



UNIVERSIDADE FEDERAL DO PARÁ
INSTITUTO DE TECNOLOGIA
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA ELÉTRICA

Adriano Madureira dos Santos

Deep Learning in Education 5.0: Proposing 3D Geometric Shapes Classification Model to Improve Learning on a Metaverse Application

DM 01/2024

Belém

2024

Adriano Madureira dos Santos

Deep Learning in Education 5.0: Proposing 3D Geometric Shapes Classification Model to Improve Learning on a Metaverse Application

DM 01/2024

Master's Thesis submitted to the Electrical Engineering Postgraduate Program examination committee at the Federal University of Pará as a partial requirement to obtain for the degree of Master of Science in Electrical Engineering with emphasis in Computational Intelligence.

Universidade Federal do Pará

Supervisor: Marcos César da Rocha Seruffo

Belém

2024

Dados Internacionais de Catalogação na Publicação (CIP) de acordo com ISBD
Sistema de Bibliotecas da Universidade Federal do Pará
Gerada automaticamente pelo módulo Ficat, mediante os dados fornecidos pelo(a) autor(a)

M178d Madureira dos Santos, Adriano.
 Deep Learning in Education 5.0: Proposing 3D Geometric
 Shapes Classification Model to Improve Learning on a Metaverse
 Application / Adriano Madureira dos Santos. — 2024.
 60 f. : il. color.

 Orientador(a): Prof. Dr. Marcos César da Rocha Seruffo
 Dissertação (Mestrado) - Universidade Federal do Pará,
 Instituto de Tecnologia, Programa de Pós-Graduação em
 Engenharia Elétrica, Belém, 2024.

 1. Metaverse. 2. Deep Learning. 3. Computer Vision. 4.
 Education 5.0. 5. Mathematics. I. Título.

CDD 006.3

Adriano Madureira dos Santos


Deep Learning in Education 5.0: Proposing 3D Geometric Shapes Classification Model to Improve Learning on a Metaverse Application

Master's Thesis submitted to the Electrical Engineering Postgraduate Program examination committee at the Federal University of Pará as a partial requirement to obtain for the degree of Master of Science in Electrical Engineering with emphasis in Computational Intelligence.

Concept: EXCELLENT

Belém, 18/01/2024

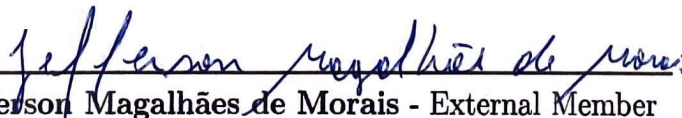
EXAMINATION COMMITTEE



Marcos César da Rocha Seruffo - Supervisor
UFPA - ITEC - PPGEE



Roberto Célio Limão de Oliveira - Internal Member
UFPA - ITEC - PPGEE



Jefferson Magalhães de Moraes - External Member
UFPA - ICEN - PPGCC

I dedicate this work to my family, friends and professors who supported me throughout the entire development process. May this work serves as an example of resilience for those who are also pursuing postgraduate studies in this field.

ACKNOWLEDGEMENTS

I firstly would like to express my gratitude to God, for the opportunity to pursue a Computational Intelligence Master's Degree. To my mother, Denise Madureira, for always being present in every moment of my life, providing strength and encouragement in my projects.

To my supervisor, Prof. Dr. Marcos Seruffo, for assisting me in my learning process, from scientific paper productions to my Master's Thesis.

To professors and friends that I had during the course, for helping and motivating me to acquire knowledge on all the topics I needed to work in this technological field that I deeply admire, especially Flávio Moura, Saulo William and Walter Júnior.

To the Inteceleri company for providing me the partner opportunity to develop strategies focusing on mitigating the country's educational challenges through Artificial Intelligence.

Finally, to the Federal University of Pará and the Post-Graduate Electrical Engineering Program for allowing me to access all the necessary resources to fulfill my dream of completing my Electrical Engineering Master of Science degree in Computational Intelligence.

In the middle of difficulty lies opportunity.
Albert Einstein

ABSTRACT

The Brazilian educational system faces significant challenges, as evidenced by low educational development assessment scores. Due to the traditional educational model employed in the country, there are difficulties in the effective transmission of complex content, leading to high rates of academic failure and subsequent school dropout. The lack of innovation, especially in basic education settings, contributes to a scenario of low mathematical proficiency among Brazilian students. In this context, this work arises as a result of an innovation built to enhance the Geometa application, developed by the Inteceleri company, through the integration of Metaverse and Artificial Intelligence technologies to create an immersive and interactive educational environment. The intention is to train Artificial Intelligence for real-time three-dimensional geometric shape recognition from real-world object images. The proposal aims to mitigate challenges faced in Brazilian basic Mathematics education by adopting innovative technological approaches aligned with Education 5.0, which can be replicated for similar technologies involving the Metaverse. Furthermore, it is also intended to create a dynamic and sustainable educational environment that not only facilitates the mathematical concepts understanding but also promotes active student participation, encouraging their creativity and autonomy in the learning process. The method used relies on the ObjectNet dataset image reclassification from objects to three-dimensional geometric shapes. The reclassified images are used to train CNN, MobileNet, ResNet, ResNeXt, ViT and BEiT Deep Learning models, which are subsequently evaluated through Machine Learning, inference time and dimension performance measures. Thus, the best-performance Artificial Intelligence model is selected for future integration into Geometa. As contributions of this work, the following were accomplished: (i) the defined models were trained for the three-dimensional geometric shapes recognition; (ii) the models were evaluated through Machine Learning, inference time and dimension performance measures; and (iii) the best-performance model was selected considering the highest assertiveness and smoothness based on models performances analysis. Concerning the obtained results, the ResNet surpassed BEiT, which was the second better performance model, in 5% Precision and 5 Inference Per Second. Finally, the ResNet model reached 84% Precision and 9 Inferences Per Second, being observed as the best-performance Artificial Intelligence for Geometa application integration flow.

Keywords: Metaverse, Deep Learning, Computer Vision, Education 5.0, Mathematics.

RESUMO

O sistema educacional brasileiro enfrenta desafios significativos, conforme evidenciado pelos baixos índices de avaliação do desenvolvimento educacional. Devido ao modelo educacional tradicional empregado no país, há dificuldades na transmissão efetiva de conteúdos complexos, levando a altos índices de fracasso escolar e consequente evasão escolar. A falta de inovação, especialmente em ambientes de educação básica, contribui para um cenário de baixa proficiência matemática entre os estudantes brasileiros. Neste contexto, este trabalho surge como resultado de uma inovação desenvolvida para aprimorar a aplicação Geometa, desenvolvida pela empresa Inteceleri, através da integração das tecnologias de Metaverso e Inteligência Artificial para criar um ambiente educacional imersivo e interativo. A intenção é refinar a Inteligência Artificial para o reconhecimento de formas geométricas tridimensionais em tempo real a partir de imagens de objetos reais. A proposta visa mitigar desafios enfrentados no ensino básico de matemática no Brasil por meio da adoção de abordagens tecnológicas inovadoras alinhadas à Educação 5.0, que possam ser replicadas para tecnologias similares envolvendo o Metaverso. Além disso, pretende-se também criar um ambiente educativo dinâmico e sustentável que não só facilite a compreensão de conceitos matemáticos, mas também promova a participação ativa dos alunos, incentivando a sua criatividade e autonomia no processo de aprendizagem. O método utilizado baseia-se na reclassificação de imagens do conjunto de dados ObjectNet de objetos para formas geométricas tridimensionais. As imagens reclassificadas são usadas para treinar os modelos CNN, MobileNet, ResNet, ResNeXt, ViT e BEiT de Aprendizado Profundo, os quais são posteriormente avaliados por meio de medidas de desempenho de Aprendizado de Máquina, tempo de inferência e dimensão. Por fim, o modelo de Inteligência Artificial de melhor desempenho é selecionado para futura integração no Geometa. Como contribuições deste trabalho foram realizados: (i) os modelos definidos foram treinados para o reconhecimento de formas geométricas tridimensionais; (ii) os modelos foram avaliados por meio de medidas de desempenho de Aprendizado de Máquina, tempo de inferência e dimensão; e (iii) o modelo de melhor desempenho foi selecionado considerando a maior assertividade e suavidade com base na análise de desempenho dos modelos. Quanto aos resultados obtidos, o ResNet superou o BEiT, modelo com o segundo melhor desempenho, em 5% de Precisão e 5 Inferência por Segundo. Por fim, o modelo ResNet atingiu 84% de Precisão e 9 Inferências por Segundo, sendo apontado como a Inteligência Artificial de melhor desempenho para fluxo de integração com a aplicação Geometa.

Palavras-chave: Metaverso, Aprendizado Profundo, Visão Computacional, Educação 5.0, Matemática.

LIST OF FIGURES

Figure 1.	Geometa application functionalities and resources	29
Figure 2.	Miritiboard VR glasses for more immersive Metaverse experiences . . .	29
Figure 3.	ResNet DL model building block	33
Figure 4.	ResNeXt DL model architecture	34
Figure 5.	ViT DL model architecture	35
Figure 6.	BEiT training and classification workflow	37
Figure 7.	Proposed method for educational application AI development and launch	43
Figure 8.	ObjectNet dataset background control, rotations and viewpoints	45

LIST OF TABLES

Table 1.	IDEB assessment scores for initial years of Brazilian elementary schools	20
Table 2.	IDEB assessment scores for final years of Brazilian elementary schools .	20
Table 3.	IDEB assessment scores for Brazilian high schools	21
Table 4.	SAEB assessment scores for initial years of Brazilian elementary schools	22
Table 5.	SAEB assessment scores for final years of Brazilian elementary schools .	23
Table 6.	SAEB assessment scores for Brazilian high schools	24
Table 7.	State-of-the-art related works proposals	42
Table 8.	Models' Precision performance results	48
Table 9.	Models' Accuracy performance results	49
Table 10.	Models' Recall performance results	49
Table 11.	Models' F1-Score performance results	50
Table 12.	Models' time and dimension evaluated performance measures	51

LIST OF ABBREVIATIONS AND ACRONYMS

AI	Artificial Intelligence
AR	Augmented Reality
b	Bias Tensor
BEiT	Bidirectional Encoder for Image Transformers
CNN	Convolutional Neural Network
DL	Deep Learning
E	Transformer Encoder
ENEM	National High School Exam
EP	Educational Proposal
FF	Feed Forward
FN	False Negative
FP	False Positive
I	Input Tensor
INEP	National Study and Research Educational Institute Anísio Teixeira
IPS	Inferences Per Second
ITDE	Inference Time and Dimension Evaluation
K	Key
K^c	Convolutional Kernel
LN	Layer Normalization
LPO	Operational Research Laboratory
MEC	Ministry of Education
MHA	Multi-Head Attention
ML	Machine Learning

MR	Mixed Reality
MS	Model Size
MTP	Model Total Parameters
N	Total Inferences Amount
O	Output Tensor
OSR	Object Shape Recognition
P	Pooling Function
PCN	National Curricular Parameters
Q	Query
ResNet	Residual Neural Network
ResNeXt	Residual Neural Network for NeXt dimension
RHAE	Human Resources in Strategic Areas Program
S	Stride
SAEB	Basic Education Evaluation System
TGSR	Three-dimensional Geometric Shapes Recognition
TIT	Total Inference Time
TN	True Negative
TP	True Positive
TPI	Time Per Inference
TRI	Item Answer Theory
ViT	Vision Transformer
VR	Virtual Reality
W	Weight Matrix

CONTENTS

1	INTRODUCTION	14
1.1	Motivations	16
1.2	General objective	17
1.3	Specific objectives	17
1.4	Thesis outline	18
2	BACKGROUND	19
2.1	The Brazilian Mathematics teaching and learning challenges	19
2.2	The technology applied for educational purposes	25
2.3	The education in Metaverse	26
2.4	The education considering Artificial Intelligence	27
2.5	GeoMeta: Learn Geometry in the Metaverse	28
2.6	The Deep Learning in object classification tasks	30
2.6.1	Convolutional Neural Network	30
2.6.2	MobileNet	32
2.6.3	Residual Neural Network	32
2.6.4	Residual Neural Network for Next dimension	33
2.6.5	Vision Transformer	34
2.6.6	Bidirectional Encoder for Image Transformers	36
2.7	Artificial Intelligence performance measures	37
3	RELATED WORKS	40
4	METHODOLOGY	43
4.1	Educational Innovation	43
4.2	Data Collection	44
4.3	Model Training	45
4.4	Model Evaluation	47
5	RESULTS	48
5.1	Machine learning performance measures evaluation	48
5.2	Inference time and dimension performance measures evaluation . .	51
6	CONCLUSION	53
	BIBLIOGRAPHY	55

1 INTRODUCTION

The Brazilian Mathematics education is a fundamental component for citizenship formation through the National Curricular Parameters (PCN, in Brazilian Portuguese), as society uses even more scientific knowledge and technological resources that must fit into the citizens' lives (BRASIL, 1997). When integrated into daily analysis and reflections, it is highlighted that technological resources have a relevant role during the teaching and learning process. The technological scarcity makes it even more challenging for Mathematics education in the traditional educational model. This is due to its planned content complexity and how it is transmitted (PREDEBON; GRITTI, 2020).

The scholar failures are among the motivations for the disapproval and scholar evasion of basic education students. Conforming indicated by Rodrigues (2017), among the factors there is the educational institute structure. This is due to the absence of an articulation and diversification planning model for inequalities and different cultural aspects. In other words, although a technological revolution has emerged in the educational field in Brazil, few technological benefits are obtained. Among the reasons, there is the teaching and learning environments' precarious structuring and lack of innovation in the country's basic schools (GONÇALVES et al., 2019).

Recently, the Students International Evaluation Program (PISA, in Brazilian Portuguese) (BRASIL, 2018) in 2018, revealed that 68.1% of 15-year-old Brazilian students do not have a basic mathematical level. In addition, the COVID-19 pandemic accentuated inequalities and inserted even more challenges in Mathematics education. Still, the National Study and Research Educational Institute Anísio Teixeira (INEP, in Brazilian Portuguese), demonstrated concerns related to the Basic Education Development Index¹ (IDEB, in Brazilian Portuguese) 2021. In the report, just 8.1% concluding students from public school elementary levels reached 6 or higher level in the Basic Education Evaluation System (SAEB, in Brazilian Portuguese) proficiency scale in Mathematics.

In the first trimester Everyone For Education organization² 2021 report also presented concerns about education. It was estimated that the students from the final years of elementary school levels would not learn 1.5 points in the Mathematics SAEB scale in an optimistic scenario. Whilst, 7.9 points would be obtained in a pessimistic scenario. Among the factors, there is student formation still based on repetitive, mechanical and individual tasks. These activities demonstrate that the actual Brazilian educational model is very similar to the Education 2.0 model.

