Gaze Estimation em Atividades de Ensino Digitais Utilizando *Convolutional Neural Networks* e *Webcam*

Jordão Frazão[1], Tardelly A. Cavalcante[1,2], Priscila Benitez[3], Kelson Aires[1], André Soares[1]

[1]PPGCC - Programa de Pós Graduação em Ciência da Computação
Universidade Federal do Piauí
Teresina – PI – Brasil

[2]IFPI - Instituto Federal do Piauí
Campo Maior - PI - Brasil

[3]CMCC - Centro de Matemática, Computação e Cognição
Universidade Federal do ABC
Santo André - SP - Brasil

4

jordaofrazao@gmail.com, tardelly.cavalcante@ifpi.edu.br

pribenitez@yahoo.com.br, {kelson,andre.soares}@ufpi.edu.br

Abstract. *The field of education has challenges due to the variety of aptitudes and restrictions of students. To innovate teaching, it is necessary to develop alternatives that enhance current proposals. One of the existing techniques in the literature is gaze estimation, which allows visualizing eye behavior when performing activities on the computer. Usually this technique is performed using commercial products or methods that are not very accessible. This work proposes a low-cost methodology to perform gaze estimation and identify regions of interest to the student's gaze in educational activities, using convolutional neural networks and webcam. The potential of the method is observed in the development of tools applicable to education.*

Resumo. *A área da educação possui desafios devido à variedade de aptidões e restrições dos estudantes. Para inovar o ensino, é necessário desenvolver alternativas que incrementem as propostas atuais. Uma das técnicas existentes na literatura é o gaze estimation, que permite visualizar o comportamento ocular ao realizar atividades no computador. Usualmente esta técnica é realizada com o uso de produtos comerciais ou métodos pouco acessíveis. Este trabalho propõe uma metodologia de baixo custo para realizar gaze estimation e identificar regiões de interesse do olhar do estudante em atividades educacionais, com uso de redes neurais convolucionais e webcam. Observa-se potencial do método no desenvolvimento de ferramentas aplicáveis a educação.*

1. Introdução

De acordo com a Lei de Diretrizes e Bases da Educação Nacional - LDB [Brazil 1997], a educação é um direito garantido para toda a população. O uso de metodologias alternativas aos métodos tradicionais de ensino é uma possibilidade para contribuir com esse

DOI: 10.5753/sbie.2022.225792

direito. Desta forma, torna-se necessária a criação de métodos de ensino acessíveis à comunidade educacional.

O desenvolvimento de métodos educacionais que possam atender a diversidade e necessidades dos estudantes é bastante desafiador [Nightingale et al. 2019]. Diante desta necessidade, o uso de métodos computacionais apresenta potencial para permitir o acesso e interação dos alunos à educação, bem como o acesso ao currículo educacional [Ahmad 2015].

Um método não intrusivo que pode ser utilizado na educação é o *eye tracking*, ou rastreamento ocular. Neste método, é identificado o comportamento ocular do usuário e estimado o olhar na tela - também conhecido como *gaze estimation*. *Gaze estimation* está relacionado à área da computação que estuda interação humano-computador usando o comportamento ocular. É um tópico de pesquisas dentro da área de visão computacional. Portanto, a análise do comportamento ocular pode revelar o foco, a atenção e estratégias cognitivas [Eckstein et al. 2017]. O uso deste método permite obter informações oculares do estudante, possibilitando identificar o olhar do usuário durante a utilização de um computador [Poole and Ball 2006].

Com o intuito de propor uma metodologia que auxilie na educação, no presente trabalho objetivou-se a utilização de *gaze estimation* em atividades de ensino realizadas no computador, com o uso de imagens obtidas por *webcam*, buscando garantir uma solução de baixo custo, tornando o método mais acessível e permitindo verificar o processamento visual e atencional dos estudantes, que é fundamental para a criação de estratégias educacionais personalizadas. Para alcançar este objetivo, foram utilizadas redes neurais convolucionais - (*Convolutional Neural Network - CNN*) para a classificação das imagens dos educandos. Essas imagens foram capturadas durante a realização de uma atividade educacional de matemática. As imagens foram classificadas em nove classes, em que cada uma delas representa um quadrante da tela, que foi dividida em 3 (três) colunas e 3 (três) linhas, informando onde o olhar estava presente na atividade.

