
AU 1	AU 2	AU 4	AU 5	AU 6	AU 7
					
Inner Brow Raiser	Outer Brow Raiser	Brow Lowerer	Upper Lid Raiser	Cheek Raiser	Lid Tightener
*AU 41	*AU 42	*AU 43	AU 44	AU 45	AU 46
					
Lid Droop	Slit	Eyes Closed	Squint	Blink	Wink
Lower Face Action Units					
AU 9	AU 10	AU 11	AU 12	AU 13	AU 14
					
Nose Wrinkler	Upper Lip Raiser	Nasolabial Deepener	Lip Corner Puller	Cheek Puffer	Dimpler
AU 15	AU 16	AU 17	AU 18	AU 20	AU 22
					
Lip Corner Depressor	Lower Lip Depressor	Chin Raiser	Lip Packerer	Lip Stretcher	Lip Funneler
AU 23	AU 24	*AU 25	*AU 26	*AU 27	AU 28
					
Lip Tightener	Lip Pressor	Lips Part	Jaw Drop	Mouth Stretch	Lip Suck

Fonte: Tian; Kanade; Cohn (2001).

Aprendizagem de máquina (*Machine Learning*) é uma área da Computação que apresenta um conjunto de métodos capaz de automaticamente detectar padrões em um conjunto de dados, podendo estes padrões ser utilizados para realizar classificação ou predição em novas entradas de dados (MITCHELL, 1997). Já a aprendizagem profunda (*Deep Learning*) é um subcampo da aprendizagem de máquina, pois o termo “profunda” vem dos algoritmos de redes neurais que podem ter um grande número de camadas de atuação (BENGIO; COURVILLE; VINCENT, 2013).

Os métodos de aprendizado profundo melhoraram drasticamente o estado da arte no reconhecimento visual e detecção de objetos, pois permite que modelos computacionais compostos por várias camadas de processamento aprendam representações de dados com vários níveis de abstração (LECUN; BENGIO; HINTON, 2015).

Transfer Learning é um processo de aproveitamento de redes neurais já treinadas para uma dada tarefa em uma outra tarefa similar, removendo a última camada de classificação e adicionando uma camada com a quantidade de neurônios necessários para

resolver o novo problema. Neste caso, somente essa camada adicionada será treinada, ou seja, o restante da rede não é afetada pela retropropagação (MENEGOLA et al., 2017).

Os conjuntos de dados (*Datasets*) são utilizados para análise de dados e estatística, são organizados em um arquivo que contém vários dados sobre um determinado assunto. Esse arquivo é gerado num formato CSV, XML, JSON ou TXT. Os *Datasets* são parte integrante da pesquisa contemporânea de reconhecimento de objetos. Alguns conjuntos de dados começaram com a captura de dados com o objetivo de representar o mundo visual (TORRALBA; EFROS, 2011). O *Kaggle*, por exemplo, é uma das comunidades populares no meio científico e é subsidiária da Google. A plataforma armazena e disponibiliza um conjuntos de dados sobre assuntos diversos e possui fóruns para troca de conhecimentos entre seus usuários.

Devido ao aumento da computação de ponta para aplicativos de inteligência e aprendizado de máquina, novas arquiteturas foram criadas para lidar com essas cargas de trabalho, em termos de tempo e energia, de maneira eficiente. Os dispositivos reduzem significativamente os recursos do dispositivo, como memória, computação e energia exigidos pelo aprendizado profundo que atualmente atuam como um forte gargalo para a adoção de dispositivos móveis. O uso da *Edge Computing* é uma estratégia conveniente para executar tarefas de aprendizagem próximo dos sensores, reduzindo a latência e melhorando os tempos de resposta (SATYANARAYANAN, 2017).

Os aceleradores de inferência USB fazem parte do *Edge Computing* e são assim chamados, pois permitem fácil integração em qualquer sistema Linux através das portas USB (Universal Serial Bus) e aumentam o poder de inferência em dispositivos de baixo desempenho, como por exemplo, um dispositivo *IOT* (Internet of Things) reduzindo o risco de exposição de dados confidenciais porque atribui localmente a capacidade de processamento (LIBUTTI et al. 2018). Assim, as empresas efetuam um controle maior da propagação de informações e conseguem estar de acordo com as políticas regulatórias, como o GDPR, conhecido como LGPD no Brasil (ESPOSITO et al., 2017).

Através de uma Revisão Sistemática da Literatura (RSL), foram selecionados alguns trabalhos relevantes e alinhados com o tema deste trabalho. Seguindo a metodologia da RSL (KITCHENHAM; CHARTERS, 2007), foram selecionadas 7 (sete) publicações, após a realização da RSL, os quais estão representados abaixo, na tabela 1, enfatizando suas principais características.

Tabela 1 – Comparativo entre os Trabalhos Relacionados.

TRABALHO	AUTOR	METODOLOGIAS	TECNOLOGIAS
A Deep Learning System for Recognizing Facial Expression in Real-Time	<u>MIAO et al.</u> (2019)	<i>Transfer Learning e Joint Supervision</i>	<i>OpenCV, Haar Cascade Classifier, TensorFlow, datasets FER-2013, CK+, JAFFE, Processador Intel Xeon E5-1603 com GPU NVIDIA Quadro K4200.</i>
Performance Analysis of	KHANAL et al. (2018)	<i>Transfer Learning</i>	<i>Nuvem da Google e da Microsoft com o dataset KDEF.</i>
Deep learning-based face analysis system for monitoring customer interest	YOLCU et al. (2020)	<i>Transfer Learning</i>	<i>Algoritmo de Viola e Jones, AdaBoost, Haar Cascade Classifier, Datasets RaFD e KDEF, Hardware Raspberry PI 2.</i>
A Survey on: Facial Emotion Recognition Invariant to Pose, Illumination and Age	BHATTACHARYA; GUPTA (2019)	<i>Transfer Learning</i>	<i>Algoritmo de Viola e Jones, AdaBoost, Active Shape Model (ASM) e Datasets CMU-MultiPIE e Amrita Emotion</i>
Real Time Emotion Recognition from Facial Expressions Using CNN Architecture	OZDEMIR et al. (2019)	<i>Transfer Learning</i>	<i>Arquitetura LeNet, Datasets JAFFE e KDEF, Haar Cascade Classifier, Keras, TensorFlow, Hardware Intel Core I7.</i>
Extended deep neural network for facial emotion recognition	JAIN; SHAMSOLMOALI; SEHDEV (2019)	<i>Transfer Learning</i>	<i>Datasets da CK+ e JAFFE Hardware Intel Core i7.</i>
A Customer Emotion Recognition through Facial Expression using POEM descriptor and SVM classifier	BOUZAKRAOUI; SADIQ; ENNEYA (2017)	<i>Transfer Learning Machine Learning</i>	<i>Haar Cascade Classifier, algoritmo POEM, (Support Vector Machine), Dataset Jaffe, Hardware Intel Core i7.</i>

Fonte: Elaborado pelo autor.



