Tecnologia assistiva para reconhecimento de cartas de baralho utilizando aprendizado profundo

Samuel Alves dos Santos[§1], Allan Duque dos Santos[§2],
Junio Cesar de Lima[§3], Fabrizzio A.A.M.N. Soares[*4], Gabriel da Silva Vieira[§*5],
[§]Instituto Federal Goiano, *Laboratório de Visão Computacional*. Urutaí - GO, Brasil.
[*]Universidade Federal de Goiás, *Laboratório Pixelab*. Goiânia - GO, Brasil.
{samuelalvesv4, allanduque.21}@gmail.com
{junio.lima, gabriel.vieira}@ifgoiano.edu.br, fabrizzio@inf.ufg.br

***Resumo*—Apresentamos neste trabalho uma aplicação inclusiva que permite pessoas com deficiência visual interagir socialmente com outros indivíduos em atividades de lazer, como jogos que envolvam carteado. Para isso, foi construído um modelo generalizável, treinado para detecção e reconhecimento de cartas de baralho através do aprendizado de máquina profundo. Neste contexto, foi projetado e implementado um sistema denominado *Smart Assistant*, baseado na API de detecção de objetos do TensorFlow. Em um local previamente definido, as cartas são colocadas dentro do campo de visão de uma câmera digital para que possam ser detectadas e classificadas em tempo real. A API SAPI, de sintetização de texto para fala (TTS), é usada para converter os rótulos das cartas detectadas (em formato de texto) em saída de áudio. Os experimentos iniciais mostram que em situações reais de jogo, a aplicação consegue identificar e classificar cartas com alta assertividade.**

***Abstract*—We present in this paper an inclusive application that allows visually impaired people to interact socially with other individuals in leisure activities such as card games. For this, we built a generalizable model that was trained for the detection and recognition of playing cards by using convolutional neural network through deep learning. A system called Smart Assistant was designed and implemented based on TensorFlow's object detection API. At a predefined location, a digital camera is used to detect cards in real-time. Then, these detected cards are sent to the classifier. After the classification, the SAPI Text to Speech Synthesis API (TTS) is used to convert the labels of recognized cards (in text format) to speech output. Experiments show that in real game situations, the application can identify and classify cards with high assertiveness.**

I. Introdução

Aprendizado profundo é um subcampo do aprendizado de máquina relacionado a algoritmos inspirados na estrutura e função do cérebro, denominados redes neurais artificiais [1]. Enquanto os algoritmos tradicionais de aprendizado de máquina são lineares, os algoritmos de aprendizagem profunda são empilhados em uma hierarquia de complexidade crescente e abstração [2]. Cada algoritmo na hierarquia aplica uma transformação não-linear em sua entrada e usa o que aprende para criar um modelo estatístico como saída.

Como a aprendizagem profunda realiza processamento intenso com resultados bastante significativos, modelos computacionais contruídos a partir de redes neurais artificiais podem ser aplicados a muitas tarefas que as pessoas realizam. A aprendizagem profunda é usada atualmente nas ferramentas mais comuns de reconhecimento de imagem, processamento de linguagem natural e software de reconhecimento de fala. Essas ferramentas aparecem em aplicativos diversos, ou seja, de carros autônomos a serviços de tradução de idiomas.

A visão é um dos sentidos humanos essenciais e desempenha o papel mais importante na percepção humana do meio ambiente [3]. Segundo a Organização Mundial da Saúde (OMS), estima-se que 285 milhões de pessoas possuem deficiência visual em todo o mundo, sendo que destas 39 milhões são cegas e 246 milhões têm baixa visão [4]. Neste contexto, o projeto propõe a construção de um sistema inclusivo, denominado *Smart Assistant*, que tem por objetivo permitir que pessoas com deficiência visual possam interagir socialmente com outros indivíduos em atividades de lazer, como jogos que envolvam carteado.

Sendo assim, o projeto propõe a elaboração de um modelo generalizável, treinado para o reconhecimento de cartas de baralho por meio do aprendizado profundo. O sistema baseia-se na API de detecção de objetos do TensorFlow, onde um modelo de rede neural convolucional (CNN, do inglês *convolutional neural network*) é treinado para detectar e classificar cartas de baralhos em imagem. Uma câmera sobreposta em um local previamente definido é inicializada para captura da imagem em tempo real. A imagem capturada é submetida ao modelo para detecção e classificação, os resultados são armazenados em um arquivo de texto, e o seu conteúdo é convertido em áudio usando a API SAPI. Por meio de um fone de ouvido a aplicação fornece informações em áudio sobre as cartas identificadas aos jogadores.