2. Trabalhos Relacionados

Eye tracking ou rastreamento ocular é uma técnica que permite que os movimentos oculares de um determinado indivíduo sejam medidos, permitindo calcular onde o usuário está olhando em um determinado momento [Poole and Ball 2006]. Para realizar este procedimento, o dispositivo utilizado com maior frequência é conhecido comumente por *eye tracker* [Duchowski and Duchowski 2017]. Para realizar a análise e identificação de características do olhar, estes dispositivos são utilizados medindo os movimentos oculares e realizando *eye tracking* de acordo com os estímulos visuais [Chen and Chen 2017].

Na maior parte dos dispositivos *eye tracker* há componentes que consistem em um conjunto de câmeras, projetores e algoritmos. Os projetores emitem luzes infravermelho nos olhos do usuário, enquanto as câmeras capturam as imagens dos olhos. As luzes emitidas nos olhos geram padrões, permitindo a execução de algoritmos para realizar processamento das imagens para capturar informações específicas sobre os olhos do usuário. Com posse desses dados, é realizado o cálculo que efetua o *gaze estimation*, ou seja, estima o olhar do usuário no ambiente de utilização [Tobii 2020]. De modo geral, o *eye tracking* é realizado via análise das imagens dos olhos capturados da pessoa [Sugano et al. 2012]. Através deste procedimento é possível efetuar o registro de diferen-

tes medidas, por exemplo, fixação, sacada e região de interesse. Regiões de interesse são áreas críticas da cena que o usuário é submetido a olhar [Holmqvist et al. 2011].

Na educação, é possível utilizar as informações obtidas através do comportamento ocular de forma vital [Halszka et al. 2017], pois é possível verificar informações importantes dos estudantes a serem avaliadas, como atenção e processamento visual. Com esse rastreio é possível traçar avaliações comportamentais, bem como obter informações sobre um usuário para avaliar condições mentais [Orsati et al. 2008].

Na esfera da pesquisa científica, são utilizadas ferramentas de *eye tracking* em diversas áreas, com destaque nas pesquisas envolvendo educação e educação inclusiva [de Araújo Cavalcante et al. 2019, Rodrigues and Rosa 2019, Strandberg 2019], psicologia e neurociência [Vargas-Cuentas et al. 2017, Chong et al. 2017, Lukasova et al. 2018, Núñez Fernández et al. 2020]. Ao realizar o *gaze estimation* durante atividades com educandos, é possível obter informações quantitativas que possibilita o especialista que o utiliza conseguir fazer análises específicas de cada indivíduo. Entretanto, apesar das diversas aplicações de *eye tracking* e *gaze estimation*, ainda é necessário torná-las tecnologias universais [Krafka et al. 2016].

Conforme [Mohri et al. 2018], a aprendizagem de máquina é conceituada amplamente como o uso de métodos computacionais que utilizam experiência obtida através de dados para melhorar o desempenho ou realizar previsões com maior acurácia. As redes neurais representam um paradigma computacional no qual a solução para um problema é aprendida a partir de um conjunto de exemplos. A inspiração para as redes neurais vem originalmente dos estudos dos mecanismos de processamento da informação nos sistemas nervosos biológicos, particularmente no cérebro humano [Bishop 1994].

Nos últimos anos, as técnicas de *deep learning* têm realizado avanços em várias áreas de aprendizado de máquina [Ponti and da Costa 2018]. Dentro do escopo do *deep learning*, as redes neurais convolucionais (CNN) são usadas em muitas tarefas de visão computacional, Entre essas tarefas a de *gaze estimation*. O primeiro método de estimativa do olhar baseado em imagens foi apresentado em [Zhang et al. 2012], este trabalho foi evoluído para uma CNN com 13 camadas baseada na rede VGG16 [Simonyan and Zisserman 2015]. Estudos apontam bons resultados na utilização de imagens dos 2 (olhos) para classificação do olhar [Fischer et al. 2018] e [Cheng et al. 2018].

Diante das limitações das ferramentas atuais que realizam *gaze estimation*, seja financeira, visto que boa parte dos dispositivos comerciais possuem custo elevado, ou técnica, pois alguns métodos exigem programação ou configuração para serem utilizados, justifica-se a proposta do presente trabalho. Visando ampliar o alcance deste método computacional, este trabalho apresenta resultados da realização de *gaze estimation* utilizando apenas o uso de um computador comum com *webcam* durante a realização de atividades educacionais, tornando o método mais acessível para uso neste contexto.