No trabalho de Miao et al. (2019) é apresentado um sistema de reconhecimento de expressão facial em tempo real que é capaz de reconhecer as expressões faciais através de uma webcam. A metodologia proposta combina uma estratégia de aprendizado de transferência (*Transfer Learning*) e supervisão conjunta (*Joint Supervision*), que é capaz de reconhecer expressões faciais de vários indivíduos simultaneamente.

No artigo de Khanal et al. (2018), os autores apresentam uma análise sobre a performance das API's disponíveis na internet para reconhecimento de emoções independente do ângulo da face. Fizeram um teste carregando as imagens de treinamento do dataset KDEF. A Microsoft não detectou quase nenhuma das imagens com o perfil completo esquerdo e direito, mas o Google detectou quase todas. A API da Microsoft apresenta um valor positivo verdadeiro médio de até 60%, enquanto que o Google apresenta o valor positivo verdadeiro máximo de 45,25%.

Em Yolcu et al. (2020), os autores explicam o funcionamento de um sistema de reconhecimento de emoções e positivamente de clientes. Basicamente cada imagem de rosto é cortada para reduzir as informações de segundo plano e reter apenas os recursos específicos de expressão e pose de cabeça, conforme o algoritmo citado por Viola e Jones (2004). Os resultados dos testes experimentais alcançaram 99,90% de precisão na estimativa de pose da cabeça e 94,61% de precisão no reconhecimento da expressão facial com um tempo de aproximadamente 0.41 segundos por cada imagem. O hardware utilizado não foi informado.

O trabalho de Bhattacharya e Gupta (2019) apresenta pesquisas sobre algoritmos que reconhecem as emoções faciais independente da posição da face, da iluminação e da idade. A ideia central do artigo é desenvolver um sistema que capture as três restrições ao mesmo tempo. Foram utilizados dois datasets o *CMU- MultiPIE* (que é um conjunto de dados com mais de 750 mil imagens de 337 pessoas em 15 pontos de vista e 19 condições de iluminação) e *Amrita Emotion* (que é uma técnica para reconhecimento de emoções em tempo real a partir de expressões faciais em imagens com variação de pose, iluminação e idade) com as emoções básicas com extração de características baseada em recursos geométricos. De acordo com os autores, o sistema funcionou bem, executando em uma Raspberry PI 2 e a acurácia foi de 95% para o *dataset CMU\_MultiPIE* e 92% para o *Amrita Emotion*.

Ozdemir et al. (2019) apresentam um estudo sobre a arquitetura *LeNet* (que é uma arquitetura simples e pequena de rede neural convolucional proposta por LeCun) para reconhecimento da expressão facial. Neste trabalho, os autores mesclaram três conjuntos de



dados, o *JAFFE* (213 imagens), o *KDEF* (4900 imagens) e um outro conjunto de dados personalizado (140 imagens) com diferentes expressões faciais. Os pixels foram removidos das expressões faciais através da biblioteca do *Haar Cascade* e, em seguida, efetuou-se o treinamento da arquitetura LeNet para a classificação dos estados emocionais. Foi utilizado o Keras e o TensorFlow para treinar a rede com um hardware Intel Core I7. O sistema atingiu uma precisão de 96,43% e validação de 91,81% para a classificação de 7 emoções diferentes por meio de expressões faciais através de uma webcam em 30 frames por segundo.

Em Bouzakraoui; Sadiq; Enneya, (2017) é apresentada uma implementação de um sistema de reconhecimento de emoções em tempo real para medir a satisfação dos clientes que frequentam uma loja. Eles explicam que posicionam uma câmera em algum lugar que capture o máximo possível do comportamento dos clientes que frequentam e testam os produtos, efetuam a filmagem pelo tempo que o estabelecimento está em funcionamento, extraem os frames da gravação e os processam de duas formas, primeiro detecta a face e depois as emoções. Para a identificação da face, utilizaram o classificador de *Haar Cascade*, para a extração das informações da face e o tratamento da imagem utilizaram o algoritmo *POEM* (Padrões de Magnitudes da Borda Orientada), que é um algoritmo para detecção do contorno da face. Para a classificação da emoção utilizaram o *SVM* (*Support Vector Machine*) através da base de dados *JAFFE* com 213 imagens faciais treinadas (6 expressões básicas e 1 neutra). A performance foi boa e a taxa de reconhecimento das emoções foi de 97% para satisfeito e por volta de 98% tanto para não satisfeito quanto para neutro.

Vários problemas evidenciados são “comuns” na área de Visão Computacional, como por exemplo, o ângulo da face, a iluminação, a definição e a idade do indivíduo. Além da qualidade da imagem, o hardware de processamento e a coleta das imagens (definição e frames por segundo) influenciam muito na tarefa de detecção da emoção; quanto melhor, mais preciso e confiável será o resultado.

Na detecção do rosto, a maioria das publicações utilizaram o algoritmo de Viola e Jones (2004). Para remover os pixels fora do contorno do rosto, utilizaram a biblioteca *Haar Cascade* e para a extração das características, alguns utilizaram o algoritmo *Adaboost* e outros o *POEM*. Para a classificação da emoção, alguns utilizaram o *POEM* (97% de acurácia), outros, o LeNet (91,81%), CK+ (96,92%) e *JAFFE* (95,24%), RaFD (94,61%), com destaque neste último para a acurácia e tempo curto para detecção, sendo possível detecção em tempo real. As APIs em nuvem não demonstraram uma boa acurácia no reconhecimento

da emoção. Para abertura do streaming e tratamento do frame, o software utilizado foi uma biblioteca chamada OpenCV, treinamento da rede, sendo que a maioria utilizou o Tensorflow e o Keras. O hardware de processamento varia entre raspberry pi 2, core i7 e um Xeon-e5 com processamento em GPU. Informações sobre as câmeras de captura e a posição de instalação não foram divulgadas, sendo que os que informaram, utilizaram uma webcam somente para testes e uma visão frontal.

