

Intensivo

# Deep Learning

Agentes de visão computacional

Para leigos



Sandeco Macedo

*Etiene, meu amor por você é profundo!*

**Copyright © 2024**

ISBN: 978-6-59942-168-6



9 786599 421686

## Prefácio

É com grande entusiasmo que apresento o livro do Prof. Sandeco Macedo, que adentra o empolgante campo do Deep Learning aplicado a criação de agentes de visão computacional.

Enquanto avançamos no século XXI, testemunhamos uma revolução na interseção entre tecnologia e medicina. O papel do aprendizado de máquina, especialmente do Deep Learning, tem sido profundamente transformador, capacitando profissionais de saúde a extrair insights valiosos e precisos de vastos conjuntos de dados de imagens.

O Prof. Sandeco, renomado cientista da computação e professor e pesquisador no Instituto Federal de Goiás (IFG) e Universidade Federal de Goiás (UFG), emerge como um guia exemplar neste campo dinâmico. Sua paixão pela inovação e dedicação ao avanço da ciência o consolidam como uma referência na convergência entre tecnologia e saúde.

Eu como médico, percebo que nesse livro, o Prof. Sandeco não apenas oferece uma visão abrangente do estado atual da aplicação do Deep Learning como por exemplo em imagens médicas que é minha de atuação, mas também compartilha insights práticos e exemplos elucidativos, beneficiando estudantes, pesquisadores e profissionais da área médica e das demais áreas de conhecimento humano.

Para mim, ao abordar os desafios e as oportunidades do uso do Deep Learning em imagens médicas, este livro se destaca como um recurso indispensável para todos os interessados em compreender e explorar todo o potencial dessa tecnologia inovadora.

Estou confiante de que este livro se tornará uma referência essencial para aqueles que buscam aprimorar seus conhecimentos na interseção entre tecnologia e medicina, contribuindo assim para o avanço da ciência e da saúde em escala global.

**Marco Aurélio Carvalho, MD**

*Intensivista Infantil pela Universidade Estadual Paulista, UNESP,*

*Faculdade de Medicina de Botucatu*

*CMIO (Chief Medical Informatics Officer) da UNIMED Piracicaba*

*Especialista em Tecnologia da Informação em Saúde pela*

*Universidade Federal de São Paulo*

*Data Scientist da Vitalyze - Oklahoma, U.S.A.*

## Sumário

Prefácio	4
1 Introdução a Deep Learning	7
1.1 Aprendizado Profundo? Vem que te explico!	9
1.2 Primeiros passos com as Redes que conseguem ver	11
2 IA sem programar – Sim, é verdade!	14
2.1 IA sem mistério: agora para qualquer um	15
2.2 Ensinando máquinas a "ver"imagens	17
2.3 Construindo nossa primeira rede neural visual	21
2.4 Vamos treinar nossa IA? Pegue seu capacete!	26
2.5 A IA no comando: fazendo previsões	28
2.6 Ensinar a IA a ser justa: o que é generalização?	30
2.7 Salvando e exportando nossa invenção	33
2.8 Sua IA online - Servidor gratuito e sem programar	36
2.9 Vamos praticar com alguns exercícios	41
3 Inteligência Artificial Visual	42
3.1 O poder milagroso dos widgets. IA no Orange Canvas	43
3.2 Adicionando superpoderes de visão ao Orange	45
3.3 Doutor e IA, eu estou bem?	47
3.4 Embeddings: A assinatura da IA em cartório	50
3.5 Mais exercícios para não perder o ritmo	55
4 O grande olho que tudo vê!	56
4.1 A revolução da convolução	57
4.2 Mapas da mina. Mapas de ativação	62
4.3 Ativando sua rede com ReLU	63
4.4 Pooling: o que realmente importa e deixando o resto de lado	64
4.5 Flatten: A IA adora comida achatada como uma pizza	66
4.6 Os neurônios da visão por IA	68
4.7 Camada de saída: Sim, entendi o que eu vi	69
4.8 Vamos praticar	70
5 Aventuras com IA nas Nuvens Gratuitas	71
5.1 A arena gratuita de treinamento de IA	72
5.2 Uma GPU grátis só pra mim? Que Maravilhoso!	74

