

INSTITUTO FEDERAL DE EDUCAÇÃO, CIÊNCIA E
TECNOLOGIA DO TRIÂNGULO MINEIRO

Vinícius Silveira Bisinoto

**Reconhecimento Automático de Valores em Dados de RPG
Utilizando Redes Neurais Convolucionais**

UBERABA - MG

2025

Vinícius Silveira Bisinoto

**Reconhecimento Automático de Valores em Dados de RPG
Utilizando Redes Neurais Convolucionais**

Trabalho de Conclusão de Curso apresentado ao Instituto Federal de Educação, Ciência e Tecnologia do Triângulo Mineiro, Campus Uberaba Parque Tecnológico, como requisito parcial para a conclusão do Curso de Engenharia da Computação.

Orientador: Prof. Dr. Marcelo Barreiro

UBERABA - MG

2025

Agradecimentos

Agradeço, primeiramente, à minha família, pelo apoio incondicional ao longo da minha trajetória acadêmica, especialmente à minha mãe Valéria, pela presença constante, e ao meu pai, que mesmo distante fisicamente, continua sendo minha inspiração e exemplo de perseverança. Agradeço à Danielly Vitória Rodrigues, amiga de longa data e profissional talentosa, pela disponibilidade e generosidade em realizar o ensaio fotográfico dos dados utilizados neste trabalho, contribuindo significativamente para a qualidade visual do dataset. Aos professores do curso de Engenharia da Computação, especialmente ao meu orientador, pelo acompanhamento atencioso e pelas valiosas sugestões ao longo da pesquisa. Por fim, agradeço aos colegas, amigos e jogadores das mesas de RPG que mestrei ao longo dos últimos anos — suas experiências práticas e comentários foram fundamentais para perceber a necessidade de uma solução como esta.

Resumo

Este trabalho apresenta o desenvolvimento de um sistema baseado em redes neurais convolucionais (CNNs) para reconhecer automaticamente valores em dados físicos de RPG a partir de imagens. Foram utilizadas técnicas de visão computacional, com foco em classificação multiclasse e pré-processamento de imagens. A arquitetura foi treinada com um conjunto próprio de imagens categorizadas. Os resultados obtidos demonstram a viabilidade da abordagem e sugerem possíveis aplicações em sessões híbridas ou automatizadas de RPG.

Palavras-chave: Visão computacional, redes neurais convolucionais, RPG, classificação de imagens.

Abstract

This work presents the development of a system based on convolutional neural networks (CNNs) to automatically recognize the values on the upper faces of physical RPG dice from images. Computer vision techniques were used, focusing on multiclass classification and image pre-processing. The architecture was trained on a custom-labeled image dataset. The obtained results demonstrate the viability of the approach and suggest possible applications in hybrid or automated RPG sessions.

Keywords: Computer vision, convolutional neural networks, RPG, image classification.

Lista de Figuras

| | | |
|---|---|----|
| 1 | Exemplos de dados poliédricos utilizados em jogos de RPG. | 9 |
| 2 | Linha do tempo simplificada e ilustrada da evolução dos RPGs | 13 |
| 3 | Fluxo de um sistema de visão computacional | 15 |
| 4 | Esquema conceitual da arquitetura de uma CNN, inspirado na LeNet-5. | 20 |
| 5 | Trecho de código-fonte com a definição da arquitetura da CNN. | 21 |

Sumário

| | | |
|----------|--|-----------|
| 1 | Introdução | 8 |
| 2 | Fundamentação Teórica | 11 |
| 2.1 | Jogos de Interpretação de Papéis (RPG) | 11 |
| 2.1.1 | História e Evolução dos RPGs | 11 |
| 2.1.2 | A Importância dos Dados Físicos no RPG | 14 |
| 2.2 | Visão Computacional | 14 |
| 2.2.1 | Definição e Objetivos | 14 |
| 2.2.2 | Etapas do Processamento de Imagens | 15 |
| 2.2.3 | Reconhecimento Óptico de Caracteres (OCR) | 16 |
| 2.3 | Aprendizado de Máquina e Aprendizado Profundo | 17 |
| 2.3.1 | Conceitos Básicos de Aprendizado de Máquina | 17 |
| 2.3.2 | Avanços com Aprendizado Profundo (Deep Learning) | 17 |
| 2.4 | Redes Neurais Convolucionais (CNNs) | 17 |
| 2.4.1 | Estrutura Geral de uma CNN | 17 |
| 2.4.2 | Funcionamento das Camadas Convolucionais | 17 |
| 2.4.3 | Pooling e Redução Dimensional | 17 |
| 2.4.4 | Funções de Ativação e Camadas Densas | 17 |
| 2.4.5 | Aplicações Práticas das CNNs | 17 |
| 2.5 | Aplicações Relacionadas: Reconhecimento de Dígitos e Objetos | 18 |
| 2.5.1 | Caso Clássico: LeNet-5 e MNIST | 18 |
| 2.5.2 | Projetos Semelhantes | 18 |
| 2.6 | Considerações Finais | 18 |
| 3 | Metodologia | 18 |
| 3.1 | Coleta e Estruturação do Conjunto de Dados | 18 |
| 3.2 | Processamento de Imagens | 19 |
| 3.3 | Divisão em Treinamento e Validação | 19 |
| 3.4 | Arquitetura da Rede Neural Convolucional | 19 |
| 3.5 | Treinamento do Modelo | 21 |