### 3 MATERIAL E MÉTODOS

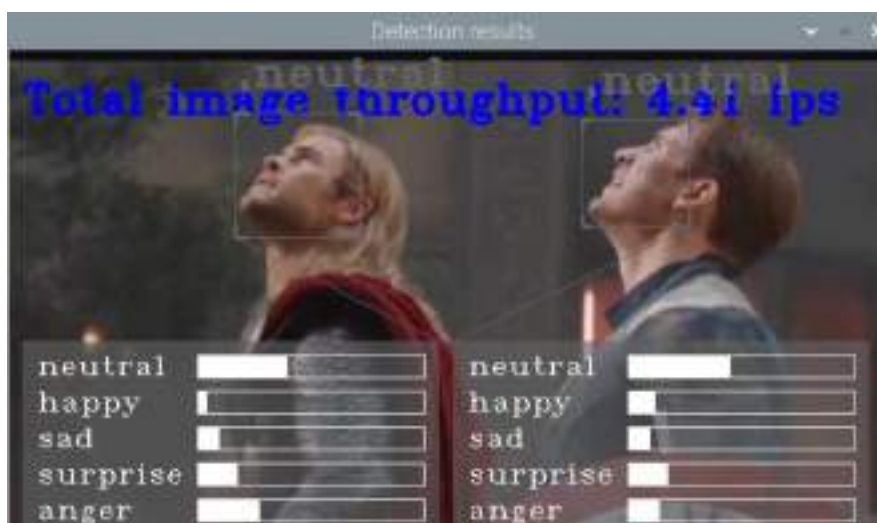
Para efetuar os testes de inferência foram definidos e selecionados alguns equipamentos que foram adquiridos baseados nos testes efetuados por Sterckval (2019) e serão listados a seguir:

1. Raspberry PI 3 model B;
2. NVIDIA Jetson Nano 4GB;
3. Google Coral USB Accelerator;
4. Intel Neural Compute Stick 2;
5. Webcam USB;
6. Câmera IMX219 HBVCAM-HPLCC-8M-130;
7. Caixa de proteção Jetson Nano + kit de suporte para câmera.

#### 3.1 CONFIGURAÇÃO DOS EQUIPAMENTOS

Através de uma Raspberry Pi, foi instalado o sistema OpenVino e um acelerador de inferência Intel Neural Stick, onde foram efetuados alguns testes de inferência, conforme mostra a Figura 2. É possível verificar o quanto da emoção *neutral* foi detectado e que o número de quadros por segundo (FPS) é de 4.41.

Figura 2 - Execução de Inferência com o OpenVino.



Fonte: Elaborado pelo autor.

Um programa em linguagem python foi usado para teste de performance em uma Raspberry Pi e a saída do programa está ilustrada na Figura 3, parte a, que mostra a emoção *sad* e o valor dos quadros por segundo (FPS) de 6.9. Outro teste de inferência sem aceleradores foi efetuado utilizando uma máquina virtual. Executado o programa através do comando `python3 videotester.py`, observa-se a saída do programa, conforme a Figura 3 b.

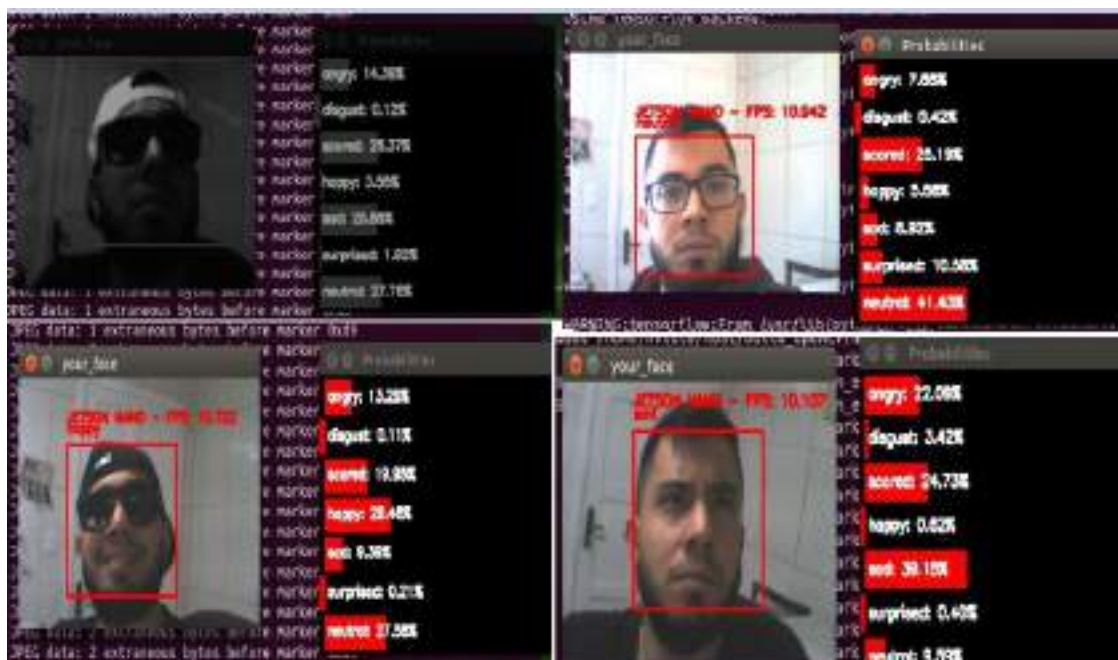
Figura 3 - Programa Python detectando a emoção *sad* e *angry*.



Fonte: Elaborado pelo autor.

A placa Jetson Nano foi utilizada para testes de inferência através de um programa em python. Os resultados estão descritos na figura 4 e o valor dos quadros por segundo (FPS) é de aproximadamente 10.

Figura 4 - Executando inferência na Jetson Nano



Fonte: Elaborado pelo autor.