3. Metodologia

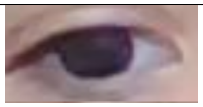
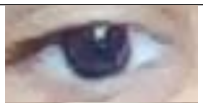
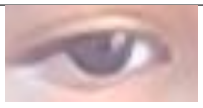
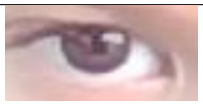
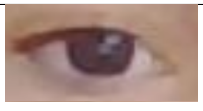
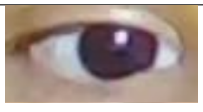
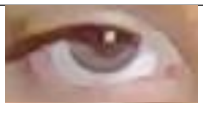
Para a realização do experimento, participaram 8 crianças de idades variadas. Devido ao cenário pandêmico atual, não foi possível expandir o número de participantes ou realizar o experimento em um grupo mais homogêneo, como uma turma escolar. Como inicialmente a proposta visa a possibilidade de utilizar estratégias lúdicas em um contexto

educacional, o experimento foi realizado apenas com crianças. Os critérios utilizados para a captação dessas crianças foram: estar matriculada no ensino regular, habilidade para utilizar mouse ou touchpad, proximidade da região dos pesquisadores e disponibilidade dos pais ou responsáveis estarem presentes acompanhando o experimento.

Para realização do experimento com as crianças foi utilizada a técnica da Psicologia denominada de *Rapport* [Rogers and Carmichael 1942]. A técnica foi utilizada com a finalidade de estabelecer uma boa relação com os educandos para reduzir a ansiedade a partir da relação de confiança e empatia, buscando garantir o entendimento e interesse para a realização das tarefas. Os experimentos foram acompanhados por uma psicóloga. Todos os responsáveis legais dos participantes assinaram o termo de consentimento livre e esclarecido.

Durante o experimento foi necessário que os participantes mantivessem o olhar direcionado para a tela durante a maior parte da atividade. Um dos participantes foi removido do experimento, pois conforme a mãe do mesmo informou, ele tinha astigmatismo, o que fez com que o usuário tirasse o olhar da tela diversas vezes, ameaçando a validade dos dados obtidos durante a sua atividade. Outros aspectos importantes que foram considerados a fim de validar o experimento foram garantir que as crianças estivessem com as suas necessidades básicas supridas, como sono e alimentação. A descrição das crianças junto com a quantidade de imagens das mesmas é realizada na Tabela 1.

Tabela 1. Descrição das Crianças.

Criança	Idade	Sexo	Base de Fine-Tuning	Base de Testes	Imagem do Olho
C1	5 anos	M	6873 imagens	2039 imagens	
C2	8 anos	M	6910 imagens	1431 imagens	
C3	6 anos	F	6807 imagens	2173 imagens	
C4	5 anos	M	6542 imagens	1474 imagens	
C5	8 anos	F	6826 imagens	2574 imagens	
C6	5 anos	F	6816 imagens	1524 imagens	
C7	7 anos	F	6818 imagens	910 imagens	

Devido a pandemia os experimentos foram realizados em ambiente domiciliar. Os experimentos foram feitos em dois tipos de ambiente: para as crianças que realizaram a atividade durante o dia, foi realizado em área externa, utilizando iluminação natural.

No caso das crianças que realizaram o experimento durante a noite, foi utilizada uma luminária para prover iluminação suficiente, de modo a melhorar a captura das imagens. A luminária tinha acoplada apenas uma lâmpada fluorescente simples, de 15W de potência, posicionada acima da *webcam*. Todos os experimentos foram realizados no mesmo computador, Acer Nitro 5, com processador i5-7300HQ e 16GB de RAM, e, portanto, usando a mesma câmera, uma *webcam* de resolução 720p e 1 Mp. O assento onde as crianças realizaram os experimentos foi ajustado conforme a altura de cada uma delas, tentando manter uma centralização da mesma com a tela e a *webcam*.

4. Procedimento

O método proposto consiste de três etapas principais. A primeira é a captura da base de imagens que são utilizadas para realizar o *fine-tuning* da CNN. Essa captura é feita enquanto o participante assiste um vídeo lúdico, personalizado baseado no seu interesse, visando manter o contato visual e a atenção. O vídeo é exibido sequencialmente, em cada uma dos 9 quadrantes que a tela foi dividida, conforme ilustra a Figura1.