1 Introdução

Os jogos de interpretação de papéis (Role-Playing Games - RPGs) têm conquistado cada vez mais espaço não só como forma de entretenimento, mas também como ferramenta pedagógica, terapêutica e social. Desde o lançamento de *Dungeons & Dragons* em 1974, o RPG se consolidou como um gênero de jogos que une imaginação, cooperação e raciocínio lógico, promovendo experiências imersivas onde os jogadores atuam em conjunto na construção de narrativas. [1, 2]

Uma das características marcantes desses jogos é o uso de dados físicos para determinar sucessos ou fracassos de determinadas ações, como ataques, testes de habilidade ou eventos aleatórios, e, diferentemente dos jogos de tabuleiro convencionais, o RPG utiliza-se de uma diversidade de dados poliédricos — como d4, d6, d8, d10, d12 e d20 — que ampliam a variação probabilística das ações dos jogadores e tornam o jogo mais dinâmico e imprevisível. Além de sua usabilidade, esses dados se tornaram símbolos da cultura RPGista, muitas vezes personalizados com diferentes materiais, cores e gravuras, ou até eternizados em tatuagens, obras artísticas e coleções, reforçando o vínculo afetivo entre jogador e dado.[2]

Apesar do crescimento de plataformas digitais que oferecem alternativas de rolagem automatizada, como *Roll20* ou *Foundry VTT*, e até comandos de rolagem em bots de Discord, muitos jogadores ainda preferem o uso dos dados físicos. Essa preferência se deve tanto ao valor simbólico e afetivo atribuído aos objetos quanto à percepção de que a aleatoriedade física é mais justa ou autêntica do que a gerada computacionalmente. O som do dado rolando, o toque, o peso e até o suspense visual da face superior prestes a ser revelada são elementos sensoriais que contribuem para a ambientação do jogo e tornam a experiência mais sensorial, orgânica e memorável. [3]

No entanto, essa escolha estética e emocional traz um desafio prático. Em sessões que envolvem múltiplas rolagens com dados, que muitas vezes podem conter múltiplos dados simultâneos — como testes de dano, ataques em área ou ações conjuntas — a contagem manual dos resultados obtidos pode se tornar demorada e propensa a erros. Esse tempo de cálculo afeta diretamente o ritmo narrativo da partida, especialmente em jogos com muitos participantes ou quando os personagens utilizados estejam com suas habilidades já em maior nível, aumentando os números e resultando que dezenas de dados acabem sendo lançados ao mesmo tempo. Mesmo entre grupos experientes, a perda de fluidez entre ação e contagem numérica pode comprometer a imersão da experiência e gerar frustrações.

A crescente demanda por ferramentas que otimizem o tempo de jogo sem abrir mão da experiência de rolagem física motivou este trabalho, que propõe uma solução baseada em visão computacional para automatizar a leitura visual dos dados físicos durante uma rolagem, mantendo o aspecto tradicional do RPG com os dados físicos enquanto melhora a fluidez narrativa das sessões através da velocidade que um cálculo automático pode trazer. [4, 5, 6]

A Figura 1 ilustra alguns exemplos de dados utilizados em sistemas de RPG, destacando

a variedade de formatos e faces numéricas. Essa diversidade, embora enriqueça o jogo, representa também uma dificuldade computacional para quem deseja automatizar a leitura desses elementos visuais.



Figura 1: Exemplos de dados poliédricos utilizados em jogos de RPG.
Foto por Danielly Vitória Rodrigues, 2025.

Partindo deste contexto, este trabalho tem como objetivo analisar a viabilidade do uso de Redes Neurais Convolucionais (CNNs) para o reconhecimento automático dos valores presentes nas faces superiores de dados físicos utilizados em jogos de RPG, a partir de um banco de imagens previamente separado e categorizado conforme o valor exibido na face superior dos dados. A proposta envolve o desenvolvimento e a avaliação experimental de um modelo de classificação multiclasse baseado em técnicas de visão computacional e aprendizado profundo, com foco na análise de seu desempenho em condições práticas simuladas. Para isso, será utilizada uma arquitetura de Rede Neural Convolucional (CNN) composta por múltiplas camadas convolucionais e de pooling, seguidas por camadas densas e uma camada de saída com função de ativação *softmax*, adequada ao reconhecimento de múltiplas classes de saída — correspondentes aos valores possíveis nas faces dos dados de RPG. As CNNs foram escolhidas por demonstrarem bastante eficiência em tarefas de reconhecimento de padrões visuais, especi-

almente em imagens onde existem cenários com variação de iluminação, ângulo, cor e textura. [7, 8]

As Redes Neurais Convolucionais (CNNs) têm seu destaque em diversas áreas da inteligência artificial, como reconhecimento facial, diagnósticos por imagem, leitura de placas de veículos e OCR (Reconhecimento Óptico de Caracteres), sendo utilizadas amplamente devido ao seu sucesso em contextos que exigem robustez, acurácia e capacidade de generalização [7, 8]. Por ter várias camadas organizadas em etapas, essa arquitetura consegue identificar detalhes visuais em diferentes níveis, o que a torna muito eficaz para reconhecer símbolos, números e formas em objetos tridimensionais. — como os dados físicos utilizados em jogos de RPG — mesmo sob condições variáveis de iluminação, ângulo ou resolução.

