

CENTRO FEDERAL DE EDUCAÇÃO TECNOLÓGICA DE MINAS GERAIS

MATTHEUS VELOSO DIAS

DESENVOLVIMENTO DE UMA REDE NEURAL DE VISÃO COMPUTACIONAL  
CONECTADA À NUVEM

Professor Orientador: Anthony Chiaratti

Professor Coorientador: Túlio Carvalho

Belo Horizonte

2019

CENTRO FEDERAL DE EDUCAÇÃO TECNOLÓGICA DE MINAS GERAIS

MATTHEUS VELOSO DIAS

DESENVOLVIMENTO DE UMA REDE NEURAL DE VISÃO COMPUTACIONAL  
CONECTADA À NUVEM

Monografia submetida para análise do orientador e da professora responsável pela disciplina Trabalho de Conclusão de Curso I como requisito para aprovação nesta e graduação no curso de Engenharia Elétrica do CEFET-MG.

Orientador: Anthony Chiaratti

Coorientador: Túlio Carvalho

Professora da Disciplina: Úrsula do Carmo

Belo Horizonte

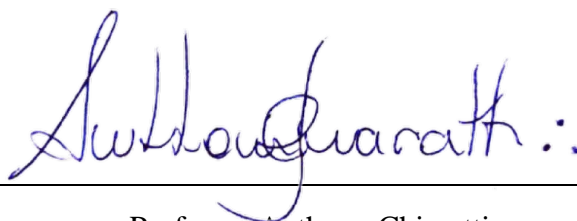
2019

## FOLHA DE APROVAÇÃO

Matheus Veloso Dias

Desenvolvimento de uma Rede Neural de Visão Computacional conectada à Nuvem

Monografia submetida para análise do orientador e da professora responsável pela disciplina Trabalho de Conclusão de Curso I como requisito para aprovação nesta e graduação no curso de Engenharia Elétrica do CEFET-MG.



---

Professor Anthony Chiaratti

Orientador / DEEB / CEFET-MG

---

Professora Úrsula do Carmo

Docente / DEE / CEFET-MG

Belo Horizonte

2019

## RESUMO

A inteligência artificial é uma tecnologia de crescente importância em todo o mundo. O Brasil, por sua vez, encontra-se defasado em relação aos países que estão na ponta do desenvolvimento desta. Desta forma, justificou-se a realização deste estudo para aprofundar o conhecimento da comunidade científica nacional sobre o tema.

Além disso, sistemas de visão computacional tem sido cada vez mais utilizados para realizar o reconhecimento de pessoas e conceder acesso a dados de vital importância. Desta forma o foco escolhido para este estudo foi compreender as inteligências artificiais focadas na realização de visão computacional para o reconhecimento facial.

Um sistema será aqui proposto para realizar tal tarefa e deve ser também conectado à nuvem, expandindo desta forma a sua aplicabilidade e inclusão em um ambiente IoT. Tal sistema deve ser desenvolvido utilizando-se como base a pesquisa já feita dentro da instituição.

Encontrou-se que a principal tecnologia utilizada para realização do reconhecimento facial é a de aplicação de uma metodologia denominada de redes neurais convolucionais, baseadas em *deep learning*. Estas funcionam de forma com que resolvem um problema de classificação em *machine learning* através da entrada de dados em formato RGB já pré-processada para isolar e centralizar o rosto a ser reconhecido.

Estes sistemas são aplicados através de sistemas na subárea de visão computacional denominada de “Reconhecimento”. Eles são auxiliados por sistemas de *machine learning* e de visão computacional clássica que preparam a imagem para ser reconhecida. Estes sistemas cuidam de problemas conhecidos como detecção, localização e rastreamento.

A partir disto analisou-se um sistema de reconhecimento facial já implementado dentro do CEFET-MG. Tal o sistema atingiu um alto grau de assertividade utilizando-se de redes neurais convolucionais unidas a um algoritmo em Haar Cascade. Ele foi implementado utilizando-se de Python, TensorFlow, OpenCV e *Raspberry Pi*.

Desta forma, o sistema foi brevemente reimplementado e foi proposta a continuidade do desenvolvimento: no TCC II será feito um aprimoramento da inteligência por trás do sistema em questão e será também desenvolvida uma conexão do sistema à nuvem.

**Palavras Chave:** Inteligência Artificial. *Machine Learning*. *Deep Learning*. Redes Neurais Convolucionais. Visão Computacional. Reconhecimento Facial. IoT.

## ABSTRACT

Artificial Intelligence is a trending technology whose importance is growing everyday. However, Brazil is still rather behind other countries who are leading cutting-edge research in this topic. Therefore, it is necessary for the Brazilian scientific community to deepen the research being done towards its artificial intelligence efforts, hence this study has been done in order to help fulfill this need.

Computer Vision systems that take care of facial recognition are being tasked with keeping vital data safe. Because of that, the focus chosen for this study was to comprehend the artificial intelligences charged with making the facial recognition software work.

A system that can successfully realize this task is put forward here. This system will also be connected to the cloud to further expand its applicability and inclusion in IoT environments. It will also be developed using a previous research already done inside the CEFET-MG institution as a starting point.

One of the findings of the present study was that the main technology used in facial recognition systems is called Convolutional Neural Network, which is deep learning based. It works by solving a machine learning classification problem after being fed RGB data (already pre-processed to isolate the face to be recognized).

These systems are applied under the computer vision subarea named “Recognition”. They are supported via the use of machine learning and classical computer vision algorithms that prepare the image to be recognized. These algorithms tasks are defined as detection, localization and tracking.

With the aforementioned acquired knowledge, a facial recognition system previously implemented inside CEFET-MG was deeply analyzed. This system reached a high recognition-rate by using a Haar Cascade algorithm together with a convolutional neural network. It was implemented using Python, TensorFlow, OpenCV and Raspberry Pi.

The software was briefly reimplemented and further development upon it is proposed. As the second part of this thesis (TCC II) an improvement shall be made in the artificial intelligence behind the facial recognition system and it shall be connected to the cloud.

**Keywords:** Artificial Intelligence. Machine Learning. Deep Learning. Convolutional Neural Networks. Computer Vision. Facial Recognition. IoT.

## SUMÁRIO

Folha de aprovação .....	3
Resumo .....	4
Abstract.....	5
Sumário.....	6
Lista de Figuras .....	9
Lista de Abreviaturas.....	10
Capítulo I    Introdução .....	11
I.1    Contextualização e relevância.....	11
I.2    Considerações sobre desafios de aceitação da tecnologia .....	13
I.3    Objetivos.....	14
I.4    Metodologia .....	14
Capítulo II    A Inteligência Artificial na Visão Computacional: Machine Learning, Deep Learning e Redes Neurais convolucionais .....	16
II.1    Introdução .....	16
II.2    Breve Histórico da I.A. ....	16
II.3    Estado da Arte.....	17
II.4    Funcionamento das redes neurais de classificação .....	18
II.4.1    Classificação versus Regressão .....	18
II.4.2    Introdução às Redes Neurais .....	18
II.4.3    O Deep Learning .....	21
II.4.4    Complexidade e métodos de prevenção .....	22
II.4.5    Tipos de Camadas.....	23
II.4.6    Tomada de decisão .....	27
II.4.7    Evolução dos dados em uma CNN .....	27
II.4.8    Retroalimentação .....	28

II.5	Aplicações na VC. ....	28
Capítulo III O Estado atual da tecnologia de Visão Computacional e do		
	Reconhecimento Facial.....	30
III.1	Introdução.....	30
III.2	O que é Visão Computacional.....	30
III.3	Aplicações da VC.....	30
III.4	A IA na Visão Computacional: Detecção, Localização, Classificação e Rastreamento .....	31
III.5	Os sistemas de Reconhecimento Facial .....	33
Capítulo IV Estudo de Implementação de uma Inteligência Artificial visando o		
	reconhecimento facial do CEFET-MG.....	35
IV.1	Introdução.....	35
IV.2	Estado Atual dos Projetos do CEFET-MG .....	35
IV.3	Tecnologias Utilizadas .....	36
IV.3.1	Python.....	36
IV.3.2	OpenCV .....	37
IV.3.3	Raspberry Pi .....	37
IV.3.4	TensorFlow;.....	37
IV.3.5	Haar Cascade;.....	37
IV.4	Implementação Básica do Projeto de (Brum & Oliveira, 2018) .....	38
IV.1	Propostas de desenvolvimento para o TCC II.....	39
IV.1.1	Compreender, com profundidade, a IA já implementada.....	39
IV.1.2	Conectar o sistema à nuvem .....	39
IV.1.3	Explicitar a certeza da tomada de decisão do softmax da IA .....	39
IV.1.4	Compreender o que causa o sistema a acusar o reconhecimento falso ....	40
IV.1.5	Aumentar a eficiência do sistema para funcionar com menos fotos na sua base de dados de treinamento .....	40

IV.1.6 Implementar retroalimentação para que o sistema possa melhorar sua performance ao longo do tempo .....	40
IV.1.7 Compreender o que seria necessário para este sistema também realizar a detecção de objetos .....	40
Capítulo V Conclusões.....	42
V.1 Considerações finais .....	42
Referências .....	43



## LISTA DE FIGURAS

Figura I.1 - Ilustração de sistema de reconhecimento facial, obtida de (Cruz, Coutinho, & Freire, 2019) .....	12
Figura II.1 - Visualização do funcionamento da regressão logística, adaptada de (Müller & Guido, 2016).....	19
Figura II.2 - Visualização das funções de ativação tangente hiperbólica (em azul) e da função <i>Relu</i> (unidade linear retificada, em vermelho), obtida de (Müller & Guido, 2016) .....	20
Figura II.3- Rede neural básica, adaptada de (Müller & Guido, 2016).....	21
Figura II.4 – Exemplo de rede neural convolucional completa, adaptada de (Saha, 2018) .....	21
Figura II.5 – Visualização dos dados presentes em cada neurônio de três camadas sequenciais de uma rede neural convolucional. Obtida de (Collins-Thompson, 2017) .	22
Figura II.6 - Representação de uma matriz 6x6x3 em formato RGB, obtida de (Ng, 2017).....	24
Figura II.7 - Convolução através de um filtro detector de quinas verticais, adaptada de (Cavaioni, 2018).....	25
Figura II.8 - <i>Max-Pooling</i> , obtida de (Cavaioni, 2018).....	26
Figura II.9 - Exemplo de CNN adaptado de (Cavaioni, 2018).....	27
Figura III.1 – Dois veículos detectados, localizados e rastreados, obtida de (Ng, 2017)	32
Figura III.2 – Frontalização das imagens de (Y. Taigman, 2014).....	34
Figura III.3 – Visualização da Rede Neural de (Y. Taigman, 2014).....	34
Figura IV.1 - Procedimento de visão computacional aplicado, obtida de (Cruz, Coutinho, & Freire, 2019) .....	35
Figura IV.2 - Vista do Raspberry Pi, obtida de (Brum, Oliveira, & Almeida, 2018) ....	37
Figura IV.3 - Filtros do Haar Cascade (Vidhya, 2019) .....	38
Figura IV.4 - Visualização numérica dos filtros da Figura IV.3 (Vidhya, 2019).....	38
Figura IV.5 - Execução do código de reconhecimento facial, autoria própria.....	38

## LISTA DE ABREVIATURAS

CEFET-MG	Centro Federal de Educação Tecnológica de Minas Gerais
CNN	Rede Neural Convolucional ( <i>Convolutional Neural Network</i> )
DL	<i>Deep Learning</i>
IA	Inteligência Artificial
IEEE	Instituto de Engenheiros Eletricistas e Eletrônicos
IoT	Internet das Coisas ( <i>Internet of Things</i> )
ML	<i>Machine Learning</i>
SVR	<i>Support Vector Regressor</i>
RGB	Forma de armazenamento de imagens em cores vermelha, verde e azul ( <i>Red, Green and Blue</i> )
TCC	Trabalho de Conclusão de Curso
VC	Visão Computacional
DCNN	Redes Neurais Convolutivas Profundas (DCNN)
ID	Identificador

## CAPÍTULO I INTRODUÇÃO

### I.1 CONTEXTUALIZAÇÃO E RELEVÂNCIA

O avanço tecnológico tem apresentado cada vez mais capacidades de melhorar a vida humana. Desde a invenção das primeiras máquinas à vapor até o presente momento tem-se visto um avanço científico de forma exponencial, de forma com que uma pessoa de classe média hoje tenha uma vida de qualidade superior àquela que possuía um bilionário um século atrás (Boudreaux, 2017).