(a) Esquema de quadrantes utilizado no trabalho.

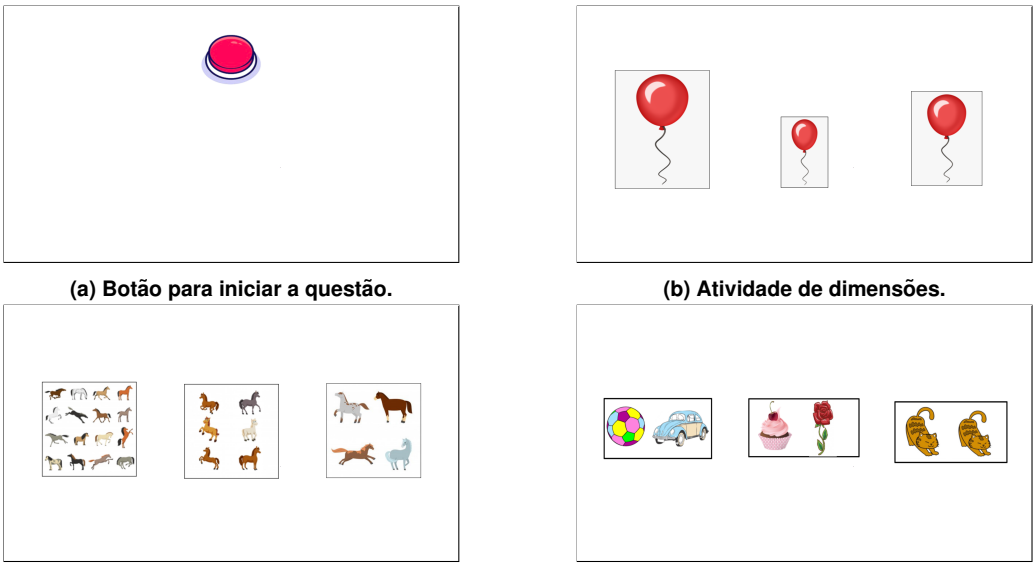
(b) Exemplo da captura de vídeo.

Figura 1. Ilustração da divisão da tela.

A segunda etapa consiste da realização da atividade educacional. Os participantes realizam uma atividade de matemática, que foi criada conforme a base nacional comum curricular - BNCC, para educação infantil, de acordo com os conhecimentos matemáticos de relações entre quantidades, dimensões e comparações [Brasil 2018]. Juntamente, foi utilizado como base o Protocolo de Registro e Avaliação das Habilidades Matemáticas - PRAHM [Costa et al. 2017], com o propósito de alinhar o experimento com o contexto educacional. Como o intuito inicial do trabalho não é avaliar habilidades acadêmicas e a princípio verificar a possibilidade de uso das medidas do comportamento ocular nessas atividades, todos realizaram o mesmo exercício. A Figura 2 exemplifica algumas das atividades que os estudantes responderam.

As atividades consistem de um botão para clicar e inicializar, enquanto cada tela contém alguns estímulos discriminativos e uma pergunta em áudio, por exemplo, “Qual o maior?”, “Qual o menor?”, para que o usuário responda clicando com o mouse. Ao clicar em uma resposta, ele avança para a próxima questão. Durante a realização dessas atividades, ocorre continuamente a captura das imagens dos estudantes através da *webcam* do computador. Essas imagens foram rotuladas em cada região utilizando um dispositivo comercial de alta precisão, permitindo comparar a classificação realizada na CNN com o resultado deste dispositivo.

A terceira etapa é a do processamento e classificação das imagens obtidas. As duas bases de imagens previamente capturadas passam inicialmente por um processamento



(a) Botão para iniciar a questão.

(b) Atividade de dimensões.

(c) Atividade de quantidade.

(d) Atividade de comparações.

Figura 2. Exemplo das Atividades.

para detecção da face e posteriormente, recorte dos dois olhos. A Figura 3 detalha o processamento.