### 3.2 COMPARAÇÃO DOS RESULTADOS DOS TESTES

Os testes foram necessários para verificar quais as melhores alternativas custo-benefício de *hardware* estavam disponíveis no mercado e, através deles, escolher qual a melhor placa para prototipagem.

Posteriormente, foi efetuado um ensaio utilizando a Nvidia Jetson Nano e, através dela, foi possível conferir uma boa performance e identificação de emoções executando de 10 a 12 quadros por segundo. Este último teste está em consonância com os testes efetuados por Sterckval (2019) e Bangash (2020) que obtiveram uma excelente performance nos testes.



Em relação aos testes realizados, observou-se que quanto melhor a qualidade de definição da imagem e o hardware de processamento, mais preciso e confiável será o resultado da inferência da emoção. Porém, existem agravantes para inferência, como por exemplo, o ângulo da face, a iluminação do ambiente, a idade do indivíduo e as oclusões da região dos olhos e da boca como uso de óculos, bonés e máscaras. Durante os testes, utilizando a Nvidia Jetson Nano, foi possível conferir uma boa performance e identificação de emoções executando de 10 a 12 quadros por segundo, sugerindo então que a melhor alternativa seria o uso da Jetson Nano.

#### **4 DESENVOLVIMENTO DO PROTÓTIPO**

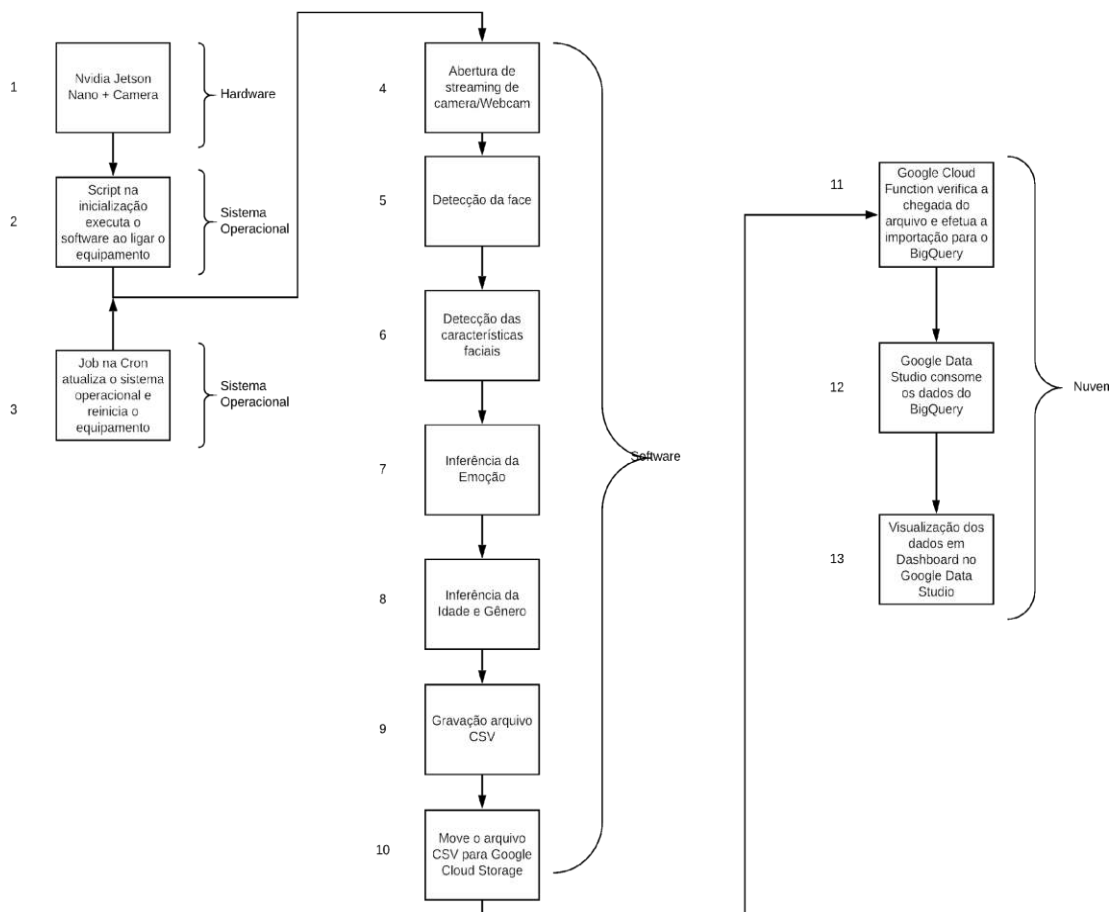
O objetivo deste capítulo é apresentar o desenvolvimento executado para a implementação de um protótipo. Motivado pelos trabalhos relacionados e conhecimentos adquiridos nas diferentes abordagens identificadas a partir da revisão sistemática da literatura e na fundamentação teórica, será demonstrado o detalhamento e a preparação do protótipo final com as respectivas considerações sobre os resultados. O fluxo completo da solução desenvolvida está ilustrado na Figura 5.

A partir do descrito anteriormente, construiu-se um fluxo completo da solução desenvolvida:

1. A Jetson Nano foi o hardware escolhido devido aos bons resultados apresentados nos testes.
2. O script faz parte de um serviço do Sistema Operacional para garantir que o software execute enquanto o equipamento estiver ligado ou for reiniciado.
3. Através do Script cadastrado na *Cron* do Sistema Operacional, será efetuado a atualização e o reinício do dispositivo para evitar que fique algum processo preso na memória RAM e na memória cache.
4. A abertura do Streaming da Câmera é uma ação básica necessária para o sistema poder capturar as imagens em tempo real.
5. A detecção da face é uma parte que seleciona apenas a região de interesse da pessoa que está na imagem.
6. A detecção das características é necessária para extrair as informações das regiões

da boca e das sobrancelhas.

Figura 5 - Fluxo completo da solução



Fonte: Adaptado de OLIVEIRA; JAQUES (2008).

7. Com as informações da região da imagem, que se encontra a boca e a sobrancelha, é efetuado a inferência através da rede neural.
8. As características faciais extraídas são utilizadas para inferir gênero e idade.
9. O tipo de gravação do arquivo que foi escolhido, foi o CSV pelo fato de ser simples de gravar localmente, por não ser necessário conexão com a internet e ser compatível com ferramentas de importação de dados para banco de dados.
10. A transferência do arquivo para a nuvem é necessária para ativar uma automação de inserção no banco de dados para ter uma centralização de todas informações.



