International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 02 Issue: 02 February 2024
Page No: 128 - 133
https://irjaeh.com

Leveraging the Jetson Nano AI Kit for Machine Learning in Quantifying Gender Bias in Image Captioning

Gururaja H S[1], Sauravi Kulkarni[2], Padmanabha J[3]*

[1]Assistant Professor, Dept. of ISE, B.M.S. College of Engineering, Bengaluru, India.

[2]UG Student, Dept. of ETE, B.M.S. College of Engineering, Bengaluru, India.

[3]Assistant Professor, Dept. of ISE, Bangalore Institute of Technology, Bengaluru, India.

Email id: *gururajhs.ise@bmsce.ac.in[1], sauravi.te20@bmsce.ac.in[2], padmanabhaj1@gmail.com[3]*

****Corresponding Author Orcid ID:*** *https://orcid.org/0000-0002-9718-4672*

Abstract

In an era where artificial intelligence intertwines with ethical concerns, this paper delves into the practical deployment of the Jetson Nano AI kit to assess gender bias in image captioning. As the AI landscape evolves, the Jetson Nano stands out as an effective tool for environments with limited resources, leveraging its edge computing capabilities to reshape AI research. The paper examines the foundational concepts of gender bias quantification and its real-world application using the Jetson Nano, highlighting its versatility in constrained environments. Throughout the exploration, the seamless integration of the Jetson Nano into machine learning processes is detailed, shedding light on crucial optimizations and adjustments. Challenges encountered are also analyzed, offering insights for researchers undertaking similar projects. Ultimately, the research underscores not just Jetson Nano's role in edge computing but also the imperative of confronting gender bias in AI-generated image descriptions. By melding ethical considerations with edge computing, this paper paves the way for a more balanced and effective AI future.

Keywords: *AI, Machine Learning, Jetson Nano, Gender Bias, Image Captioning.*

1. Introduction

The swift progression of artificial intelligence now intersects with heightened ethical awareness, prompting a distinct intersection that warrants investigation. This study sets forth a comprehensive exploration, aiming to harness the tangible capabilities of the Jetson Nano AI kit. Central to this endeavor is the aspiration to shed light on quantifying gender bias in image captioning, thereby enriching the ongoing conversation surrounding ethical AI practices [10,11]. The research is situated within a landscape defined by the rapid advancements of AI technologies. In this context, the Jetson Nano stands out as a noteworthy and practical asset, especially within environments with limited resources. Beyond its inherent functionalities, the Jetson Nano acts as a pivotal force for driving innovation, given its compact design and edge computing prowess, presenting transformative prospects for both researchers and budding enthusiasts.

At the core of this research initiative lies the project titled 'Quantifying Gender Bias in Image Captioning', representing a harmonious blend of its objectives with the capabilities of the Jetson Nano [9,12]. This fusion of intent and technology becomes the arena where both the foundational theories of gender bias quantification and their pragmatic implementations are scrutinized [8,13]. The exploration into the seamless incorporation of the Jetson Nano within the machine learning framework unveils a promising avenue for future investigations. Its streamlined design and intuitive features mark it as an indispensable asset for both students and scholars navigating the intricacies of machine learning and deep learning. This trajectory is further illuminated by spotlighting key optimizations and essential adjustments crucial for its efficacious deployment. The research not only accentuates the Jetson Nano's capabilities in edge computing but also

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 02 Issue: 02 February 2024
Page No: 128 - 133
https://irjaeh.com

underscores the pressing need to address gender bias in AI-powered image descriptions [7]. Additionally, it beckons emerging enthusiasts and researchers in the realms of machine learning and deep learning. Armed with the Jetson Nano, there exists a formidable instrument poised to revolutionize the AI domain, aspiring for a more encompassing, just, and streamlined future.