É neste contexto que tecnologias passam da fase de descoberta até a fase de desenvolvimento e enfim a fase de ampla aplicação. Apesar de as duas primeiras duas fases recebem bastante atenção midiática (como por exemplo, na recente supremacia quântica supostamente alcançada pelo Google (Rincon, 2019)) é a terceira fase, a aplicação na vida das pessoas, que de fato impacta o mundo.

Por este motivo faz sentido buscar tecnologias já descobertas e em estágio avançado de desenvolvimento para que estas passem a ter um impacto maior ainda na sociedade. É justamente nesta situação que se encontra a tecnologia de Inteligência Artificial (IA), principalmente em termos de Brasil.

Ora, em países mais avançados tecnologicamente como Estados Unidos e China a IA já é aplicada por uma grande parcela das empresas de grande porte para resolver uma série de problemas e otimizar processos produtivos. Google, Alibaba, Baidu, Daimler, Bosch e Facebook são algumas das empresas que estão na ponta da pesquisa neste campo.

Entretanto, o mesmo não se nota na maior economia da América Latina. No Brasil faltam profissionais com conhecimento suficiente para realizar a implementação desta com qualidade. Além disso, empresas se mostram, em grande parte, desinteressadas pois ainda não compreenderam as vantagens que esta tecnologia pode vir a ter em seu negócio.

O mais interessante é que a IA tem a capacidade de ser aplicada em praticamente qualquer setor econômico: há hoje aplicações já em uso que vão de gerenciamento de campanhas de marketing (Tran, 2018) à execução de vitais funções da automação industrial (Kroll, Schaffranek, Schriegel, & Niggemann, 2014). Isso demonstra parte do que pode ser alcançado com esta.

Desse modo fica evidente a necessidade de expansão dos conhecimentos da comunidade científica nacional para com esta tecnologia. Além dela, é valioso explorar as outras tecnologias

que a ela estão interligadas. Isto se dá visto que o uso de algumas tecnologias que recentemente alcançaram um nível alto de maturidade comercial, como a Internet das Coisas (IoT) e a visão computacional (VC), pode proporcionar a criação de soluções antes inviáveis e revolucionar diversas indústrias.

Entre estas destacam-se a as aplicações dos sistemas de visão computacional em duas áreas principais: confirmação de identidade (exemplificada na Figura I.1) e veículos autônomos. Enquanto a primeira já se encontra em ampla e crescente utilização (Apple, 2019), a segunda ainda está em estágios anteriores de prontidão comercial, sendo de fato implementada em micro ecossistemas nos EUA e China (The Economist, 2019).

A confirmação de identidade é feita via uma aplicação da VC conhecida como reconhecimento facial. Hoje em dia, boa parte dos dados do mundo estão protegidos por estes. Celulares de topo de linha, por exemplo, já conseguem alcançar um grande nível de assertividade no reconhecimento de quem é e de quem não é dono de um celular. Estes prometem ser suficientemente seguros para proteger até mesmo informações em aplicativos que guardam dados de grande sensibilidade (Apple, 2019).



Figura I.1 - Ilustração de sistema de reconhecimento facial, obtida de (Cruz, Coutinho, & Freire, 2019)

Este é um motivo pelo qual este trabalho foi realizado. O objetivo não foi o de realizar uma pesquisa de ponta, mas sim de contribuir para o aprimoramento do conhecimento das comunidades local (dentro da instituição em que a pesquisa foi realizada) e nacional sobre os temas de IA, VC e reconhecimento facial.

No Capítulo I contextualiza-se o tema, são listados os objetivos práticos que desejou-se obter e é definida a metodologia a ser utilizada. No Capítulo II, por sua vez, é apresentada a

forma como funcionam as inteligências artificiais modernas, aprofundando-se no tópico de redes neurais convolucionais (CNN). Já no Capítulo III é explorado o tema da visão computacional e o funcionamento de sistemas de reconhecimento facial modernos. Enfim, no capítulo IV é feita uma análise exploratória sobre um sistema de reconhecimento facial baseado em VC e IA desenvolvido por alunos do ensino técnico do CEFET-MG e consequentemente são listadas propostas de continuidade do projeto para o Trabalho de Conclusão de Curso (TCC) II que continuará o trabalho que for aqui iniciado.

## **I.2 CONSIDERAÇÕES SOBRE DESAFIOS DE ACEITAÇÃO DA TECNOLOGIA**

Antes de, de fato, prosseguir-se com o desenvolvimento da proposta do presente trabalho, faz-se necessário desmistificar alguns conceitos que pairam no imaginário popular referente à inteligência artificial e seus possíveis impactos sociais. É evidente, desde o movimento Ludista, que as pessoas têm resistência para aceitação de novas tecnologias no mercado de trabalho por medo de serem substituídas (Bartlett, 2018).

É também notório que, em mais de dois séculos de revolução industrial, a automatização que surgiu contribuiu para a destruição de diversos cargos e funções, principalmente aquelas que exigem menos escolaridade e representam um trabalho mais manual e menos intelectual. A IA segue também neste rumo, sendo capaz de realizar tarefas não antes pensadas possíveis de serem feitas por um computador (como exemplificará o Capítulo II ).

É neste contexto que está surgindo um movimento neoludista de resistência ao desenvolvimento da IA (Bartlett, 2018) ou até mesmo está sendo proposta a criação de uma renda básica universal (Baer & Weller, 2019). Entretanto, não há evidências históricas que de fato suportam a ideia de que, desta vez, a automação vai de fato acabar com a maioria dos empregos e da empregabilidade da população (Globerman, 2019).

Ora, os Estados Unidos da América, o país que está mais avançado nesta tecnologia encontra-se com uma taxa de desemprego que está em seu mínimo histórico (The White House, 2019). Além de que o desenvolvimento da IA ainda está muito distante de conseguir replicar uma inteligência geral que seria capaz de, de fato, realizar pensamento analítico e possivelmente tornar a mão-de-obra humana obsoleta (Joshi, 2019).

Desta forma, como afirmam (Thiel & Masters, 2014), a melhor forma de se realizar trabalho será com a utilização conjunta daquilo que as pessoas fazem de melhor com aquilo que a IA e as demais máquinas fazem de melhor. Tal fato ocasionará em maior produtividade econômica, aumentando, no longo prazo, o bem-estar da população (Geller, 2019).

O único, porém, será que esta tecnologia reforçará, como já mencionado, a tendência de reduzir-se a demanda por trabalhadores com pouca escolaridade e expandir a demanda por aqueles que tem acesso à aplicação das técnicas científicas mais atuais. Desta forma justifica-se a elaboração deste trabalho para expandir a capacitação da comunidade científica nacional sobre este assunto.

### **I.3 OBJETIVOS**

- Avaliar os atuais métodos visão computacional visando o reconhecimento facial;
- Compreender o avanço recente de tecnologias de Inteligência Artificial e sua aplicação em visão computacional;
- Realizar um estudo crítico e aprofundado sobre projetos de VC já desenvolvidos no CEFET-MG.
- Implementar o sistema desenvolvido e estudado em ambiente real de utilização dentro do CEFETMG.
- Desenvolver uma inteligência artificial em alto nível seja capaz de realizar, com alto grau de assertividade, a identificação de diferentes pessoas e outras características importantes;
- População dos dados do reconhecimento facial em servidor em nuvem para utilização em plataforma IoT;

### **I.4 METODOLOGIA**

A fim de cumprir todos os objetivos do presente trabalho, será realizado um estudo aprofundado das principais técnicas de engenharia mencionadas. Em específico: desejou-se compreender com profundidade as inteligências artificiais que realizam o reconhecimento facial em sistemas de visão computacional.

A principal descoberta da pesquisa sobre a IA é de que os sistemas mais avançados e eficientes no reconhecimento facial são inteligências artificiais baseadas em *Machine Learning (ML)* e *Deep Learning (DL)*: redes neurais convolucionais (Y. Taigman, 2014). Deste modo seu funcionamento será amplamente estudado, explorado e explicado.

Serão também estudados, via revisão bibliográfica, a visão computacional como um todo e os sistemas que são responsáveis por realizar o reconhecimento facial. A VC possui 13 subáreas em que o reconhecimento está entre estas (Szeliski, 2010).

Reconhecimento este que pode ser facial ou de objetos. Um sistema de VC de reconhecimento facial possui quatro atribuições principais para realizar a identificação:

detecção, localização, rastreamento e classificação. O problema do reconhecimento facial é um problema de classificação e possui um algoritmo de DL dedicado para realizá-lo. Mesmo assim, haverão outros algoritmos responsáveis pelo restante das funções de VC.

Após realizada tal pesquisa, cujos resultados são os capítulos Capítulo II e Capítulo III, será feito um estudo da implementação de reconhecimento facial desenvolvida no CEFET-MG (Brum, Oliveira, & Almeida, 2018). Deste modo, será explorado o código, as ferramentas utilizadas e será feita uma nova implementação prática no âmbito do presente trabalho. Nota-se que tal sistema, mesmo sendo altamente assertivo e robusto, foi implementado a partir de uma metodologia de caixa preta, o que mostra espaço para haver continuidade desta pesquisa.

Desta forma, é proposto que se realize a continuidade deste projeto de pesquisa científica no âmbito do TCC II para tornar prática a contribuição no tema de IA, uma vez que a parte teórica já está, em grande parte, concluída no presente trabalho. Além disso uma exploração prática abre espaço para incluir o trabalho na crescente área de IoT, abrindo espaço até mesmo para o desenvolvimento de novas aplicações não antes pensadas.