A principal motivação para o desenvolvimento deste estudo nasce da vivência pessoal do autor como mestre de RPG, especialmente ao observar, ao longo de diversas sessões, a forte preferência dos jogadores pelo uso de dados físicos. Mesmo com a ampla oferta de ferramentas digitais que simulam rolagens de forma prática e eficiente, é comum notar que muitos jogadores ainda optam pelos dados reais, seja pelo apego emocional, a sensação de adrenalina e expectativa ao rolar um dado, ou pela conexão simbólica que esses objetos representam dentro da cultura do RPG. No entanto, essa escolha, apesar de compreensível e até esperada, pode acabar trazendo desafios durante o jogo, principalmente em momentos em que múltiplos dados são rolagados ao mesmo tempo. Nessas situações, a contagem manual dos resultados tende a consumir um tempo precioso, além de estar sujeita a erros que podem comprometer a precisão e, por consequência, a imersão e o ritmo narrativo da partida. Dessa forma, automatizar esse processo de leitura dos valores nos dados físicos surge como uma solução equilibrada: respeita e preserva a tradição dos jogos de mesa, mas aproveita os recursos da tecnologia para tornar a experiência mais fluida, dinâmica e eficiente para todos os participantes.

Essa proposta também se alinha com os avanços mais recentes no uso de sistemas de visão computacional, com destaque para o aumento da sua adesão em tecnologias assistivas. Um bom exemplo disso são os sistemas de reconhecimento óptico de caracteres (OCR), que hoje já são amplamente utilizados na conversão de documentos físicos em versões digitais editáveis, com uma precisão invejável. De forma semelhante, desenvolver um sistema capaz de identificar automaticamente os números presentes nas faces superiores dos dados físicos representa uma aplicação com potencial promissor. Tal recurso pode funcionar como uma importante ferramenta de apoio tanto para mestres quanto para jogadores, especialmente para aquelas sessões realizadas de forma online ou híbrida, onde a visualização e validação das rolagens físicas continua sendo valorizada como parte da experiência principal.

Com base nas necessidades identificadas ao longo da experiência prática com jogos de RPG de mesa, este trabalho tem como objetivo principal o desenvolvimento e a avaliação de um modelo computacional, fundamentado em redes neurais convolucionais (CNNs), capaz de reconhecer os valores exibidos nas faces superiores de dados físicos. Essa identificação será realizada a partir da análise de um conjunto de imagens previamente coletado e devidamente

categorizado.

Para atingir esse objetivo, propõe-se então a execução das seguintes etapas:

- Construção de um banco de dados composto por imagens de diversos tipos de dados de RPG, capturadas sob diferentes ângulos de visão, exibindo valores diversos em sua face superior com o intuito de refletir a variabilidade presente em situações reais;
- Projeção e treinamento de uma arquitetura de rede neural convolucional voltada especificamente para a tarefa de classificação multiclasse dos valores numéricos visíveis nas imagens;
- Avaliação do desempenho do modelo utilizando métricas quantitativas, como acurácia e taxa de erro, considerando diferentes cenários simulados;
- Discussão voltada tanto à avaliação da aplicabilidade prática do sistema em sessões de RPG, quanto à identificação de pontos de melhoria e caminhos possíveis para o aprimoramento e a evolução do projeto em versões futuras.

A proposta citada parte, portanto, da oportunidade avaliada pela observação de uma limitação no uso dos dados físicos em sessões de RPG, propondo o uso de técnicas de aprendizado profundo como uma solução inovadora. Ao longo do trabalho, serão apresentados os fundamentos teóricos que embasam o projeto, a metodologia adotada para construção e validação do modelo, os resultados experimentais obtidos e, por fim, as considerações finais e sugestões para futuras melhorias.

2 Fundamentação Teórica

Este capítulo apresenta os fundamentos teóricos que sustentam o desenvolvimento do projeto, abrangendo os temas de jogos de RPG, visão computacional, aprendizado profundo, redes neurais convolucionais e reconhecimento de padrões visuais. O objetivo é fornecer ao leitor o embasamento necessário para compreender a relevância, os desafios e as soluções técnicas envolvidas no reconhecimento automático de valores em dados físicos de RPG.

2.1 Jogos de Interpretação de Papéis (RPG)

2.1.1 História e Evolução dos RPGs

Os jogos de interpretação de papéis, popularmente conhecidos como RPGs (do inglês, Role-Playing Games), constituem um dos gêneros lúdicos mais complexos e influentes da cultura popular contemporânea. Sua origem remonta a uma série de influências históricas, culturais e mecânicas que se entrelaçaram ao longo do século XX, até culminar na criação do primeiro

sistema formal de RPG: *Dungeons & Dragons* (D&D), publicado em 1974 por Gary Gygax e Dave Arneson.