11. A Google Cloud Function (Função na Nuvem) é uma ferramenta que efetua a automação de inserção do arquivo CSV no banco de dados Google BigQuery.
12. O Google Data Studio é uma ferramenta gratuita para montagem de painéis e visualização de dados com uma interface amigável e integração nativa com o BigQuery.

Google Data Studio possibilita a visualização dos dados em qualquer lugar, a qualquer momento, sendo possível a visualização tanto no smartphone quanto num navegador de internet. As placas de acrílico adquiridas vieram desmontadas e com um adesivo protetor, que foram removidos e após montados intuitivamente, visto que o produto não vem com manual de instrução, formando uma caixa. Os parafusos menores foram usados para fixar a Jetson Nano na base de acrílico e os parafusos maiores foram usados para fechar a caixa. Esta proteção de acrílico evita que os contatos fiquem expostos e facilita a fixação em algum local.

O protótipo finalizado, conforme a Figura 6 parte a, seria fixado no topo superior direito da vitrine da loja, com a câmera direcionada para os transeuntes. A simulação de como ficaria disposto na loja está descrito na Figura 6 parte b.

Figura 6 – Simulação de Disposição na Loja.



Fonte: Elaborado pelo autor.

Para ser possível executar e obter o pleno funcionamento do Software, é necessário efetuar a instalação e a configuração do Sistema Operacional, conforme anunciado pela



fabricante da placa, e a instalação das bibliotecas do Python. Para garantir que o Sistema Operacional se mantenha atualizado e seguro, foi necessário criar um script que efetua a atualização e reinicia o equipamento, efetua o agendamento na “cron” para a execução do *script* em um horário determinado, neste caso, às 9 horas da manhã.

De maneira geral, o algoritmo recebe os dados contidos em uma pasta, efetua a abertura do vídeo em tempo real da câmera, prepara o conjunto de dados, realiza um laço de repetição quadro a quadro do vídeo, interpreta os resultados obtidos, efetua a predição de face, idade, gênero, emoção e grava um arquivo com os resultados. O algoritmo de transferência utiliza a biblioteca do *google-cloud-storage*, *Linha5*, para efetuar a transferência dos arquivos. É efetuado o carregamento da chave com a extensão “.JSON”, *Linha 7*, e é instanciado um cliente do “*storage*”, *Linha 9*. Efetuado a declaração do nome do diretório criado anteriormente, através da variável “*bucketName*”, *Linha 12*, a variável “*localFolder*” recebe o caminho do sistema operacional onde está situado o próprio algoritmo, *Linha 13*. Para todos os arquivos “.CSV” que estão no diretório é efetuado o *upload* para o *bucket* através da função “*upload\_from\_filename*”, *Linha 24*, e a exclusão do arquivo, como ilustrado na *Figura 7*.

O *Google Cloud Storage* é um serviço de armazenamento de arquivos. Além disso, ele combina o desempenho e a escalabilidade da nuvem do Google com recursos avançados de compartilhamento e segurança. Foi efetuado a criação de um diretório (*bucket*) para armazenamento dos arquivos em formato *CSV*.

As Funções em Nuvem (*Google Cloud Functions*) são serviços que oferecem maneiras de executar um código, desenvolvido em uma linguagem específica, em que não há necessidade de provisionar servidores, gerenciar, corrigir ou atualizar. As funções são dimensionadas automaticamente e são altamente disponíveis e tolerantes a falhas. Para efetuar a importação dos arquivos do *Google Cloud Storage* para o *Google Bigquery* foi efetuada a criação de uma *Cloud Function*, como a que segue: - *deploy datatransfer --runtime python37 --trigger-resource ml\_datatransfer --trigger-event google.storage.object.finalize*

Figura 7 - Programa que efetua a transferência dos arquivos para o *Storage*.

```

1 import os
2 import sys
3 from os import listdir
4 from os.path import isfile, join
5 from google.cloud import storage
6
7 os.environ["GOOGLE_APPLICATION_CREDENTIALS"]="/home/gabriel/Downloads/people_analytics/people/corporativo-212121-0c3780f29940.json"
8
9 client = storage.Client()
10 bucketName = "ml_datatransfer"
11 bucketFolder = "fotos"
12 bucket = client.get_bucket(bucketName)
13 localFolder = os.getcwd()
14
15 def upload_files(bucketName):
16     """Upload files to GCP bucket."""
17     try:
18         files = [f for f in listdir(localFolder) if isfile(join(localFolder, f))]
19         uploaded = []
20         for file in files:
21             if ".csv" in file:
22                 localFile = localFolder + file
23                 blob = bucket.Blob(file)
24                 blob.upload_from_filename(localFile)
25                 uploaded.append(file)
26                 os.system('rm -f {}'.format(file))
27         return f'Uploaded {uploaded} to '{bucketName}' bucket'
28     except:
29         print('Error on upload')
30
31 if __name__ == '__main__':
32     retorno = upload_files(bucketName)
33     print(retorno)
34
35

```

Fonte: Elaborado pelo autor.

O *Google Bigquery* é um serviço de armazenamento de dados utilizado para a análise de grandes volumes de dados. As tabelas de dados são organizadas pelo *BigQuery* em unidades chamadas de conjuntos de dados. Os conjuntos de dados (*dataset*) podem ser consultados usando uma versão de consultas padrão SQL (*Structured Query Language*), que permite a análise abrangente de um grande conjunto de dados usando consultas manuais ou ferramentas de visualização. Foi criado um dataset para armazenamento dos dados e os campos criados foram nomeados como:

- 1.<Data>: armazena a data e hora exata da gravação (*timestamp* - Tempo independente de qualquer fuso horário ou calendário, representado em nanossegundos).
- 2.<Pessoas>: armazena o número de pessoas detectadas.
- 3.<Emocao>: armazena qual a emoção detectada.
- 4.<Percentemocao>: armazena o percentual de certeza que o modelo teve para identificar a emoção.

5.<Genero>: armazena o gênero da pessoa detectada.

6.<Percentgenero>: armazena o percentual de certeza que o modelo teve para detectar o gênero da pessoa.