¹ https://download.inep.gov.br/institucional/apresentacao_saeb_ideb_2021.pdf

² <https://relatorio-trimestral.todospelaeducacao.org.br/relatorio-trimestral/>

The Education 2.0 is an obsolete and inefficient model due to its low contributions to individual formation. This is due to the prioritization of memorization, reading and repetition approaches, whilst aiming to prevent the students from making mistakes during the learning process (FELCHER; FOLMER, 2021b). Despite this, the Brazilian expectations to adjust to a superior educational model are low. The reason is the fact that technology is not an integral component of the actual Brazilian educational process and pedagogical methodologies. Among the educational models available, Education 2.0 is the second-generation approach presented in society's history, being Education 5.0 the most up-to-date model.

For example, when aiming to establish a teaching method based on Education 3.0, it would be necessary to start dealing with low teaching technology usage. This includes the technological device availability, automation and scientific knowledge systematization (MELLO; NETO; PETRILLO, 2021). In this educational model, the teachers are responsible for digital technology usage as an educational resource. The goal is to encourage autonomy, critical thinking, participation and elaboration of opportunities to release the students' creativity.

Similarly, the teaching technologies scarcity turns the traditional teaching models even distant from the technology-based educational models, such as the Education 4.0 approach. In this educational model, it would be necessary a high-quality technology presence in teaching methods, for example, Artificial Intelligence (AI), Metaverse and Cloud Computing (MORAN, 2018). In addition, Education 5.0 also emerges as a revolutionary solution for the educational field by promising to insert digital technologies into human benefits. In this sense, the goal is to include the human being in the center of innovation and technological transformation (FELCHER; FOLMER, 2021a).

The Education 5.0 also leverages the idea that digital and technological knowledge is important, but socioemotional skills also must be considered during the teaching and learning process. This is because these skills enable individuals to use technologies healthily and productively while building relevant solutions to solve general society challenges. The Metaverse technologies, through Virtual Reality (VR), Augmented Reality (AR) and Mixed Reality (MR) have been vastly projected to a variety of fields when Education 5.0 is considered.

Metaverse offers the users interaction with virtual environments and objects, involving people in new simulated realities (MYSTAKIDIS, 2022). Thus, it is a technology that would provide attractiveness and efficiency to the Education 5.0 concepts. Nowadays, immersion is a component of these technology categories, specifically when considering VR, AR and MR, which were also used to improve the teaching and learning educational process, although these technologies have been little explored (FAKHOURI; MURGO; SISCOOTTO, 2022).

Additionally, AI has a fundamental role in immersive systems development. In Monterubbianesi et al. (2022), some Metaverse immersive systems applications conducted in dental disciplines are summarized. In addition, it is argued that these tools are also effective in educational technology, by their capability to improve the students learning and clinical training.

This observation is similar to the Three-dimensional Geometric Shapes Recognition (TGSR) task. This is because various virtual and real objects related to simulated environments could be used to improve the teaching and learning process of basic Mathematics. In this context, object classification could help in pattern recognition and improve the students' performance understanding. Whilst, this technology also provides personalized instructions to meet the students' specific needs (KASHIVE; POWALE; KASHIVE, 2020).

When integrated into Metaverse technologies, AI can be efficient in assisting innovation development to mitigate Brazilian educational challenges. Considering this, the Inteceleri company has developed the Geometa application, proposing the use of Metaverse environments for education. The company's goal is to enhance the plane and spatial geometries content teaching and learning for basic education students. Thus, this work was elaborated in partnership with the company, aiming to improve Geometa through future AI integration.

In this sense, the main contribution of this work is the Deep Learning (DL) models' training and evaluation for TGSR task through real-world object images. The goal is to propose an AI model to integrate into the Geometa application. The AI model evaluation involved the use of Machine Learning (ML), inference time and dimension performance measures. The intention was to analyze and select the highest assertiveness and smoothness model for TGSR, whilst ensuring the highest quality in future user experience through the application. The future integration of the best AI model into Geometa aims to further facilitate the teaching and learning of geometry contents through the application.

1.1 Motivations

This research is part of a project developed by the Human Resources in Strategic Areas Program (RHAÉ, in Brazilian Portuguese), which has the goal of developing and improving the Geometa³ mobile application. The RHAÉ proposal differential is the offer to users the technology selection to learn Mathematics. In other words, it is proposed the use of VR and AR integrated with AI to perform TGSR in an MR environment. The application was developed by the Inteceleri⁴ company, which is a project partner that builds educational solutions. The company has been developing gamification applications

³ <https://play.google.com/store/apps/details?id=com.Inteceleri.Geometa&pli=1>

⁴ <https://www.inteceleri.com.br>

and methodologies to help students in the basic Mathematics education process.

The improvement to future Geometa versions is under development in partnership with the Operational Research Laboratory (LPO, in Brazilian Portuguese) from the Federal University of Pará. The main goal is to use AI to perform real-time geometric shape recognition through VR and AR environments built from objects and landscapes of Amazon. The Geometa version involves the students in a VR Metaverse, proposing plane and spatial geometries activities to turn the teaching and learning process simpler and more practical.

In this context, this work proposes a TGSR model development for images to be implemented on Geometa allying Metaverse and AI technologies. These components are related to Education 5.0, by inducing the application of an interactive and practical methodology of innovation in favor of education. This is a teaching experiential and immersive methodology, in which the Metaverse technology will be used to facilitate the teaching and learning process. The methodology's goal is to improve Brazilian Mathematics teaching and learning, in addition to mitigating the existing educational challenges.

1.2 General objective

- The best performance DL model selection for TGSR through real-world object images for launch into future Geometa application versions. The launch involves an innovative Education 5.0 approach for teaching and learning plane and spatial geometries contents. The aim is to improve students' performances in Mathematics through the application.

1.3 Specific objectives

- Train DL models for TGSR task through real-world object images dataset.
- Evaluate the TGSR models according to ML, time and dimension performance measures. Posterior to the evaluation, analysis the models' reliability and smoothness for mobile application integration.
- Select the best performance model using an evaluation criteria mainly focused on models' precision and inference performance quality. In sequence, launch the found model for launch into future Geometa application versions.

1.4 Thesis outline

This Master Thesis is composed of seven chapters and bibliographical references. In addition, to the introductory chapter, the remainder of this work is structured as follows:

- Chapter 2 presents the background considering the educational contexts, AI concepts and specific details that sustain the employed method.
- Chapter 3 presents a literature review involving the related works to the proposed method.
- Chapter 4 presents the adopted methodology for classification model development and tool implementation.
- Chapter 5 presents the results and discussions about the developed model and the proposed tool.
- Chapter 6 presents the final considerations and future work.

2 BACKGROUND

In this section, the main work concepts will be presented. The aim is to emphasize the existing advances and gaps in literature through this process. Furthermore, it is also intended to outline the work's relevancy in contributing to the knowledge advance in Mathematics teaching and learning process. In this context, the goal is to provide an intellectual and solid baseline involving the topic, in addition to justifying its importance and impact potential in the educational field.

2.1 The Brazilian Mathematics teaching and learning challenges

The Brazilian Mathematics education faces significant challenges found by IDEB and SAEB statistics. These indicators are crucial instruments for teaching and learning quality assessment in Brazil, providing objective data about student's performance and learning. In 2007, INEP built the IDEB as a national index conducted to measure learning quality and establish teaching improvement goals. The indicator allows for effective monitoring of the education quality, being calculated through two main components: the scholar approval rate and the mean SAEB exam performances. These exams are performed with questions involving the Portuguese and Mathematics disciplines.

The IDEB interpretation allows to understand the increase in the learning results, assessing the balance over the approval rate. Thus, high proficiency would not ensure a high IDEB and simple approval without achieving adequate learning would result in low test performances. The last educational-related statistics revealed a challenging scenario, in which various Brazilian schools were not able to achieve the established goals for Mathematics.

The IDEB and SAEB indicators still face significant challenges in their implementation in schools. This is due to the adequate structure scarcity to leads to educational problems. Table 1 presents the last realized exam, showing the general IDEB scores obtained by the Brazilian students from initial years of elementary schools, in addition to the Ministry of Education (MEC) established goals.

Even though the goals have been reached in the majority of cases for the initial years of elementary school during the years, it was not observed any significant improvements in IDEB during 16 years from its formulation. Although this is the school level that has higher education quality, there was a 0.1-point decrease in IDEB 2021 concerning the previous assessment. In addition to this, the goal was not achieved. In general, the IDEB additional obtained points by year had shown a positive behavior. Nevertheless, these

Table 1. IDEB assessment scores for initial years of Brazilian elementary schools

IDEB	Score	Goal
2005	3.8	-
2007	4.2	3.9
2009	4.6	4.2
2011	5.0	4.6
2013	5.2	4.9
2015	5.5	5.2
2017	5.8	5.5
2019	5.9	5.7
2021	5.8	6.0

Source: Inep (2022)

scores started to decrease from the evaluation realized between 2013 to 2021.

The IDEB counts with 3.8 initial points in 2005, reaching a value up to 34.48% higher in 2021, since the expectations were 36.67%, that is, a value 2.19% points lower than what was expected in the same year. This reveals that although the values were close to the goal, these were distant from the ideal expectations values in the IDEB main educational scale. In sequence, Table 2 presents the IDEB scores information and the goals for the final years of elementary school.

Table 2. IDEB assessment scores for final years of Brazilian elementary schools

IDEB	Score	Goal
2005	3.5	-
2007	3.8	3.5
2009	4.0	3.7
2011	4.1	3.9
2013	4.2	4.4
2015	4.5	4.7
2017	4.7	5.0
2019	4.9	5.2
2021	5.1	5.5

Source: Inep (2022)

There are even major differences when considering the final years of elementary schools since it was verified that the intended goal for 2013 was not reached. Additionally, an improvement of just 0.1 points concerning the previous assessment was observed in 2013, whilst the goal consists of an attempt to achieve a value of 0.5 points or higher. At this school level, the IDEB had an initial score of 3.5 points in 2005, reaching a maximum value of 31.37% higher in 2021.

The difference is about 5% points over the score intended for the same year. This creates a vulnerable and concerning situation for Brazilian students' education, since means

that the teaching and learning process quality was not able to reach the expectations of advances during the years. Finally, Table 3 presents the IDEB scores and goals for Brazilian high schools.

Table 3. IDEB assessment scores for Brazilian high schools

IDEB	Score	Goal
2005	3.4	-
2007	3.5	3.4
2009	3.6	3.5
2011	3.7	3.7
2013	3.7	3.9
2015	3.7	4.3
2017	3.8	4.7
2019	4.2	5.0
2021	4.2	5.2

Source: Inep (2022)

In the high school case, an even more concerning situation due to the absence of IDEB improvements was observed. This school level also presents the lowest assessment in the indicator when compared to the others. The score maintained the same value between 2011 and 2015, presenting little improvement of just 0.1 points in 2017. However, one of the highest and rarest increases in IDEB was observed in 2019 when the three levels were compared, counting with 0.4 more points than in the previous assessment. Still, the IDEB maintained the same score in 2021 when compared to 2019, that is, there were no improvements in the teaching and learning quality.

Counting with an initial score of 3.4 points in 2005, the IDEB reached a maximum value of 19% higher in 2021, even though the value expectations were 34.62% points higher in the score. This shows that the difference of IDEB is around 15.62% concerning the intended to the same year. Thus, in general, revealing a concern about the improvement of the country's education quality. the IDEB leverages into consideration the assessments obtained by the SAEB exam, which are the tests employed every 2 years in high and elementary schools since 1995.

The SAEB¹ is an assessment system that involves tests and questionnaires, proposed by INEP and conducted to examine Brazilian education quality, allowing the investigation and improvement of educational policy. In general, the SAEB aims to assess the student's abilities in Portuguese and Mathematics disciplines, turning able to identify whether the students are learning the transmitted contents in the scholar environment.

Similarly to the National High School Exam (ENEM, in Brazilian Portuguese), the SAEB exam assesses the students through obtained scores from a statistical model

¹ <https://www.gov.br/inep/pt-br/areas-de-atuacao/avaliacao-e-exames-educacionais/saeb>

denominated Item Answer Theory (TRI, in Brazilian Portuguese)². Based on the obtained score, the SAEB is capable of indicating the students learn in a discipline-specific content in the SAEB proficiency scale³. This scale is divided into established levels by score ranges obtained through the exam.

It is noteworthy that the scale levels are not the same among the different scholar levels. This is because the initial years of elementary school count with a level 0 to 10 scale, which starts from a score of values that are lower than 125 points and higher than 350 points. For the final years of elementary school is considered a score stratification from level 1 to 9, counting with a minimum value of 200 points to 400 points. At least, the score division is between level 1 to 10 for high school, being a minimum score of 225 points to values higher than 450 points.

The score scale stratification and the evaluated contents in each level on the proficiency exam could be verified in Inep (2021). In an attempt to present the scores and levels obtained by the students for each assessment employed during the last years, since the SAEB exam conception, the SAEB's Mathematics exam results are shown in Table 4 for the initial years of elementary school.

Table 4. SAEB assessment scores for initial years of Brazilian elementary schools

SAEB	Score	Proficiency
1995	191	Level 3
1997	191	Level 3
1999	181	Level 3
2001	176	Level 3
2003	177	Level 3
2005	182	Level 3
2007	193	Level 3
2009	204	Level 4
2011	210	Level 4
2013	211	Level 4
2015	219	Level 4
2017	224	Level 4
2019	228	Level 5
2021	217	Level 4

Source: Inep (2021)

Concerning the score obtained in the exam, it was verified for the initial years of elementary school that 191 was the first obtained score, reaching a minimum score of 7.85% less points and a maximum of 16.23% more points when compared to the first obtained

² <https://www.gov.br/secom/pt-br/assuntos/noticias/2021/10/veja-como-funciona-a-metodologia-da-teoria-de-resposta-ao-item-tri>

³ <https://www.gov.br/inep/pt-br/centrais-de-conteudo/acervo-linha-editorial/publicacoes-institucionais/avaliacoes-e-exames-da-educacao-basica/escalas-de-proficiencia-do-saeb>

score. The last score obtained was 217 points in 2021, being 11.98% bigger than the first applied assessment. In a preliminary proficiency analysis, the initial years of elementary school were maintained at Level 3 from 1995 to 2009, when the Proficiency Level turned to 4 holding until 2019. In this year, this school level was able to obtain proficiency Level 5, decreasing again into Level 4 in 2021. The score results for the SAEB Mathematics exam for the final years of elementary school are presented in Table 5.

Table 5. SAEB assessment scores for final years of Brazilian elementary schools

SAEB	Score	Proficiency
1995	253	Level 3
1997	250	Level 3
1999	246	Level 2
2001	243	Level 2
2003	245	Level 2
2005	240	Level 2
2007	247	Level 2
2009	249	Level 2
2011	253	Level 3
2013	252	Level 3
2015	256	Level 3
2017	258	Level 3
2019	263	Level 3
2021	256	Level 3

Source: Inep (2021)

For the final years of elementary school, the first score of 253 was verified, achieving a minimum with a decrease of 5.14% less points and a maximum of 3.8% more points concerning the first assessment score. The last score obtained was 256 in 2021, being 1.17% bigger than the first assessment employed. Still, the proficiency started from Level 3 in 1995, which maintained the score until 1999, when a decrease to Level 2 occurred and held until 2011. From this year, just Level 3 was obtained until 2021, that is, the year when the last SAEB assessment exam was employed. The score results of the SAEB Mathematics exam are presented in Table 6 for the high school.