Segundo Peterson (2012), os RPGs não surgiram isoladamente, mas como consequência direta da evolução dos jogos de guerra (wargames), que eram amplamente praticados por entusiastas militares e civis ao longo dos séculos XIX e XX. Esses jogos utilizavam miniaturas, mapas e regras complexas para simular combates históricos, sendo voltados principalmente para recreações táticas. A grande inovação de Arneson, e posteriormente de Gygax, foi substituir o controle de exércitos pelo controle de personagens individuais com papéis específicos e atributos únicos — introduzindo, assim, o elemento narrativo e subjetivo que diferencia os RPGs dos jogos de tabuleiro e estratégia.

Peterson também enfatiza o papel dos clubes de wargames norte-americanos no desenvolvimento das primeiras versões do que viria a se tornar o RPG moderno. Em particular, o cenário fictício de Blackmoor, criado por Dave Arneson, foi um dos primeiros mundos de campanha onde os jogadores podiam assumir papéis de personagens fixos e participar de aventuras continuadas.

A formalização dos RPGs como gênero distinto se deu com a publicação da primeira edição de *Dungeons & Dragons*, que introduziu conceitos centrais como atributos numéricos (Força, Destreza, Inteligência, etc.), rolagem de dados poliédricos, níveis de experiência (XP), classes (como guerreiro, mago, ladrão) e sistemas de combate que integravam chance e estratégia. A proposta do jogo era que o Mestre (ou *Dungeon Master*, DM) assumisse o papel de narrador e árbitro, enquanto os jogadores assumiam personagens fictícios e colaboravam na construção de uma história interativa.

Hitchens e Drachen (2007) classificam os RPGs em quatro categorias principais: RPGs de mesa (como D&D), RPGs ao vivo (*Live Action Role-Playing Games*, LARP), RPGs eletrônicos (*Computer RPGs*) e MUDs (*Multi-User Dungeons*). Essa taxonomia revela como o conceito de RPG se expandiu para diferentes mídias e contextos, mantendo como elemento unificador a ideia de interpretar um papel em um mundo imaginário sob certas regras.

A evolução dos RPGs também passou por uma transformação estética e temática significativa. Durante os anos 1980 e 1990, sistemas alternativos ao D&D surgiram, como GURPS (*Generic Universal RolePlaying System*), *Vampire: The Masquerade* (que inaugurou o gênero *Storyteller*), e *Call of Cthulhu*, baseado na obra de H. P. Lovecraft. Esses sistemas enfatizavam aspectos como interpretação dramática, dilemas morais e ambientações sombrias, ampliando o alcance narrativo dos jogos.

Nos anos 2000, os RPGs testemunharam uma popularização global com o advento de plataformas digitais, fóruns de discussão e, mais recentemente, com a ascensão de *actual plays* (sessões transmitidas ao vivo por streamers e criadores de conteúdo) e podcasts especializados. Títulos como *Critical Role* e canais como o *Ordem Paranormal* ajudaram a impulsionar uma nova geração de jogadores e expandiram o apelo do RPG como forma de entretenimento participativo.

A Figura 2 ilustra uma linha do tempo simplificada da evolução dos RPGs, desde suas origens nos wargames até sua presença contemporânea em plataformas digitais.



Figura 2: Linha do tempo simplificada e ilustrada da evolução dos RPGs, desde os wargames até os RPGs modernos.

Fonte: Silva, Gabriela Prates da; Prates da Silva, Eduardo (2020). *História do RPG: linha do tempo*. figshare. Figure.

<https://doi.org/10.6084/m9.figshare.12493895.v5>

Em síntese, os RPGs representam uma forma única de jogo baseada na colaboração narrativa, na improvisação criativa e na imersão ficcional. Sua evolução demonstra não ape-

nas uma transformação nas mecânicas de jogo, mas também um reflexo de mudanças culturais, sociais e tecnológicas. Ao longo das décadas, o RPG consolidou-se como ferramenta educacional, artística e terapêutica, além de manter seu lugar como um dos pilares do entretenimento alternativo.

2.1.2 A Importância dos Dados Físicos no RPG

Discussão sobre a função simbólica e prática dos dados físicos, incluindo o impacto da rolagem real na imersão e dinâmica do jogo.

- Referência: [2]

2.2 Visão Computacional

2.2.1 Definição e Objetivos

A Visão Computacional é um campo interdisciplinar da ciência e engenharia que busca desenvolver sistemas capazes de extrair, interpretar e compreender informações a partir de imagens ou vídeos do mundo real. Inspirada nos mecanismos da visão humana, essa área utiliza algoritmos e técnicas matemáticas para permitir que computadores reconheçam padrões, identifiquem objetos e tomem decisões com base em dados visuais [4].

Dentro do contexto da Engenharia da Computação, a Visão Computacional representa um elo entre o hardware e o software, exigindo conhecimento sólido em áreas como álgebra linear, processamento de sinais, inteligência artificial e arquitetura de computadores. A integração de câmeras com sistemas embarcados, o uso de GPUs para paralelismo em algoritmos de processamento de imagem e a aplicação de redes neurais profundas são exemplos de como esse campo dialoga com os fundamentos da engenharia [9].