## CAPÍTULO II A INTELIGÊNCIA ARTIFICIAL NA VISÃO COMPUTACIONAL: MACHINE LEARNING, DEEP LEARNING E REDES NEURAIS CONVOLUCIONAIS

### II.1 INTRODUÇÃO

O objetivo deste capítulo é contextualizar o leitor quanto ao funcionamento e ao estado da arte da Inteligência Artificial aplicada à visão computacional. Inicialmente será feito um breve histórico desta tecnologia, em seguida serão apresentadas as formas mais avançadas da IA; será discutido então o funcionamento de redes neurais de classificação e, por fim, as aplicações na VC.

### II.2 BREVE HISTÓRICO DA I.A.

O interesse e a filosofia por trás da inteligência de máquinas são tópicos antigos que datam do tempo do Iluminismo a partir de René Descartes (Bringsjord & Govindarajulu, 2018). Para a sociedade, há muito valor em fazer com que máquinas exibam o comportamento animal de aprender ao longo de sua vida útil, pois tal feito possibilita que tarefas antes feitas exclusivamente por pessoas possam ser automatizadas.

Dentro deste contexto é possível notar diferentes momentos na história em que tal criação filosófica foi tomando forma, como quando máquinas começaram a oferecer ajustes do último uso do usuário, quando a instrumentação pneumática foi inventada e até mesmo o momento de advento da eletrônica analógica e digital, permitindo que computadores fizessem a memorização digital de informações.

Junto a estas invenções, avanços filosóficos que trataram de pensar na criação de uma inteligência artificial multipropósito surgiram (Bringsjord & Govindarajulu, 2018). A ideia foi de fazer mímica à inteligência humana e, quem sabe um dia, ultrapassar até mesmo as capacidades de raciocínio abstrato do cérebro humano.

A partir desta ideia foram desenvolvidas formas determinísticas de utilizar-se de um número enorme de cálculos matemáticos, recentemente possibilitados devido aos constantes avanços da computação e da lei de Moore, para ultrapassar a capacidade humana em alguns aspectos. O mais notável marco desta fase foram as seguidas vitórias de um computador sobre o melhor jogador de Xadrez do mundo (Bringsjord & Govindarajulu, 2018).

No século XX, foram desenvolvidas novas formas de fazer com que um computador possa realizar uma compreensão abstrata de informações definidas por pessoas. Desta forma surgiu a subárea da Inteligência Artificial denominada de *machine learning* ou aprendizado de



máquina. Com grandes quantidades de informações e um processo denominado de aprendizado supervisionado (em que os dados são rotulados por pessoas e então processados) foi possível que a área de IA realizasse tarefas anteriormente categorizadas como exclusivamente biológicas, como o reconhecimento de pessoas em imagens (Collins-Thompson, 2017).

Ainda dentro da área de *machine learning*, nos anos 60 foi proposto utilizar-se de redes de aprendizado com múltiplos níveis de processamento interconectados. Tal tática fez com que os algoritmos criados desta forma fossem muito superiores em performance em relação aos aprendizados com apenas uma camada. Assim nasceram as redes neurais e o conhecido *deep learning* ou aprendizado profundo, que representa o estado da arte da tecnologia de inteligência artificial.

### **II.3 ESTADO DA ARTE**

As tecnologias mais avançadas de inteligência artificial tratam-se de utilizar-se de grandes bases de dados para treinamento das redes neurais em aprendizado profundo. Tais bases de dados foram disponibilizadas apenas recentemente devido à uma constante e volumosa geração de dados principalmente na internet.

Tais tecnologias são capazes de alcançarem alto rendimento em alguns tipos de aplicações como reconhecimento de objetos, reconhecimento de pessoas, automação de tarefas manuais, automação de escrita, de conversa e até mesmo de produção musical.

Em algumas tarefas as máquinas conseguem alcançar desempenhos praticamente iguais ao desempenho humano (Y. Taigman, 2014) e até mesmo superior (David Silver A. H., 2016). (Razavian, 2014), por exemplo, propõe uma forma de imediata e fácil implementação de uma inteligência artificial aplicada à mesma finalidade deste trabalho.

Estas máquinas não aprendem apenas com a grande quantidade de dados que se utilizam para treinar, mas também a partir de toda nova entrada de informação. Para esta assimilam uma relevância e utilizam-na via retroalimentação na rede. Tal prática permite que um dos grandes objetivos desta tecnologia seja alcançado: o sistema se adapta e apresenta melhores resultados ao longo do tempo de acordo com o ambiente em que é utilizado.

É possível traçar um paralelo entre tal aprendizado e a forma como as crianças aprendem: estas realizam observações sobre como os adultos ao seu redor realizam tarefas como falar e andar. Ao longo do tempo, tentam imitá-los, cometendo erros e sendo corrigidas pelos pais até o ponto em que conseguem, de fato, realizar a atividade em questão.

Uma forma de exemplificar esta tarefa é citar o famoso caso de 2016 em que o software AlphaGo (David Silver A. H., 2016), desenvolvido pelo Google, foi capaz de vencer o melhor jogador de Go do mundo. Go é considerado um jogo praticamente impossível de ser programado discriminadamente uma vez que a quantidade de possibilidade de movimentos (e, conseqüentemente, de cálculos necessários) cresce aceleradamente à medida que o jogo avança. Tal vitória formou-se como um verdadeiro marco do avanço da tecnologia de *deep learning*.

Não bastasse tal avanço, a tecnologia utilizada em 2016 foi treinada em diversos jogos contra jogadores humanos amadores e profissionais. Entretanto, é mais prático fazer a Inteligência praticar contra si mesma. Desta forma é mostrado em (David Silver J. S., 2017) que, iniciando o treinamento com nenhum conhecimento sobre o jogo, o AlphaZero foi capaz de vencer jogadores profissionais tendo treinado somente contra si mesmo a partir de zero conhecimento do jogo e de aprendizados obtidos através da prática do jogo.

Torna-se possível, então, compreender que as inteligências artificiais alcançaram um nível em que são capazes de atravessar milênios de técnicas desenvolvidas e de conhecimentos sobre um jogo jogando apenas contra si mesma milhares de vezes. Extrapola-se tal fato para diversos outros campos do conhecimento humano e é possível enxergar que a tecnologia *deep learning* possibilita uma nova grande fase de automação de tarefas no mercado de trabalho mundial.

## **II.4 FUNCIONAMENTO DAS REDES NEURAIS DE CLASSIFICAÇÃO**

### *II.4.1 CLASSIFICAÇÃO VERSUS REGRESSÃO*

Em *machine learning*, existem dois tipos de problemas que podem ser solucionados: classificação e regressão (Collins-Thompson, 2017). O primeiro decide a qual saída discreta se tratam os dados de entrada no sistema, como na diferenciação entre uma maçã e uma laranja ou entre o rosto de uma e outra pessoa. O segundo se trata de oferecer um número contínuo como saída do sistema.

Fica claro, então, que para a aplicação a ser aqui discutida afundo, a visão computacional, a aplicação de algoritmos de classificação conseguirá resolver uma série de problemas. Desta forma, é rico que seja feito um aprofundamento sobre como funcionam os algoritmos que exercem esta função.

### *II.4.2 INTRODUÇÃO ÀS REDES NEURAIS*

Existem diversas formas diferentes de resolver o problema de classificação via IA (Collins-Thompson, 2017), entre os quais podemos citar os métodos *K Nearest Neighbors*,

*Support Vector Machines*, *Kernelized Support Vector Machines*, *Decision Trees* e regressão logística. Eles realizam a tarefa de classificação utilizando dados disponíveis sobre algum tipo de objeto ou indivíduo para treinamento. Este consiste em ajuste dos fatores matemáticos internos às suas funções de classificação. O modelo gerado por este treinamento é então aplicado à dados não vistos anteriormente para realizar a classificação destes.

Estes métodos são de grande importância e resolvem uma série de problemas difíceis para algoritmos determinísticos. O foco deste estudo, porém, se dará nas redes neurais que representam a maior parte dos avanços da área, como mencionado na seção II.3 .

Para compreender o funcionamento de redes neurais na classificação, inicia-se com a visualização do funcionamento do algoritmo de regressão logística.

Na Figura II.1 entende-se que  $x[n]$  representam as entradas, enquanto que  $w[n]$  representam os pesos dados para cada entrada e  $y$  é a saída discreta como desejada para algoritmos de classificação. O treinamento destes modelos refere-se ao ajuste dos valores que são atribuídos aos diferentes pesos.

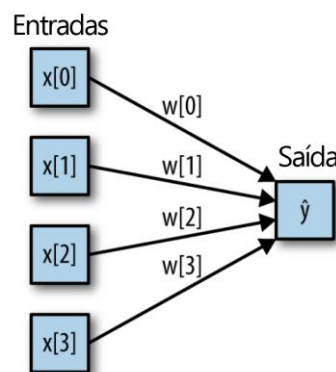


Figura II.1 - Visualização do funcionamento da regressão logística, adaptada de (Müller & Guido, 2016)

A saída aplica uma função denominada de função de ativação ao resultado da soma de todos valores multiplicados pelos seus respectivos pesos. Tais funções servem para retirar a linearidade das relações encontradas e tornar os modelos mais poderosos ainda (Collins-Thompson, 2017). Um exemplo destas funções é a tangente hiperbólica, que tende a fazer com que os valores fiquem próximos de 1 se forem altos e -1 se forem baixos (Figura II.2).

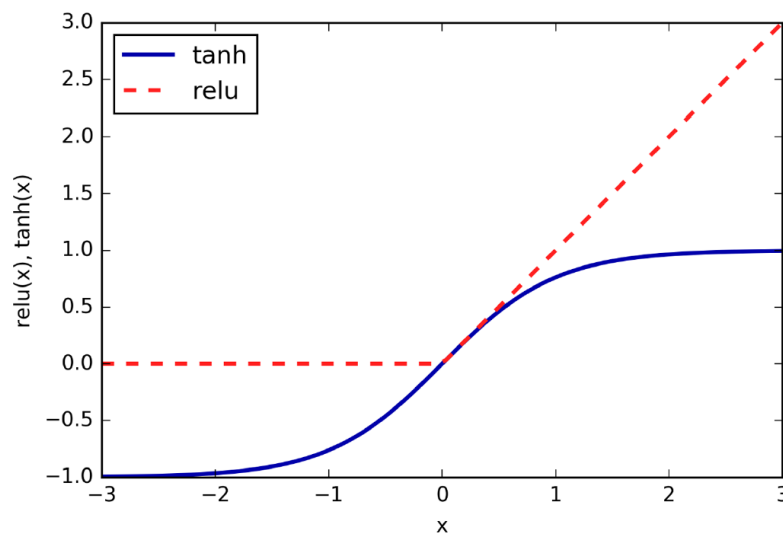


Figura II.2 - Visualização das funções de ativação tangente hiperbólica (em azul) e da função *Relu* (unidade linear retificada, em vermelho), obtida de (Müller & Guido, 2016)

Separam-se parte dos dados disponíveis para treinamento e parte para teste, de forma com que seja possível testar a acurácia do modelo em dados que ele não tenha visto anteriormente, mas que já estão em posse do engenheiro. Um importante passo a ser feito é comparar-se o desempenho do modelo tanto para classificar os dados de treinamento como os dados de teste, para verificar a presença ou não de *overfitting*.