In the high school case, an initial score of 282 was obtained, reaching a minimum of 5.32% less points and a maximum of 2.42% more points when compared to the first score. The last obtained score was 270 in 2021, being 4.26% points lower than the first assessment applied. The Level 3 was maintained from 1995 until 2005 when it decreased to Level 2 and was maintained until 2009. In this year, the high school Proficiency Level started to present a considerable variation among Levels 2 and 3, maintaining Level 3 from 2009 until 2011 and in 2019, in addition to Level 2 from 2013 until 2017 and maintaining this proficiency in 2021, that is, in the last applied assessment.

Table 6. SAEB assessment scores for Brazilian high schools

SAEB	Score	Proficiency
1995	282	Level 3
1997	289	Level 3
1999	280	Level 3
2001	277	Level 3
2003	279	Level 3
2005	271	Level 2
2007	273	Level 2
2009	275	Level 3
2011	275	Level 3
2013	270	Level 2
2015	267	Level 2
2017	270	Level 2
2019	277	Level 3
2021	270	Level 2

Source: Inep (2021)

In this context, it was also verified that the proficiency levels had been around 3, 4 and 5 for the initial years of elementary school, whilst 2 and 3 for the final years of elementary school, finally being 2 and 3 to the high school. Thus, it was verified that among the most complex contents that Brazilian students learn from Mathematics are Spaces and Shapes, Quantities and Measures, in addition to Algebra and Functions.

According to Santos, Nunes and Ferreira (2022), the detailed IDEB and SAEB analysis must be essential aspects when considering pedagogical meetings at school, to face the challenges related to the performance levels through the school pedagogical proposals. The goal is to adequate the activities to allow the students to develop tasks that can consolidate the necessary content knowledge for their lives.

Although this, the actual challenges faced by Brazilian teachers and students in Mathematics education are just being approached in the scientific literature. Conforming Ortigão et al. (2018) and Laubenstein et al. (2019), it is emphasized as challenges: the absence of adequate formation to teachers; the need for efficient didactic resources; and the necessity of innovative pedagogical strategies. It is also noteworthy that the funds for teacher training and high-quality didactic materials could be differential to mitigate the problems related to education quality, although it does not include its complete solution.

The Mathematics education inefficiency entails serious consequences to society, being scholar evasion one of the most concerning. The students who are facing difficulties in Mathematics are susceptible to internalized aversion to school, which results in high scholar abandon indices. According to Penteado et al. (2019) and Bianchini et al. (2018), it is evidenced the relation between disinterest in Mathematics and scholar evasion as one

of the causes, which extends better pedagogical approaches necessities to build interest and comprehension to students.

The resolution of teaching and learning Mathematics challenges in Brazil is a complex task, although it is essential for the country's educational development. The IDEB and SAEB statistical analysis allied to scientific evidence, highlights the concrete actions urgency to improve the Mathematics teaching quality, aiming to reach the quantitative goals and also the formation of citizens to be able to solve daily problems and complex challenges. The data and analysis integration provides a comprehensive vision of the actual Brazilian educational landscape. This underlines the necessity of strategic actions conducted to mitigate the educational problems, attempting to provide prosperous environments for the full development of Brazilian students.

2.2 The technology applied for educational purposes

The rise of technology has undeniably transformed the educational landscape, providing unprecedented opportunities for both teachers and students. This era of technological advancement demands a deep understanding and strategic application of digital tools in the teaching and learning process. In this subsection, the theoretical concepts involving the use of technology in the contemporary educational scenario are outlined.

A paradigm often associated with technology in education is the concept of Education 5.0. This term denotes an educational approach that seeks to explore the potential of technologies such as AI, VR, AR, IoT and Big Data, aiming to provide a personalized, collaborative and immersive learning experience (COŞKUN, 2022). The Education 5.0 approach proposes the integration of these innovative technologies into the educational curriculum to prepare students for the challenges and demands of contemporary society.

The VR emerges as a technology of special interest in the educational context. Through the creation of simulated and immersive environments, the VR solution provides students the opportunity to explore and interact with content engagingly and interactively. This solution ensures immersive educational experiences, enabling students to explore real-world situations in a safe and controlled way (LAMPROPOULOS et al., 2022). Actual research has consistently indicated that the integration of VR into the teaching-learning process can enhance students' understanding, retention, and application of knowledge (AKÇAYIR; AKÇAYIR, 2017).

Additionally, AR emerges as another prominent technology in the educational field. In comparison with VR, which creates completely immersive virtual environments, the AR overlays virtual objects and information onto the real environment. This allows students to interact contextually with digital content, making learning more tangible and interactive. The AR can be applied in various disciplines, including History, Science, Mathematics and

Arts, providing new perspectives and exploration opportunities for students (HIDAYAT; SUKMAWARTI; SUWANTO, 2021).

Finally, the AI technology plays a vital role in modern education. This innovation can be used to personalize the teaching process, in addition to adapting contents and methodologies according to the students' needs (LUAN et al., 2020). Additionally, AI facilitates automated assessment, identification of learning patterns, and the development of virtual assistants to support students throughout the educational process (ŞAHİN; YURDUGÜL, 2020).

It is imperative to emphasize that the incorporation of technology into teaching would not be considered a goal in itself, but rather a tool to enhance existing pedagogical practices. In this scenario, the teachers have a central role in the selection and appropriation of technologies, ensuring that they align with educational goals and promote meaningful and critical learning.

These technologies offer innovative resources and possibilities for the educational process, enabling more immersive, personalized and collaborative learning experiences. The exploration of VR, AR and AI in the educational context has been the subject of research, highlighting their potential to enhance education and prepare students to face the challenges of their daily lives.

2.3 The education in Metaverse

The intersection between Mathematics education and emerging technologies, especially the Metaverse, has stood out as an innovative research field. This occurs because the Metaverse is a persistent, three-dimensional virtual environment shared by users, capable of transforming how Mathematics is taught and learned.

The use of the Metaverse in Mathematics educational contexts has become a subject of study due to its ability to create interactive and immersive environments. In the Metaverse, abstract Mathematics concepts can be visualized tangibly, providing a deeper and more concrete understanding for students (ŞEYMA; ÖZDEMİR, 2022). Furthermore, the Metaverse allows real-time collaboration, enabling students and teachers to interact and solve complex Mathematics problems together, regardless of their physical locations (DÍAZ; SALDAÑA; AVILA, 2020).

The Mathematics concepts' visual representation in the Metaverse can be particularly beneficial for students who learn better through visual and interactive experiences (ALTHANI; MADGE; POESIO, 2022). For example, a variety of factors can be dynamically explored, allowing students to manipulate variables in real time and observe the resulting changes in the teaching process (MUSTAFA, 2022).

Moreover, the Metaverse also provides opportunities for the creation of customized and challenging educational scenarios. The teachers can design narrative-based mathematical activities that involve real-world problems, encouraging students to apply mathematical concepts to solve challenges within the virtual environment (HERRERA; PÉREZ; ORDÓÑEZ, 2019). These activities not only increase student engagement but also promote the practical application of mathematical knowledge (AKMAN; ÇAKIR, 2023).

Nevertheless, it is crucial to address accessibility and equity issues when implementing the Metaverse in Mathematics education. The reason is that there are students who do not have equal access to technological devices or high-speed internet connections, which can create disparities in access to Metaverse resources (KADDOURA; HUSSEINY, 2023). Additionally, it is essential to ensure that the design of virtual environments is inclusive, taking into account the needs of students with different abilities and learning styles (ZALLIO; CLARKSON, 2022).

2.4 The education considering Artificial Intelligence

The digital revolution has brought innovations to the field of education with AI emerging as a powerful and transformational tool. The AI concept refers to the ability of machines to learn and perform tasks that would previously require human intelligence. In the educational context, AI can be applied to roles from virtual classroom assistants to advanced adaptive learning systems (BAKER et al., 2010). These systems have the unique ability to analyze large sets of educational data to identify learning patterns and improve the educational methods aiming to meet the specific needs of the students (HOLMES; BIALIK; FADEL, 2023).

Teaching and learning adaptability is one of the key benefits provided by AI. Through the real-time analysis of students' performance, AI systems can adjust content and the optimize pedagogical approach aiming at the understanding and retention of the students (AZEVEDO et al., 2022). This creates a highly adaptive learning experience where the students can maximize their educational potential.

In this sense, there is also the automation of administrative tasks among the relevant applications of AI, proposing teachers' free time for direct teaching. The AI systems can manage assessments, provide instant feedback to students and even assist in creating educational content (CHEN; CHEN; LIN, 2020). This not only increases efficiency but also allows teachers to dedicate more time to individualized support for students, thus strengthening the student-teacher connection (AL-MALAH; JINAH; ALRIKABI, 2020).

Despite the benefits, the use of AI in education has challenges. The points are facts such as data privacy, which is a central issue, as AI systems require access to sensitive student information to work effectively (NGUYEN et al., 2023). Additionally, concerns

about equity arise, with the possibility that students with less access to technology or technological training may have trouble learning (BAIDOO-ANU; ANSAH, 2023). These ethical and social issues must be carefully considered when implementing AI systems for scholarly environments.

2.5 GeoMeta: Learn Geometry in the Metaverse

The "GeoMeta: Learn Geometry in the Metaverse" is an application currently under development by Inteceleri Technology for Education. The proposal aims to provide students and teachers with a simpler way to teach and learn plane and spatial geometries through the Metaverse. The application utilizes the Metaverse to simulate and replicate three-dimensional virtual learning environments through VR and AR. These environments are created from scenes and objects from everyday life, as well as landscapes and contexts from the Pará Amazon.

The scenes and objects constructed in the application are connected by regular geometric relationships, allowing the implementation of often complex mathematical concepts in a simple manner. The goal of this application is to provide users with immersive and meaningful learning experiences to deepen their understanding of geometry in the real world. The application is already available for download on Android and iOS platforms, providing a variety of implemented features aiming a meaningful learning of complex Mathematics concepts.

In this sense, the application provides environments, objects and activities, as illustrated in Figure 1. Among the available activities, the user must initially customize an avatar (Figure 1.1), which will represent their identity in the application. Subsequently, in a virtual classroom environment, the user must identify plane and/or spatial geometric shapes to answer quizzes about the chosen forms (Figure 1.2). Additionally, the user can also explore real environments related to the identified object in a previous task, using images extracted from Google Earth⁴ (Figure 1.3).

The expeditions can be conducted through the world map available in the classroom virtual environment (Figure 1.4). During the expeditions, the users must identify the geometric shapes hidden throughout the environments accessible in the application (Figure 1.5). The main goal of the educational game is to identify and respond to the highest number of quizzes possible related to the available geometric shapes. Finally, users can learn about other concepts, such as planning and the cartesian plane (Figure 1.6).

The application also incentivizes the correct answers in quizzes provided by users with a small robot as a reward. These rewards serve as a way to ensure the continuity of learning, keeping the user immersed in the possibilities of achievements within the

⁴ <https://www.google.com/earth/>

Figure 1. Geometa application functionalities and resources



Source: Inteceleri company website

environment. For a more immersive experience in the application, it is necessary to use VR glasses. Therefore, the Inteceleri company provides the Miritiboard VR glasses, which are characterized by being an accessible and easy-to-use resource. Figure 2 shows the design of the MiritiBoard VR.

Figure 2. Miritiboard VR glasses for more immersive Metaverse experiences



Source: Inteceleri company website

The Miritiboard VR is made from palm tree fibers known as Maurita Flexuosa, commonly referred to as Miriti or Buriti, which is a raw material native to the Amazon. This component enhances interaction with the Metaverse environment and can be used with the Metaverse and other VR applications, as demonstrated in the company's video⁵. Moreover, this educational tool aims to provide a dynamic and sustainable learning system, enabling the immersion of students and teachers in various virtual environments worldwide.

⁵ <https://www.youtube.com/watch?v=Wx79bwiWPvE>

This facilitates the learning of geometry elements, activity creation and geography, among other subjects.

2.6 The Deep Learning in object classification tasks

The image recognition task has become a fundamental research area in Computer Vision, largely due to the DL field. The DL is a subfield of ML that involves Deep Artificial Neural Networks, composed of multiple layers of processing units. Through an iterative training process on large sets of labeled image data. The Deep Neural Networks can learn complex hierarchical representations of the visual features of images (SEWAK; SAHAY; RATHORE, 2020).

In the context of image recognition, the DL has been applied to solve a variety of societal problems. Among the reliable DL applications, there is object recognition, where Deep Neural Networks can classify different categories of objects through images (DHILLON; VERMA, 2020). Moreover, DL has been successfully applied to tasks such as object detection, image segmentation and even automatic generation of image descriptions (JIAO et al., 2019; STEFANINI et al., 2022).

The interconnection between DL and Computer Vision involves the use of Deep Neural Networks to extract features from images and subsequently analyze these features to identify objects and patterns. For instance, the use of deep convolutions allows the networks to identify complex features at different scales and orientations, capturing essential details for object recognition (CRUTTWELL et al., 2022). Additionally, advances in object detection and recognition in videos and real-time scenarios were also driven by DL. This subfield has significant applications in tasks such as video surveillance, autonomous vehicle driving and medical image analysis (REDMON et al., 2016; CAO et al., 2017).

Despite the impressive advancements, it is emphasized that image recognition still faces significant challenges, such as object recognition in adverse lighting conditions or identifying objects in highly polluted images. Additionally, ethical issues, such as bias in facial recognition algorithms, have also received considerable attention (BUOLAMWINI; GEBRU, 2018). Finally, the next subsections will present the details of the models' architectural constructions, which were used in this research.

2.6.1 Convolutional Neural Network

The Convolutional Neural Network (CNN) is a traditional DL architecture composed of three primary types of layers: the Convolutional Layer, the Pooling Layer and the Fully Connected Layer (LECUN; BENGIO; HINTON, 2015). The Convolutional Layer, situated at the inception of the network is responsible for able the model to discern image regions

and intrinsic features until it ultimately achieves object recognition. This layer is the fundamental of CNNs and performs the majority of computations.