Os objetivos da Visão Computacional variam conforme o domínio de aplicação, mas incluem tarefas como:

- **Reconhecimento de padrões:** identificar rostos, placas de veículos, objetos ou caracteres;
- **Segmentação de imagem:** separar regiões de interesse com base em características como cor, textura ou bordas;
- **Rastreamento:** acompanhar o movimento de objetos ao longo do tempo em sequências de vídeo;
- **Reconstrução tridimensional:** obter modelos 3D a partir de múltiplas imagens 2D;
- **Análise de cenas:** compreender o contexto global de uma imagem, como a contagem de pessoas ou a detecção de anomalias.

O avanço das redes neurais convolucionais (CNNs), a disponibilidade de grandes conjuntos de dados rotulados e o aumento no poder computacional possibilitaram resultados significativos nos últimos anos, com aplicações em áreas como saúde (diagnóstico por imagem), indústria (inspeção automatizada), segurança (monitoramento por câmeras) e entretenimento (realidade aumentada).

Em projetos como este TCC, a Visão Computacional assume papel central, possibilitando a interface entre os dados físicos do mundo real (os dados de RPG) e o sistema computacional responsável por interpretá-los automaticamente.

2.2.2 Etapas do Processamento de Imagens

O processo de construção de um sistema de visão computacional envolve diversas etapas encadeadas, que vão desde a aquisição dos dados até a inferência dos resultados. Cada uma dessas etapas tem o seu papel no desempenho do modelo, exigindo que sejam feitas decisões técnicas pensadas com base nos objetivos do problema em questão.

A Figura 3 ilustra o fluxo completo do trabalho em um sistema de visão computacional, trazendo destaque para os pontos de decisão e os vieses que podem influenciar negativamente o resultado final.

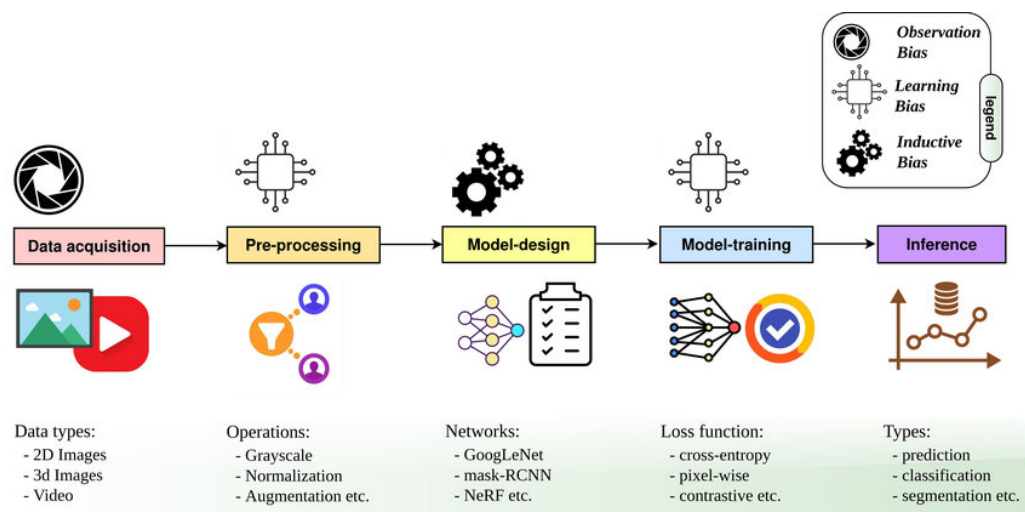


Figura 3: Fluxo de um sistema de visão computacional.

Fonte: *Physics-Informed Computer Vision: A Review and Perspectives - Scientific Figure on ResearchGate.*

Available from: https://www.researchgate.net/figure/Computer-vision-pipeline-showing-different-biases-and-different-points-of-physics_fig3_371136308 [accessed 5 Jun 2025]

Esse fluxo é composto por cinco etapas, sendo elas:

- **Aquisição de dados:** A etapa onde ocorre a coleta das imagens ou vídeos que serão utilizados no treinamento do modelo. Os dados podem vir de fotos (2 dimensões), sensores

com profundidade (3 dimensões) ou streamings em tempo real. A qualidade e a diversidade dos dados capturados influenciam diretamente a capacidade do modelo de fazer generalização.

- **Pré-processamento:** A etapa na qual são aplicadas transformações que preparam os dados para serem utilizados pela rede neural. Dentre as operações mais comuns estão conversão para tons de cinza, normalização dos valores dos pixels, e técnicas de aumento de dados (*data augmentation*). No contexto deste trabalho, essa etapa incluiu o corte das bordas, redimensionamento padronizado das imagens e conversão de canais de cor RGB.
- **Projeto do modelo:** A etapa responsável pela arquitetura da rede convolucional que será utilizada, como o número de camadas, quais filtros, suas funções de ativação e dimensões de entrada/saída. Modelos populares, como GoogLeNet, ResNet e Mask-RCNN, oferecem diferentes estratégias para resolver tarefas como classificação, detecção ou segmentação.
- **Treinamento do modelo:** A etapa que o modelo é exposto às imagens rotuladas e ajusta seus pesos internos utilizando funções de perda como a entropia cruzada (*cross-entropy*) ou erro quadrático médio. O desempenho do modelo é medido com base em sua acurácia e na capacidade de generalizar para dados não vistos.
- **Inferência:** A etapa após o treinamento. Aqui o modelo é utilizado para fazer previsões sobre novos dados. O tipo de inferência pode variar entre classificação, segmentação, detecção de objetos ou regressão, dependendo do objetivo do projeto.