O *overfitting* é um fenômeno matemático que ocorre quando o modelo é demasiadamente viciado nas características dos dados de treinamento que lhe foram passados, mas é bem menos capaz de ser assertivo com dados nunca anteriormente vistos.

Partido deste raciocínio, é possível que sejam alterados diversos aspectos deste modelo para fazer com que ele melhor se adapte a diferentes tipos de dados, como a função de ativação ou o aumento algébrico do número de entradas (por exemplo, utilizando  $x[0]^2, x[1]^2, x[2]^2, x[3]^2$  como informações de entrada).

Estes métodos são amplamente utilizados em redes neurais, entretanto, a principal diferença da regressão logística para uma rede neural é a presença de uma ou mais camadas ocultas, denominadas de camadas de neurônios. Destaca-se que cada neurônio aplica a função de ativação à sua respectiva soma. Visualiza-se o modelo na Figura II.3.



Neste exemplo, as camadas possuem diferentes tipos, a serem explorados posteriormente. De momento o mais importante é que seja compreendido que diferentes neurônios de diferentes camadas se especializam, via treinamento, a identificar padrões em diferentes combinações de características que sejam úteis para aumentar a assertividade da classificação.

Em algumas camadas pode haver compreensão humana de quais fatores o algoritmo está levando em conta para tomar decisões, como na Figura II.5 na qual é possível ver, nas camadas 2 e 3 (da esquerda para direita), que são separadas as diferentes partes do rosto de uma pessoa para que sejam comparadas e para realizar a verificação de identidade:

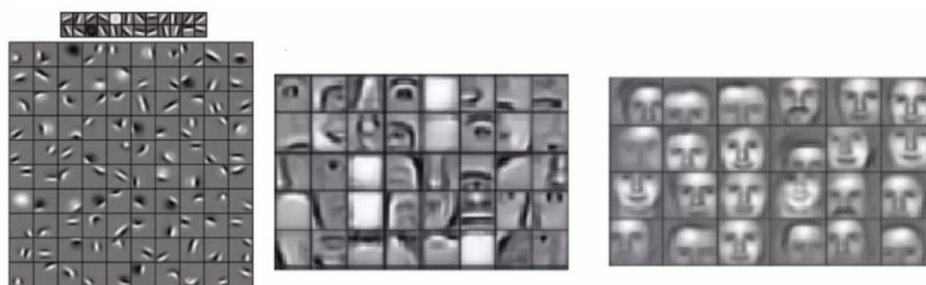


Figura II.5 – Visualização dos dados presentes em cada neurônio de três camadas sequenciais de uma rede neural convolucional. Obtida de (Collins-Thompson, 2017)

Um fator importante do DL é que em camadas muito profundas é altamente provável que os fatores que o algoritmo leva em consideração para realizar suas decisões sejam completamente incompreensíveis para os engenheiros responsáveis pelo seu funcionamento, sendo possível compreender sua acurácia apenas empiricamente.

Tal fato traz à tona uma interessante questão sobre a partir de qual ponto uma decisão deve ser confiada à uma IA, visto o nível de estrago que uma decisão errada tomada por algoritmo, muitas vezes por motivos incompreensíveis e imprevisíveis, pode vir a causar. Esta reflexão ilustra a necessidade de compreensão profunda sobre o assunto do profissional responsável.

#### II.4.4 COMPLEXIDADE E MÉTODOS DE PREVENÇÃO

É válido fazer um parêntesis, neste momento, sobre a crescente complexidade e demora no treinamento que é prevista para este tipo de modelo, a partir do momento em que são adicionadas mais camadas. Nota-se, comparando a Figura II.3 com a Figura II.1, que o número de pesos a serem calculados cresce de forma exponencial com a profundidade da CNN.

Parte deste problema é resolvido à bruta força via utilização de mais poder computacional, utilizando-se do já mencionado aumento da disponibilidade deste em computadores modernos.

Entretanto não há lei de Moore que seja suficientemente poderosa para fazer com que algoritmos exponencialmente mais complexos consigam ser processados em tempo hábil. O algoritmo *DeepFace*, mencionado em (Y. Taigman, 2014), aplica diversos métodos para conter esta complexidade e mesmo assim demorou 3 dias inteiros para realizar seu treinamento no seu enorme banco de imagens.

É muito importante, então, aplicar métodos que reduzam a complexidade da rede, minimizando a perda de desempenho e o *overfitting*. Um método bastante utilizado para este fim é o de aplicar uma penalidade algébrica que puna algoritmos que tenham uma grande quantidade de parâmetros. Desta forma a maior parte dos pesos que entram e saem de cada entrada e neurônio tendem a zero, fazendo com que o algoritmo seja capaz de ignorar informações de baixa relevância.

#### II.4.5 TIPOS DE CAMADAS

Visando compreender o funcionamento de uma rede neural de classificação, é vital que fiquem claros os diferentes tipos de camadas que são aplicadas aos dados. Utiliza-se, para tal,

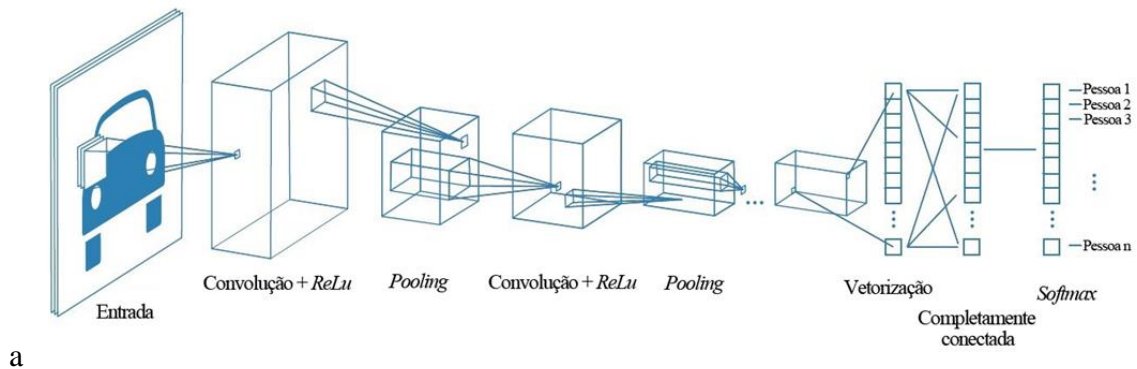


Figura II.4 como base, na qual destacam-se 4 tipos de camadas:

- Camadas convolucionais;
- Camadas de subamostragem ou *pooling*;
- Camadas completamente conectadas;
- Camadas localmente conectadas.

Estes tipos de camadas são os mesmos daqueles utilizados em aplicações de visão computacional avançadas, como (Y. Taigman, 2014) e (Levi & Hassner, 2015), mostrando que este modelo representa o que há de mais avançado na tecnologia.

Especificamente sobre cada tipo de camada, entende-se por convolucionais as camadas nas quais são aplicadas matematicamente as operações de convolução às matrizes de entrada com uma matriz-filtro (chamada por alguns autores de *kernel*) para que haja extração das características. É este tipo de camada que forma a base de algoritmos conhecidos como redes neurais convolucionais ou redes neurais convolucionais profundas (DCNN), que são aplicadas em (Y. Taigman, 2014) e em (Levi & Hassner, 2015).

A justificativa da necessidade de utilizar-se de camadas convolucionais deve-se ao fato pelo qual imagens em alta resolução possuem uma quantidade enorme de informações que se tornam inviáveis de serem processadas quando aplicadas diretamente em camadas completamente conectadas (Ng, 2017) como na Figura II.3.

Tal problema é exemplificado por uma conta simples: Uma imagem de média resolução de 1MP (megapixel) pode ser representada por uma matriz tridimensional 1000x1000x3 (na Figura II.6 têm-se uma versão simplificada desta matriz).

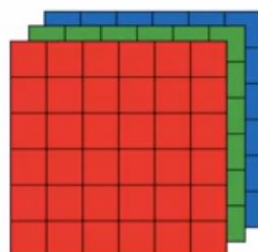


Figura II.6 - Representação de uma matriz 6x6x3 em formato RGB, obtida de (Ng, 2017)

Enquanto que 1000x1000 são os valores de claridade de cada pixel que, juntos, formarão a imagem quando projetados numa tela, o número 3 vem da representação das 3 matrizes que se referem às cores vermelho, azul e verde que são conjuntamente projetadas para formar as cores da imagem. Apenas até este momento já se nota que foram obtidas 3 milhões de características.

Ao utilizar-se de 1000 neurônios na primeira camada, as 3 milhões de características atravessarão 3 bilhões de pesos apenas entre a primeira e segunda camada. Este número crescente inviabiliza este tipo de aplicação, principalmente tendo em vista os fatos nos quais



deseja-se utilizar múltiplas camadas e as fotos de celulares modernos facilmente atingem tamanhos superiores a 20MP.

A operação matemática é feita então através da convolução matricial de um certo grupo sequencial dentro de uma matriz através de um filtro de dimensões desejadas. Cada filtro conseguirá extrair diferentes tipos de características da matriz total, que serão transferidas para os neurônios. Tal operação matricial pode ser vista na Figura II.7, sendo que a camada 1 da Figura II.5 representa o resultado de uma camada convolutiva.

Nota-se que há uma redução de tamanho na matriz de saída. Como pode não ser vantajoso perder informações, podem ser adicionadas linhas e colunas extras à matriz inicial para que o resultado da convolução possua as mesmas dimensões da matriz inicial (tal procedimento é denominado como *padding*). Outra consideração a ser aqui feita é que a operação de convolução utilizada em redes neurais convolucionais não realiza a transposição da matriz-filtro antes de multiplicar os valores, diferenciando-a assim da convolução matricial normalmente encontrada em definições matemáticas.

Além disso, os valores da matriz-filtro do exemplo da Figura II.7 representam um detector de quinas verticais definido pelo engenheiro responsável. Entretanto, como já de costume neste tópico, estes valores podem ser pesos que o programa aprenderá sozinho via treinamento e tentativa-e-erro. Tal etapa extra de treinamento é um dos motivos da necessidade pelo aumento da quantidade de dados disponíveis para treino a fim de evitar o *overfitting*.