The convolution process entails the movement of a kernel or filter across the image's receptive fields, assessing the presence of specific features within the image. The kernel is responsible for traversing the entire image, calculating a dot product between the input pixels and the filter after each iteration. The cumulative output of these dot products constitutes a feature map. Consequently, this layer transforms the image into numerical values, enabling the CNN to interpret the image and extract pertinent patterns. Let I be the input tensor, K^c be the convolutional kernel and O be the output tensor, the Convolutional Layer function is expressed in Equation 1.

$$O_{i,j} = \sum_m \sum_n I_{i+m,j+n} \cdot K_{m,n}^c \quad (1)$$

The Equation 1 presents $O_{i,j}$ in which is represented the output element at position (i,j) . The $K_{m,n}^c$ is the convolutional kernel setups at position (m,n) . The summation is performed over all relevant positions in the input tensor. Besides this, the Pooling Layer, similarly to the Convolutional Layer, employs a kernel or filter to scan the input image. However, the difference is that the Pooling Layer serves to reduce parameter count and incurs some information loss. Nevertheless, this reduction in complexity enhances CNN's efficiency. Let P be the pooling function and S the stride, the pooling operation can be expressed as presented by Equation 2.

$$O_{i,j} = P(I_{i+m \cdot S, j+n \cdot S}) \quad (2)$$

The Equation 2 presents $O_{i,j}$ as the pooling result at position (i,j) . The operation P is applied to the local neighborhood defined by the pooling window. The stride S adjusts indices of the input I within the pooling window. Meanwhile, the Fully Connected Layer assumes the responsibility of image classification based on the features extracted from preceding layers. In this context, fully connected implies that all inputs or nodes from one layer establish connections with every activation unit or node in the subsequent layer. Let W be the weight matrix, b be the bias tensor and σ be the activation function. The Y is the Fully Connected Layer operation output tensor, which is presented in Equation 3.

$$Y = \sigma(W \cdot I + b) \quad (3)$$

The Equation 3 presents the application of a σ activation function over the features processed by previous layers to perform the classification process. The operational principle of a CNN revolves around each layer employing filters or kernels to process the image, progressively enhancing the level of detail and complexity. In the initial layers, these filters

begin with rudimentary features, while successive layers increase in complexity, allowing for the identification of unique object representations. The output of each Convolutional Layer, representing a partially recognized image, serves as the input for the subsequent layer. Ultimately, in the Fully Connected Layer, the CNN achieves image or object recognition.

2.6.2 MobileNet

The MobileNet model is a CNN that serves as the foundation for training highly efficient classifiers in terms of size and speed (HOWARD et al., 2017). The network's computational performance and power consumption are directly proportional to the number of multiply-accumulate operations, signifying fused multiplication and addition operations.

This model uses depthwise convolution, which represents channel-wise spatial convolutions. Besides this, it also uses pointwise convolution, which corresponds to convolutions designed to change the dimensional characteristics of the data. This convolution employs a kernel and merges features generated by depthwise convolution with the output channels mirroring the input channels.

The MobileNet convolutions use filters to extract specific pixel subsets from images, yielding predictive outputs based on image characteristics. The traditional CNN utilizes pre-defined filters, conducting convolutions with image pixel matrices, thereby filtering objects within the image. A comparison is then made against a repository of pre-defined objects to identify matches, enabling image prediction.

However, these filters demand substantial GPU resources for efficient performance across vast datasets, beyond the capacity of most mobile devices. In this sense, the MobileNet presents convolution approaches differently from traditional CNNs, through the use of depthwise and pointwise convolution. This enhances efficiency, facilitating mobile system integration. The reduced computation and recognition times yield rapid responses and turn the MobileNet into a viable solution for image recognition models designed for mobile applications. These models are also compact, low-latency and energy-efficient.

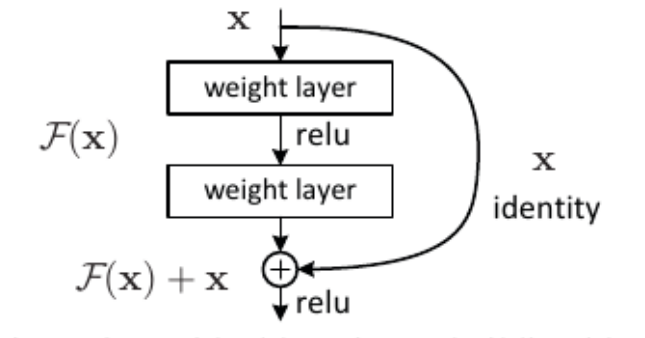
2.6.3 Residual Neural Network

The Residual Neural Network (ResNet) presents an innovative solution to the vanishing gradient problem through the integration of skip connections (HE et al., 2016). These connections involve the stacking of multiple identity mappings, which are essentially convolutional layers with negligible initial impact. The architecture bypasses these layers and reutilizes activations from the preceding layer, thereby expediting initial training by condensing the network into a reduced number of layers.

The ResNet architecture concept of residual blocks was proposed to address the vanishing or exploding gradient problem. The residual block of a ResNet model architecture

is presented in Figure 3. The skip connections presented in the blocks link the activations of one layer to subsequent layers, skipping intermediate layers. Residual networks are constructed by cascading the residual blocks.

Figure 3. ResNet DL model building block



Source: He et al. (2016)

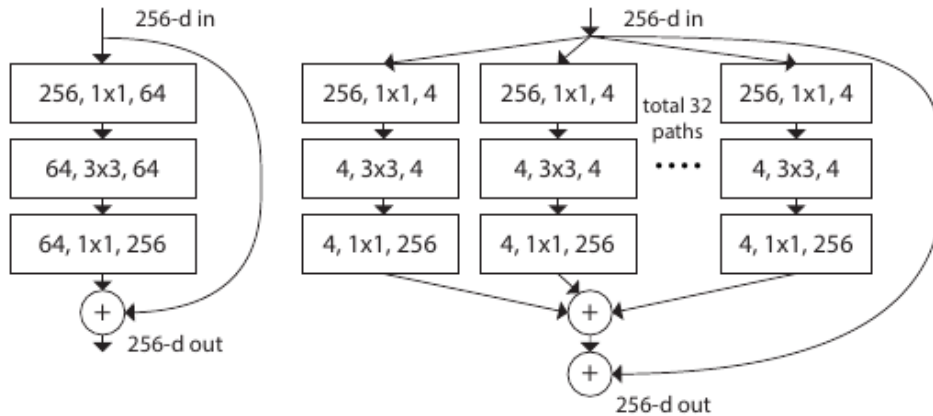
The underlying principle behind this architecture is the shift from expecting layers to learn the fundamental mapping to permitting the network to adapt to the residual mapping. Rather than defining an initial mapping, the network learns to approximate it. The inclusion of skip connections offers a valuable advantage: if any layer impairs architectural performance, regularization can bypass it. Consequently, very deep neural networks can be trained without succumbing to the issues associated with vanishing or exploding gradients.

2.6.4 Residual Neural Network for Next dimension

The Residual Neural Network for NeXt dimension (ResNeXt) model is very similar to the ResNet model, the main difference is the presence of another dimension called the cardinality (XIE et al., 2017). In contrast to the “Network-in-Network” approach, it is “Network-in-Neuron” and expands along a new dimension. The ResNeXt and ResNet building blocks are compared in Figure 4, presenting its residual layers and considering its input channels, filter size and output channels.

These two models have the same complexity although the ResNeXt model has one more dimension concerning the ResNet model. This additional dimension proposes parallel layers in the model’s architecture. Instead of a linear function in a simple neuron, a nonlinear function is performed for each path. The cardinality dimension controls the number of complex transformations. It is increased directly, added together and also added with the skip connection path. Unlike ResNet, in ResNeXt, the neurons at one path will not be connected to the neurons at other paths.

Figure 4. ResNeXt DL model architecture



Source: Xie et al. (2017)

2.6.5 Vision Transformer

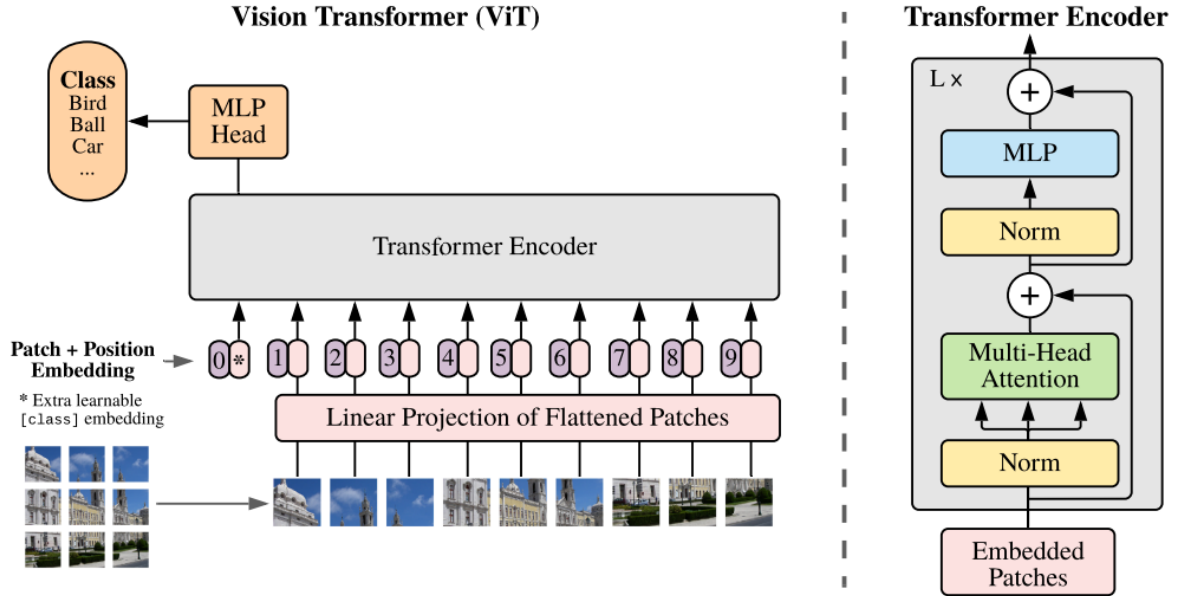
The Vision Transformer (ViT) model operates by partitioning an image into fixed-size patches, embedding each patch accurately and introducing positional embeddings as an input to the transformer encoder (DOSOVIISKIY et al., 2010). Within the ViT architecture, the self-attention layer assumes a central role, enabling the global embedding of information across the entire image. Additionally, the model is trained to encode the relative spatial relationships between image patches, facilitating the reconstruction of the image's underlying structure.

In recent years, the field of computer vision has witnessed a paradigm shift with the introduction of Image Transformers. Unlike traditional CNNs, Image Transformers leverage self-attention mechanisms to capture contextual dependencies across the input image, making them highly effective for various vision tasks such as image classification, object detection, and image generation.

The self-attention mechanism plays a crucial role in capturing contextual information within input data. It empowers the ViT model to focus its attention on different regions of the input data, prioritizing those most pertinent to the given task. Consequently, the self-attention mechanism calculates a weighted sum of the input data, with weights determined by the similarity between input features. This enables the model to assign greater importance to relevant input features, enhancing its capacity to capture informative representations of the input data.

In essence, self-attention serves as a foundational computational primitive, quantifying pairwise interactions among entities, thereby facilitating the learning of hierarchical structures and alignments within input data. Notably, attention mechanisms have proven to be instrumental in enhancing the robustness of vision networks. The ViT model architecture and its training process are presented in Figure 5.

Figure 5. ViT DL model architecture



Source: Dosovitskiy et al. (2021)

In this sense, the ViT architecture treats images as sequences, making independent learning of image structure possible. The input images are treated as sequences of patches, each patch flattened into a single vector through channel concatenation and linear projection to the desired input dimension. These image patches serve as sequence tokens, similar to words in natural language processing. The ViT is a model based on Image Transformers, which also employs attention mechanisms and transformer architectures to capture long-range dependencies. Whilst, the architecture generates context-aware representations for image understanding through Transformer Encoder.

The main component of the Transformer Encoder is the Attention Layer, which enables the model to weigh different parts of the input image differently. The weighing process is based on the relevance of the parts to the task at hand. The self-attention mechanism calculates a set of attention scores that determine the importance of each element in the input sequence concerning every other element. This mechanism is mathematically expressed in Equation 4:

$$\text{Attention}(Q, K, V) = \text{softmax} \left(\frac{QK^T}{\sqrt{d_k}} \right) V \quad (4)$$

The Equation 4 presents the Attention Layer function applied over Q , K and V which represents the query, key and value matrices, respectively. These factors are the fundamental aspects of the architecture. The division by $\sqrt{d_k}$ is a normalization factor and the softmax function ensures that the attention scores sum to 1. The resulting attention

scores are then used to weigh the values, producing the final attended output.

The Transformer Encoder is responsible for processing the input image and extracting meaningful representations. It consists of multiple layers, each comprising a combination of sub-layers, including Multi-Head Attention and Feed Forward networks. The output of each sub-layer is passed through a residual connection and Layer Normalization. The mathematical representation of the Transformer Encoder is presented in Equation 5.

$$E(x) = \text{LN}(x + \text{MHA}(x) + \text{FF}(x)) \quad (5)$$

The Equation 5 shows the Transformer Encoder function E over x representing the input features, the Layer Normalization function as LN , the Multi-Head Attention function as MHA and the Feed Forward function as FF . The Layer Norm ensures stable training and adaptability to variations among training images. Meanwhile, the Multi-Head Attention network generates attention maps from embedded visual tokens, guiding the network's focus to critical image regions, such as objects. Finally, the Multi-Layer Perceptrons act as a two-layer classification network. The final network block, often referred to as the Multi-Layer Perceptron head, serves as the transformer's output.

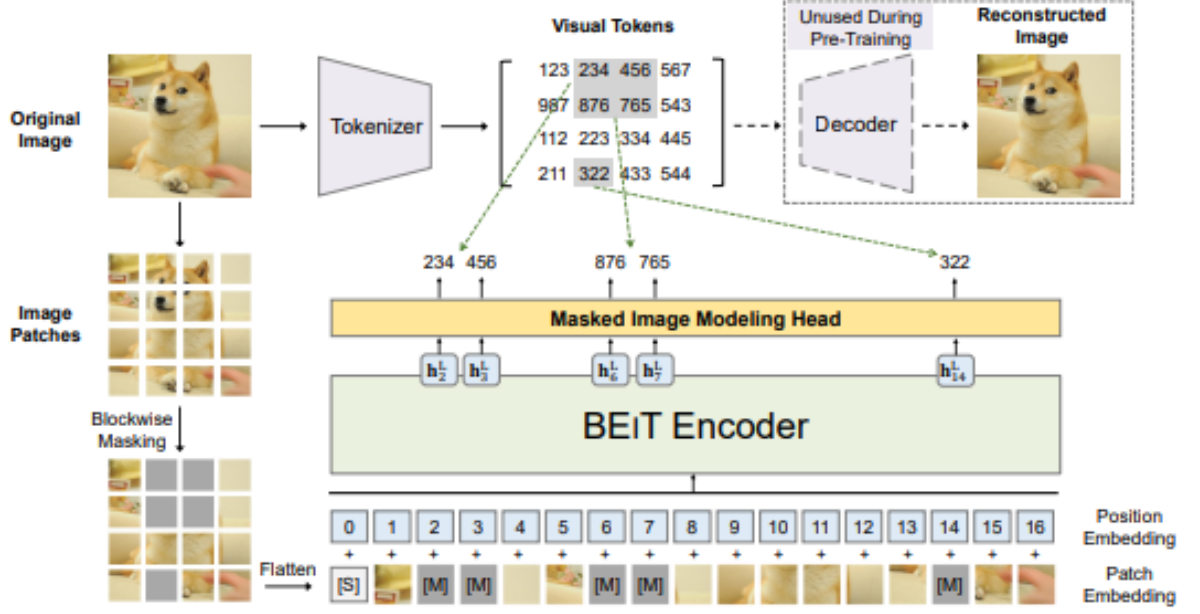
2.6.6 Bidirectional Encoder for Image Transformers

The Bidirectional Encoder for Image Transformers (BEiT) is a ViT state-of-the-art image classification model that includes a pre-training task namely Masked Image Modeling (BAO et al., 2021). This method uses two views for each image, i.e. image patches and visual tokens. In this model, the image patch is masked as a block rather than masking randomly. About 40% of the masks need to be masked in blocks and a minimum of 16 patches constitute a block. The model learns to recover the visual tokens of the original image, instead of the raw pixels of masked patches. In the case of the visual tokens, the image is represented as a sequence of discrete tokens obtained by an image tokenizer, instead of raw pixels. The process of pre-training the BEiT model is presented in Figure 6.