O restante do texto está organizado como se segue. A seção II descreve os trabalhos relacionados que lidam com processamento de imagens para auxiliar pessoas com deficiência visual. A seção III descreve o funcionamento do sistema proposto. O experimento realizado é apresentado na seção IV, onde também se discute os resultados iniciais desta pesquisa. Concluímos nosso texto na seção V.

II. Trabalhos Relacionados

No contexto de visão computacional, é possível identificar pessoas, lugares e objetos. Para reconhecer imagens, os computadores podem utilizar tecnologias de visão de máquina e algoritmos de inteligência artificial. Segundo Ghellere [5] a detecção de objetos em imagens permanece como um dos

maiores desafios dentro da área de visão computacional, pois os objetos contidos nelas podem estar sob as mais variadas perspectivas e transformações de escala e rotação, o que torna mais complexa a tarefa de detectá-los.

No contexto de detecção de objetos em imagens para auxílio a pessoas com deficiência visual, diferentes estratégias são apresentados na literatura. Em Nishajith at al. um boné inteligente é proposto para auxiliar cegos e pessoas com baixa visão a navegarem livremente em ambientes domésticos [6]. Poggi e Mattoccia [7] desenvolveram um dispositivo que orienta usuários por meio de mensagens de áudio e feedback tátil, permitindo perceber informações cruciais relacionadas ao ambiente circundante e, portanto, evitar obstáculos ao longo do caminho. Hollinger e Ward [8] descrevem um programa de computador capaz de reconhecer todos os membros de um baralho utilizando segmentação de cores combinada com a Transformação de Círculos de Hough. Entretanto, nenhum destes trabalhos utiliza o aprendizado profundo como técnica de classificação para auxílio a deficientes visuais na detecção de cartas de um baralho.

III. Descrição do *Smart Assistant*

O fluxograma mostrado na Figura 1 representa e descreve o processo do funcionamento do *Smart Assistant*. Esse fluxograma consiste de 5 passos interdependentes.

No primeiro passo, uma câmera captura imagens das cartas em tempo real e submete ao modelo já treinado. No segundo passo, ocorre a identificação das cartas presentes em cenas. Os resultados da detecção de cada carta são armazenados em formato de texto, no terceiro passo. O quarto passo, converte o texto, que representa a carta identificada, em áudio usando a API SAPI. Por fim, no quinto passo, a aplicação fornece informações em áudio por meio do fone de ouvido utilizado pelos usuários. Ao final do quinto passo, inicia-se um laço de repetição para identificação das cartas durante a execução do jogo. Esse laço é continuado até que a aplicação seja encerrada. Durante a execução deste processo, diferentes softwares são utilizados, dentre eles a API do TensorFlow para a detecção e classificação de objetos, e API SAPI, para realizar conversão de texto em áudio. Essas APIs são discutidas a seguir.

A. API do TensorFlow

A API do TensorFlow é amplamente usada no campo de detecção de objetos. A API é treinada usando um conjunto de dados definidos em uma das fases do projeto. Neste artigo, o conjunto de dados contém 10.400 imagens das 52 cartas encontradas no baralho. Sendo assim, um modelo é treinado com imagens que contêm cartas de baralho em conjunto com rótulos que tipificam a classe de cada uma delas, como uma Dama de Ouro, um Valete de Espada ou um Ás de Paus, por exemplo.

Quando uma imagem é submetida ao modelo treinado para reconhecimento das cartas de baralho, ele produz um número determinado de resultados, como mostra a Tabela I.