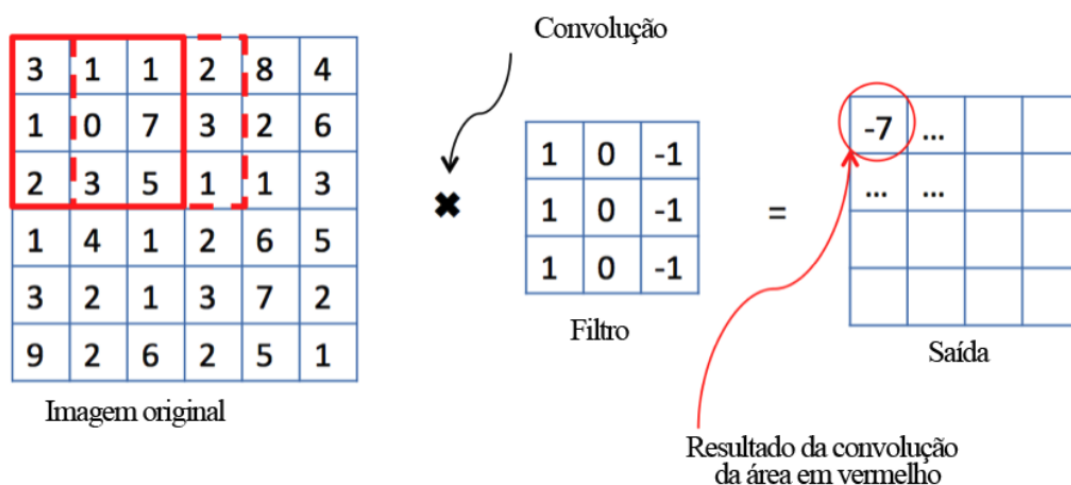


Figura II.7 - Convolução através de um filtro detector de quinas verticais, adaptada de (Cavaioni, 2018)

Finalmente, as matrizes comuns de entrada de uma CNN são tridimensionais (devido às cores vermelho, verde e azul). Por este motivo uma camada de uma rede real aplica filtros tridimensionais (por exemplo  $3 \times 3 \times 3$ ), alcançando uma saída bidimensional a cada neurônio convolucional (em que se aplica convolução entre a matriz original e um filtro, soma-se um valor de viés e aplica-se uma função de ativação ao resultado).

Desta forma, a quantidade de parâmetros desta rede neural independe do tamanho da imagem e é um resultado apenas do número de filtros. Cada filtro  $3 \times 3 \times 3$  possui 27 parâmetros e 1 parâmetro de viés, então com 1000 neurônios o resultado são 28000 parâmetros. Observa-se que este valor é muito menor do que o resultado encontrado para o mesmo número de neurônios em camadas completamente conectadas.

Camadas de subamostragem ou *pooling* são camadas que possuem uma função de reduzir a quantidade de dados e consolidá-los após uma camada convolucional. Tal operação é feita via a aplicação de um filtro à matriz de entrada com uma operação normalmente associada à estatística. Desta forma os filtros não são matrizes, mas sim operações que calculam, por exemplo, o valor máximo ou a média dos valores em uma certa região da matriz. Uma visualização da forma mais aplicada de *pooling* pode ser vista na Figura II.8.

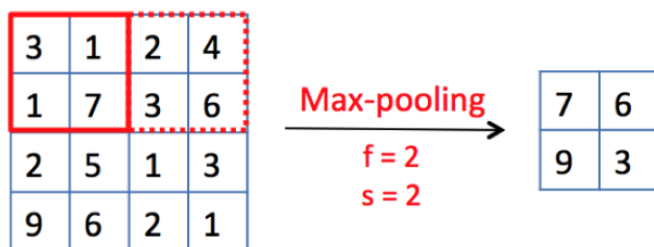


Figura II.8 - *Max-Pooling*, obtida de (Cavaioni, 2018)

Camadas completamente conectadas são o tipo de camada mencionado na Figura II.3 em que todos os neurônios de uma camada estão conectados à camada seguinte. A conversão da saída de camadas convolucionais ou camadas de *pooling* para estas é feita de forma em que se transforma a matriz final das camadas anteriores em um vetor linear e então conecta-se estes aos diferentes neurônios com seus respectivos pesos (Figura II.9 - Exemplo de CNN adaptado de ).

Camadas localmente conectadas operam de maneira semelhante às convolucionais, com a diferença de que para cada neurônio há um filtro específico. Nota-se que, por motivos

semelhantes aos das camadas completamente conectadas, este tipo de camada possui um alto custo operacional.

#### II.4.6 TOMADA DE DECISÃO

A saída da última camada completamente conectada é então aplicada através de uma camada denominada *softmax* (também completamente conectada) que analisa os parâmetros finais calcula as probabilidades de todos os objetos conhecidos serem o objeto em questão (Ng, 2017).

Cada neurônio desta camada representa uma possível opção de classificação na saída. O que a diferencia das outras camadas é que sua função de ativação funciona através de um cálculo no qual o número “e” é elevado aos valores resultantes da soma com pesos e divide-se cada valor pela soma de todas as exponenciais. Desta forma a soma total dos valores finais é 1, funcionando o resultado individual de cada neurônio como uma probabilidade atribuída pela rede neural para aquela opção de classificação de acordo com o que foi encontrado nas características. A classe com maior probabilidade é definida como a saída.

#### II.4.7 EVOLUÇÃO DOS DADOS EM UMA CNN

A entrada de dados em uma CNN de VC é, como já mencionado, normalmente o conjunto uma matriz tridimensional, referenciando o formato RGB para cada pixel individual da imagem. Com isso, tem-se que esta matriz atravessará diversas camadas, alterando-se profundamente a forma como os dados são organizados. A Figura II.9 exemplifica tal fato.

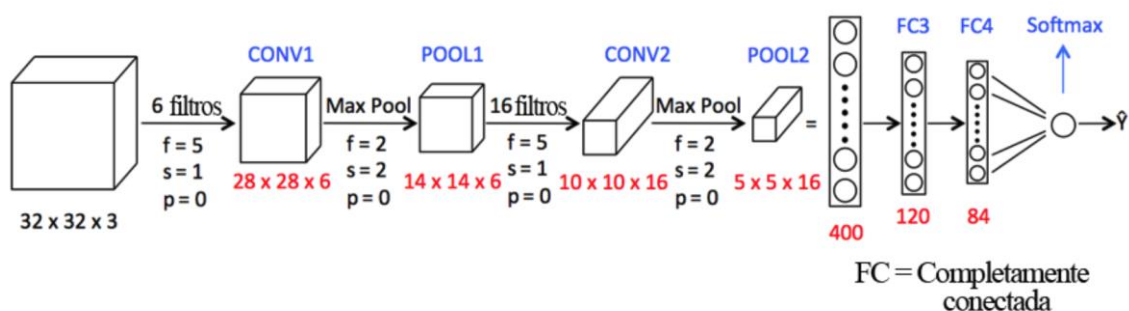


Figura II.9 - Exemplo de CNN adaptado de (Cavaioni, 2018)

Nota-se, na Figura II.9, que a matriz inicial sofreu diversas alterações pelas camadas convolucionais e pelas de subamostragem, deixando seu formato original de 32x32x3 até alcançar o formato 5x5x16 antes de ser transformada em um vetor de 400 neurônios.

Tudo isto acontece devido aos números que estão envolvidos na elaboração dos filtros: f, s e p. Enquanto ‘f’ representa as dimensões vertical e horizontal dos filtros tridimensionais

(ou seja com  $f = 5$  há de se ter um filtro  $5 \times 5 \times 3$  na camada 1, com a terceira dimensão sempre igual à dimensão de profundidade dos dados), ‘s’ representa a propriedade *stride*, que rege a forma como o filtro realiza sua iteração ao longo da imagem:  $s=1$  ele realiza a operação de um em um pixel,  $s=2$  de dois em dois pixels e assim por diante. Já o ‘p’, por sua vez, representa a quantidades de linhas e colunas a serem adicionadas via *padding*.

Em geral, os dados se tornarão menos longos e compridos e mais profundos com a aplicação de um crescente número de filtros à medida que a rede neural se aprofunda. Tal fator contribui para que a camada final alcance o estado visível na Figura II.9, que é capaz de representar as características de alto nível que são visíveis na camada final da Figura II.5.

#### II.4.8 RETROALIMENTAÇÃO

Como já mencionado neste texto, uma das características fundamentais da inteligência artificial é ser capaz de retroalimentar-se de acordo com o nível de assertividade de suas previsões. Tal método é comumente implementado através de aprendizagem reforçada (*reinforcement learning*).

Aprendizagem reforçada é uma terceira via entre aprendizagem supervisionada e não-supervisionada na qual os algoritmos reutilizam dados de sua utilização para complementar seu treinamento. Estes funcionam de forma com que as diversas variações da inteligência são premiadas matematicamente por alcançarem o seu objetivo de forma mais assertiva. Tal resultado é então computado (Heess, 2017).

O mais importante é entender que o sistema avalia se sua previsão foi assertiva para um determinado caso (se a informação da classificação verdadeira para este caso estiver disponível) e se utiliza desta informação para ajustar os parâmetros da rede numa forma de pós-treinamento.

É a partir destes elementos com múltiplas conexões e pesos entre entradas, tipos de camadas e saídas que funcionam os algoritmos de classificação.

### II.5 APLICAÇÕES NA VC.

A partir da discussão sobre o funcionamento de redes neurais convolucionais fica evidente que estas podem ser amplamente utilizadas em visão computacional uma vez que uma entrada que pode ser muito bem trabalhada por este sistema é uma imagem em forma de matriz com valores para as cores vermelho, azul e verde. É assim, por exemplo, que funcionam as aplicações de (Y. Taigman, 2014) e (Levi & Hassner, 2015).

A aplicações mais comuns são: o reconhecimento facial, o rastreamento de objetos, a automação de tarefas e a condução veicular. O foco deste trabalho é compreender a fundo o funcionamento dos sistemas que realizam reconhecimento facial e realizar então uma implementação prática deste.

## CAPÍTULO III O ESTADO ATUAL DA TECNOLOGIA DE VISÃO COMPUTACIONAL E DO RECONHECIMENTO FACIAL

### III.1 INTRODUÇÃO

Este capítulo tem como foco o tema da visão computacional. Nesta parte é expandida a compreensão de como esta tecnologia funciona, além de aprofundar-se na discussão sobre o estado da arte e das aplicações mais importantes. Ao final, são apresentados os sistemas de reconhecimento facial e seu funcionamento.

### III.2 O QUE É VISÃO COMPUTACIONAL

VC é a área da inteligência artificial responsável por interpretar e utilizar imagens, vídeos e objetos 3D para orientar ações de um computador. Ela reúne os diferentes métodos e tecnologias que permitem que as máquinas consigam emular o sentido da visão, analisando o ambiente e reagindo aos estados e mudanças deste.

É uma área que tem apresentado um crescimento substancial nas últimas décadas, influenciado principalmente pela melhoria dos sensores fotossensíveis e do aumento da disponibilidade de processamento (Szeliski, 2010).

### III.3 APLICAÇÕES DA VC

A visão computacional possui pelo menos 50 anos de história (Szeliski, 2010), tendo desenvolvido neste tempo diversas aplicações, que vão de áreas avançadas como reconhecimento facial, detecção, classificação e rastreamento de objetos a áreas mais simples, como detecção de cores e segmentação de uma imagem.