5.3	Vários processos de uma vez? Sim, por favor!	76
6	Deep Learning - Ao infinito e além	79
6.1	Um problema sério a ser resolvido que IA pode ajudar	80
6.2	Carregando imagens para a festa no Colab	82
6.3	Carregando Dados, o combustível de toda IA	85
6.4	Validação: Toda IA também tem seu Enem	87
6.5	Montando uma rede neural: vamos construir juntos	88
6.6	Afinando a rede: os ajustes finais	90
6.7	O grande show: treinando nosso IAluno	91
6.8	Resultados: olha como a nossa IA é boa!	92
6.9	Hora de exercitar: vamos nessa!	95
7	Otimizando em Deep Learning	96
7.1	Eu tenho poucos dados, e agora? Vamos resolver!	97
7.2	Transplante de cérebro digital	101
7.3	Uma IA aprendendo com outra IA	102
7.4	A mágica dos grandes modelos prontos de visão	103
7.5	Transfer learning com Keras: Meu Deus como é fácil!	105
7.6	Fine-tuning: aquele toque final que faz diferença	107
7.7	Callbacks: Os assistentes secretos	109
7.8	Empacotando tudo	111
7.9	Pratique o aprendizado: hora dos exercícios	112

# Introdução a Deep Learning

A inteligência artificial (IA) é uma das inovações tecnológicas mais transformadoras do século XXI. Seu impacto pode ser sentido em uma ampla variedade de setores, desde a saúde até o transporte, passando por finanças, educação e entretenimento. A IA tem o potencial de revolucionar a maneira como vivemos e trabalhamos, automatizando tarefas repetitivas e oferecendo novas formas de resolver problemas complexos. Graças aos avanços na IA, estamos começando a ver sistemas que podem aprender, adaptar-se e executar tarefas de forma autônoma, tornando-se cada vez mais parte integrante de nossas vidas diárias.

Na área da saúde, a IA está proporcionando avanços significativos no diagnóstico e tratamento de doenças. Algoritmos de aprendizado de máquina são utilizados para analisar imagens médicas, identificar padrões que podem escapar ao olho humano e prever o desenvolvimento de doenças com alta precisão. Além disso, assistentes virtuais baseados em IA estão ajudando médicos a acessar informações médicas e históricos de pacientes de maneira mais eficiente, permitindo uma tomada de decisão mais informada e rápida.

No setor de transporte, a IA está no coração do desenvolvimento de veículos autônomos. Esses veículos utilizam uma combinação de sensores, aprendizado de máquina e processamento de dados em tempo real para navegar e tomar decisões no trânsito. Empresas como Tesla, Waymo e Uber estão na vanguarda dessa tecnologia, prometendo reduzir acidentes de trânsito, melhorar a eficiência do transporte e transformar a mobilidade urbana.

O impacto da IA no mercado financeiro também é notável. Algoritmos de negociação automatizada analisam grandes volumes de dados em frações de segundo para executar operações no mercado de ações com alta precisão. Além disso, sistemas de IA são utilizados para detectar fraudes, prever tendências de mercado e fornecer recomendações personalizadas de investimentos, ajudando investidores a tomar decisões mais informadas e seguras.

Na educação, a IA está sendo usada para personalizar a experiência de aprendizagem. Plataformas de aprendizado adaptativo ajustam o conteúdo e o ritmo de ensino com base nas necessidades individuais dos alunos, tornando a educação mais eficaz e acessível. Assistentes virtuais ajudam estudantes com dúvidas, fornecem feedback instantâneo e facilitam a administração de cursos, melhorando a eficiência tanto para alunos quanto para educadores.

O entretenimento também foi profundamente transformado pela IA. Serviços de streaming como Netflix e Spotify utilizam algoritmos de recomendação para sugerir conteúdos baseados nas preferências dos usuários, aumentando o engajamento e a satisfação do cliente. Além disso, a IA está sendo usada na criação de conteúdo, desde composições musicais geradas por máquinas até roteiros de filmes e jogos de vídeo, abrindo novas fronteiras para a criatividade e a produção artística.