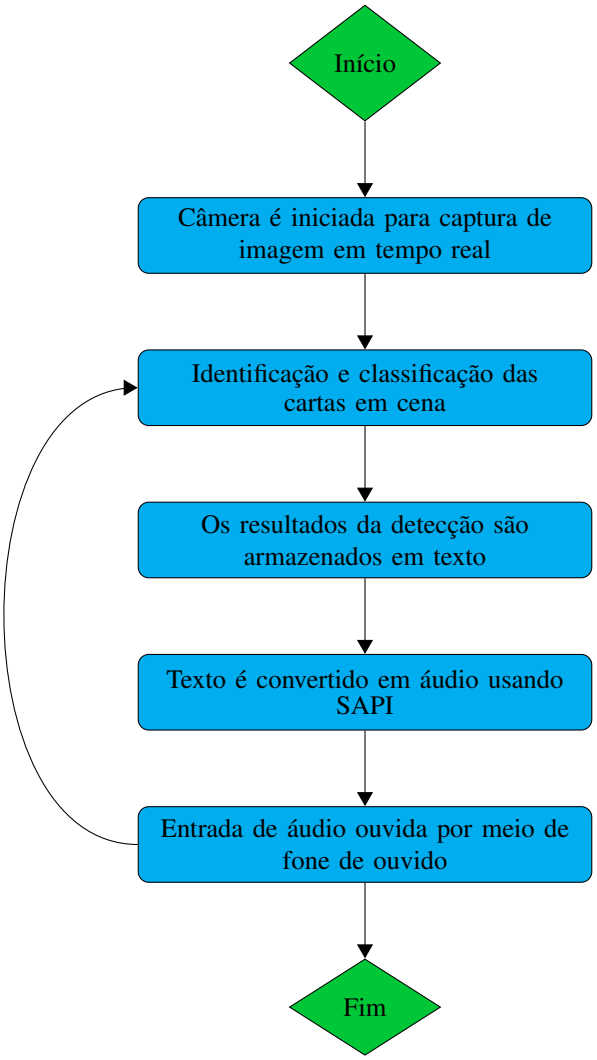

Figura 1. Fluxograma que descreve o processo do funcionamento do *Smart Assistant*.

1) Pontuação de confiança: a pontuação de confiança é um número entre 0 e 1 que indica a confiança de que o objeto foi genuinamente detectado. Quanto mais próximo o número estiver de 1, mais confiante é o modelo. Para interpretar esses resultados, podemos observar a pontuação e a localização de cada objeto detectado.
2) Localização: para cada objeto detectado, o modelo retornará uma matriz de quatro números representando um retângulo delimitador que envolve sua localização.

Tabela I
Saída do modelo *Faster RCNN Inception v2*.

Classe	Ponto	Localização
Ás copas	0.92	[18, 21, 57, 63]
Dama Paus	0.88	[100, 30, 180, 150]
Valete Ouros	0.87	[7, 82, 89, 163]
Sete Copas	0.23	[42, 66, 57, 83]
Dois Ouros	0.11	[6, 42, 31, 58]

Este é o processo do algoritmo de detecção de objetos TensorFlow. Depois de detectar o objeto, suas localizações são importantes para iniciar o processo de rastreamento. Na abordagem proposta neste artigo é utilizado o rastreamento de objetos por rede neural convolucional, ao invés de usar algoritmos convencionais baseados na visão por computador.

Nos últimos anos, as redes neurais convolucionais têm sido aplicadas a algoritmos de processamento de imagem e vídeo [9] [10] [11] [12]. As redes de classificação de imagens baseadas em redes neurais superam os algoritmos convencionais, bem como a visão humana em certas situações [10]. Um dos recursos mais desafiadores na detecção de objetos é a localização do objeto em um quadro de imagem. Por outro lado, a classificação de objetos é uma tarefa relativamente simples, pois o único trabalho da rede é classificar determinada imagem [13]. A API do TensorFlow fornece diferentes modelos de classifição de objetos, o modelo *Faster RCNN Inception v2* é utilizado em razão de apresentar alta precisão na detecção de objetos [14] [15] [16].

B. API SAPI

SAPI (*Speech Application Programming Interface*) é uma API desenvolvida pela Microsoft que permite a conversão de texto em áudio. Essa API foi utilizada para conversão de texto em áudio, como os resultados da classificação das cartas serem apresentados no formato de texto, fez-se necessária a apresentação dos resultados em áudio para que o jogador com deficiência visual pudesse realizar suas jogadas, independentemente de auxílio de terceiros.