In this sense, the original image is divided into image patches and these are masked randomly, through a blockwise masking process. In sequence, the image patch is flattened and transformed into a vector and the algorithm also obtains positional embeddings and patch embeddings from these image patches. The obtained embeddings are inserted in the encoder architecture. Finally, the model's goal is to predict the masked image tokens.

The evaluation process is performed through an image tokenizer. The image patches also correspond to visual tokens. This allows the comparison between the real image token and the predicted token related to an image patch. Finally, the image can be reconstructed using these tokens. Once the pre-training is done, it can be applied to the classification tasks through a training process.

Figure 6. BEiT training and classification workflow



Source: Bao et al. (2022)

2.7 Artificial Intelligence performance measures

The ML field has a variety of performance measures to evaluate the trained models (NASER; ALAVI, 2021). Every classification task has True Positives (TP), True Negatives (TN), False Positives (FP) and False Negatives (FN) as fundamental performance measures to assess the model's precision and other evaluations. In the context of the TGSR task, TP is the correctly classified instance representation as the desired geometric shape, i.e., when the model correctly identifies the shape present in the image. Similarly, TN are also correctly classified instances but identified as not being the geometric shape in question.

In contrast, FP are cases where the model incorrectly classifies a shape as the desired geometric shape, whilst false negatives FN occur when the model fails to recognize the actual geometric shape in the image. These performance measures are crucial to evaluate the models' effectiveness in differentiating the image classes, providing valuable insights related to the predictions' reliability.

In this sense, ML performance measures are intrinsically linked to the ability of AI models to recognize patterns and make accurate classifications. The Precision is presented in Equation 6 and reflects the model's ability to distinguish geometric shapes during the classification process in the TGSR task.

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (6)$$

The Precision is calculated based on TP and FP classified samples. In a TGSR task context, the lower the number of incorrectly classified images as geometric shapes compared to ground-truth labels, the higher the precision of the model. In addition to Precision, the Accuracy is also evaluated to provide an overall view of the model’s performance when all geometric shapes are considered and presented in Equation 7.

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{N}} \quad (7)$$

The Accuracy considers TP and TN classified samples, providing the rate between correctly classified geometric shapes and Total Inferences Amount (N) made. The performance measure evaluation is also performed individually for each specific geometric shape, allowing the identification of the model’s performance for each class. In addition, Recall is also used to identify how often the model makes correct inferences of a specific geometric shape when just a geometric shape class is considered and presented in Equation 8.

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (8)$$

The Recall considers TP and FN in its calculation, highlighting the model’s ability to correctly identify a specific geometric shape, minimizing the occurrence of FN. Finally, the F1-Score, according to Equation 9, is also evaluated to identify the model’s quality by considering both Precision and Recall together. This performance measure provides a stability measure among the two mentioned, which makes it useful when seeking a balance between the measures.

$$\text{F1-Score} = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \quad (9)$$

The ML performance measures are crucial for assessing the capabilities of AI models in recognizing geometric shapes and making accurate decisions. Even so, evaluating the inference time and dimension of models also plays a crucial role in choosing the appropriate smoothness model for integration with the application (OGDEN; GUO, 2019). This evaluation aims to meet user experience criteria and application flow during model classifications. The main performance measure in this context is the number of Inferences Per Second (IPS), which reflects the model’s ability to perform inferences within a specific time window and is presented in Equation 10.

$$\text{IPS} = \frac{\text{N}}{\text{TIT}} \quad (10)$$

The IPS calculation considers N and the Total Inference Time (TIT), which corresponds to the time needed for the model to make inferences on all test set samples. Additionally, the Time Per Inference (TPI) is also evaluated as an alternative performance

measure based on TIT, which is presented in Equation 11. This measure estimates the necessary time to make an inference for a test set sample in milliseconds.

$$\text{TPI} = \frac{\text{TIT}}{N} \quad (11)$$

Finally, the Model Total Parameters (MTP) in million parameters and Model Size (MS) in Megabytes on disk performance measures were also evaluated in this work. In general, the goal was to evaluate the models' reliability and smoothness through these performance measures.

3 RELATED WORKS

This section will present the related works, focusing on showing the shape recognition task through Computer Vision and DL methods. These works allow to evaluate the techniques employed and ensure their state-of-the-art applicability. This research term search protocol adopted was centered on four key concepts: “3D shapes recognition”, “3D shapes classification”, “object classification” and “educational applications”. The setups considered were IEEE Xplore and Scopus. In this sense, it was firstly highlighted that although the TGSR task is similar to general Object shape recognition (OSR), there are strict differences between the tasks.

Among the object recognition tasks, it is emphasized mainly the data restrictions considered for the TGSR task when compared to the OSR task. In other words, these two tasks have contrast in the used data, once the TGSR needs just the object images representing three-dimensional geometric shapes employed in basic Mathematics education. Examples of the classes related to the TGSR task are parallelepiped, sphere, cylinder, cone and surface (which are three-dimensional representations of bidimensional geometric shapes). In this sense, due to the similarities among the tasks, the OSR methods previously used will be presented aiming to propose a comparison concerning the TGSR strategy.

A variety of algorithms were developed to perform the OSR on literature, as presented in Jin and Li (2022), which proposes a framework to learn the image characteristics obtained through 3D objects. The focus was on identifying representative visualizations of these objects. The Zanuttigh and Minto (2017) explore the deep map sets development, rendering the input 3D shapes through different viewpoints; in sequence, these deep maps are inserted in a multiple ramifications neural network.

In Wang et al. (2019), the idea is to perform three-dimensional image processing through bidimensional preprocessing, considering the image fragmentation and employing CNN models. Barbu et al. (2019) developed a highly automated platform that can collect controlled datasets through crowdsourcing image capture and annotation; in this sense, an object classifiers performance investigation was performed in the built dataset (i.e. ObjectNet¹), using an image annotation task.

For other cases, the edge-pixel values approach also presents benefits, such as the execution time and high precision; in contrast, it also has noise data processing complexity disadvantages (CHEN et al., 2010). Similarly, the fuzzy logic methods have disadvantages in low recognition rates when compared to other low-precision and traditional methods (JORGE; FONSECA, 1999).

¹ <https://objectnet.dev>

Despite these methods being relevant to the field, the neural network models for OSR methods present the best results, in addition to being the most used in state-of-the-art (MUZAHID et al., 2020). The neural networks are among the most used techniques in shape classification research, once they present high capability to learn patterns and can lead with real-world asymmetries (DHILLON; VERMA, 2020).

Additionally, in Zhou et al. (2019), a polar vision strategy was used, in which the main 3D shape characteristics and their internal structures are reflected, exploring DL techniques for object classification. Another research example that performs the OSR is Qi et al. (2017), planning a neural network to consume cloud points directly; the proposal concerns the input points invariance permutation and proposes a unified architecture among object classification, component segmentation and scene semantic analysis applications.

In addition, in Feng et al. (2018), various advanced CNN models are used for 3D objects panoramic visualization, including the identification of more points when compared to traditional CNN models. The CNNs fusion strategy for OSR is also used in other research, such as Xu and Todorovic (2016). The work employed a CNN-based cluster search to identify the ideal model architecture to estimate the best network parameters.

Conforming to Dirik and Yanardag (2022), the ViT architecture could be used to apply 2D vision transformations aiming to solve OSR tasks, being also shown the 3D reconstruction task through 2D images using this architecture. Whilst in Gupta and Khan (2022), an efficient three-dimensional objects localization approach is presented to mobile devices using the MobileNet model, in which the neural network is trained in a big annotated database using an unsupervised and supervised techniques combination.

Liu et al. (2022), it is approached the complex efficient 3D cloud-point representation challenge for ML tasks, in which the proposed technique involves the BEiT model. In this work, the autoencoders masked are used to learn 3D point-cloud representations which can be used in OSR and shape reconstruction tasks. In An et al. (2018), it is established the human face recognition challenge in different poses and illumination conditions. For this, a transfer learning approach and 3D morphological modeling were proposed using the ResNet model.

Besides, in Balagopal et al. (2018), a trained CNN through a big computerized tomography image database is presented to learn the automatic organ image segmentation. The work includes images of the prostate, bladder, rectum and pelvic bones for the segmentation process. In addition, the architecture consists of a localization 2D organ volume network followed by a 3D segmentation network for prostate, bladder, rectum and femoral heads volumetric segmentation, using the ResNeXt DL model.

Finally, Açikgöl and Şad (2021) described the students' acceptance level related to Mathematics learning through mobile technology usage. The study found that the majority

of the students are favorable to mobile technology usage allied to Mathematics learning. In Coutinho, Almeida and Jatobá (2021), technological proposals to help in Mathematics learning in classrooms and smartphone usage as a mediation instrument between the students and Mathematics concepts are also presented. The Table 7 summarizes the related works discussed and presents the presence and absence of OSR, TGSR, DL, Educational Proposal (EP) and Inference Time and Dimension Evaluation (ITDE).

Table 7. State-of-the-art related works proposals

Related Work	OSR	TGSR	DL	EP	ITDE
Jorge and Fonseca (1999)	✓	✗	✗	✗	✗
Chen et al. (2010)	✓	✗	✗	✗	✗
Xu and Todorovic (2016)	✓	✗	✓	✗	✗
Zanuttigh and Minto (2017)	✓	✗	✓	✗	✗
Qi et al. (2017)	✓	✗	✓	✗	✗
Feng et al. (2018)	✓	✗	✓	✗	✗
An et al. (2018)	✓	✗	✓	✗	✗
Balagopal et al. (2018)	✗	✗	✗	✓	✗
Wang et al. (2019)	✓	✗	✓	✗	✗
Barbu et al. (2019)	✓	✗	✓	✗	✗
Zhou et al. (2019)	✓	✗	✓	✗	✗
Muzahid et al. (2020)	✓	✗	✓	✗	✗
Dhillon and Verma (2020)	✓	✗	✓	✗	✗
Açikgöl and Şad (2021)	✗	✗	✗	✓	✗
Coutinho, Almeida e Jatobá (2021)	✗	✗	✓	✗	✗
Jin and Li (2022)	✓	✗	✓	✗	✗
Dirik and Yanardag (2022)	✓	✗	✓	✗	✗
Gupta and Khan (2022)	✓	✗	✓	✗	✗
Liu et al. (2022)	✓	✗	✓	✗	✗
Proposed Method	✓	✓	✓	✓	✓

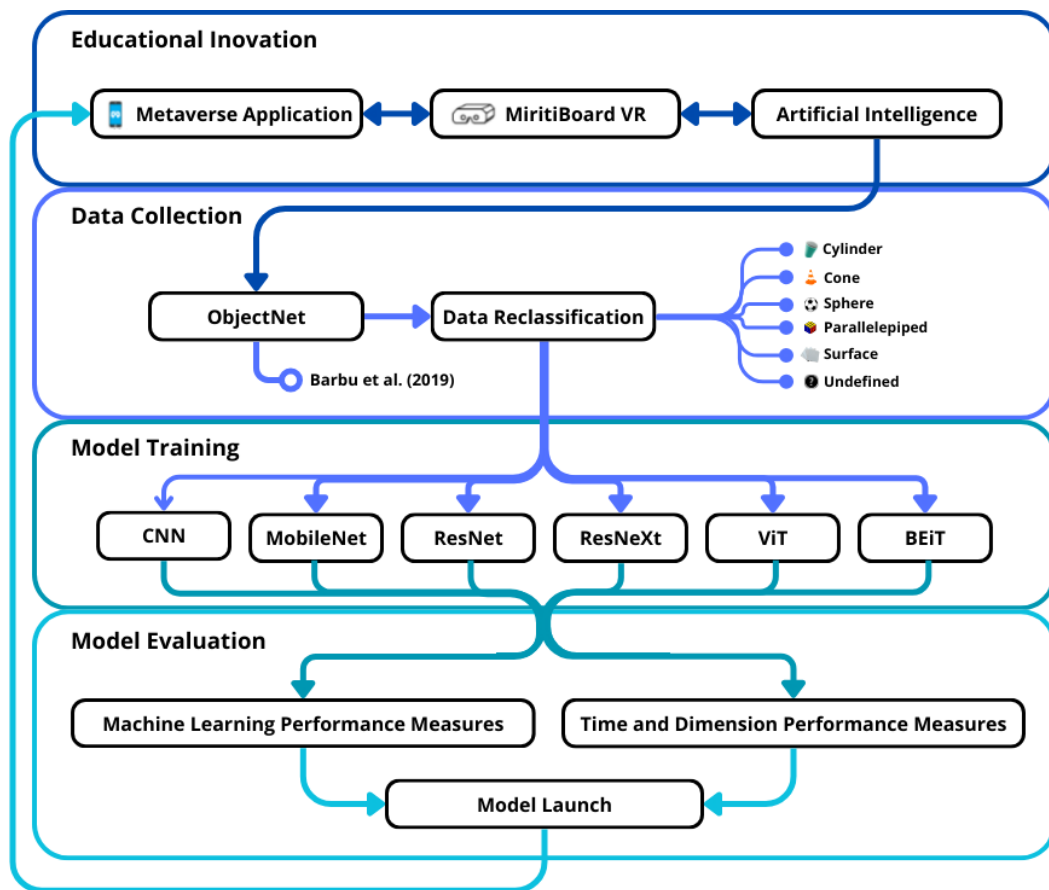
Source: Author

Nowadays, in conformity to the discussed related works, there were no found approaches that: focus on integrating trained AI models with an application to be used in basic scholar environments; training Deep Learning models for TGSR and not general OSR; and evaluating ITDE performance measures to analyze the models' smoothness for launch into applications. Thus, no works were found that replicate exactly what is proposed in this work. This occurs mainly because the found and mentioned works focus on just recognizing the general objects' shapes. This work proposes the TGSR task as an EP considering the use of DL models to evaluate ML and ITDE performance measures through real-world object images. The focus is the selection of the best-performance model for launch into future Geometa versions.

4 METHODOLOGY

The goal of this proposal is to train an AI model for integration building the second version of the Geometa educational application. In this sense, the method was organized into several steps presented in Figure 7. The method includes implementation blocks involving educational innovation, real-world image collection, training of AI models and fitted models performance evaluation. The aim is to select the best performance model for integration with the application. The implementation blocks will be presented in the following subsections.

Figure 7. Proposed method for educational application AI development and launch



Source: Author

4.1 Educational Innovation

Initially, the intention is to enhance the application through AI model integration for TGSR using VR and AR modules. The goal is to enable real-world object pattern recognition as three-dimensional geometric shapes in real-time, further simplifying the

teaching of complex Mathematics concepts. In this context, DL models were trained on images from the ObjectNet dataset (BARBU et al., 2019) and the models' performances were evaluated through defined performance measures. The details about the next steps will be provided in the following subsections.