Embora os benefícios da IA sejam vastos, também é importante considerar os desafios e implicações éticas que acompanham essa tecnologia. Questões de privacidade, segurança, viés algorítmico e impacto no emprego são preocupações significativas que precisam ser abordadas. Regulamentações apropriadas e uma governança responsável são essenciais para garantir que a IA seja desenvolvida e utilizada de maneira ética e benéfica para a sociedade.

No entanto, para entender completamente o impacto e o potencial da IA, é crucial explorar suas subáreas, como o deep learning. O deep learning, uma técnica avançada de machine learning, está por trás de muitos dos avanços mais impressionantes da IA. Utilizando redes neurais profundas, o deep learning permite que sistemas de IA processem e analisem grandes volumes de dados com uma precisão sem precedentes, abrindo caminho para inovações contínuas em diversos setores.

Imagine que você é um chefe de cozinha encarregado de preparar um prato extremamente complexo, como um suflê de chocolate perfeito. Para alcançar a excelência, você deve seguir várias etapas específicas: selecionar os ingredientes certos, misturar nas proporções exatas, bater as claras em neve na consistência ideal e ajustar a temperatura do forno precisamente. Agora, suponha que você tenha uma equipe de assistentes de cozinha, cada um especializado em uma dessas tarefas.

No entanto, ao invés de seguir uma receita rígida, cada assistente é treinado para aprender e melhorar continuamente com a prática. Inicialmente, cada assistente começa com um conhecimento básico, mas conforme eles trabalham repetidamente, ajustam suas técnicas com base no feedback recebido. Por exemplo, se o suflê não cresceu como esperado, o assistente responsável por bater as claras em neve ajustará sua técnica na próxima tentativa. Esse processo de aprendizado contínuo e ajuste fino é análogo ao que ocorre em uma rede neural profunda (deep learning).

No deep learning, cada camada da rede neural atua como um assistente especializado, responsável por extrair diferentes níveis de características dos dados brutos. A camada inicial pode identificar características simples, como bordas em uma imagem, semelhante a um assistente de cozinha identificando os ingredientes básicos. Camadas subsequentes combinam essas características simples para formar representações mais complexas, como texturas e formas, assim como os assistentes de cozinha combinam ingredientes e técnicas para preparar elementos mais complexos do prato. Finalmente, as camadas superiores da rede neural agregam todas essas características para tomar uma decisão final, como um chef combinando todos os elementos para criar o suflê perfeito.

Assim como um chef de cozinha e sua equipe podem aprimorar continuamente suas habilidades através da prática e do feedback, uma rede neural profunda se refina iterativamente durante o treinamento, ajustando os pesos e os parâmetros para minimizar os erros. Esse processo permite que o deep learning alcance resultados impressionantes em tarefas como reconhecimento de imagem, processamento de linguagem natural e muitas outras aplicações complexas.

## 1.1 APRENDIZADO PROFUNDO? VEM QUE TE EXPLICO!

Imagine que você é um chefe de cozinha encarregado de preparar um prato extremamente complexo, como um suflê de chocolate perfeito. Para alcançar a excelência, você deve seguir várias etapas específicas: selecionar os ingredientes certos, misturar nas proporções exatas, bater as claras em neve na consistência ideal e ajustar a temperatura do forno precisamente. Agora, suponha que você tenha uma equipe de assistentes de cozinha, cada um especializado em uma dessas tarefas.

No entanto, ao invés de seguir uma receita rígida, cada assistente é treinado para aprender e melhorar continuamente com a prática. Inicialmente, cada assistente começa com um conhecimento básico, mas conforme eles trabalham repetidamente, ajustam suas técnicas com base no feedback recebido. Por exemplo, se o suflê não cresceu como esperado, o assistente responsável por bater as claras em neve ajustará sua técnica na próxima tentativa. Esse processo de aprendizado contínuo e ajuste fino é análogo ao que ocorre em uma rede neural profunda (deep learning).