2. Related Work

In this section, prior research is reviewed and contextualized relevant to the study, with an emphasis on topics such as societal bias in image captioning, performance evaluation of the Nvidia Jetson Nano, practical applications of the Jetson Nano in machine learning, and its role in embedding AI in DIY projects. The seminal work in paper [1] investigates the amplification of societal bias in image captioning models. The authors introduce LIC, a novel metric for studying captioning bias amplification. Their study emphasizes the importance of considering the entire context in bias evaluation and highlights potential challenges in bias mitigation efforts. The authors in paper [2] provide insights into the performance evaluation of the Nvidia Jetson Nano using a real-time machine learning application. It sheds light on practical aspects of using the Jetson Nano for machine learning tasks, which is pertinent to the research involving the Jetson Nano. The research in [3] presents a practical application of the Jetson Nano in the development of a portable sign language translation platform. It showcases the Jetson Nano's role in enabling real-world, deep learning-based applications and highlights its relevance to the exploration of practical implementations. The article in [4] offers hands-on insights into embedding AI using the Nvidia Jetson Nano for DIY projects. It underscores the Jetson Nano's compact yet potent capabilities for AI integration in real-world projects, aligning to highlight its utility in various contexts. By encapsulating and framing these interconnected studies, the research is situated within the expansive tapestry of image captioning bias, evaluations of Jetson Nano's capabilities, and its pragmatic implications in the domains of machine learning and DIY AI endeavors. Together, these scholarly works furnish invaluable perspectives, anchoring the importance and relevance of the current research.

3. Methodology

3.1 Data Collection

Following the methodology delineated in the CVPR paper [1], this study employed a segment of the Microsoft Common Objects in Context (MSCOCO) captions dataset [5]. The images chosen for experimentation were specifically selected based on binary gender annotations, as detailed in [6]. The gender annotations were bifurcated into the categories of "female" and "male". These annotations were accessible for images within the validation subset of the MSCOCO dataset. The gender-specific dataset comprised a total of 10,780 images, while the race-related dataset contained 10,969 images. To maintain an equitable distribution, the data partitioning approach mirrored that of [1]. This involved ensuring an equivalent count of images for each distinct protected attribute value, culminating in 5,966 images allocated for gender training and 662 designated for gender testing.

3.2 Evaluation Metrics

3.2.1 Bias Assessment Metrics

Consistent with the approach outlined in [1], bias in image captioning was evaluated using an extensive range of metrics. These metrics provided a multifaceted assessment of bias within the generated captions Utilizing LIC (Language-Image Consistency), in conjunction with its derivatives LICD and LICM as detailed in [1], allowed for a quantitative assessment of the alignment between language descriptions and corresponding image representations in the captions. The following are the gender bias metrics [6]:

1. **Ratio:** This metric assesses the ratio of protected attribute occurrences in the generated captions.
2. **Error:** Error measures the accuracy of gender attribute predictions.
3. **Bias Amplification (BA):** BA evaluates the degree of bias amplification in the captions.
4. **Directional Bias Amplification:**
   - DBAG (Directional Bias Amplification - Gender to Object): DBAG quantifies bias amplification from gender to object descriptions.
   - DBAO (Directional Bias Amplification - Object to Gender): DBAO assesses bias

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 02 Issue: 02 February 2024
Page No: 128 - 133
https://irjaeh.com

amplification from object descriptions to gender. These metrics collectively provide a comprehensive understanding of bias in image captioning, allowing us to analyze both the direction and magnitude of bias amplification.

3.3 Caption Generation Model Selection

Within the study, a range of caption generation models were considered, encompassing NIC (Neural Image Captioning), SAT, FC, Att2in, UpDn, Transformer, OSCAR, NIC+, and NIC+Equalizer, as previously outlined in [1]. Yet, the primary emphasis is placed on the NIC model (Neural Image Captioning) for caption generation [14]. This selection stems from NIC's commendable computational efficiency, making it particularly apt for real-time applications and scenarios with constrained computational capabilities. The deliberate choice to focus solely on the NIC model resonates with the overarching goal of resource-optimized captioning, specifically designed for integration with the Jetson Nano AI kit.

3.3.1 BERT Classifier and Implementation

For the implementation of LIC using the BERT classifier in the study, the following software and library versions were utilized, ensuring alignment with the constraints and compatibility requirements of the Nvidia Jetson Nano:

- **Python Version:** Python 3.6.9 (as it was the compatible version on the Nvidia Jetson Nano)
- **PyTorch:** Version 1.8 (utilized for neural network operations)
- **NumPy:** Version 1.19.2 (used for numerical computations)
- **Transformers:** Version 4.3.2 (applied for BERT-based natural language processing)
- **Spacy:** Version 2.4.2 (utilized for text preprocessing)
- **Scikit-learn (sklearn):** Version 0.24 (for machine learning and evaluation tasks)
- **NLTK:** Version 3.6.3 (employed for natural language processing functions)