Além das etapas técnicas, a imagem também destaca três tipos de vieses que podem afetar os resultados: **viés de observação** (dados enviesados), **viés de aprendizado** (modelo limitado em sua capacidade de generalização) e **viés indutivo** (pressupostos errados no design do modelo).

O entendimento de todas essas etapas é essencial para engenheiros da computação que desejam aplicar visão computacional a problemas do mundo real, como o reconhecimento automático de valores em dados físicos de RPG proposto neste trabalho.

2.2.3 Reconhecimento Óptico de Caracteres (OCR)

Breve explicação do OCR tradicional e comparação com abordagens baseadas em aprendizado profundo.

- Referência: [4, 9]

2.3 Aprendizado de Máquina e Aprendizado Profundo

2.3.1 Conceitos Básicos de Aprendizado de Máquina

Aprendizado supervisionado, não supervisionado e por reforço; problemas de classificação.

- Referência: [9]

2.3.2 Avanços com Aprendizado Profundo (Deep Learning)

O surgimento das redes neurais profundas e suas vantagens em problemas com grande volume de dados e complexidade visual.

- Referência: [8]

2.4 Redes Neurais Convolucionais (CNNs)

2.4.1 Estrutura Geral de uma CNN

Apresentação da arquitetura típica: camadas convolucionais, pooling, flatten e densas.

- Referência: [7, 8]

2.4.2 Funcionamento das Camadas Convolucionais

Filtro, stride, padding e extração de características espaciais.

- Referência: [7, 6]

2.4.3 Pooling e Redução Dimensional

Funções de pooling (max, average) e sua função na redução de sobreajuste.

- Referência: [6]

2.4.4 Funções de Ativação e Camadas Densas

Comparação entre ReLU, Sigmoid e Softmax. Papel das camadas densas e da saída.

- Referência: [8]

2.4.5 Aplicações Práticas das CNNs

Aplicações em reconhecimento facial, OCR, leitura de placas, diagnóstico por imagem.

- Referência: [8, 7]

2.5 Aplicações Relacionadas: Reconhecimento de Dígitos e Objetos

2.5.1 Caso Clássico: LeNet-5 e MNIST

Modelo LeNet-5 de LeCun e sua aplicação no MNIST para reconhecimento de dígitos manuscritos.

- Referência: [7]

2.5.2 Projetos Semelhantes

Análise de projetos com reconhecimento de dados, OCR em contextos semelhantes, projetos com CNNs aplicadas a objetos tridimensionais e datasets customizados. (Adaptar conforme trabalhos encontrados na revisão de literatura.)

2.6 Considerações Finais

Síntese dos principais conceitos abordados, reforçando o embasamento teórico que sustenta a metodologia aplicada no presente trabalho.

3 Metodologia

Este capítulo descreve os materiais e os métodos empregados no desenvolvimento do presente trabalho, com foco na estruturação do conjunto de dados, no processamento das imagens, no modelo de rede neural convolucional utilizado e nas estratégias de treinamento e validação do classificador. As decisões adotadas ao longo do processo são justificadas com base na realidade do projeto e nas limitações observadas durante a execução experimental.

3.1 Coleta e Estruturação do Conjunto de Dados

Para a realização dos testes, foi construído um conjunto de imagens próprio contendo fotografias de dados físicos utilizados em jogos de RPG. As imagens foram capturadas em estúdio fotográfico, com auxílio da profissional Danielly Vitória Rodrigues, garantindo qualidade adequada de resolução, iluminação e enquadramento.

As imagens foram organizadas em pastas segundo os valores visíveis nas faces superiores dos dados, totalizando inicialmente 10 classes: *0_zero* até *9_nine*. Cada uma dessas pastas continha subpastas correspondentes aos diferentes tipos de dados (ex.: d4, d6, d8, d10). Essa estrutura permitiu manter a diversidade de formatos dentro de cada valor, simulando condições reais de uso.

Entretanto, durante a análise da distribuição dos dados, foi observado que as classes *0_zero* até *9_nine* possuíam um número significativamente inferior de imagens em relação às

demais. Essa escassez comprometia a representatividade estatística e a robustez do treinamento, além de aumentar o risco de viés ou sobreajuste da rede.

Por essa razão, optou-se por restringir o escopo do modelo inicial às classes **1 a 8**, que apresentavam uma quantidade razoável de imagens (acima de 80 amostras por classe no total). As classes excluídas foram mantidas como possibilidade para expansão futura do projeto, mediante a ampliação da base.