IV. Experimentos e Resultados

Para validação da proposta foi construído um protótipo, onde o modelo treinado é aplicado no reconhecimento de cartas em situações reais de uso. O experimento contou com a participação de 10 voluntários. Os participantes são alunos de cursos técnicos e graduação, com faixa etária até 25 anos. Como nenhum dos participantes é deficiente visual, durante o experimento seus olhos foram vendados. Entre as diversas opções de jogos de cartas, foi selecionada o *Black Jack*, também conhecido como 21, devido ser um jogo simples e bastante conhecido. Um ambiente para realização foi planejado e preparado de forma simples e intuitiva para facilitar a utilização pelos participantes, como mostra a Figura 2.

Para criação da base de dados foi utilizado um baralho contendo 52 cartas. Foram capturadas 200 imagens de cada carta, com variedades de fundos, ângulos e condições de iluminação, sendo que desse total 70% foram separadas para treinamento e 30% para teste do modelo, conforme as boas práticas apresentadas na literatura. Um *script* em Python foi implementado para o pré-processamento dessas imagens em blocos no qual as imagens foram redimensionadas para o tamanho 619×1024. Em seguida, todas foram rotuladas manualmente usando o software ImgLabel [17]. A Figura 3 mostra o ponto estratégico para leitura das cartas durante a execução do jogo. Além disso, sobre esse mesmo local uma câmera é inicializada para realizar o processo de leitura.

Figura 2. Ambiente para execução do experimento.

Figura 3. Ponto estratégico para detecção e classificação das cartas.

Durante a execução do experimento, a aplicação fornece informações sobre as cartas identificadas aos jogadores através do uso de um fone de ouvido, conforme mostra a Figura 4. As cartas são reconhecidas uma por vez e ao término do reconhecimento, o fone de ouvido é repassado ao adversário e o processo se repete até que haja um ganhador.

O modelo *Faster RCNN Inception V2* da API do TensorFlow foi utilizado para treinamento na detecção das cartas. No treinamento foram realizadas 900.000 iterações do modelo usando a base de dados. O modelo detecta com sucesso as cartas e classifica de acordo com o seu tipo. Dado o resultado do treinamento, foi percebido um erro estatístico na faixa de 0.01, como mostrado na Figura 5, que está abaixo de erro aceitável de 0.05 [6].

Ao final do experimento, foi distribuído um questionário com 8 perguntas em escala e uma aberta a fim de se avaliar a proposta e identificar falhas e/ou oportunidades de melhorias da aplicação. Com o questionário foi possível avaliar o nível de dificuldade e desempenho da aplicação. De acordo com as respostas, 70% das pessoas avaliaram o nível de dificuldade da aplicação como médio e 30% como fácil.

Figura 4. Execução do experimento.

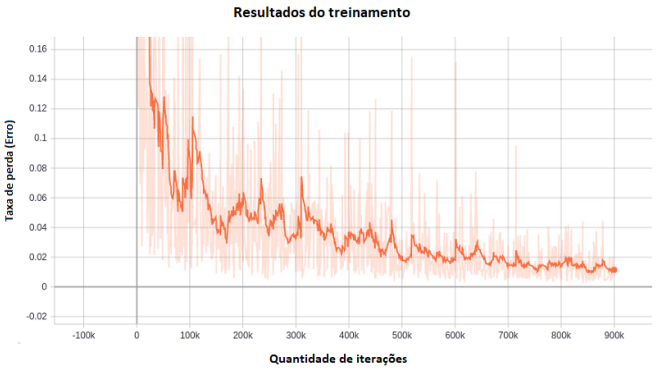

Figura 5. Faixa de erro por iteração do modelo treinado.

Esses dados são explicados na questão aberta onde muitos percebem a necessidade de um fone de ouvido para cada jogador. Devido o fone ser único e compartilhado, os jogadores sentiram dificuldades no momento do repasse. Segundo esses participantes a necessidade do fone individual irá fazer com que os jogadores promovam estratégias no decorrer do jogo eliminando a necessidade da espera pelo fone de ouvido. Referente ao desempenho, cerca de 90% das pessoas avaliaram a aplicação como boa e viável.