No deep learning, cada camada da rede neural atua como um assistente especializado, responsável por extrair diferentes níveis de características dos dados brutos. A camada inicial pode identificar características simples, como bordas em uma imagem, semelhante a um assistente de cozinha identificando os ingredientes básicos. Camadas subsequentes combinam essas características simples para formar representações mais complexas, como texturas e formas, assim como os assistentes de cozinha combinam ingredientes e técnicas para preparar elementos mais complexos do prato. Finalmente, as camadas superiores da rede neural agregam todas essas características para tomar uma decisão final, como um chef combinando todos os elementos para criar o suflê perfeito.

Assim como um chef de cozinha e sua equipe podem aprimorar continuamente suas habilidades através da prática e do feedback, uma rede neural profunda se refina iterativamente durante o treinamento, ajustando os pesos e os parâmetros para minimizar os erros. Esse processo permite que o deep learning alcance resultados impressionantes em tarefas como reconhecimento de imagem, processamento de linguagem natural e muitas outras aplicações complexas.

O Deep Learning, também conhecido como aprendizado profundo, é uma subárea do aprendizado de máquina que tem se desenvolvido de forma exponencial nas últimas décadas. Sua origem remonta às primeiras redes neurais artificiais na década de 1940, mas foi somente com o advento de poderosos processadores e grandes volumes de dados que essa tecnologia alcançou seu potencial. O desenvolvimento de algoritmos de aprendizado profundo foi impulsionado por avanços significativos em áreas como processamento de imagens e reconhecimento de voz, culminando em uma revolução tecnológica que transformou diversas indústrias.

Conforme a Figura 1.1, deep learning é uma subárea do machine learning, que por sua vez é uma subárea da inteligência artificial. Este diagrama visualiza a hierarquia e a relação entre esses três campos. A Inteligência Artificial, representada pelo círculo maior, é a área abrangente que inclui qualquer técnica que permita às máquinas imitar o comportamento humano. Dentro desse grande círculo, encontra-se o machine learning, que utiliza métodos estatísticos para permitir que as máquinas aprendam e melhorem com a experiência. Finalmente, dentro do círculo de machine learning, está o deep learning, que usa redes neurais profundas para interpretar grandes volumes

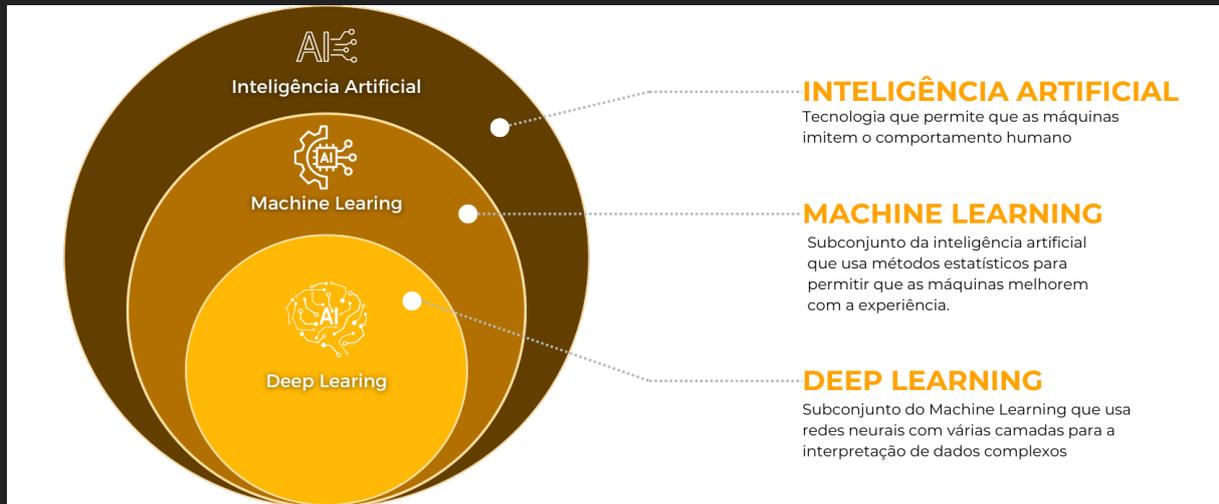


Figura 1.1: Diagrama mostrando a relação entre Inteligência Artificial, Machine Learning e Deep Learning.

de dados e extrair características complexas.