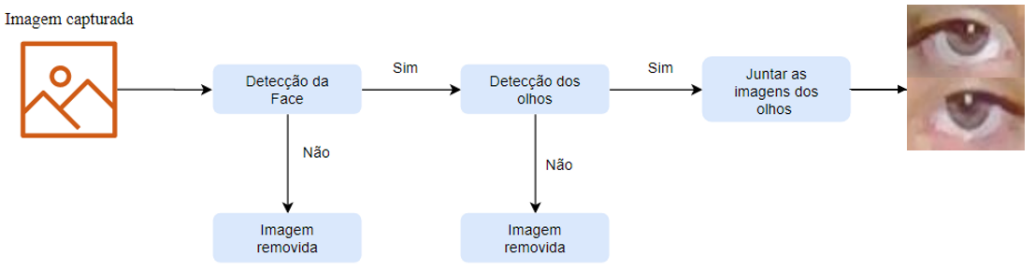


Figura 3. Processamento das imagens.

Com as imagens processadas, é realizada a etapa de treinamento e classificação utilizando CNN. A CNN utilizada foi a InceptionResNetV2 [Szegedy et al. 2017], pois nos estudos preliminares, em comparação com outras arquiteturas, ela obteve os melhores resultados. As imagens foram redimensionadas para 299x299 para utilização na CNN. São utilizados 4 *datasets* nesta etapa. Os dois primeiros para realização de *transfer learning*, inicialmente treinando os pesos da rede com o *dataset* ImageNet [Deng et al. 2009], para extração das características das imagens.

Após o treinamento dos pesos, as camadas iniciais da rede são congeladas e é realizado o *fine-tuning* do modelo utilizando o *dataset* MPIIGaze [Zhang et al. 2015], que é uma base de 213.659 imagens de 15 usuários criada para a realização de *gaze estimation*. O *dataset* foi processado e rotulado no formato utilizado neste trabalho, que é a divisão

da tela em 9 quadrantes. Deste modo, o modelo é treinado para o problema específico de *gaze estimation*. O *dataset* é dividido em 80% treino e 20% validação, alcançando uma acurácia máxima de classificação na base de validação de 76%.

Com esse modelo obtido, é feita a classificação da base de testes, que consiste nas imagens obtidas durante a realização da atividade para verificação da acurácia inicial. Em seguida, é utilizada a base de *fine-tuning* individual, que é formada pelas imagens obtidas durante a visualização do vídeo, visando melhorar a classificação das imagens para cada um. Com o modelo treinado, as imagens são classificadas e são removidas as imagens com nível de confiança da classificação abaixo de 95%, visando melhorar a acurácia do modelo.

5. Resultados

Com todas as bases das crianças obtidas e realizados *fine-tuning* e classificação, foram obtidos os resultados apresentados na Tabela 2.

Tabela 2. Resultados da Acurácia da Classificação das Imagens Utilizando o Modelo Proposto.

Criança	Sem *Fine-Tuning*	Com *Fine-Tuning*	Imagens com Nível de Confiança >95%
C1	51,40%	74,00%	81,50%
C2	48,00%	85,12%	98,81%
C3	35,00%	78,14%	95,81%
C4	34,26%	73,27%	73,70%
C5	57,50%	81,08%	96,35%
C6	38,39%	78,81%	90,50%
C7	47,14%	84,18%	97,66%

É possível identificar que em todas as crianças a classificação das imagens utilizando apenas a CNN com o treinamento do *dataset* MPIIGaze tem resultados baixos, conseguindo uma acurácia de classificação máxima de apenas 57,50%. Pode-se identificar que mesmo ao utilizar bases de imagens grandes, ainda não se consegue obter bons resultados na realização de *gaze estimation*.

Ao realizar o *fine-tuning* com as imagens obtidas na captura de vídeo, pode-se identificar um aumento na acurácia da classificação para todas as crianças, em algumas inclusive chegando a ter um aumento maior que o dobro da acurácia sem o *fine-tuning*, como é possível verificar na C3, C4 e C6. Estes números indicam que o uso de uma captura lúdica e personalizada para obter imagens da criança previamente à atividade auxilia no treinamento do modelo, usando essa base para realizar *fine-tuning* de um modelo treinado com o *dataset* MPIIGaze.

Verifica-se a acurácia da classificação das imagens que a CNN classificou com nível de confiança acima de 95%, observa-se que a acurácia alcança valores bem satisfatórios em boa parte dos usuários, conseguindo uma acurácia acima de 90% nas imagens de 5 das 7 crianças. A maioria, inclusive, alcança acurácia acima dos 76% obtidos com a base de validação da MPIIGaze, que foi classificada previamente. No entanto, nos casos das crianças C1 e C4, não foi possível observar um grande aumento na acurácia nesta parte, observa-se uma acurácia mais baixa em duas das crianças mais novas do

experimento, o que indica que o modelo ainda é sensível para maiores movimentos do estudante, que normalmente ocorre em crianças mais novas.