4.2 Data Collection

The image collection process was initially performed considering that the ideal type of images would include: centralized everyday objects; a background that contains the fewest possible non-related objects; and good lighting conditions for proper object visualization. Thus, a search for existing datasets of objects was performed to simplify the process of obtaining the corresponding images, making it possible to perform just an object-to-geometric shape reclassification process.

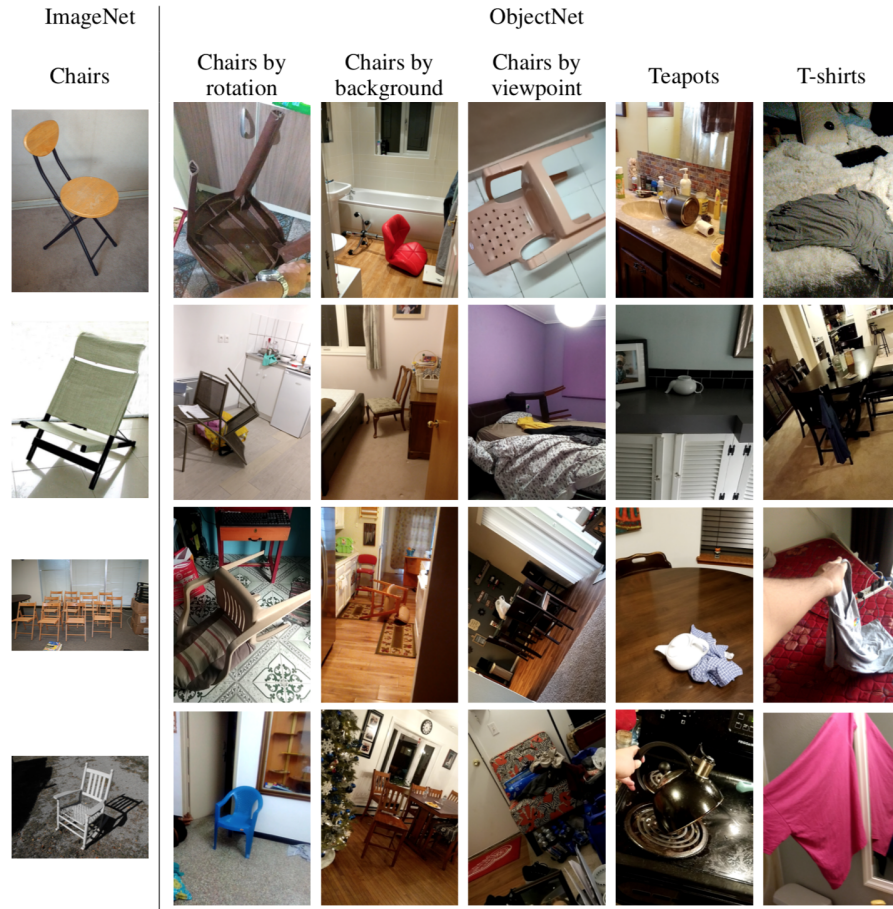
For this reason, the dataset selected for this proposal was ObjectNet. This dataset contains real-world object images involving background control, rotations and viewpoints (BARBU et al., 2019). In general, these factors collectively allow the image to emphasize the main object, underlining the patterns for the geometric shape classification process. They also provide better flexibility and adaptation of the models since they present objects in a variety of different contexts. The Figure 8 compares images from the ImageNet and ObjectNet datasets, presenting the background control, rotations and object viewpoints.

Since the ObjectNet dataset was initially labeled for object classification task by its authors, just close object classes to six main three-dimensional geometric shapes were selected, which were: Cylinder, Cone, Sphere, Parallelepiped, Undefined, and Surface. The choice of these regular three-dimensional geometric shapes is justified by the fact that are strongly used in Brazilian basic education. The object labels were modified to geometric shape labels for each original ObjectNet image-related class. This allowed the dataset construction intended for TGSR using ObjectNet images.

In this sense, images from 313 classes of real objects from ObjectNet were transformed into the six classes of three-dimensional geometric shapes mentioned above through the reclassification process. Considering this, 4409 images were selected and distributed among the classes as follows: 1001 images for cylinder; 145 for cone; 496 for sphere; 895 for parallelepiped; 1125 as undefined; and 747 for surface. Finally, the samples were divided for the training, validation and testing stages, keeping 80% of the images for training, 10% for validation and 10% for testing the AI.

The reason for the unbalanced data is mainly due to the low amount of object images similar to specific geometric shapes in the ObjectNet dataset, such as cones and spheres. Posteriorly, the datasets were preprocessed through the composition of transformations. The random horizontal flipping, random resized crop and normalization techniques were

Figure 8. ObjectNet dataset background control, rotations and viewpoints



Source: Barbu et al. (2019)

performed for the training set. In the case of validation and test sets, the images were resized to maintain the same size as the training set and were normalized. This provided images of a fixed size of 224 x 224 pixels, along with the preprocessing of all images for the subsequent training of TGSR models.

4.3 Model Training

The models were trained through the Keras¹ and Huggingface² and PyTorch³ libraries, which have different methods for a variety of DL tasks, including Computer Vision. The CNN, ViT, and ResNet architectures were selected focusing on validating the capabilities of different state-of-the-art model architectures for the TGSR task. The CNN, MobileNet, ViT, BEiT, ResNet and ResNeXt were selected from the libraries. The selection of these models was due to reasonable performance observed in related works (FENG

¹ <https://keras.io>

² <https://huggingface.co>

³ <https://pytorch.org>

et al., 2018; GUPTA; KHAN, 2022; AN et al., 2018; BALAGOPAL et al., 2018; DIRIK; YANARDAG, 2022; LIU; CAI; LEE, 2022). In this sense, this work aims to evaluate performances from the lowest to highest complexity models. Thus, a brief comparison between the architectures will be presented below.

In the DL field, the CNNs (FUKUSHIMA, 1980) are notorious for their performance in image classification. These models are used to extract features and perform pattern recognition through data, including low computational cost. Due to their low complexity, CNNs can learn patterns with a good representation of raw pixels internally without exhausting processing (BHATT et al., 2021).

The MobileNet was selected for being one of the traditional CNN successors, which were originally designed for mobile devices and embedded computer vision applications integration (HOWARD et al., 2017). Additionally, this model performs depthwise separable convolutions to produce lightweight deep neural networks with low response time for mobile device applications.

An advanced CNN-model category is the ResNet architecture, which has been used in human face recognition challenges considering different poses and lighting conditions (AN et al., 2018). In this architecture, the ResNet is underlined, which was introduced as a residual learning-based model (HE et al., 2016). Its main contrast concerning the other models is the ability to facilitate the deep neural networks training by increasing the number of layers while keeping low computational cost and satisfactory performance (KOONCE; KOONCE, 2021).

The ResNeXt model was developed posterior to the ResNet architecture, whose difference is the presence of repeated ResNet architecture building blocks, which gather a set of transformations involving the same topology (XIE et al., 2017). When compared to ResNet, this model has the cardinality dimension representing the size of the set of transformations in addition to the default depth and width model dimensions.

The last and most complex selected architecture was ViT, which has also shown promising performance results for image classification tasks. The ViT model was the first proposed ImageNet⁴ pre-trained model, involving the Transformer architecture for the task (CARON et al., 2021). When trained, this model was able to present comparable results to state-of-the-art CNNs.

Posterior to the ViT model, the Microsoft company built the BEiT model, which is a state-of-the-art ViT model that differs by presenting the pre-training process through semi-supervised learning (BAO et al., 2021). During pre-training, images have randomly masked regions and the goal of the algorithm is to recover the original image from corrupted images while continually recognizing the image patterns.

⁴ <https://www.image-net.org>

The computational modeling was done using Python 3.6 programming language and the HuggingFace, Keras and PyTorch libraries. The experimental setup involved using a computer with an Intel(R) Core(TM) i3-10100F CPU 3.60 GHz, NVIDIA GeForce RTX 2060 12GB VRAM, 16GB RAM DDR4, 1TB HD. Finally, the results were collected during a week of training and processing.

For the experiments, the selected models the following hyperparameters were found as better among the models: learning rate of 10^{-3} value; a total of 100 epochs; and the AdamW optimizer. The remaining model-specific settings were kept as the defaults defined by each used library. This hyperparameter setup was chosen to focus on maintaining low processing costs when considering the comparison to deep hyperparameter tuning. This choice ensured the best performances and shortest training for TGSR models.

4.4 Model Evaluation

The AI model selection for integration into the Geometa application requires careful evaluations based on ML, time and dimension performance measures. This is due to users' experience concerning the application, considering the TGSR task precision, time cost and dimension requirements on mobile devices.

In this context, four ML performance measures were defined for the models' evaluation, which are Precision, Accuracy, Recall and F1-Score. Precision was defined as the main performance measure in this evaluation, due to the assessment of models' capabilities in distinguishing the different three-dimensional geometric shape classes. The other performance measures were considered as alternatives for tiebreaker criteria.

When considering time and dimension, it was defined five inference-based and size performance measures, which are IPS, TIT, TPI, MTP and MS. The IPS was defined as the main performance measure in this evaluation, because of the assessment of the models' smoothness for a better user experience for the Geometa end-users. Similarly, the other performance measures were defined as alternatives. This approach aims to ensure the selection of a model that satisfies the highest quality in assertiveness and flexibility in deployment into the application.

The model launch in future Geometa versions will be conducted considering the AI deployment into a classification server. This process will be performed by turning up an inference API for Geometa application requests. The first goal will be to turn the application able to send object image classification requests to the server. Posteriorly, the server would be able to process the user image, recognize the three-dimensional geometric shape class and return the classification to the application. Thus, the application will be able to use the AI through real-world object images to show geometric shape concepts related to the AI classification.

5 RESULTS

This Master Thesis involves the definition of the most efficient model for the TGSR task considering two evaluations according to different performance measures categories. The purpose of these evaluations was to identify the model with the highest precision, the lowest complexity and response time for the Geometa application. In this context, the ML performance measures evaluation can indicate the highest precision model, capable of providing high rigor in class decisions and minimizing inference error. Conversely, the inference time and dimension performance measures evaluation can indicate the lower complexity and response time model for the AI integration. The following subsections will present each of these evaluations.

5.1 Machine learning performance measures evaluation

The ML performance measures evaluation involved the model performance analysis in inferring the defined geometric shapes classes, as well as overall performance across all classes. The primary goal in this stage was to identify the model with the highest precision in the performed inferences. Accordingly, other ML performance measures were assessed to provide a used models' performance overview and tiebreaker. In this context, the models' Precision was initially evaluated and presented in Table 8.

Table 8. Models' Precision performance results

Class	CNN	MobileNet	ViT	BEiT	ResNet	ResNeXt
Cylinder	0.00	0.29	0.80	0.81	0.80	0.60
Cone	0.00	0.00	0.78	0.65	1.00	0.67
Sphere	0.09	0.00	0.86	0.90	0.98	0.84
Undefined	0.24	0.33	0.74	0.80	0.76	0.66
Parallelepiped	0.16	0.24	0.72	0.75	0.71	0.69
Surface	0.21	0.00	0.67	0.81	0.77	0.53
All	0.12	0.14	0.76	0.79	0.84	0.66

Source: Author

In the Precision, ResNet yielded the best result with 84%, followed by BEiT with 79%. Except for the Cone and Sphere classes, ResNet exhibited an average 3.25% reduction when other classes were compared concerning BEiT's performance. Despite this, BEiT demonstrated 8% lower classification performance for Spheres and 35% lower for Cones. CNN, MobileNet, ResNeXt and ViT models were surpassed by ResNet in all classes, achieving an overall improvement of up to 72% in general Precision compared to CNN.

Subsequently, the Accuracy of each model was evaluated for the task and presented in Table 9.

Table 9. Models' Accuracy performance results

Class	CNN	MobileNet	ViT	BEiT	ResNet	ResNeXt
Cylinder	0.76	0.73	0.90	0.91	0.91	0.83
Cone	0.97	0.97	0.97	0.98	0.98	0.97
Sphere	0.81	0.89	0.96	0.98	0.99	0.96
Undefined	0.42	0.68	0.88	0.91	0.89	0.84
Parallelepiped	0.66	0.42	0.89	0.90	0.89	0.86
Surface	0.79	0.83	0.89	0.91	0.90	0.84
All	0.21	0.26	0.75	0.80	0.78	0.65

Source: Author

The BEiT model achieved the 80% Accuracy, followed by ResNet with 78%. Except for the Sphere class, where BEiT showed an average 1% reduction in Accuracy, and the Cylinder and Cone classes, where the same Accuracy was obtained when compared to ResNet, BEiT outperformed ResNet by an average of 1.3% across other classes. CNN, MobileNet, ResNeXt, and ViT models were surpassed by BEiT in all classes, achieving an overall improvement of up to 59% in general Accuracy compared to CNN. Subsequently, the Recall of the models for TGSR was evaluated and presented in Table 10.

Table 10. Models' Recall performance results

Class	CNN	MobileNet	ViT	BEiT	ResNet	ResNeXt
Cylinder	0.00	0.12	0.74	0.81	0.82	0.71
Cone	0.00	0.00	0.50	0.93	0.43	0.29
Sphere	0.08	0.00	0.86	0.94	0.90	0.82
Undefined	0.62	0.25	0.85	0.87	0.86	0.72
Parallelepiped	0.15	0.83	0.71	0.78	0.79	0.51
Surface	0.08	0.00	0.65	0.59	0.61	0.56
All	0.16	0.20	0.72	0.82	0.73	0.60

Source: Author

In this context, BEiT demonstrated the best performance, achieving 82% Recall, followed by ResNet with 73%. Except for the Cylinder, Parallelepiped and Surface classes, in which there was an average drop in BEiT's Recall of about 4% when compared to ResNet, MobileNet and ViT models, BEiT exhibited the best Recall for all other classes. CNN, MobileNet, ResNet, ResNeXt and ViT models were surpassed by BEiT in all classes, achieving an overall improvement of up to 66% in general Recall compared to CNN. Finally, the F1-Score of the models for the task was evaluated and shown in Table 11.

Once again, BEiT exhibited the best performance with a 79% F1-Score, followed by ResNet with 76%. Except for the Sphere class, where BEiT showed a 2% reduction in

Table 11. Models' F1-Score performance results

Class	CNN	MobileNet	ViT	BEiT	ResNet	ResNeXt
Cylinder	0.00	0.17	0.77	0.81	0.81	0.65
Cone	0.00	0.00	0.60	0.76	0.60	0.40
Sphere	0.09	0.00	0.86	0.92	0.94	0.83
Undefined	0.35	0.28	0.79	0.83	0.80	0.69
Parallelepiped	0.15	0.37	0.72	0.76	0.75	0.58
Surface	0.12	0.00	0.66	0.68	0.68	0.54
All	0.15	0.14	0.73	0.79	0.76	0.62

Source: Author

F1-Score, and the Cylinder and Surface classes, where the same F1-Score was obtained compared to ResNet, BEiT demonstrated the best F1-Score for all other classes. CNN, MobileNet, ResNeXt and ViT models were surpassed by BEiT in all classes, achieving an overall improvement of up to 64% in general F1-Score compared to CNN.