Em (Szeliski, 2010) o autor defende que é possível subdividir o campo em 13 áreas, que são listadas abaixo:

- Formatação de Imagem;
- Processamento de imagem;
- Detecção e Correspondência de Características;
- Segmentação;
- Alinhamento baseado em Características;
- Estruturas criadas à Base de Movimento;
- Estimativa Densa de Movimento;
- Costura de Imagem;
- Fotografia Computacional;

- Correspondência Estéreo;
- Reconstrução 3D;
- Renderização à base de Imagens;
- Reconhecimento.

Esta gama de aplicações é ampla, o que é motivado pelo fato de que a visão biológica realiza um grande processamento contínuo de dados, orientando diversos sistemas biológicos de tomada de decisão. A capacidade de um computador de imitar tal comportamento, portanto, é de grande valia para que diversas tarefas possam ser automatizadas.

Este trabalho discutirá diretamente apenas as aplicações presentes na área definida por Szeliski como “Reconhecimento”. Nesta se encontram as aplicações mais avançadas da área, que são:

- Detecção de Objetos;
- Reconhecimento Facial;
- Reconhecimento de Instância;
- Categorização;
- Compreensão de Cena e Contexto.

Apesar de distintas, as subáreas de “Reconhecimento” têm sido constantemente trabalhadas via utilização de redes neurais convolucionais (Ng, 2017). Estas aplicações representam uma forma de resolver grande parte dos problemas relacionados, por exemplo, à área de veículos autônomos.

#### **III.4 A IA NA VISÃO COMPUTACIONAL: DETECÇÃO, LOCALIZAÇÃO, CLASSIFICAÇÃO E RASTREAMENTO**

Como já mencionado nas seções II.5 e III.3 , Redes Neurais Convolucionais tem sido uma importante ferramenta para a resolução de problemas na visão computacional. Mais especificamente, tem-se que alguns dos problemas que são resolvidos por este tipo de IA na VC são problemas que tratam de maneira concomitante detecção, classificação, localização, e rastreamento de pessoas e objetos (Ng, 2017).

O problema de classificação já foi trabalhado no Capítulo II . Os outros problemas acima citados podem ser considerados problemas de regressão e classificação cuja função específica é detectar, localizar e rastrear objetos em imagens e vídeos.

Na verdade, em contextos de visão computacional é difícil pensar que pode haver um problema puro de classificação, pois para que os algoritmos de classificação de objeto ou de reconhecimento atuem, faz-se necessário que a imagem já esteja pré-trabalhada para isolar a área que se refere ao objeto a ser classificado.

Desta forma, o primeiro passo é separar, numa imagem, o que é objeto e o que é plano de fundo. Tal passo é denominado de detecção e é realizado, nas CNN's, através da alocação de uma saída do sistema para este fim. O valor de saída é 1 se houver, em uma determinada imagem, frame ou vídeo, algum objeto ou pessoa. O valor de saída é zero caso a imagem seja composta apenas de plano de fundo.

Realizada a detecção, é necessário que sejam localizados na imagem os objetos detectados. Este procedimento é feito de forma com que a rede neural analisa diferentes partes da imagem e define retângulos que se referem a um objeto, como pode ser visto nos carros da Figura III.1.



Figura III.1 – Dois veículos detectados, localizados e rastreados, obtida de (Ng, 2017)

Uma vez feitas a detecção e localização é possível que sejam realizadas as classificações desejadas, como definir se os objetos na Figura III.1 são carros ou pessoas. Em último lugar, tratando-se de vídeos, é interessante rastrear o constante movimento dos objetos na cena em questão.

Para tal, são atribuídos ID's para os diferentes objetos na imagem para localizá-los novamente e, à medida que eles são reconhecidos em locais diferentes nos frames subsequentes, o algoritmo detecta que o mesmo objeto se moveu para um lugar diferente. Destaca-se que, como normalmente o algoritmo tem acesso à diferença de tempo entre os frames, com um modelo 3D do lugar em questão torna-se possível utilizar o rastreamento para medir grandezas físicas de movimento dos objetos.



Um outro método de detecção, localização e rastreamento é denominado Haar Cascade, sendo este um método baseado em *machine learning* que é utilizado por (Brum & Oliveira, 2018) e explorado no Capítulo IV .

### **III.5 OS SISTEMAS DE RECONHECIMENTO FACIAL**

Sistemas de reconhecimento facial são sistemas responsáveis por utilizar-se de imagens e vídeos para entender se uma determinada pessoa se encontra num determinado local. O problema do reconhecimento facial é um problema de classificação cujas classes são definidas como as diferentes identidades das pessoas.

O foco deste estudo é compreender como é realizado o reconhecimento facial através da aplicação de Redes Neurais Convolucionais a imagens e vídeos, justificando-se, portanto, a compreensão das diferentes técnicas que podem ser utilizadas para este fim.

Além dos métodos já mencionados nas seções e capítulos anteriores, é possível realizar aplicações de VC que não dependam de *deep learning* nem de *machine learning*. Métodos projetados manualmente oferecem formas determinísticas de realizar a mesma tarefa. Na verdade, antes dos anos 2000 toda a área de Visão Computacional dependia exclusivamente destes métodos (Szeliski, 2010).

A maior vantagem apresentada por estes métodos é que os programas criados desta maneira podem ser depurados de maneira transparente em todas as suas fases. Como dito no Capítulo II , o fato de algoritmos de *deep learning* funcionarem muitas vezes como caixas-pretas é um grande impedimento para fazer com que os programadores compreendam os motivos que os fazem falhar.

Uma maneira interessante de se trabalhar o reconhecimento facial é criando uma terceira via entre estes métodos: utiliza-se de métodos convencionais aliados a técnicas de *machine learning* para realizar as etapas pré-classificação e de CNN's para realizar esta. É desta forma que funciona o algoritmo de (Y. Taigman, 2014).

Esta metodologia é muito útil pois outros métodos além de CNN's podem ser bastante eficazes na realização de detecção, localização e rastreamento. Na metodologia do DeepFace (Y. Taigman, 2014) a face a ser reconhecida é detectada, localizada e passa por uma operação matricial de frontalização para facilitar o funcionamento do algoritmo em *deep learning* (Figura III.2).

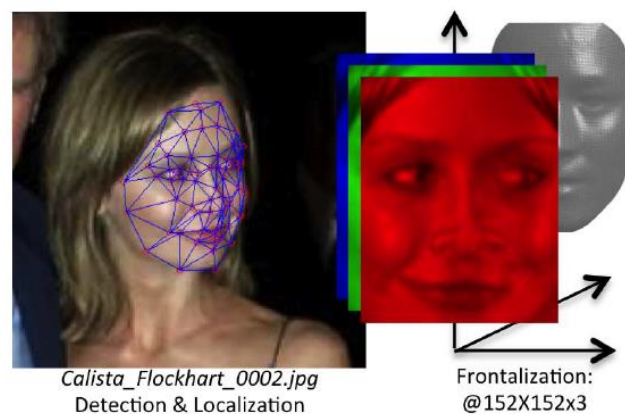


Figura III.2 – Frontalização das imagens de (Y. Taigman, 2014)

Percebe-se, na Figura III.2, que são mapeados (localizados) 67 pontos de interesse na face de uma pessoa detectada de maneira tridimensional. Tal procedimento é realizado via utilização de um *Support Vector Regressor (SVR)* treinado em múltiplas iterações para refinar seus resultados. SVR é uma forma de aplicação para problemas de regressão (neste caso, para localização) da tecnologia de *Support Vector Machines*, mencionada anteriormente na seção II.4.2, mas não discutida afundo neste trabalho.

É a imagem já frontalizada que é a entrada das CNN's. O algoritmo, então, procede a tratar as matrizes RGB somente da região de interesse da imagem para realizar o reconhecimento facial. A rede neural completa que realiza este procedimento pode ser vista na Figura III.3.

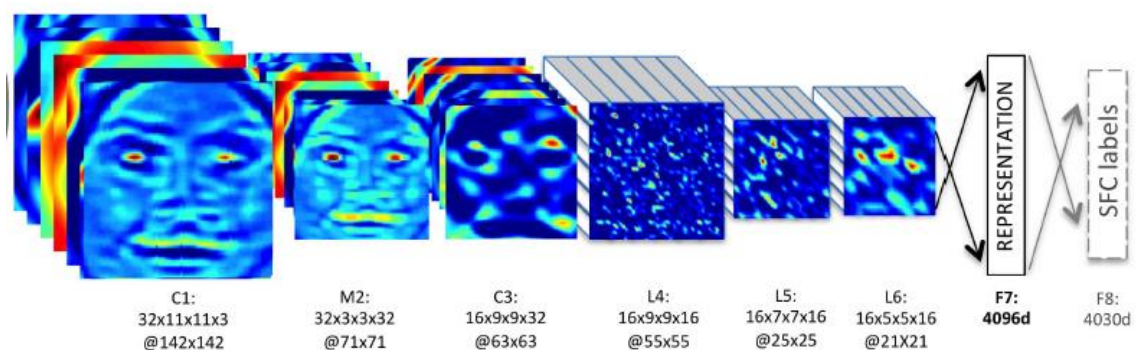


Figura III.3 – Visualização da Rede Neural de (Y. Taigman, 2014)

Numa breve descrição da rede neural acima, nota-se que C1 e C3 são camadas convolucionais, M2 é uma camada de subamostragem na forma de *max-pooling*, L4, L5 e L6 são camadas localmente conectadas, F7 e F8 são camadas totalmente conectadas e a saída do vetor-representação de F8 passa por um *softmax*.

## CAPÍTULO IV ESTUDO DE IMPLEMENTAÇÃO DE UMA INTELIGÊNCIA ARTIFICIAL VISANDO O RECONHECIMENTO FACIAL DO CEFET-MG

### IV.1 INTRODUÇÃO

Este capítulo apresentará os projetos preliminares referentes ao tema em questão já desenvolvidos no CEFET-MG, além de caracterizar as ferramentas utilizadas, mostrar resultados preliminares de uma implementação prática e realizar ponderações sobre o que pode ser desenvolvido como continuidade no TCC II.

### IV.2 ESTADO ATUAL DOS PROJETOS DO CEFET-MG

O CEFET-MG é uma instituição que engloba múltiplos níveis educacionais e possui uma série de projetos desenvolvidos por alunos dos distintos ensinos e cursos. O presente trabalho explorará o projeto de reconhecimento facial desenvolvido por alunos do curso técnico de Eletrônica via iniciação científica (Brum & Oliveira, 2018).

Em sumário, a proposta do trabalho foi de criar um sistema de reconhecimento facial, utilizando-se de técnicas convencionais de VC e de *machine learning*. Na aplicação feita pelos alunos, o sistema utilizou-se das informações obtidas de uma câmera conectada a uma unidade *Raspberry Pi*, que analisava a imagem da câmera em seu banco de dados, reconhecia ou não as pessoas e abria ou mantinha fechada uma porta (Brum & Oliveira, 2018). O procedimento pode ser melhor visto na Figura IV.1.