6. Discussão

Analisando os resultados com a classificação das imagens obtidas durante a realização de uma atividade educacional, verifica-se que a metodologia proposta alcança acurácia de até 98% em algumas crianças e uma acurácia geral de 90,65%, considerando todas as imagens da Tabela 2. Observa-se uma acurácia maior do que trabalhos apresentados na literatura, como [George and Routray 2016] e [Núñez Fernández et al. 2020], visto que os mesmos apresentam acurácia de 89,81% e 89,54%, em cenários que classificam as imagens em 7 e 3 classes, respectivamente.

Além disso, verificamos que o uso de uma captura prévia à atividade que será realizada é importante, visto que ao realizar o *fine-tuning* do modelo de CNN treinado previamente com a base de dados MPIIGaze, temos como resultado uma melhora na acurácia para todas as crianças que realizaram o experimento. Desta forma, possibilita utilizar uma captura personalizável, conforme o interesse de cada criança, tornando a atividade mais interessante e aumentando a probabilidade da criança realizá-la com sucesso. Esta captura pode substituir o processo de calibração que é utilizado normalmente em alguns dispositivos, e em muitos casos, cansativo para ser realizado pelas crianças.

Através da realização da atividade, é possível obter alguns dados do olhar dos estudantes, considerando que o método informe corretamente o olhar. Por exemplo, é possível verificar a porcentagem do tempo de cada atividade que a criança ficou olhando para determinada região ou estímulo. De acordo com a Figura 4, podemos observar a porcentagem do tempo que as crianças C2 e C5, respectivamente, ficaram olhando para os estímulos exibidos nas regiões ilustradas - considerando nestes exemplos apenas as três regiões da linha central, pois nas outras não tinham estímulos, por isso as porcentagens não estão exibidas na imagem. Desta forma é possível, por exemplo, identificar viés de região, caso determinada região tenha mais tempo do olhar durante a atividade, ou até mesmo algum estímulo que a criança evitou ou teve mais foco, permitindo uma personalização futura, alterando um estímulo ou posições das atividades.

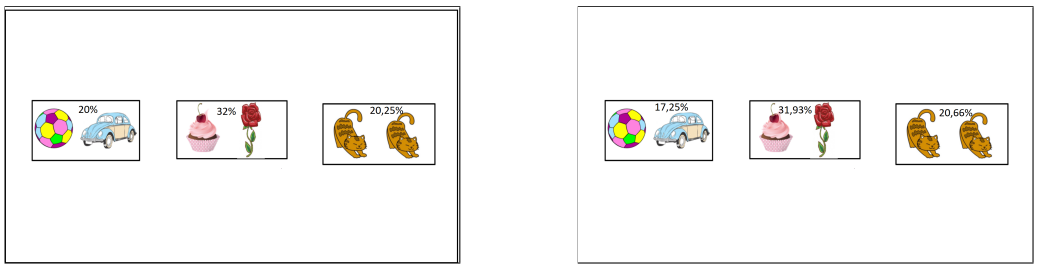


Figura 4. Exemplo de Resultado Obtido da Análise do Olhar.

É possível observar na Figura 4 em que a pergunta era “Qual é o igual” que, apesar da resposta correta estar no estímulo mais a direita, houve um foco maior no estímulo central, possibilitando observar um certo viés de posição.

Outro ponto fundamental nos resultados obtidos foi o uso de iluminação apropriada. Nos dois cenários utilizados, com luz natural e com luminária, foi possível identificar que os resultados foram positivos, melhorando tanto a captura das imagens como a

acurácia da classificação das mesmas. No entanto, há a necessidade de se verificar alternativas para aperfeiçoar as imagens com processamento computacional, visando encontrar outra alternativa para esse problema.

Vale ressaltar que, os resultados das duas crianças mais novas a participarem do experimento, criança 1 e 4, foram mais baixos que o restante, 81,53% e 73,70%, respectivamente. Estes resultados apresentam uma tendência menor devido a maior mobilidade dessas crianças e menor concentração durante toda a atividade, indicando que é necessário realizar adequações nas bases de imagens capturadas para tentar reduzir o impacto do movimento da cabeça durante a captura das mesmas, além de otimizar as atividades de acordo com as habilidades que são esperadas para cada idade.