3.2 Processamento de Imagens

Todas as imagens passaram por uma etapa de pré-processamento com o objetivo de remover espaços em branco ao redor dos dados e padronizar o formato das entradas. O processamento incluiu:

- **Corte absoluto:** Remoção de bordas fixas nas extremidades superior, inferior, esquerda e direita da imagem, mantendo foco no dado centralizado;
- **Redimensionamento:** Ajuste da imagem para dimensões fixas de 160x160 pixels, facilitando a entrada na rede convolucional;
- **Conversão RGB:** Padronização do canal de cor para compatibilidade com bibliotecas de visão computacional;

Esse tratamento foi realizado utilizando a biblioteca OpenCV. As imagens processadas foram salvas em pastas separadas para treino e validação.

3.3 Divisão em Treinamento e Validação

Para cada uma das classes (1 a 8), as imagens foram embaralhadas e divididas em 80% para treinamento e 20% para validação. A divisão foi feita com semente fixa para garantir reprodutibilidade. Devido ao tamanho limitado do dataset, optou-se por não subdividir o conjunto de validação em um conjunto adicional de teste. Essa decisão visa manter uma quantidade mínima de imagens por classe nas avaliações, evitando séries com menos de cinco amostras.

Para futuros trabalhos, recomenda-se a coleta de um novo conjunto independente de imagens para testes finais e validação cruzada do modelo.

3.4 Arquitetura da Rede Neural Convolucional

A arquitetura da Rede Neural Convolucional (CNN) utilizada neste trabalho foi projetada para equilibrar capacidade de extração de características e simplicidade estrutural, considerando a dimensão relativamente pequena do conjunto de dados e o objetivo de classificação multiclasse (valores de 1 a 8). A estrutura final consiste em:

- Uma camada de reescalonamento (Rescaling) que normaliza os valores dos pixels das imagens de entrada para o intervalo $[0,1]$, dividindo por 255. Essa etapa é essencial para acelerar a convergência durante o treinamento.
- Três blocos compostos por:
 - Uma camada Conv2D com filtros 3×3 , ativação **ReLU** e preenchimento 'same', com 16, 32 e 64 filtros, respectivamente;
 - Uma camada *MaxPooling2D* após cada convolução, reduzindo as dimensões espaciais e promovendo generalização.
- Uma camada de *Flatten*, que transforma o mapa de características extraído em um vetor unidimensional.
- Uma camada densa *Dense* com 128 neurônios e ativação **ReLU**, responsável por combinar as características extraídas.
- Uma camada de saída *Dense* com 8 neurônios e ativação **softmax**, adequada para classificação multiclasse com rótulos inteiros de 0 a 7 (associados às classes 1 a 8).

O modelo foi compilado com otimizador Adam e função de perda *Sparse Categorical Crossentropy*, uma vez que os rótulos são inteiros e não codificados em one-hot, utilizando o otimizador Adam com taxa de aprendizado **0.0005**. A métrica de avaliação escolhida foi a acurácia.

A Figura 4 apresenta um diagrama ilustrativo inspirado na arquitetura LeNet-5, um dos modelos de redes neurais convolucionais mais clássicos, proposto por LeCun et al. (1998) para reconhecimento de dígitos manuscritos. Embora a estrutura exata da rede implementada neste trabalho difira em termos de parâmetros, profundidade e dimensionalidade das entradas, o diagrama é representativo do fluxo de processamento típico de uma CNN: camadas de convolução seguidas por subamostragem (pooling), camadas densas totalmente conectadas e uma camada final com função de ativação *softmax* para classificação multiclasse.

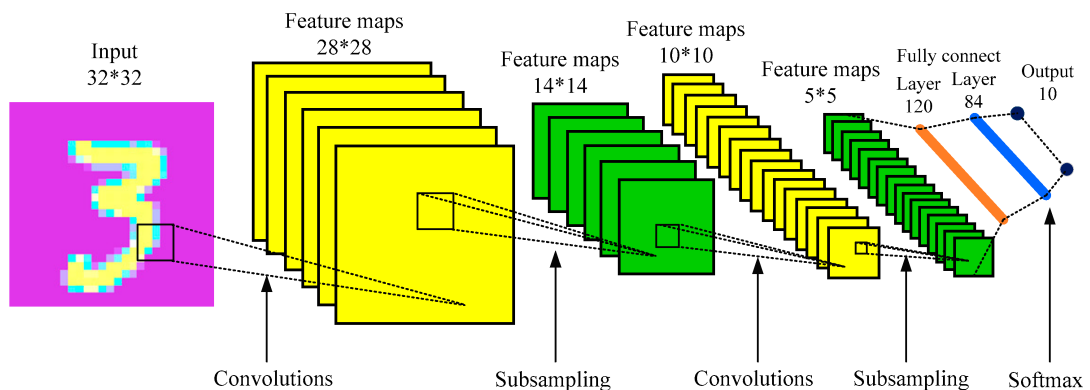


Figura 4: Esquema conceitual da arquitetura de uma CNN, inspirado na LeNet-5.