Observe como cada camada interna é um subconjunto da camada externa. A Inteligência Artificial é o campo mais amplo, englobando todas as técnicas que permitem às máquinas realizar tarefas que normalmente requerem inteligência humana. Machine Learning é um subconjunto da IA que se concentra em algoritmos que permitem às máquinas aprender a partir dos dados. Deep Learning é uma abordagem específica dentro do machine learning que utiliza redes neurais profundas com várias camadas para lidar com a complexidade dos dados.

Essa representação é crucial para entender a progressão do geral ao específico, mostrando que o deep learning é uma técnica avançada dentro do campo do machine learning, que por sua vez é parte essencial da inteligência artificial. A hierarquia ilustrada ajuda a compreender a importância e o papel específico de cada uma dessas áreas no desenvolvimento de tecnologias inteligentes.

No cerne do Deep Learning estão as redes neurais artificiais, inspiradas na estrutura e funcionamento do cérebro humano. Essas redes são compostas por camadas de neurônios artificiais que processam dados de entrada, aprendem padrões complexos e fazem previsões ou classificações. Diferentemente dos métodos tradicionais de aprendizado de máquina, que muitas vezes requerem engenharia manual de características, as redes neurais profundas podem aprender representações de dados de forma automática e hierárquica, o que as torna extremamente eficazes para tarefas complexas.

Comparado a outras técnicas de aprendizado de máquina, o Deep Learning se destaca pela sua capacidade de lidar com dados não estruturados e por sua robustez em aprender representações de alta dimensão. Enquanto métodos como máquinas de vetores de suporte ou árvores de decisão podem ser limitados pela necessidade de extração manual de características, as redes neurais profundas podem aprender essas características diretamente dos dados brutos, como imagens, áudio ou texto, permitindo um desempenho superior em muitas aplicações.

As aplicações do Deep Learning são vastas e abrangem várias áreas. Na visão computacional, ele é usado para reconhecimento de objetos, detecção de anomalias e segmentação de imagens. Na

área de processamento de linguagem natural, é empregado em tarefas como tradução automática, análise de sentimento e geração de texto. Em saúde, o Deep Learning está revolucionando o diagnóstico médico ao permitir a análise automática de imagens de raios-X e ressonância magnética. Em transporte, é um componente crucial no desenvolvimento de veículos autônomos. Esses exemplos ilustram como o Deep Learning está transformando indústrias e criando novas possibilidades.

A importância do Deep Learning na atualidade não pode ser subestimada. Ele está na vanguarda da inteligência artificial, impulsionando inovações que antes eram consideradas ficção científica. Sua capacidade de processar e interpretar grandes volumes de dados em tempo real está permitindo avanços significativos em áreas como segurança, saúde, entretenimento e negócios. À medida que os algoritmos se tornam mais sofisticados e os dados mais abundantes, o impacto do Deep Learning continuará a crescer.

As redes neurais, componentes essenciais do Deep Learning, são compostas por neurônios organizados em camadas. Cada neurônio recebe entradas, aplica uma função de ativação e transmite o resultado para a próxima camada. As redes podem ser treinadas usando algoritmos de otimização, como o gradiente descendente, que ajusta os pesos das conexões entre neurônios para minimizar a diferença entre as previsões da rede e os valores reais. Esse processo de treinamento é iterativo e exige grandes volumes de dados e poder computacional significativo.

As Redes Neurais Convolucionais (CNNs) representam uma arquitetura específica de redes neurais projetada para processar dados com uma grade topológica, como imagens. Inspiradas pela organização do córtex visual dos animais, as CNNs são altamente eficazes para tarefas de visão computacional. Elas utilizam camadas convolucionais que aplicam filtros para extrair características locais dos dados de entrada, preservando a relação espacial entre os pixels. Isso permite que as CNNs aprendam características hierárquicas, desde bordas simples até padrões complexos, em diferentes níveis de abstração.