7. Conclusões

Este trabalho apresentou uma metodologia de baixo custo para realizar *gaze estimation* e identificar regiões de interesse do olhar do estudante em atividades educacionais, com uso de redes neurais convolucionais e *webcam*. Esta metodologia identifica a região onde o estudante olhou, com uso de CNN. Com base nos resultados obtidos, identificou-se que esta metodologia permite realizar esse processo de forma promissora, evidenciados com taxas de até 98,81% de acurácia na classificação das imagens com emprego de *fine-tuning* individual, através da captura de imagens utilizando vídeos lúdicos e utilizando as imagens com nível de confiança acima de 95%.

Como propostas de trabalhos futuros, planeja-se evoluir o método, buscando mapear as posições da cabeça para tentar ajustá-las e reduzir o efeito causado pela sua movimentação, bem como aplicar técnicas de processamento de imagem para possibilitar melhorias nas imagens. Outra característica necessária de avaliação é o uso de óculos, para verificar o seu impacto nos resultados de uma captura. Pretende-se também evoluir a metodologia buscando aumentar a quantidade de regiões utilizadas na tela, para verificar diferentes cenários de aplicação. Também é desejado, devido a maior acessibilidade, explorar este método utilizando dispositivos móveis como celulares e tablets. Para melhor visualização dos dados pelos profissionais educadores, há o intuito de criar uma interface onde seja mais fácil interpretar os dados dos estudantes, para melhor tomada de decisão.

Além disso, há o intuito de incrementar tanto a captura e classificação das imagens, visando evoluir a metodologia para uma ferramenta educacional, tornando essa tecnologia social viável para escolas públicas, sendo possível utilizar no ambiente da sala de aula. Deste modo, seria possível o uso de profissionais educadores para realização de suas atividades pensadas e elaboradas de acordo com o que esses profissionais desejam avaliar ou ensinar para os seus educandos.

Referências

Ahmad, F. K. (2015). Use of assistive technology in inclusive education: making room for diverse learning needs. *Transcience*, 6(2):62–77.

Bishop, C. M. (1994). Neural networks and their applications. *Review of scientific instruments*, 65(6):1803–1832.

Brasil (2018). *Base Nacional Comum Curricular*. Brasília.

Brazil (1997). *Lei de diretrizes e bases da educação nacional*. Conselho de Reitores das Universidades Brasileiras.

Chen, X. and Chen, Z. (2017). Exploring visual attention using random walks based eye tracking protocols. *Journal of Visual Communication and Image Representation*, 45:147–155.

Cheng, Y., Lu, F., and Zhang, X. (2018). Appearance-based gaze estimation via evaluation-guided asymmetric regression. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 100–115.

Chong, E., Chanda, K., Ye, Z., Southerland, A., Ruiz, N., Jones, R. M., Rozga, A., and Rehg, J. M. (2017). Detecting gaze towards eyes in natural social interactions and its use in child assessment. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 1(3):1–20.

Costa, A. B. d., Picharillo, A. D. M., and Elias, N. C. (2017). Avaliação de habilidades matemáticas em crianças com síndrome de down e com desenvolvimento típico. *Ciência & Educação (Bauru)*, 23:255–272.

de Araújo Cavalcante, T., Frazão, J., Paiva, A., Maia, I. M., Benitez, P., and Soares, A. (2019). Eye tracking como estratégia de ensino e avalização na educação inclusiva: Aplicação com alunos com autismo. In *Brazilian Symposium on Computers in Education (Simpósio Brasileiro de Informática na Educação-SBIE)*, volume 30, page 1221.

Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., and Fei-Fei, L. (2009). Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition*, pages 248–255. Ieee.

Duchowski, A. T. and Duchowski, A. T. (2017). *Eye tracking methodology: Theory and practice*. Springer.

Eckstein, M. K., Guerra-Carrillo, B., Miller Singley, A. T., and Bunge, S. A. (2017). Beyond eye gaze: What else can eyetracking reveal about cognition and cognitive development? *Developmental Cognitive Neuroscience*, 25:69–91. Sensitive periods across development.