Based on the overall analyzed models' performances, BEiT outperformed ResNet in alternative measures evaluation. Compared to ResNet in alternative performance measures, BEiT showed an improvement of 2% in Accuracy, 9% in Recall, and 3% in F1-Score. Nevertheless, these performance values are not sufficient to achieve the benefits of 5% improvement in the ResNet's Precision when compared to BEiT, which can provide higher rigor in the TGSR task, ensuring more certainty in inferences for the application end-users. Moreover, except for BEiT, ResNet was able to outperform other models in all ML performance measures for the task.

The ViT, ResNeXt, MobileNet and CNN models, in that order, showed the lowest performance in all performance measures compared to the higher-performing ResNet and BEiT models. Especially for the CNN and MobileNet models, the worst results were obtained, presenting zero Precision, Recall and F1-Score values in identifying Cylinders and Cones through real-world objects. Thus, in this assessment of precision and rigor in TGSR, ResNet was selected as the most suitable model for AI integration in the application, followed by BEiT, which showed the second-best performance.

It is assumed that factors contributing to the lower performance of the other models are mainly related to the learning complexity of ObjectNet images. In Barbu et al. (2019), a performance analysis of models trained on ImageNet¹ when tested on the ObjectNet database showed a decrease in Accuracy in the range of 40 to 45% for all classes. ObjectNet is complex due to the intersection of real-world images and controls, making it challenging for models to generalize.

¹ <https://www.image-net.org>

5.2 Inference time and dimension performance measures evaluation

Since the models were evaluated using ML performance measures, they were further assessed in terms of response time and complexity for integration with the final application. In this stage, just the ResNet, ResNeXt, ViT and BEiT models were considered, since the CNN and MobileNet models showed the lowest performance for the task. In this context, the TIT in seconds for each model was initially evaluated and presented in Table 12.

Table 12. Models' time and dimension evaluated performance measures

Class	ViT	BEiT	ResNet	ResNeXt
TIT	116	123	51	55
TPI	264	279	115	126
IPS	4	4	9	8
MTP	86	86	26	25
MS	327	330	90	88

Source: Author

According to the obtained results, the ResNet model showed a TIT of 4 seconds less than ResNeXt; 65 seconds less compared to ViT; and 72 seconds less than BEiT. This performance result also indicated that the BEiT model exhibited inferior time performance compared to the other models, taking 123 seconds to conclude all inferences on the test set. Posterior to evaluating the TIT, the TPI in milliseconds was assessed for each model.

In this performance measure, the ResNet model showed an improvement of 11 milliseconds compared to ResNeXt; 149 milliseconds compared to ViT; and 164 milliseconds compared to BEiT. The BEiT model again exhibited inferior performance in the measure compared to the remaining models, taking 279 milliseconds to perform an inference. Subsequently, the evaluation of the IPS was conducted.

Conforming the obtained performances, the ResNet model surpassed the others, showing a difference of 1 IPS greater than ResNeXt; and 5 IPS higher when compared to ViT and BEiT. Among the inferior performance models, ViT and BEiT showed the lowest performances, with a value of just 4 IPS in the measure. Later, the MTP in million parameters was evaluated for each model.

Consonant to the obtained values, the ResNeXt model proved to be the most compact, surpassing the other models by a difference in MTP of 1 million less compared to ResNet; and 61 million less relative to BEiT and ViT. Among the less compact models, the ViT and BEiT models exhibited the largest dimensions, presenting 86 million parameters. Finally, the MS was analyzed for each model on disk in megabytes.

The ResNeXt model was verified as the model that occupies less disk space than the others, surpassing them by a difference of 2 megabytes less compared to ResNet; 239 megabytes less than ViT; and 242 megabytes less relative to BEiT. Among the models that

occupy more disk space, BEiT exhibited the largest dimension, presenting 330 megabytes of occupied space.

Finally, ResNet was observed as the better performance model among the evaluated models, reaching up to 9 IPS, whilst maintaining proximity in MTP and MS performance measures when compared to ResNeXt, proving to be the fastest and lightest model. In general, the ResNeXt, ViT and BEiT models in that order, obtained the lowest results in this evaluation and the ResNet model was able to present 59% improvement in TIT and TPI, 56% in IPS, 70% in MTP, and 73% in MS when compared to BEiT, the model that proved to be the least stable in all these performance measures.

Regarding Accuracy, ResNet showed the best results compared to other models, surpassing BEiT in both precision and stability in inference performance. In this context, ResNet proved to be the most viable alternative for solutions of this nature, which aim to integrate an AI model in streaming with Metaverse technologies. Due to its performance, ResNet was chosen for future implementation with the Geometa application.

6 CONCLUSION

This research is connected to the scientific paper entitled "Artificial Intelligence in Education 5.0: A Methodology for Three-dimensional Geometric Shape Classification for an Educational Tool". This manuscript has been accepted for publication at the international event IEEE Latin American Conference on Computational Intelligence (IEEE LA-CCI 2023). Although, it will just be published posterior to this master thesis presentation. The scientific paper research addresses the AI application for TGSR task through real-world object images, in which the focus is the AI and future Geometa versions integration. In this sense, the pre-acceptance for publication at the LA-CCI event highlights the relevance and contribution of the research to the academic community.

For this work, the implementation blocks were sequentially applied for the methodology execution, which provides: (i) the refinement of TGSR models, in which the CNN, MobileNet, ViT, BEiT, ResNet, and ResNeXt models were trained; (ii) the ML, time and dimension performance measures evaluation, in which ResNet and BEiT models presented the best results; and (iii) the ResNet model selection, which achieved the highest result according to the evaluation criteria through its precision and inference performance.

In general, this work involved the AI models' training for the TGSR task through real-world object images. The state-of-the-art models had their performances compared to validate the best alternatives for the task. In this sense, the ResNet model proved to be the best model for integration with Metaverse mobile applications, such as Geometa; since its performance was satisfactory for the task and can enable gamification activities in the application. This proposes more interactivity to users in the process of immersive and meaningful learning, deepening their understanding of geometry and the real world.

As limitations, there is a scarcity of image databases of objects centered in different perspectives. Due to this, the ObjectNet database was chosen for the development of the research, since it includes variations in rotation, viewpoints, backgrounds, difficulty and presence of elements. Another limitation based on this dataset was the lack of object-related images in ObjectNet to some of the defined geometric shape classes, such as Cones and Spheres, which turned the dataset unbalanced.

The future work will propose the following for Geometa project workflows:

- **Robust Geometa dataset construction**

A crucial step for advancing the AI capabilities in Geometa involves the creation of a dedicated balanced dataset containing a diverse range of object-related images to three-dimensional geometric shapes. This dataset will serve as the foundation

for training and evaluating future AI model pipelines, ensuring robustness and applicability across the Geometa educational context.

- **Preprocessing techniques exploration**

Investigate and compare different preprocessing methods to enhance the input data quality and improve the overall performance of the AI models. This includes assessing the impact of techniques such as image augmentation, normalization and noise reduction on model precision and generalization.

- **DL state-of-the-art models training and assessment**

There will be proposed new experiments considering other state-of-the-art Deep Learning models. The goal will be to identify the most effective architecture for the Geometa application and other Metaverse applications. Evaluate models based on their ability to precisely recognize three-dimensional geometric shapes from real-world images. This will consider factors such as model complexity, training time and resource requirements.

- **Pruning and optimization strategies**

Pruning and optimization techniques implementations will be developed focusing on reducing the trained models' computational cost without compromising their precision. Explore methods such as weight pruning, quantization and model compression to create more efficient and deployable AI models suitable. The goal will be to future launch lightweight and best-performance AI models into the Geometa application and other Metaverse applications.

- **AI model integration through classification service API**

Develop and deploy a classification API service that seamlessly integrates the trained AI model into the Geometa application. This involves creating an efficient and scalable infrastructure for real-time shape recognition, ensuring a smooth user experience and precision quality. This would propose the best end-user experience to accomplish the educational objectives of the Geometa application.

The intention in addressing these future activities is to assist the Geometa project in AI field advances in education, whilst enhancing the practical utility and accessibility of the application. The final goal is to contribute to the future improvement of education for Brazilian students at the basic scholar levels.

BIBLIOGRAPHY

AÇIKGÜL, K.; ŞAD, S. N. High school students' acceptance and use of mobile technology in learning mathematics. *Education and Information Technologies*, Springer, v. 26, n. 4, p. 4181–4201, 2021.

AKÇAYIR, M.; AKÇAYIR, G. Advantages and challenges associated with augmented reality for education: A systematic review of the literature. *Educational research review*, Elsevier, v. 20, p. 1–11, 2017. Directly cited on page 25.

AKMAN, E.; ÇAKIR, R. The effect of educational virtual reality game on primary school students' achievement and engagement in mathematics. *Interactive Learning Environments*, Taylor & Francis, v. 31, n. 3, p. 1467–1484, 2023. Directly cited on page 27.

AL-MALAH, D. K. A.-R.; JINAH, H. H. K.; ALRIKABI, H. T. S. Enhancement of educational services by using the internet of things applications for talent and intelligent schools. *Periodicals of Engineering and Natural Sciences*, v. 8, n. 4, p. 2358–2366, 2020. Directly cited on page 27.

ALTHANI, F.; MADGE, C.; POESIO, M. Less text, more visuals: Evaluating the onboarding phase in a gwap for nlp. In: *Proceedings of the 9th Workshop on Games and Natural Language Processing within the 13th Language Resources and Evaluation Conference*. [S.l.: s.n.], 2022. p. 17–27. Directly cited on page 26.

AN, Z.; DENG, W.; YUAN, T.; HU, J. Deep transfer network with 3d morphable models for face recognition. In: IEEE. *2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018)*. [S.l.], 2018. p. 416–422. Directly cited on page 46.

AZEVEDO, R.; BOUCHET, F.; DUFFY, M.; HARLEY, J.; TAUB, M.; TREVORS, G.; CLOUDE, E.; DEVER, D.; WIEDBUSCH, M.; WORTH, F. et al. Lessons learned and future directions of metatutor: Leveraging multichannel data to scaffold self-regulated learning with an intelligent tutoring system. *Frontiers in Psychology*, Frontiers, v. 13, p. 813632, 2022. Directly cited on page 27.

BAIDOO-ANU, D.; ANSAH, L. O. Education in the era of generative artificial intelligence (ai): Understanding the potential benefits of chatgpt in promoting teaching and learning. *Journal of AI*, İzmir Academy Association, v. 7, n. 1, p. 52–62, 2023. Directly cited on page 28.

BAKER, R. S.; D'MELLO, S. K.; RODRIGO, M. M. T.; GRAESSER, A. C. Better to be frustrated than bored: The incidence, persistence, and impact of learners' cognitive-affective states during interactions with three different computer-based learning environments. *International Journal of Human-Computer Studies*, Elsevier, v. 68, n. 4, p. 223–241, 2010. Directly cited on page 27.

BALAGOPAL, A.; KAZEMIFAR, S.; NGUYEN, D.; LIN, M.-H.; HANNAN, R.; OWRANGI, A.; JIANG, S. Fully automated organ segmentation in male pelvic ct images. *Physics in Medicine & Biology*, IOP Publishing, v. 63, n. 24, p. 245015, 2018. Directly cited on page 46.

BAO, H.; DONG, L.; PIAO, S.; WEI, F. Beit: Bert pre-training of image transformers. *arXiv preprint arXiv:2106.08254*, 2021. Directly cited 2 times on page 36 and 46.

BARBU, A.; MAYO, D.; ALVERIO, J.; LUO, W.; WANG, C.; GUTFREUND, D.; TENENBAUM, J.; KATZ, B. Objectnet: A large-scale bias-controlled dataset for pushing the limits of object recognition models. *Advances in neural information processing systems*, v. 32, 2019. Directly cited on page 44.

BHATT, D.; PATEL, C.; TALSANIA, H.; PATEL, J.; VAGHELA, R.; PANDYA, S.; MODI, K.; GHAYVAT, H. Cnn variants for computer vision: history, architecture, application, challenges and future scope. *Electronics*, MDPI, v. 10, n. 20, p. 2470, 2021. Directly cited on page 46.

BIANCHINI, B. L.; NASSER, L.; ONUCHIC, L.; IGLIORI, S. B. C. Mathematics education at university level: Contributions from brazil. In: _____. *Mathematics Education in Brazil : Panorama of Current Research*. Cham: Springer International Publishing, 2018. p. 85–101. ISBN 978-3-319-93455-6. Disponível em: <https://doi.org/10.1007/978-3-319-93455-6_5>.

BRASIL, M. d. E. Parâmetros curriculares nacionais para o ensino fundamental. *Brasília, MEC/SEF*, 1997. Directly cited on page 14.

BRASIL, M. d. E. *Relatório Brasil no Pisa 2018*. 2018. Url<https://www.gov.br/inep/pt-br/areas-de-atuacao/avaliacao-e-exames-educacionais/pisa/resultados>. Directly cited on page 14.

BUOLAMWINI, J.; GEBRU, T. Gender shades: Intersectional accuracy disparities in commercial gender classification. In: PMLR. *Conference on fairness, accountability and transparency*. [S.l.], 2018. p. 77–91. Directly cited on page 30.

CAO, Z.; SIMON, T.; WEI, S.-E.; SHEIKH, Y. Realtime multi-person 2d pose estimation using part affinity fields. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. [S.l.: s.n.], 2017. p. 7291–7299. Directly cited on page 30.

CARON, M.; TOUVRON, H.; MISRA, I.; JÉGOU, H.; MAIRAL, J.; BOJANOWSKI, P.; JOULIN, A. Emerging properties in self-supervised vision transformers. In: *Proceedings of the IEEE/CVF international conference on computer vision*. [S.l.: s.n.], 2021. p. 9650–9660. Directly cited on page 46.

CHEN, L.; CHEN, P.; LIN, Z. Artificial intelligence in education: A review. *Ieee Access*, Ieee, v. 8, p. 75264–75278, 2020. Directly cited on page 27.

CHEN, W.; YAO, L.; ZHOU, J.; DONG, H. A fast geometry figure recognition algorithm based on edge pixel point eigenvalues. In: CITESEER. *Proceedings of the Third International Symposium on Computer Science and Computational Technology (ISCSCT'10)*. [S.l.], 2010. p. 14–15. Directly cited on page 40.

COŞKUN, A. E. Conceptions of society and education paradigm in the twenty-first century. In: *Educational Theory in the 21st Century: Science, Technology, Society and Education*. [S.l.]: Springer Nature Singapore Singapore, 2022. p. 141–171. Directly cited on page 25.

COUTINHO, W. A.; ALMEIDA, V. E. d.; JATOBÁ, A. Aplicativos móveis em sala de aula: Uso e possibilidades para o ensino da matemática na eja. *ETD Educação Temática Digital*, UNICAMP, v. 23, n. 1, p. 20–43, 2021.