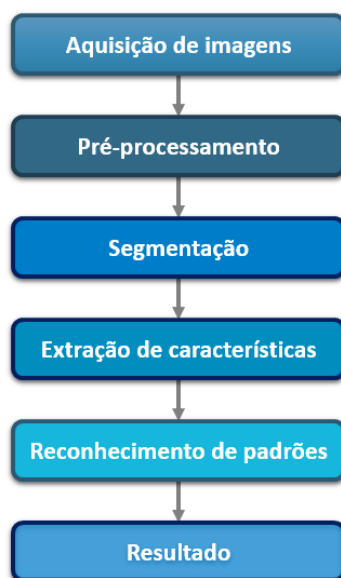


Figura IV.1 - Procedimento de visão computacional aplicado, obtida de (Cruz, Coutinho, & Freire, 2019)

O projeto utilizou-se de aproximadamente 90 imagens de cada pessoa a ser reconhecida. Em seguida foi feito o treinamento da rede neural para tais imagens e os resultados foram considerados satisfatórios, não tendo ocorrido nenhum erro de reconhecimento nos testes realizados.

Foram também incluídas algumas propostas de melhoria do sistema, como a necessidade de diferenciar entre uma pessoa real e uma foto desta e a implementação de um sistema de retroalimentação à medida com que uma pessoa seja reconhecida com alto grau de assertividade.

Estas questões não serão exploradas no presente trabalho, entretanto destaca-se o fato de que a IA foi aplicada a partir de uma metodologia de “caixa preta”. Faz sentido, então, compreender e propor uma expansão do escopo deste robusto trabalho de nível técnico com os conhecimentos apresentados nos capítulos anteriores. Propõe-se compreender e redesenhar a tecnologia por trás do reconhecimento facial.

### **IV.3 TECNOLOGIAS UTILIZADAS**

Antes de implementar o projeto e alterar a forma como este funciona, faz sentido compreender a abrangência das tecnologias utilizadas por este. As principais que foram utilizadas para possibilitar a execução do sistema de reconhecimento são explicadas ao longo desta seção.

#### *IV.3.1 PYTHON*

Python é a linguagem de programação mais utilizada no mundo (Cruz, Coutinho, & Freire, 2019). Ela é um código desenvolvido de forma *open source* que une leveza na execução, simplicidade em sua sintaxe e uma grande gama de bibliotecas capazes de expandir suas capacidades. É hoje uma linguagem de grande importância nos meios de análise de dados e inteligência artificial. Além de tudo isto, é também uma das linguagens aceitas no Raspberry Pi.

Algumas das bibliotecas mais importantes desta linguagem para as aplicações de Inteligência Artificial e VC são: NumPy, Pandas, Matplotlib, SciKit-Learn, Cython e TensorFlow. Além disso é também importante salientar a plataforma Anaconda que realiza a execução dos pacotes.

#### IV.3.2 *OPENCV*

*OpenCV* ou *Open Source Computer Vision Library* é uma biblioteca que pode ser utilizada com o Python possibilitando que este tenha a capacidade de realizar executar programas de visão computacional.

#### IV.3.3 *RASPBERRY PI*

Raspberry Pi é um módulo desenvolvido pela fundação Raspberry que une um sistema aberto de desenvolvimento em Linux, capaz de suportar aplicações em Python, a um hardware poderoso e versátil.



Figura IV.2 - Vista do Raspberry Pi, obtida de (Brum, Oliveira, & Almeida, 2018)

#### IV.3.4 *TENSORFLOW*;

TensorFlow é uma plataforma *open source* de *ML* e *DL* que pode ser instalada como pacote para Python. É hoje uma das ferramentas mais utilizadas por cientistas de dados para aplicações de inteligência artificial.

#### IV.3.5 *HAAR CASCADE*;

Haar Cascade é um algoritmo que trabalha uma imagem ou vídeo num sistema de visão computacional visando a realização de detecção, localização e rastreamento de objetos ou pessoas. Este método funciona a partir de utilização de filtros e convolução, assim como as CNN's. A diferença é de que os filtros são todos predefinidos pelos engenheiros, tornando cada um destes aptos a reconhecer somente um tipo específico de padrão, não tendo sua funcionalidade alterada pelo treinamento. Alguns filtros do Haar Cascade estão ilustrados na Figura IV.3, assim como seu modelo matemático (Figura IV.4).

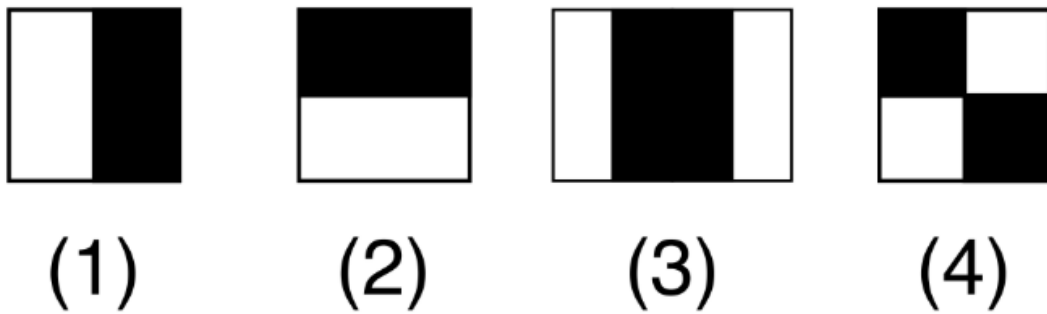


Figura IV.3 - Filtros do Haar Cascade (Vidhya, 2019)

-1	-1	5	5	5	-1	5	-1	5	-1	-1
-1	-1	5	-1	-1	-1	5	-1	-1	5	-1
-1	-1	5	-1	-1	-1	5	-1	-1	-1	5

Figura IV.4 - Visualização numérica dos filtros da Figura IV.3 (Vidhya, 2019)

É notável que Haar Cascade pode ser utilizado em união a redes neurais convolutivas (Nair, Mansoori, Moghe, Shah, & Talele, 2019), numa forma de mescla de diferentes tecnologias para alcançar melhor performance, assim como previsto que poderia ocorrer pela seção III.5.

#### IV.4 IMPLEMENTAÇÃO BÁSICA DO PROJETO DE (BRUM & OLIVEIRA, 2018)

Após compreendido o projeto de (Brum & Oliveira, 2018) e realizados os estudos cujo resultado são o Capítulo II e o Capítulo III do presente trabalho, decidiu-se então realizar uma implementação básica do sistema já desenvolvido em um *Notebook* Windows.

Os códigos para execução do programa foram cortesia de (Cruz, Coutinho, & Freire, 2019). O resultado da implementação pode ser visto na Figura IV.5.

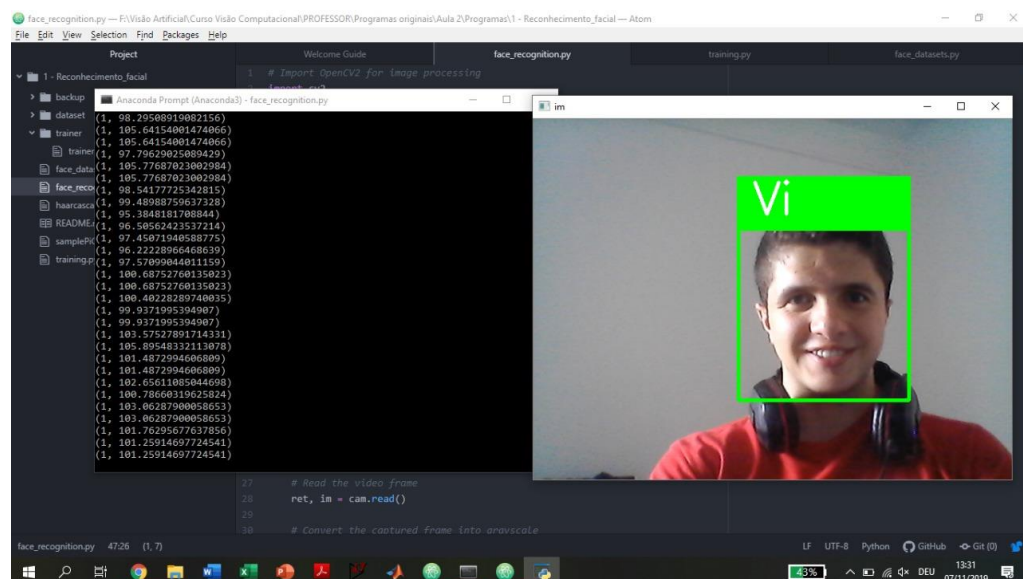


Figura IV.5 - Execução do código de reconhecimento facial, autoria própria

Na Figura IV.5 nota-se que o sistema reconheceu o autor do presente trabalho como “Vi” ou Vinicius, que era o único usuário que estava já cadastrado na base de dados e cujo a IA estava treinada para reconhecer. Outro ponto importante de ser salientado é de que a partir dos estudos do material de (Brum & Oliveira, 2018) e (Brum, Oliveira, & Almeida, 2018) não ficou claro se o sistema utilizado é baseado em redes neurais convolucionais ou não, sendo tal informação mencionada apenas em (Cruz, Coutinho, & Freire, 2019).

#### **IV.1 PROPOSTAS DE DESENVOLVIMENTO PARA O TCC II**

A partir dos conhecimentos obtidos nos Capítulo II e no Capítulo III , além do sistema já em funcionamento de (Brum & Oliveira, 2018) apresentado no Capítulo IV , propõe se realizar um estudo experimental de diferentes formas de aplicação desta tecnologia e a sua eficácia no reconhecimento facial.

Mais especificamente, propõe-se que sejam abordados aspectos específicos como:

##### *IV.1.1 COMPREENDER, COM PROFUNDIDADE, A IA JÁ IMPLEMENTADA*

Como foi mencionado ao fim da seção IV.4 , o relatório do desenvolvimento do sistema de reconhecimento facial estudado não evidencia o uso de CNN's ou de qual tecnologia de *machine learning* está sendo aplicada para realização do reconhecimento facial. Faz sentido, então, proceder com um estudo aprofundado do código, disponibilizado por cortesia da instituição, para compreender e explicar exatamente como funciona este algoritmo.

##### *IV.1.2 CONECTAR O SISTEMA À NUVEM*

Dado o crescimento de aplicações em IoT mencionado no Capítulo I tem-se que um ponto que pode ampliar a aplicabilidade de um sistema de reconhecimento facial é fazer com que este se comunique a um sistema em nuvem. Propõe-se que a decisão da IA em relação ao reconhecimento ou não de certo usuário seja um dado enviado para a nuvem, aguardando-se então a resposta desta para agir.