Fischer, T., Chang, H. J., and Demiris, Y. (2018). Rt-gene: Real-time eye gaze estimation in natural environments. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 334–352.

George, A. and Routray, A. (2016). Real-time eye gaze direction classification using convolutional neural network. In *2016 International Conference on Signal Processing and Communications (SPCOM)*, pages 1–5. IEEE.

Halszka, J., Holmqvist, K., and Gruber, H. (2017). Eye tracking in educational science: Theoretical frameworks and research agendas. *Journal of eye movement research*, 10(1).

Holmqvist, K., Nyström, M., Andersson, R., Dewhurst, R., Jarodzka, H., and Van de Weijer, J. (2011). *Eye tracking: A comprehensive guide to methods and measures*. OUP Oxford.

Krafka, K., Khosla, A., Kellnhofer, P., Kannan, H., Bhandarkar, S., Matusik, W., and Torralba, A. (2016). Eye tracking for everyone. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2176–2184.

Lukasova, K., Nucci, M. P., Neto, R. M. d. A., Vieira, G., Sato, J. R., and Amaro Jr, E. (2018). Predictive saccades in children and adults: A combined fmri and eye tracking study. *PloS one*, 13(5):e0196000.

Mohri, M., Rostamizadeh, A., and Talwalkar, A. (2018). *Foundations of machine learning*. MIT press.

Nightingale, K. P., Anderson, V., Onens, S., Fazil, Q., and Davies, H. (2019). Developing the inclusive curriculum: Is supplementary lecture recording an effective approach in supporting students with specific learning difficulties (splds)? *Computers & Education*, 130:13–25.

Núñez Fernández, D., Barrientos Porras, F., Gilman, R. H., Vittet Mondonedo, M., Sheen, P., and Zimic, M. (2020). A convolutional neural network for gaze preference detection: A potential tool for diagnostics of autism spectrum disorder in children. *arXiv e-prints*, pages arXiv–2007.

Orsati, F. T., Schwartzman, J. S., Brunoni, D., Mecca, T., and de Macedo, E. C. (2008). Novas possibilidades na avaliação neuropsicológica dos transtornos invasivos do desenvolvimento: Análise dos movimentos oculares. *Avaliaçao Psicologica: Interamerican Journal of Psychological Assessment*, 7(3):281–290.

Ponti, M. A. and da Costa, G. B. P. (2018). Como funciona o deep learning. *arXiv preprint arXiv:1806.07908*.

Poole, A. and Ball, L. J. (2006). Eye tracking in hci and usability research. In *Encyclopedia of human computer interaction*, pages 211–219. IGI Global.

Rodrigues, P. and Rosa, P. J. (2019). Eye-tracking as a research methodology in educational context: A spanning framework.

Rogers, C. R. and Carmichael, L. (1942). Counseling and psychotherapy: Newer concepts in practice.

Simonyan, K. and Zisserman, A. (2015). Very deep convolutional networks for large-scale image recognition.

Strandberg, A. (2019). Eye movements during reading and reading assessment in swedish school children: a new window on reading difficulties. In *Proceedings of the 11th ACM Symposium on Eye Tracking Research & Applications*, pages 1–3.

Sugano, Y., Matsushita, Y., and Sato, Y. (2012). Appearance-based gaze estimation using visual saliency. *IEEE transactions on pattern analysis and machine intelligence*, 35(2):329–341.

Szegedy, C., Ioffe, S., Vanhoucke, V., and Alemi, A. A. (2017). Inception-v4, inception-resnet and the impact of residual connections on learning. In *Thirty-first AAAI conference on artificial intelligence*.

Tobii (2020). What is eye tracking?

Vargas-Cuentas, N. I., Roman-Gonzalez, A., Gilman, R. H., Barrientos, F., Ting, J., Hidalgo, D., Jensen, K., and Zimic, M. (2017). Developing an eye-tracking algorithm as a potential tool for early diagnosis of autism spectrum disorder in children. *PloS one*, 12(11):e0188826.

Zhang, J., Zhuo, L., Li, Z., and Zhao, Y. (2012). An approach of region of interest detection based on visual attention and gaze tracking. In *2012 IEEE International Conference on Signal Processing, Communication and Computing (ICSPCC 2012)*, pages 228–233. IEEE.

Zhang, X., Sugano, Y., Fritz, M., and Bulling, A. (2015). Appearance-based gaze estimation in the wild. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4511–4520.