O Deep Learning representa um avanço significativo na capacidade das máquinas de aprender e interpretar dados complexos. Com uma base histórica sólida e uma evolução contínua, ele se estabeleceu como uma ferramenta indispensável em diversas áreas da tecnologia moderna, transformando a maneira como interagimos com o mundo digital e abrindo novas fronteiras para a inovação.

## 1.2 PRIMEIROS PASSOS COM AS REDES QUE CONSEGUEM VER

Imagine que você é um fotógrafo tentando capturar a melhor imagem possível de uma paisagem. Para obter uma foto perfeita, você precisa de uma série de filtros e ajustes que destacam diferentes aspectos da cena. Primeiramente, você pode usar um filtro que realça as bordas dos objetos na imagem, tornando os contornos mais nítidos. Em seguida, você pode aplicar um filtro que ajusta o brilho e o contraste, ajudando a destacar áreas específicas da paisagem. Finalmente, você ajusta a saturação das cores para tornar a imagem mais vibrante e atraente. Cada um desses filtros contribui para a melhoria da imagem final, destacando diferentes características e detalhes.

Da mesma forma, uma Rede Neural Convolucional (CNN) usa várias camadas de filtros, conheci-

dos como convoluções, para analisar e melhorar os dados de entrada, que geralmente são imagens. Cada camada convolucional aplica um conjunto de filtros que extrai características específicas dos dados. As primeiras camadas podem identificar características simples, como bordas e texturas, semelhante ao filtro do fotógrafo que realça contornos. À medida que os dados passam por camadas mais profundas da rede, as características extraídas se tornam progressivamente mais complexas e abstratas, como formas, padrões e objetos inteiros, tal como os ajustes subsequentes feitos pelo fotógrafo para melhorar a imagem.

Esses filtros em uma CNN são ajustados durante o processo de treinamento, onde a rede aprende quais características são mais importantes para a tarefa em questão, seja reconhecimento de objetos, classificação de imagens ou qualquer outra aplicação de visão computacional. Assim como o fotógrafo ajusta seus filtros e configurações com base na iluminação e nas condições da cena, uma CNN ajusta seus filtros internos para otimizar o desempenho na tarefa de análise de imagem, resultando em uma compreensão profunda e detalhada dos dados visuais.

Essa analogia ajuda a entender como as CNNs decompõem e analisam imagens de maneira hierárquica, utilizando filtros convolucionais para capturar características em vários níveis de complexidade, aprimorando a capacidade da rede de interpretar e classificar imagens com alta precisão.

As Redes Neurais Convolucionais (CNNs) são uma classe de redes neurais artificiais projetadas especificamente para processar dados que possuem uma estrutura de grade, como imagens. A estrutura de uma CNN é composta por várias camadas, incluindo camadas convolucionais, camadas de pooling e camadas totalmente conectadas. As camadas convolucionais aplicam filtros, ou kernels, aos dados de entrada para extrair características locais, preservando a relação espacial entre os pixels.

Os filtros nas camadas convolucionais são matrizes de pesos que deslizam sobre a entrada, realizando operações de convolução que resultam em mapas de características. Cada filtro é capaz de detectar diferentes características, como bordas, texturas ou padrões, em várias posições da imagem. Após a aplicação dos filtros, os mapas de características passam por uma função de ativação não linear, como ReLU (Rectified Linear Unit), para introduzir não linearidades no modelo.

As camadas de pooling são usadas para reduzir a dimensionalidade dos mapas de características, mantendo as informações mais relevantes. O pooling mais comum é o *max pooling*, que seleciona o valor máximo em uma janela de tamanhos fixos, reduzindo a resolução espacial dos mapas de características e, conseqüentemente, o número de parâmetros e a carga computacional da rede. Além disso, camadas de normalização, como Batch Normalization, são frequentemente utilizadas para estabilizar e acelerar o treinamento, normalizando as ativações das camadas anteriores.