CRUTTWELL, G. S.; GAVRANOVIĆ, B.; GHANI, N.; WILSON, P.; ZANASI, F. Categorical foundations of gradient-based learning. In: SPRINGER INTERNATIONAL PUBLISHING CHAM. *European Symposium on Programming*. [S.l.], 2022. p. 1–28. Directly cited on page 30.

DHILLON, A.; VERMA, G. K. Convolutional neural network: a review of models, methodologies and applications to object detection. *Progress in Artificial Intelligence*, Springer, v. 9, n. 2, p. 85–112, 2020. Directly cited 2 times on page 30 and 41.

DÍAZ, J.; SALDAÑA, C.; AVILA, C. Virtual world as a resource for hybrid education. *International Journal of Emerging Technologies in Learning (iJET)*, International Journal of Emerging Technology in Learning, v. 15, n. 15, p. 94–109, 2020. Directly cited on page 26.

DIRIK, A.; YANARDAG, P. 3d-latentmapper: View agnostic single-view reconstruction of 3d shapes. *arXiv preprint arXiv:2212.02184*, 2022. Directly cited on page 46.

DOSOVITSKIY, A.; BEYER, L.; KOLESNIKOV, A.; WEISSENBORN, D.; ZHAI, X.; UNTERTHINER, T.; DEHGHANI, M.; MINDERER, M.; HEIGOLD, G.; GELLY, S. et al. An image is worth 16x16 words: Transformers for image recognition at scale. arxiv 2020. *arXiv preprint arXiv:2010.11929*, 2010. Directly cited on page 34.

FAKHOURI, A. G.; MURGO, C. S.; SISCOOTTO, R. A. Sesvr: Realidade virtual como auxiliadora no processo de ensino/aprendizagem de habilidades socioemocionais. *Revista Brasileira de Informática na Educação*, v. 30, p. 471–493, 2022. Directly cited on page 15.

FELCHER, C. D. O.; FOLMER, V. Educação 5.0: Reflexões e perspectivas para sua implementação. *Revista Tecnologias Educacionais em Rede (ReTER)*, p. e5–01, 2021. Directly cited on page 15.

FELCHER, C. D. O.; FOLMER, V. EducaÇÃo 5.0: ReflexÕes e perspectivas para sua implementaÇÃo. *Revista Tecnologias Educacionais em Rede (ReTER)*, v. 2, n. 3, p. e5/01–15, out. 2021. Disponível em: <<https://periodicos.ufsm.br/reter/article/view/67227>>. Directly cited on page 15.

FENG, Y.; ZHANG, Z.; ZHAO, X.; JI, R.; GAO, Y. Gvcnn: Group-view convolutional neural networks for 3d shape recognition. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. [S.l.: s.n.], 2018. p. 264–272. Directly cited on page 46.

FUKUSHIMA, K. Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. *Biological cybernetics*, Springer, v. 36, n. 4, p. 193–202, 1980. Directly cited on page 46.

GONÇALVES, M. J. R.; CARVALHO, A. L. M. de; SILVA, M. J. da; ARAÚJO, M. F. de; NASCIMENTO, S. B. d. S. L.; ALVES, Y. L. de O. A evolução da tecnologia na educação. *Revista Processus de Estudos de Gestão, Jurídicos e Financeiros*, v. 10, n. 37, p. 21–34, 2019. Directly cited on page 14.

- GUPTA, N.; KHAN, N. M. Efficient and scalable object localization in 3d on mobile device. *Journal of Imaging*, MDPI, v. 8, n. 7, p. 188, 2022. Directly cited on page 46.
- HE, K.; ZHANG, X.; REN, S.; SUN, J. Deep residual learning for image recognition. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. [S.l.: s.n.], 2016. p. 770–778. Directly cited 2 times on page 32 and 46.
- HERRERA, L. M.; PÉREZ, J. C.; ORDÓÑEZ, S. J. Developing spatial mathematical skills through 3d tools: augmented reality, virtual environments and 3d printing. *International Journal on Interactive Design and Manufacturing (IJIDeM)*, Springer, v. 13, p. 1385–1399, 2019. Directly cited on page 27.
- HIDAYAT, H.; SUKMAWARTI, S.; SUWANTO, S. The application of augmented reality in elementary school education. *Research, Society and Development*, v. 10, n. 3, p. e14910312823–e14910312823, 2021. Directly cited on page 26.
- HOLMES, W.; BIALIK, M.; FADEL, C. Artificial intelligence in education. In: . [S.l.]: Globethics Publications, 2023. Directly cited on page 27.
- HOWARD, A. G.; ZHU, M.; CHEN, B.; KALENICHENKO, D.; WANG, W.; WEYAND, T.; ANDREETTO, M.; ADAM, H. Mobilenets: Efficient convolutional neural networks for mobile vision applications. *arXiv preprint arXiv:1704.04861*, 2017. Directly cited 2 times on page 32 and 46.
- INEP. *Escalas de Proficiência do SAEB*. 2021. Acessado em: 10 de jan. de 2024. Disponível em: <<https://www.gov.br/inep/pt-br/centrais-de-conteudo/acervo-linha-editorial/publicacoes-institucionais/avaliacoes-e-exames-da-educacao-basica/escalas-de-proficiencia-do-saeb>>.
- INEP. *Resultados do IDEB*. 2022. Acessado em: 10 de jan. de 2024. Disponível em: <<https://www.gov.br/inep/pt-br/areas-de-atuacao/pesquisas-estatisticas-e-indicadores/ideb/resultados>>.
- JIAO, L.; ZHANG, F.; LIU, F.; YANG, S.; LI, L.; FENG, Z.; QU, R. A survey of deep learning-based object detection. *IEEE access*, IEEE, v. 7, p. 128837–128868, 2019. Directly cited on page 30.
- JIN, X.; LI, D. Rotation prediction based representative view locating framework for 3d object recognition. *Computer-Aided Design*, v. 150, p. 103279, 2022. ISSN 0010-4485. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0010448522000586>>.
- JORGE, J. A.; FONSECA, M. J. A simple approach to recognise geometric shapes interactively. In: SPRINGER. *GREC*. [S.l.], 1999. v. 99, p. 266–274. Directly cited on page 40.
- KADDOURA, S.; HUSSEINY, F. A. The rising trend of metaverse in education: challenges, opportunities, and ethical considerations. *PeerJ Computer Science*, PeerJ Inc., v. 9, p. e1252, 2023. Directly cited on page 27.
- KASHIVE, N.; POWALE, L.; KASHIVE, K. Understanding user perception toward artificial intelligence (ai) enabled e-learning. *The International Journal of Information and Learning Technology*, Emerald Publishing Limited, v. 38, n. 1, p. 1–19, 2020. Directly cited on page 16.

KOONCE, B.; KOONCE, B. Resnet 50. *Convolutional Neural Networks with Swift for Tensorflow: Image Recognition and Dataset Categorization*, Springer, p. 63–72, 2021. Directly cited on page 46.

LAMPROPOULOS, G.; KERAMOPOULOS, E.; DIAMANTARAS, K.; EVANGELIDIS, G. Augmented reality and gamification in education: A systematic literature review of research, applications, and empirical studies. *applied sciences*, MDPI, v. 12, n. 13, p. 6809, 2022. Directly cited on page 25.

LAUBENSTEIN, D.; GUTHÖHRLEIN, K.; LINDMEIER, C.; SCHEER, D.; SPONHOLZ, D. The ‘learning office’ as an approach for inclusive education in mathematics: Opportunities and challenges. In: _____. *Inclusive Mathematics Education: State-of-the-Art Research from Brazil and Germany*. Cham: Springer International Publishing, 2019. p. 107–121. ISBN 978-3-030-11518-0. Disponível em: <https://doi.org/10.1007/978-3-030-11518-0_9>.

LECUN, Y.; BENGIO, Y.; HINTON, G. Deep learning. *nature*, Nature Publishing Group UK London, v. 521, n. 7553, p. 436–444, 2015. Directly cited on page 30.

LIU, H.; CAI, M.; LEE, Y. J. Masked discrimination for self-supervised learning on point clouds. In: SPRINGER. *Computer Vision–ECCV 2022: 17th European Conference, Tel Aviv, Israel, October 23–27, 2022, Proceedings, Part II*. [S.l.], 2022. p. 657–675. Directly cited on page 46.

LUAN, H.; GECZY, P.; LAI, H.; GOBERT, J.; YANG, S. J.; OGATA, H.; BALTES, J.; GUERRA, R.; LI, P.; TSAI, C.-C. Challenges and future directions of big data and artificial intelligence in education. *Frontiers in psychology*, Frontiers Media SA, v. 11, p. 580820, 2020. Directly cited on page 26.

MELLO, C. d. M.; NETO, J. R. M. d. A.; PETRILLO, R. P. Educação 5.0: educação para o futuro. *Rio de Janeiro: Freitas Bastos*, 2021. Directly cited on page 15.

MONTERUBBIANESI, R.; TOSCO, V.; VITIELLO, F.; ORILISI, G.; FRACCASTORO, F.; PUTIGNANO, A.; ORSINI, G. Augmented, virtual and mixed reality in dentistry: A narrative review on the existing platforms and future challenges. *Applied Sciences*, v. 12, n. 2, 2022. ISSN 2076-3417. Disponível em: <<https://www.mdpi.com/2076-3417/12/2/877>>.

MORAN, J. Metodologias ativas para uma aprendizagem mais profunda. *Metodologias ativas para uma educação inovadora: uma abordagem teórico-prática*. Porto Alegre: Penso, p. 02–25, 2018. Directly cited on page 15.

MUSTAFA, B. Analyzing education based on metaverse technology. *Technium Soc. Sci. J.*, HeinOnline, v. 32, p. 278, 2022. Directly cited on page 26.

MUZAHD, A.; WAN, W.; SOHEL, F.; WU, L.; HOU, L. Curvenet: Curvature-based multitask learning deep networks for 3d object recognition. *IEEE/CAA Journal of Automatica Sinica*, IEEE, v. 8, n. 6, p. 1177–1187, 2020. Directly cited on page 41.

MYSTAKIDIS, S. *Metaverse. Encyclopedia*, 2, 486–497. 2022. Directly cited on page 15.

NASER, M.; ALAVI, A. H. Error metrics and performance fitness indicators for artificial intelligence and machine learning in engineering and sciences. *Architecture, Structures and Construction*, Springer, p. 1–19, 2021. Directly cited on page 37.

NGUYEN, A.; NGO, H. N.; HONG, Y.; DANG, B.; NGUYEN, B.-P. T. Ethical principles for artificial intelligence in education. *Education and Information Technologies*, Springer, v. 28, n. 4, p. 4221–4241, 2023. Directly cited on page 27.

OGDEN, S. S.; GUO, T. Characterizing the deep neural networks inference performance of mobile applications. *arXiv preprint arXiv:1909.04783*, 2019. Directly cited on page 38.

ORTIGÃO, M. I. R.; SANTOS, J. R. V. dos; DALTO, J. O. Assessment and mathematics education: Possibilities and challenges of brazilian research. In: _____. *Mathematics Education in Brazil : Panorama of Current Research*. Cham: Springer International Publishing, 2018. p. 171–192. ISBN 978-3-319-93455-6. Disponível em: <https://doi.org/10.1007/978-3-319-93455-6_9>.

PENTEADO, M. G.; MARCONE, R. Inclusive mathematics education in brazil. In: _____. *Inclusive Mathematics Education: State-of-the-Art Research from Brazil and Germany*. Cham: Springer International Publishing, 2019. p. 7–12. ISBN 978-3-030-11518-0. Disponível em: <https://doi.org/10.1007/978-3-030-11518-0_2>.

PREDEBON, F. T.; GRITTI, P. O que desmotiva os alunos para aprender matemática? *CONTRAPONTO: Discussões científicas e pedagógicas em Ciências, Matemática e Educação*, v. 1, n. 1, p. 79–94, 2020. Directly cited on page 14.

QI, C. R.; SU, H.; MO, K.; GUIBAS, L. J. Pointnet: Deep learning on point sets for 3d classification and segmentation. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. [S.l.: s.n.], 2017. p. 652–660.

REDMON, J.; DIVVALA, S.; GIRSHICK, R.; FARHADI, A. You only look once: Unified, real-time object detection. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. [S.l.: s.n.], 2016. p. 779–788. Directly cited on page 30.

RODRIGUES, A. O fracasso escolar e suas implicações no processo de ensino e de aprendizagem. 2017.

ŞAHİN, M.; YURDUGÜL, H. Educational data mining and learning analytics: past, present and future. *Bartın University Journal of Faculty of Education*, Bartın University, v. 9, n. 1, p. 121–131, 2020. Directly cited on page 26.

SANTOS, A. C. P.; NUNES, S. M. L.; FERREIRA, A. A. O ideb e o saeb: uma análise e interpretação dos seus resultados. *Pesquisa e Debate em Educação*, v. 12, n. 2, p. 1–e34598, 2022.

SEWAK, M.; SAHAY, S. K.; RATHORE, H. An overview of deep learning architecture of deep neural networks and autoencoders. *Journal of Computational and Theoretical Nanoscience*, American Scientific Publishers, v. 17, n. 1, p. 182–188, 2020. Directly cited on page 30.

ŞEYMA, E.; ÖZDEMİR, E. The metaverse in mathematics education: The opinions of secondary school mathematics teachers. *Journal of Educational Technology and Online Learning*, Gürhan DURAK, v. 5, n. 4, p. 1041–1060, 2022. Directly cited on page 26.

STEFANINI, M.; CORNIA, M.; BARALDI, L.; CASCIANELLI, S.; FIAMENI, G.; CUCCHIARA, R. From show to tell: A survey on deep learning-based image captioning. *IEEE transactions on pattern analysis and machine intelligence*, IEEE, v. 45, n. 1, p. 539–559, 2022. Directly cited on page 30.

WANG, C.; CHENG, M.; SOHEL, F.; BENNAMOUN, M.; LI, J. Normalnet: A voxel-based cnn for 3d object classification and retrieval. *Neurocomputing*, Elsevier, v. 323, p. 139–147, 2019.

XIE, S.; GIRSHICK, R.; DOLLÁR, P.; TU, Z.; HE, K. Aggregated residual transformations for deep neural networks. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. [S.l.: s.n.], 2017. p. 1492–1500. Directly cited 2 times on page 33 and 46.

XU, X.; TODOROVIC, S. Beam search for learning a deep convolutional neural network of 3d shapes. In: IEEE. *2016 23rd International Conference on Pattern Recognition (ICPR)*. [S.l.], 2016. p. 3506–3511.

ZALLIO, M.; CLARKSON, P. J. Designing the metaverse: A study on inclusion, diversity, equity, accessibility and safety for digital immersive environments. *Telematics and Informatics*, Elsevier, v. 75, p. 101909, 2022. Directly cited on page 27.

ZANUTTIGH, P.; MINTO, L. Deep learning for 3d shape classification from multiple depth maps. In: IEEE. *2017 IEEE International Conference on Image Processing (ICIP)*. [S.l.], 2017. p. 3615–3619.

ZHOU, Y.; ZENG, F.; QIAN, J.; HAN, X. 3d shape classification and retrieval based on polar view. *Information Sciences*, Elsevier, v. 474, p. 205–220, 2019.