7.<Faixaetaria>: armazena a faixa etária do indivíduo.

O *Google Cloud Data Studio* é uma ferramenta gratuita para visualização dos dados. Através dela é possível transformar os dados em relatórios e painéis informativos, fáceis de ler e de compartilhar e totalmente personalizável. O painel criado para visualização dos dados está ilustrado na Figura 8.

Figura 8 - Painel criado para visualização dos dados.



Fonte: Elaborado pelo autor.

## 5 ANÁLISE DE RESULTADOS

Devido à pandemia por COVID-19, com o isolamento social e o uso de máscaras para proteção, não foi possível efetuar a instalação do equipamento em alguma loja física e testar, pois além de não ter uma boa assertividade para detectar a emoção com máscara (oclusão da região da boca), os resultados seriam distorcidos da realidade e, por isso, o foco mudará para simulação. Para futuras implementações e experimentação do projeto, foram iniciadas conversas com alguns interessados, mas, por enquanto, será efetuado a análise de vídeos de

câmeras de segurança como forma de pré-validação do projeto. A seguir será demonstrado o funcionamento do Software.

Por meio do comando “*cv2.VideoCapture(0)*” é efetuada a abertura do *Streaming* (ao vivo) da câmera, conforme ilustrado na Figura 9 parte a. A detecção da face é um ponto muito importante a ser levado em consideração, pois alguns algoritmos apenas detectavam uma face e, desta maneira, através do *Yolo* (algoritmo para detecção de objetos baseado em redes neurais convolucionais), é possível detectar todas as faces do quadro da imagem e efetuar o tratamento. Para comprovar que o modelo detecta somente o rosto de pessoas, foi testado com a foto de um cachorro, conforme ilustrado na Figura 9, parte b.

Figura 9 - Abertura do Streaming da Câmera e Detecção.



Fonte: Elaborado pelo autor.

A Figura 10, parte a, ilustra os testes da emoção “surpreso” e “bravo”, juntamente com o percentual de certeza que o modelo obteve para tal finalidade. Para o teste de inferência de gênero e idade, a imagem utilizada está ilustrada na Figura 10, parte b.

Figura 10 - Inferência das Emoções Surpreso e Bravo, Idade e Gênero.



Neste teste, o modelo de detecção de gênero obteve uma acurácia de 90%. O Sistema acertou a identificação do gênero com taxas de 87% no caso *female* e 96% no *male*, e acertou também a faixa de idade através da imagem, já que o homem possui 25 anos e a mulher, 40 anos.

Os objetivos do trabalho foram buscar uma solução que integrasse a detecção facial de pessoas, a identificação de emoções com a possibilidade de escalar e obter os resultados independente da localização geográfica do equipamento, contando com uma boa luminosidade do local. Utilizando o hardware Jetson Nano, foi possível obter bons resultados de acurácia, conforme foi detalhado nas seções anteriores. Os experimentos demonstraram um bom resultado, o que era esperado no quesito pessoas e emoção. É necessário melhorar a acurácia na identificação de idade, mas como não é o foco deste trabalho, não terá muita ênfase nesta fase. O protótipo demonstrou estar no limite da sua capacidade de processamento para aplicações em tempo real e para melhorar os resultados, seria necessário um hardware mais potente.

## 6 CONSIDERAÇÕES FINAIS

Este trabalho identificou a necessidade de atualização dos métodos de satisfação do cliente em redes de varejo, avaliando a possibilidade da substituição de métodos convencionais de avaliação de satisfação por métodos visuais, usando técnicas de visão



computacional e análise de emoções faciais. Neste sentido, é proposto um protótipo de aplicativo para avaliar a satisfação dos clientes por meio da coleta e análise de dados através de câmeras de vídeo. O aplicativo parte de um fluxo de execução para coleta e análise de imagens para a identificação de características de clientes em relação a um produto com base em expressões faciais, gênero e faixa etária.

Um dos objetivos específicos era o uso de um equipamento que fosse o mais resiliente possível as interrupções de internet, de sistema e até de energia para que as informações sejam mais reais e condizentes com a realidade. O uso da Jetson Nano possibilita o uso de uma conexão a internet permanente e dedicada, evitando que as informações trafeguem pela rede e estejam sujeitas a interceptação, visto que as imagens analisadas possuem informações sensíveis e permite que o equipamento seja conectado a um banco de baterias ou gerador de energia.

Nas lojas virtuais existem várias métricas, como por exemplo, a taxa de conversão, que é utilizada para entender a eficiência da comunicação do site, que mostra onde está o ponto fraco e onde se deve atuar. Um exemplo disso é um cruzamento de dados de pessoas que visitaram o site da loja virtual versus quantas realmente compraram um produto.

Com esta solução, a partir do cruzamento das informações de quantidade de vendas do dia e de quantas pessoas passaram pela frente da loja, obtém-se uma métrica de taxa de conversão da loja física, sendo possível identificar onde está o ponto fraco da estratégia. Com estas informações em mãos, fica mais tangível os objetivos e as estratégias de vendas.

Para trabalhos futuros, visando uma proposta de expansão, são necessárias algumas mudanças no protótipo e no ambiente de nuvem, como por exemplo, a adição de um campo de identificação de cada dispositivo para efetuar uma filtragem por dispositivo. Objetivando uma melhoria nos resultados, é necessário um equipamento com maior poder de processamento, visando efetuar o rastreamento das pessoas e garantir uma métrica mais precisa de emoção por pessoa e, para aumentar a assertividade na inferência da idade e emoção, é necessário mais treinamento do conjunto de dados. De caráter utópico, o rastreamento de íris seria ideal para a verificação de qual produto mais chamou atenção dos transeuntes, mas, atualmente, só é possível através de um equipamento, como por exemplo, um óculos de realidade aumentada.



## REFERÊNCIAS

BANGASH, Imran. **NVIDIA Jetson Nano vs Google Coral vs Intel NCS: A Comparison**. 2020. Disponível em: <https://towardsdatascience.com/nvidia-jetson-nano-vs-google-coral-vs-intel-ncs-a-comparison-9f950ee88f0d>. Acesso em: 22 jul. 2020.