In the study, the initial plan was to fine-tune the BERT classifier, aligning with the recommendations from the foundational paper. However, given the inherent hardware and software restrictions of the Nvidia Jetson Nano platform, an alternative approach was devised. Rather than embarking on fine-tuning, a pre-trained BERT model was leveraged to capitalize on its established linguistic comprehension capabilities. The choice to utilize a pre-trained BERT model was primarily influenced by Jetson Nano's computational and software constraints. While fine-tuning might promise enhanced optimization, the pre-trained model facilitated the attainment of significant outcomes within Jetson Nano's limited computational framework. [1] had advised an initial batch size of 64 for the BERT classifier, as illustrated in the provided bert_leakage.py code. Yet, during the actualization of the Nvidia Jetson Nano, memory limitations became evident. Efforts to execute the BERT classifier with batch sizes ranging from 64 to 24 were thwarted by memory saturation, leading to premature termination of processes. To strike a balance between memory conservation and model efficacy, the batch size was curtailed to 16. This adjustment ensured a successful execution of the Jetson Nano, affirming its adaptability to the platform's inherent resource constraints. The foundational guidance from [1] also indicated a standard training span of 20 epochs for the BERT classifier, as embedded in the bert_leakage.py code. However, real-world application on the Nvidia Jetson Nano unveiled challenges tied to the platform's hardware limitations. To circumvent potential operational hitches on the Jetson Nano, the training epochs were pruned from the prescribed 20 down to 5. This pre-emptive measure was instituted to avert potential system instabilities or memory overflows associated with an extended training regimen. Notwithstanding the truncated training cycles, the derived outcomes remained both relevant and impactful, seamlessly aligning with the Jetson Nano's resource envelope. Conclusively, the BERT classifier underwent training, calibrated to a reduced epoch count of 5, meticulously tailored to resonate with the Nvidia Jetson Nano's resource parameters.

3.4 Training Duration and Computational Efficiency

The training phase, employing the pre-trained BERT model on Nvidia’s Jetson Nano, was executed with a truncated epoch count of 5, tailored to the device's hardware limitations. The initial training iteration, encompassing 380 batches, concluded in roughly 42 minutes. This equates to an average processing rate of about 9.047 batches per minute, underscoring the

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 02 Issue: 02 February 2024
Page No: 128 - 133
https://irjaeh.com

commendable computational efficacy of the implementation on the Jetson Nano platform.

3.5 Results

Upon completing 10 iterations of the specified batches, an accuracy rate of 58.26% was ascertained. Throughout the training phase, several pivotal performance metrics were monitored:

- **Train Loss:** Post the culmination of 5 epochs, the training loss plateaued, settling around the 4.2 mark. This metric is indicative of the model's proficiency in minimizing discrepancies between predicted and actual values during the training trajectory.
- **Train Accuracy:** The NIC image captioning model demonstrated a training accuracy of 58.26% following the 5 epochs. This percentage encapsulates the model's prowess in accurately assigning and forecasting labels within the confines of the training dataset.
-

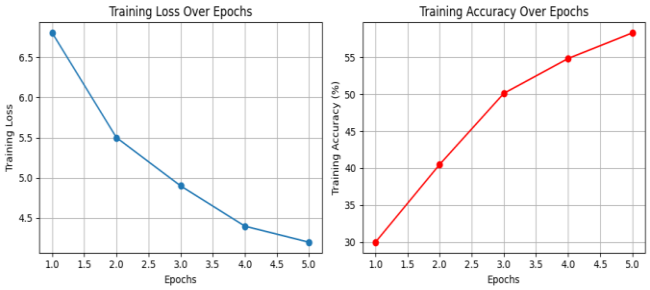

Figure 1 The NIC model's training loss and training accuracy over epochs

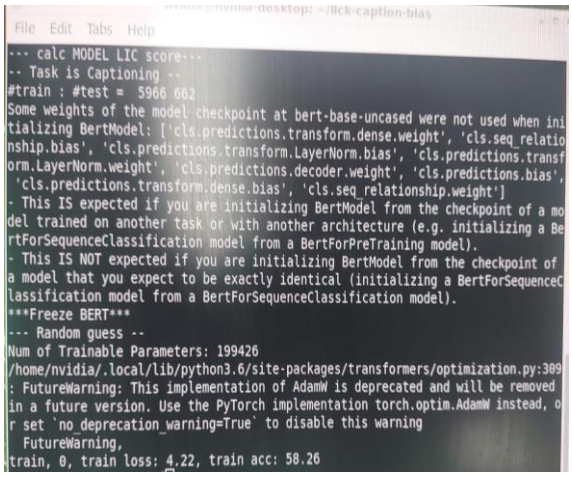

Figure 2 The NIC model's train accuracy on Jetson Nano after 10 iterations of all the batches