Para ilustrar, considere uma rede convolucional simples composta por uma camada convolucional com um conjunto de filtros, seguida por uma camada de pooling e uma camada totalmente conectada que gera a saída final. Essa estrutura básica pode ser expandida para redes muito mais profundas e complexas, como as arquiteturas VGG, ResNet e Inception, que são amplamente utilizadas em aplicações modernas.

As CNNs oferecem várias vantagens para a classificação de imagens em comparação com redes totalmente conectadas tradicionais. A principal vantagem é a capacidade de capturar características espaciais hierárquicas devido à aplicação local dos filtros. Além disso, a redução de parâmetros proporcionada pelas camadas convolucionais e de pooling torna as CNNs mais eficientes e menos propensas ao *overfitting*.

As aplicações práticas das CNNs são vastas e incluem a classificação de imagens, a detecção de objetos, a segmentação de imagens e o reconhecimento de ações em vídeos. Em saúde, as CNNs são usadas para analisar imagens médicas, como raios-X e ressonâncias magnéticas, auxiliando no diagnóstico de doenças. Em segurança, são aplicadas em sistemas de vigilância para detectar comportamentos suspeitos ou identificar indivíduos.

Para melhorar o desempenho e a generalização das CNNs, diversas técnicas de regularização são empregadas, como *dropout*, *weight decay* e *data augmentation*. Essas técnicas ajudam a prevenir o *overfitting* e melhorar a robustez do modelo.

O treinamento de uma CNN envolve a minimização de uma função de perda, como a entropia cruzada, usando algoritmos de otimização, como o gradiente descendente estocástico (SGD). Durante o treinamento, o modelo ajusta os pesos dos filtros para reduzir a diferença entre as previsões e os valores reais. A avaliação de uma CNN é feita usando métricas como acurácia, precisão, recall e F1-score, aplicadas a um conjunto de dados de validação.

Em resumo, as Redes Neurais Convolucionais são uma poderosa ferramenta no arsenal de técnicas de aprendizado profundo, oferecendo uma abordagem eficiente e eficaz para a análise de dados visuais. Sua capacidade de aprender representações hierárquicas de dados as torna indispensáveis em diversas aplicações práticas.

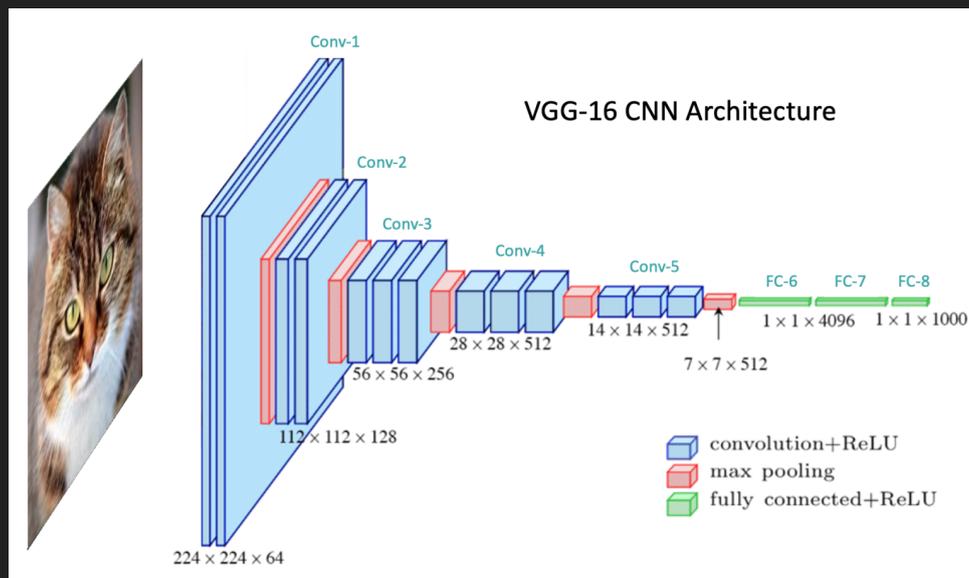


Figura 1.2: Arquitetura VGG-16 mostrando as várias camadas convolucionais, camadas de pooling e camadas totalmente conectadas.