A Figura 5 apresenta uma captura de tela do trecho de código-fonte responsável pela definição da arquitetura final da rede neural convolucional utilizada neste trabalho. A imagem evidencia a sequência de camadas adotadas no modelo, implementadas com a biblioteca Keras.

```
model = tf.keras.Sequential([
    tf.keras.layers.Rescaling(
        1. / image_color_channel_size,
        input_shape=image_shape
    ),
    tf.keras.layers.Conv2D(16, 3, padding='same', activation='relu'),
    tf.keras.layers.MaxPooling2D(),
    tf.keras.layers.Conv2D(32, 3, padding='same', activation='relu'),
    tf.keras.layers.MaxPooling2D(),
    tf.keras.layers.Conv2D(64, 3, padding='same', activation='relu'),
    tf.keras.layers.MaxPooling2D(),
    tf.keras.layers.Flatten(),
    tf.keras.layers.Dense(128, activation='relu'),
    tf.keras.layers.Dense(8, activation='softmax')
])
```

Figura 5: Trecho de código-fonte com a definição da arquitetura da CNN.

3.5 Treinamento do Modelo

O treinamento da rede neural convolucional (CNN) foi conduzido de forma iterativa, passando por diversas fases de ajuste de parâmetros, avaliação de desempenho e análise qualitativa dos resultados. O objetivo principal era obter um modelo capaz de classificar corretamente os valores nas faces superiores de dados físicos de RPG com alta acurácia e boa generalização.

Inicialmente, o modelo foi treinado por 20 épocas, com um *learning rate* de 0.001, sem aplicação de técnicas de *data augmentation*. Essa primeira configuração foi escolhida por simplicidade e por estar alinhada com abordagens comuns na literatura introdutória de CNNs. Apesar da acurácia de treino apresentar crescimento constante, a acurácia de validação mostrava valores baixos, sugerindo dificuldade de generalização do modelo.

Tentativas subsequentes foram feitas para aumentar a robustez do treinamento por meio de técnicas de aumento de dados. Utilizou-se uma camada de *data augmentation* composta por rotações, espelhamentos e zoom aleatórios. No entanto, todos os testes com essa abordagem resultaram em queda expressiva de desempenho, com redução da acurácia de validação para menos da metade do valor anteriormente obtido. Tais alterações possivelmente introduziram variações que descaracterizaram os padrões dos números nas faces dos dados, prejudicando o aprendizado.

Também foram avaliadas estratégias de tratamento de imagem, como conversão para escala de cinza, equalização de histograma e aplicação de *thresholding*. Embora visualmente eficazes em destacar o contraste entre o fundo branco e o dado, esses tratamentos acabaram por remover detalhes importantes dos números em certos casos. A perda de informação visual com-

prometeu a capacidade de discriminação do modelo, levando à decisão de manter as imagens em sua versão RGB original, apenas cortadas e redimensionadas para 160×160 pixels.

Para melhorar a capacidade de aprendizado, o número de épocas foi aumentado para 40 e, posteriormente, 50. Com isso, a acurácia de treino superou os 90%, mas a acurácia de validação permaneceu significativamente inferior. Essa discrepância levantou a hipótese de um desequilíbrio na divisão dos conjuntos. O conjunto de validação, com apenas 10% das amostras (proporção 90/10), continha um número muito limitado de imagens por classe, o que impactava negativamente a avaliação do modelo.

A divisão foi então ajustada para 80/20 (treinamento/validação), o que resultou em melhoria expressiva. Em uma das execuções com 50 épocas, observou-se que a acurácia de validação estabilizava por volta da 40ª época, formando um platô. Por isso, optou-se por reduzir o número de épocas para 40.

Além disso, o *learning rate* foi ajustado para 0.0006, com o intuito de proporcionar um aprendizado mais estável e menos sujeito a oscilações abruptas nos pesos da rede. Essa configuração final resultou em uma acurácia de treinamento de aproximadamente 84% e acurácia de validação de 72%, com perda de validação (val_loss) estabilizada em torno de 0.95, o que indica um bom nível de generalização para a tarefa de classificação de oito classes distintas (valores de 1 a 8).

Referências

- [1] PETERSON, J. *Playing at the world: A history of simulating wars, people and fantastic adventures, from chess to role-playing games*. San Diego, CA: Unreason Press, 2012.
- [2] HITCHENS, M.; DRACHEN, A. The many faces of role-playing games. *International Journal of Role-Playing*, v. 1, p. 3–21, 2007.
- [3] ZAGAL, J. P.; DETERDING, S. *Role-playing game studies: Transmedia foundations*. Routledge, 2017.
- [4] SZELISKI, R. *Computer vision: Algorithms and applications*. London: Springer, 2010.
- [5] IBM. O que são redes neurais convolucionais? <https://www.ibm.com/br-pt/think/topics/convolutional-neural-networks>, 2024. Acesso em: 25 maio 2025.
- [6] DATACAMP. Introdução às redes neurais convolucionais (cnns). <https://www.datacamp.com/pt/tutorial/introduction-to-convolutional-neural-networks-cnns>, 2023. Acesso em: 25 maio 2025.

- [7] LECUN, Y.; BOTTOU, L.; BENGIO, Y.; HAFFNER, P. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, v. 86, n. 11, p. 2278–2324, 1998.
- [8] GOODFELLOW, I.; BENGIO, Y.; COURVILLE, A. *Deep learning*. Cambridge, MA: MIT Press, 2016. Tradução recomendada: *Aprendizado profundo*. Editora Elsevier, 2018.
- [9] BISHOP, C. M. *Pattern recognition and machine learning*. New York: Springer, 2006.