V. Conclusão

Este trabalho propõem um sistema inclusivo capaz de auxiliar deficientes visuais em jogos de baralho. O sistema possui uma arquitetura simples que em tempo real detecta e classifica cartas, e, que por meio de áudio, informa os resultados da detecção e classificação aos jogadores. A API de detecção do TensorFlow é usada para treinar o modelo *Faster RCNN Inception V2*, que pode detectar e classificar com sucesso as 52 cartas de acordo com o sua classe. Os experimentos iniciais mostram uma alta acurácia na detecção das cartas, o que permite classificar a aplicação *Smart Assistant* como viável em ambientes reais.

Como trabalho futuro é planejado aumentar o número de participantes, inclusive adicionando participantes com decifiência visual, além de permitir que cada participante tenha seu próprio fone de ouvido. Nesse novo cenário, um novo questionário será criado, bem como adição de novos tipos de jogos de baralho.

Referências

[1] M. S. d. Melo, "Dual scaling: uma implementação em gpu com o tensorflow."

[2] R. A. d. S. Lopes and V. G. d. M. Braga, "Um sistema para o aprendizado automático de jogos eletônicos baseado em redes neurais e q-learning usando interface natural," 2017.

[3] H. Jabnoun, F. Benzarti, and H. Amiri, "Object detection and identification for blind people in video scene," in *2015 15th International Conference on Intelligent Systems Design and Applications (ISDA)*, Dec 2015, pp. 363–367.

[4] C. C. Kulpa, F. G. Teixeira, and R. P. d. Silva, "Um modelo de cores na usabilidade das interfaces computacionais para os deficientes de baixa visão," *Design & tecnologia [recurso eletrônico]. Porto Alegre, RS. Vol. 1, n. 1 (2010), p. 66-78*, 2010.

[5] J. S. Ghellere, "Detecção de objetos em imagens por meio da combinação de descritores locais e classificadores," B.S. thesis, Universidade Tecnológica Federal do Paraná, 2015.

[6] A. Nishajith, J. Nivedha, S. S. Nair, and J. Mohammed Shaffi, "Smart cap - wearable visual guidance system for blind," in *2018 International Conference on Inventive Research in Computing Applications (ICIRCA)*, July 2018, pp. 275–278.

[7] M. Poggi and S. Mattoccia, "A wearable mobility aid for the visually impaired based on embedded 3d vision and deep learning," in *2016 IEEE Symposium on Computers and Communication (ISCC)*, June 2016, pp. 208–213.

[8] G. Hollinger, N. Ward, and E. C. Everbach, "Introducing computers to blackjack: Implementation of a card recognition system using computer vision techniques."

[9] P. Dhar, S. Guha, T. Biswas, and M. Z. Abedin, "A system design for license plate recognition by using edge detection and convolution neural network," in *2018 International Conference on Computer, Communication, Chemical, Material and Electronic Engineering (IC4ME2)*. IEEE, 2018, pp. 1–4.

[10] T. Weyand, I. Kostrikov, and J. Philbin, "Planet-photo geolocation with convolutional neural networks," in *European Conference on Computer Vision*. Springer, 2016, pp. 37–55.

[11] P. Mlynarski, H. Delingette, A. Criminisi, and N. Ayache, "3d convolutional neural networks for tumor segmentation using long-range 2d context," *Computerized Medical Imaging and Graphics*, vol. 73, pp. 60–72, 2019.

[12] D. Roblek, C. Szegedy, and J. S. Jurewicz, "Object detection using neural network systems," Jan. 17 2019, uS Patent App. 15/650,790.

[13] M. Peker, "Comparison of tensorflow object detection networks for licence plate localization," in *2019 1st Global Power, Energy and Communication Conference (GPECOM)*, June 2019, pp. 101–105.

[14] W. Liu, S. Liao, and W. Hu, "Perceiving motion from dynamic memory for vehicle detection in surveillance videos," *IEEE Transactions on Circuits and Systems for Video Technology*, pp. 1–1, 2019.

[15] J. Guo, J. Lu, Y. Qu, and C. Li, "Traffic-sign spotting in the wild via deep features," in *2018 IEEE Intelligent Vehicles Symposium (IV)*, June 2018, pp. 120–125.

[16] M. Zanfir, A. Popa, A. Zanfir, and C. Sminchisescu, "Human appearance transfer," in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, June 2018, pp. 5391–5399.

[17] T. Lin, "Labelimg," 2015.