BENGIO, Yoshua; COURVILLE, Aaron; VINCENT, Pascal. Representation Learning: a review and new perspectives. **IEEE Transactions On Pattern Analysis And Machine Intelligence**, [s. l], v. 35, p. 1798-1828, ago. 2013. Mensal. Disponível em: <https://ieeexplore.ieee.org/document/6472238>. Acesso em: 22 jul. 2020.

BHATTACHARYA, Saswati; GUPTA, Mousumi. A Survey on: Facial Emotion Recognition Invariant to Pose, Illumination and Age. *In: SECOND INTERNATIONAL CONFERENCE ON ADVANCED COMPUTATIONAL AND COMMUNICATION PARADIGMS (ICACCP)*. 2019, Índia. **Anais eletrônicos...** Índia, 25-28 fev. 2019, p. 1-6. Disponível em: <https://ieeexplore.ieee.org/document/8883015>. Acesso em: 22 jul. 2020.

BOUZAKRAOUI, Moulay Smail; SADIQ, Abdelalim; ENNEYA, Nourddine. A Customer Emotion Recognition through Facial Expression using POEM descriptor and SVM classifier. *In: BDCA'17: PROCEEDINGS OF THE 2ND INTERNATIONAL CONFERENCE ON BIG DATA, CLOUD AND APPLICATIONS*. 2017, Marrocos. **Anais eletrônicos...** Marrocos, mar. 2017, p. 1-6. Disponível em: <https://dl.acm.org/doi/10.1145/3090354.3090436>. Acesso em: 22 jul. 2020.

CARNEGIE MELLON UNIVERSITY (Pittsburgh). **CMU-MultiPIE**. Disponível em: <http://www.cs.cmu.edu/afs/cs/project/PIE/MultiPie/Multi-%20Pie/home.html>. Acesso em: 28 jun. 2020.

CECCACCI, Silvia et al. An Emotion Recognition System for monitoring Shopping Experience. *In: PETRA '18: PROCEEDINGS OF THE 11TH PERVASIVE TECHNOLOGIES RELATED TO ASSISTIVE ENVIRONMENTS CONFERENCE*. 2018, Grécia. **Anais eletrônicos...** Grécia, jun. 2018, p. 102-103. Disponível em: <https://dl.acm.org/doi/10.1145/3197768.3201518>. Acesso em: 23 jul. 2020.

EKMAN, Paul; FRIESEN, Wallace V.. Measuring facial movement. **J Nonverbal Behav** 1, 1976, p. 5 6–75. Disponível em: <https://link.springer.com/article/10.1007/BF01115465>. Acesso em: 25 jul. 2020.

ESPOSITO, Christian et al. Challenges of Connecting Edge and Cloud Computing: A Security and Forensic Perspective. 2017. *IEEE Cloud Computing*, 4(2), p. 13–17. Disponível em: [https://www.researchgate.net/publication/316501480\\_Challenges\\_of\\_Connecting\\_Edge\\_and\\_Cloud\\_Computing\\_A\\_Security\\_and\\_Forensic\\_Perspective](https://www.researchgate.net/publication/316501480_Challenges_of_Connecting_Edge_and_Cloud_Computing_A_Security_and_Forensic_Perspective). Acesso em: 25 jul. 2020.





GONZALEZ, Rafael C.; WOODS, Richard E.. **Processamento digital de imagens**. 3. ed. São Paulo: Pearson, 2010.

**GOOGLE**. Google Bigquery para usuários de armazenamento de dados. **Disponível em:** <https://cloud.google.com/solutions/bigquery-data-warehouse?hl=pt-br>. **Acesso em: 25 jun. 2020.**

**GOOGLE**. **Google Cloud Functions**. Disponível em: <https://developers.google.com/learn/topics/functions>. Acesso em: 25 jun. 2020.

**GOOGLE**. **Google Cloud Storage**. Disponível em: <https://cloud.google.com/storage>. Acesso em: 24 jun. 2020.

**GOOGLE**. **Google Cloud Data Studio**. Disponível em: <https://support.google.com/datastudio/answer/6283323?hl=pt-BR>. Acesso em: 24 jun. 2020.

**GOOGLE**. **Timestamp**. Disponível em: <https://developers.google.com/android/reference/com/google/firebase/Timestamp>. Acesso em: 25 jun. 2020.

JAIN, Deepak Kumar; SHAMSOLMOALI, Pourya; SEHDEV, Paramjit. **Extended deep neural network for facial emotion recognition**. Pattern Recognition Letters, V. 120, 2019, p. 69-74, ISSN 0167-8655. Disponível em: <https://www.sciencedirect.com/science/article/abs/pii/S016786551930008X?via%3Dihub>. Acesso em: 23 jun. 2020.

**KAGGLE**. **Kaggle**: your machine learning and data science community. your machine learning and data science community. 2020. Disponível em: <https://www.kaggle.com/>. Acesso em: 30 jun. 2020.

KHANAL, Salik Ram et al. Performance analysis of Microsoft's and Google's Emotion Recognition API using pose-invariant faces. *In: DSAI 2018: proceedings of the 8th international conference on software development and technologies for enhancing accessibility and fighting info-exclusion*. **Proceedings...** Association for Computing Machinery, New York, p. 172-178. Disponível em: <https://dl.acm.org/doi/10.1145/3218585.3224223>. Acesso em: 27 jul. 2020.

KITCHENHAM, Barbara; CHARTERS, Stuart. **Guidelines for performing Systematic Literature Reviews in Software Engineering**. Reino Unido: Kitchenham, 2007, p. 65. Relatório conjunto da Universidade Keele e da Universidade Durham. Disponível em: <https://www.bibsonomy.org/bibtex/aed0229656ada843d3e3f24e5e5c9eb9>. Acesso em: 28 jul. 2020.

LECUN, Yann; BENGIO, Yoshua; HINTON, Geoffrey. Deep learning. **Nature** **521**, Reino Unido, maio 2015, p. 436-444. Disponível em: <https://www.nature.com/articles/nature14539#citeas>. Acesso em: 08 dez. 2020.

LECUN, Yann et. al. **Gradient-Based Learning Applied to Document Recognition**. 1998. Disponível em: <http://yann.lecun.com/exdb/publis/pdf/lecun-01a.pdf>. Acesso em: 25 jun. 2020.