##### *IV.1.3 EXPLICITAR A CERTEZA DA TOMADA DE DECISÃO DO SOFTMAX DA IA*

Seria de grande valia, para trabalhos futuros sobre a aplicação de (Brum & Oliveira, 2018) em questão, compreender qual é o grau de certeza da tomada de decisão da unidade *softmax* que se espera encontrar na camada final rede neural convolucional. Esta probabilidade pode ser usada como métrica para aprimorar o sistema.

#### *IV.1.4 COMPREENDER O QUE CAUSA O SISTEMA A ACUSAR O RECONHECIMENTO*

##### *FALSO*

A implementação do sistema de reconhecimento facial no presente trabalho se mostrou ineficiente em sua tarefa principal. Mesmo que este ainda seja um resultado preliminar, faz sentido realizar um estudo do porquê o sistema acusou uma pessoa desconhecida de ser um usuário cadastrado. Como este erro não ocorreu em (Brum & Oliveira, 2018), há uma investigação a ser feita sobre como este problema pode ser evitado.

#### *IV.1.5 AUMENTAR A EFICIÊNCIA DO SISTEMA PARA FUNCIONAR COM MENOS FOTOS NA SUA BASE DE DADOS DE TREINAMENTO*

O presente sistema necessita de uma forma de funcionar assertivamente com menos fotos. Tal avanço tecnológico torna-se necessário uma vez que, em aplicações comerciais a necessidade de obter-se 90 fotos de cada usuário pode diminuir a atratividade do produto.

Pode-se citar como exemplo o cadastro de identidade em sistemas de reconhecimento facial utilizado em celulares modernos, que duram poucos minutos para realizar o cadastro e o sistema funciona de maneira considerada segura (Apple, 2019).

#### *IV.1.6 IMPLEMENTAR RETROALIMENTAÇÃO PARA QUE O SISTEMA POSSA MELHORAR SUA PERFORMANCE AO LONGO DO TEMPO*

Como mencionado no Capítulo II , uma das maiores vantagens de uma aplicação em *machine learning* é a sua capacidade de retroalimentação, ou seja, de utilizar-se de dados obtidos durante sua execução para melhorar a assertividade de seu sistema. Propõe-se que seja elaborado um sistema que reforce seu conhecimento sobre determinado usuário sempre que ele for reconhecido com alto grau de certeza.

Uma possibilidade para resolver tal problema é pedir para a câmera armazenar uma foto toda vez que o *softmax* encontrar uma alta probabilidade de um usuário ser aquele que já está cadastrado no sistema. O armazenamento desta foto e o constante treinamento da CNN podem levar a tecnologia de reconhecimento a níveis ainda mais assertivos e simplificar sua implementação.

#### *IV.1.7 COMPREENDER O QUE SERIA NECESSÁRIO PARA ESTE SISTEMA TAMBÉM REALIZAR A DETECÇÃO DE OBJETOS*

Como mencionado no Capítulo I , uma das crescentes áreas da VC é a área de veículos autônomos. Mesmo esta não sendo o foco da presente pesquisa, é interessante que se explore as possibilidades do sistema de VC via CNN desenvolvido também realizar a detecção e



rastreamento de objetos, visando também avançar a pesquisa no sentido de promover o desenvolvimento destes.

## CAPÍTULO V CONCLUSÕES

Tendo como parâmetros os objetivos definidos para este estudo, nota-se que houve execução com sucesso de alguns destes. Em específico, o estudo foi capaz de explorar de maneira profunda o principal algoritmo utilizado para VC: as redes neurais convolucionais (Y. Taigman, 2014).

Em referência ao segundo objetivo mencionado no Capítulo I resulta deste estudo uma melhor compreensão referente aos avanços recentes de sistemas de inteligência artificial aplicados em visão computacional, tendo sido abordados artigos recentes que representam onde está o estado da arte desta tecnologia.

O terceiro e quarto objetivos foram também cumpridos com sucesso, tendo o projeto criado em (Brum, Oliveira, & Almeida, 2018) sido estudado e implementado no âmbito deste trabalho.

Nota-se, porém, que este TCC I não foi capaz de cumprir os 2 últimos objetivos descritos no Capítulo I. Desta forma é proposto aqui que o desenvolvimento de uma IA, a população dos dados em nuvem a avaliação de diferentes códigos que podem ser implementados aconteça no TCC II.

### V.1 CONSIDERAÇÕES FINAIS

Finalizado este estudo preliminar da inteligência artificial aplicada ao reconhecimento facial pode-se concluir que houve um necessário avanço em relação à obtenção dos conhecimentos necessários para compreender e implementar um sistema deste tipo. A revisão bibliográfica aqui presente foi essencial para tal, possibilitando o avanço de pesquisas neste sentido dentro do CEFET-MG.

Uma outra contribuição que pode ser mencionada é a de unificar conhecimentos de IA, ML, DL, VC e reconhecimento facial num só lugar. Nota-se que estas informações estão disponíveis principalmente em língua estrangeira. Neste sentido, houve continuidade dos esforços de (Brum, Oliveira, & Almeida, 2018) e (Cruz, Coutinho, & Freire, 2019) em facilitar o acesso a estes materiais na comunidade científica nacional.

## REFERÊNCIAS

- Apple. (2019). **About Face ID advanced technology**. Retrieved from <https://support.apple.com/en-us/HT208108>
- Baer, D., & Weller, C. (2019). **Democratic candidate Andrew Yang just promised to give 10 American families \$12,000 over a year of 'universal basic income.' Here's how the radical policy plan would actually work**. Business Insider. Retrieved from <https://www.businessinsider.com/what-is-basic-income-2016-8>
- Bartlett, J. (2018). **Will 2018 be the year of the neo-luddite?** The Guardian. Retrieved from <https://www.theguardian.com/technology/2018/mar/04/will-2018-be-the-year-of-the-neo-luddite>
- Boudreaux, D. (2017). **Seu padrão de vida hoje é muito maior do que o de um magnata americano há 100 anos**. Instituto Mises Brasil. Retrieved from <https://www.mises.org.br/Article.aspx?id=2672&ac=209170>
- Bringsjord, S., & Govindarajulu, N. S. (2018). **Artificial Intelligence**. (E. N. Zalta, Ed.) *The Stanford Encyclopedia of Philosophy (Fall 2018 Edition)*. Retrieved from <https://plato.stanford.edu/archives/fall2018/entries/artificial-intelligence/>
- Brum, A. L., & Oliveira, M. C. (2018). **FACE UNLOCK: Uma aplicação de Visão Computacional**. Iniciação Científica, CEFET-MG, DEEB.
- Brum, A. L., Oliveira, M. C., & Almeida, A. d. (2018). **Interferência da maquiagem e acessórios no reconhecimento facial**. *Febrar*.
- Cavaioni, M. (2018). **DeepLearning series: Convolutional Neural Networks**. Medium. Retrieved from <https://medium.com/machine-learning-bites/deeplearning-series-convolutional-neural-networks-a9c2f2ee1524>
- Collins-Thompson, K. (2017). **Applied Machine Learning in Python**. *University of Michigan*. Coursera.
- Cruz, E. N., Coutinho, L. P., & Freire, V. B. (2019). **Curso de Visão Computacional**.
- David Silver, A. H. (2016). **Mastering the game of Go with deep neural networks and tree search**. *Nature*.

David Silver, J. S. (2017). **Mastering the game of Go without human knowledge.** *Nature*.

Geller, A. P. (2019). **O segredo do enriquecimento econômico – e por que os países em desenvolvimento continuam atrasados.** Retrieved from <https://www.mises.org.br/article/3108/o-segredo-do-enriquecimento-economico-e-por-que-os-paises-em-desenvolvimento-continuam-atrasados>

Globerman, S. (2019). **Artificial Intelligence will kill jobs—and create them.** Vancouver, British Columbia, Canadá. Retrieved from <https://www.fraserinstitute.org/article/artificial-intelligence-will-kill-jobs-and-create-them>

Heess, N. (2017). **Emergence of Locomotion Behaviours in Rich Environments.** *Cornell University*. Retrieved from <https://arxiv.org/abs/1707.02286>

Joshi, N. (2019). **How Far Are We From Achieving Artificial General Intelligence?** *Forbes*. Retrieved from <https://www.forbes.com/sites/cognitiveworld/2019/06/10/how-far-are-we-from-achieving-artificial-general-intelligence/#723a1b9b6dc4>

Kroll, B., Schaffranek, D., Schriegel, S., & Niggemann, O. (2014). **System modeling based on machine learning for anomaly detection and predictive maintenance in industrial plants.** *Proceedings of the 2014 IEEE Emerging Technology and Factory Automation (ETFA)*.

Levi, G., & Hassner, T. (2015). **Age and Gender Classification using Convolutional Neural Networks.** *2015 IEEE Conference on Computer Vision and Pattern Recognition Workshops*.

Müller, A. C., & Guido, S. (2016). **Introduction to Machine Learning with Python.** O'Reilly.

Nair, A., Mansoori, S., Moghe, R., Shah, P., & Talele, K. (2019). **Driver Assistant System using Haar Cascade and Convolutional Neural Networks (CNN).** *Proceedings of the Third International Conference on Trends in Electronics and Informatics (ICOEI 2019)*.

Ng, A. (2017). **Convolutional Neural Networks Course.** *Deeplearning.ai*. Coursera.

- Razavian, A. (2014). **CNN Features Off-the-shelf: an Astounding baseline for Recognition.** *2014 IEEE Conference on Computer Vision and Pattern Recognition Workshops.*
- Rincon, P. (2019). **Google claims 'quantum supremacy' for computer.** BBC.  
Retrieved from <https://www.bbc.com/news/science-environment-50154993>
- Saha, S. (2018). **A Comprehensive Guide to Convolutional Neural Networks — the ELI5 way.** Towards Data Science. Retrieved from <https://towardsdatascience.com/a-comprehensive-guide-to-convolutional-neural-networks-the-eli5-way-3bd2b1164a53>
- Szeliski, R. (2010). *Computer Vision: Algorithms and Applications.* Springer. Fonte: <http://szeliski.org/Book/>
- The Economist. (2019). **Driverless cars are stuck in a jam.** The Economist. Retrieved from <https://www.economist.com/leaders/2019/10/10/driverless-cars-are-stuck-in-a-jam>
- The White House. (2019). **U.S. Unemployment Rate Falls to 50-Year Low.** Washington, DC, Estados Unidos. Retrieved from <https://www.whitehouse.gov/articles/u-s-unemployment-rate-falls-50-year-low/>
- Thiel, P., & Masters, B. (2014). *From Zero to One.*
- Tran, K. (2018). **Google is capitalizing on AI in marketing.** Business Insider.  
Retrieved from <https://www.businessinsider.com/google-uses-ai-to-enhance-ad-campaigns-2018-7>
- Vidhya. (2019). **What are Haar Features used in Face Detection ?** Medium.  
Retrieved from <https://medium.com/analytics-vidhya/what-is-haar-features-used-in-face-detection-a7e531c8332b>
- Y. Taigman, M. Y. (2014). **DeepFace: Closing the Gap to Human-Level Performance in Face Verification.** *2014 IEEE Conference on Computer Vision and Pattern Recognition.*