LIBUTTI, Leandro Ariel et al. **Benchmarking Performance and Power of USB Accelerators for Inference with MLPerf**. 2018. Disponível em: [http://workshops.inf.ed.ac.uk/accml/papers/2020/AccML\\_2020\\_paper\\_4.pdf](http://workshops.inf.ed.ac.uk/accml/papers/2020/AccML_2020_paper_4.pdf). Acesso em: 08 dez. 2020.

MARKET, MARKETS AND MARKETS. **Market Search Report**. Junho, 2019. Disponível em: <https://www.marketsandmarkets.com/Market-Reports/facial-recognition-market-995.html>. Acesso em: 30 jun. 2020.

MEHRABIAN, Albert; RUSSELL, James A.. **An Approach to Environmental Psychology**. Cambridge, MA: MIT Press, 1974.

MENEGOLA, Afonso et al. Knowledge transfer for melanoma screening with deeplearning. *In: IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017)*, 2017, Austrália. **Anais...** Austrália: IEEE, 2017, p. 297-300. Disponível em: <https://ieeexplore.ieee.org/document/7950523>. Acesso em: 29 jul. 2020.

MIAO, Yu et al. A Deep Learning System for Recognizing Facial Expression in Real-Time. **ACM Transactions on Multimedia Computing, Communications, and Applications**, 2019, New York: Association for Computing Machinery, Appl. 15, 2, Article 33, June 2019. Disponível em: <https://dl.acm.org/doi/fullHtml/10.1145/3311747>. Acesso em: 29 jul. 2020.

MITCHELL, Tom M. Does machine learning really work?. **AI magazine**, v. 18, n. 3, p. 11-11, 1997. Disponível em: <https://ojs.aaai.org/index.php/aimagazine/article/view/1303#:~:text=Abstract-,Does%20machine%20learning%20really, Yes.&text=Newer%20research%20is%20beginning%20to,and%20life%20long%20cumulative%20learning>. Acesso em: 20 jul. 2020.

NVIDIA CORPORATION (Brasil). **NVIDIA JETSON NANO**: levando o poder da IA moderna a milhões de dispositivos. Levando o Poder da AI Moderna a Milhões de Dispositivos. 2020. Disponível em: <https://www.nvidia.com/pt-br/autonomous-machines/embedded-systems/jetson-nano/>. Acesso em: 30 jun. 2020.

OLIVEIRA, Eduardo de; JAQUES, Patrícia Augustin. Inferindo as emoções do usuário pela face através de um sistema psicológico de codificação facial. *In: IHC: PROCEEDINGS OF THE VIII BRAZILIAN SYMPOSIUM ON HUMAN FACTORS IN COMPUTING*

SYSTEMS. Porto Alegre, 2008. **Anais eletrônicos...** Porto Alegre, out. 2008, p. 156-165. Disponível em: <https://dl.acm.org/doi/10.5555/1497470.1497488>. Acesso em: 23 jul. 2020.

OZDEMIR, Mehmet Akif et al. **Real Time Emotion Recognition from Facial Expressions Using CNN Architecture**. 2019. Medical Technologies Congress (TIPTEKNO), Izmir, Turkey, 2019, p. 1-4. Disponível em: <https://ieeexplore.ieee.org/document/8895215>. Acesso em: 23 jul. 2020.

REDMON, Joseph. **You Only Look Once: Unified, Real-Time Object Detection**. 2015. Disponível em: <https://arxiv.org/abs/1506.02640>. Acesso em: 19 jul. 2020.

SATYANARAYANAN, Mahadev. **The emergence of edge computing**. Computer 50(1), p. 30-39, 2017. Disponível em: <https://ieeexplore.ieee.org/document/7807196>. Acesso em: 22 jul. 2020.

STERCKVAL, Sam. **Google Coral Edge TPU vs NVIDIA Jetson Nano: A quick deep dive into EdgeAI performance**. 2019. Disponível em <https://blog.usejournal.com/google-coral-edge-tpu-vs-nvidia-jetson-nano-a-quick-deep-dive-into-edgeai-performance-bc7860b8d87a>. Acesso em: 22 jul. 2020.

SZELISKI, Richard. **Computer Vision: Algorithms and Applications**. 1. ed. Berlin, Heidelberg: Springer-Verlag, 2010. Disponível em: [http://szeliski.org/Book/drafts/SzeliskiBook\\_20100903\\_draft.pdf](http://szeliski.org/Book/drafts/SzeliskiBook_20100903_draft.pdf). Acesso em: 08 dez. 2020.

TECNOLOGIA, Fernando K (org.). **Inteligência Artificial IOT com Jetson Nano**. Disponível em: <https://www.fernandok.com/2019/05/inteligencia-artificial-iot-com-jetson.html>. Acesso em: 25 jun. 2020.

TIAN, Ying-li; KANADE, Takeo; COHN, Jeffrey F. **Recognizing action units for facial expression analysis**. IEEE Transactions On Pattern Analysis And Machine Intelligence, [S.l.], v. 23, n. 2, fev. 2001, p. 97-115. Disponível em: <https://ieeexplore.ieee.org/document/908962>. Acesso em: 25 jun. 2020.

TORRALBA, Antonio; EFROS, A. Alexei. **Unbiased look at dataset bias**. Providence, RI, 2011, p. 1521-1528. Disponível em: <https://ieeexplore.ieee.org/document/5995347>. Acesso em: 20 jun. 2020.

VERMA, Madhushi et al. **Real Time Age, Gender And Emotion Detection using Deep Learning Techniques**. Bennett University, Winter Internship Project Report. India, 2018. Disponível em: <https://github.com/LeadingIndiaAI/Real-Time-Age-Gender-and-Emotion-Detection-using-Deep-Learning-Techniques/blob/master/Project%20Report.pdf>. Acesso em: 09 nov. 2020.



VIOLA, Paul; JONES, Michael J.. **Robust real-time face detection**. Int. J.Comput. Vis. 57, 2004, p. 137-154. Disponível em: <https://doi.org/10.1023/B:VISI.0000013087.49260.fb>. Acesso em: 19 jun. 2020.

YOLCU, Gozde et al. **Deep learning-based face analysis system for monitoring customer interest**. J Ambient Intell Human Comput 11, 2020, p. 237- 248. Disponível em: <https://doi.org/10.1007/s12652-019-01310-5>. Acesso em: 19 jun. 2020.