The results in Figure 1 and Figure 2 emphasize the viability of utilizing pre-trained BERT models for actual deployment on devices with limited resources. This showcases promising prospects for executing real-time and streamlined natural language processing operations.

3.6 Ethical Considerations

The paramount significance of ethical considerations in AI research, especially in addressing bias and fairness within machine learning models, is duly recognized.

- **Data Collection:** The dataset utilized in this study was procured and meticulously curated with a steadfast commitment to upholding privacy and consent standards. Every data source was vetted to ensure compliance with ethical protocols, and any data that could potentially be sensitive or personally identifiable underwent appropriate anonymization processes.
- **Bias Mitigation:** Throughout the model's training phase, time-honoured techniques for bias mitigation were integrated to counteract and reduce any inherent biases within the data. A concerted effort was made to steer the model's predictions and subsequent outputs towards a trajectory of fairness and impartiality.
- **Fairness:** Integral to the methodology was a thorough assessment of model fairness. This involved scrutinizing its performance across diverse demographic categories, thereby enabling the identification and subsequent rectification of any discernible discrepancies in predictions.

4. Summary

4.1 Replication of CVPR Methodology

The research methodology closely aligns with the framework outlined in [1]. Guided by the insights from [1], the foundational principles and experimental setup were adopted as the cornerstone for evaluating bias in image captioning.

4.2 Dataset and Model Consistency

To ensure the study's consistency and relevance, the identical image and captioning datasets from the CVPR paper were selected. Specifically, the Microsoft Common Objects in Context (MSCOCO) captions dataset, previously used to evaluate societal bias escalation in image captioning, was employed. This

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 02 Issue: 02 February 2024
Page No: 128 - 133
https://irjaeh.com

continuity with the dataset aimed to align with the methodology of the CVPR paper, facilitating direct comparisons and building upon the groundwork established by prior studies.

4.3 Model Selection

Adhering to the CVPR's approach, the study centered on the Neural Image Captioning (NIC) model for caption generation [15-17]. The choice to utilize NIC resonates with the efficiency demands of real-time applications and settings constrained by computational resources, thereby directly correlating with the focus on the Nvidia Jetson Nano within this research.

4.4 Adaptations for Resource Constraints

In the effort to replicate the CVPR methodology, challenges emerged stemming from the hardware limitations of the Nvidia Jetson Nano platform. To address these constraints, necessary adjustments were made, including refining the batch size and reducing the number of training epochs, ensuring alignment with the resource constraints of the Jetson Nano [18].

Conclusion

In a period characterized by the intricate interplay of artificial intelligence and ethical nuances, this research has embarked on a transformative journey, spotlighting the Jetson Nano AI kit's unparalleled potential. Central to this exploration was the endeavor to quantify gender bias in image captioning, thereby enriching the broader dialogue on ethical AI considerations. The AI domain, with its ever-expanding repertoire of tools and methodologies, has witnessed the emergence of numerous innovations. Notably, the Jetson Nano distinguishes itself as a pivotal instrument, specially tailored for environments constrained by resources. Beyond its tangible attributes, its form factor and intrinsic edge computing prowess position it as a catalyst, beckoning students and researchers to navigate the intricacies of machine learning and deep learning with newfound simplicity. At the nexus of this research lies the initiative, 'Quantifying Gender Bias in Image Captioning.' This endeavor epitomizes the harmonious fusion of overarching objectives with the Jetson Nano's inherent capabilities. As the intricacies of integrating the Jetson Nano into the machine learning paradigm were unveiled, a beacon was illuminated, paving the way for others to embark upon similar odysseys. Throughout this expedition, the iterative process of optimization was punctuated with insights and challenges, all shared transparently, serving as guiding lights for peers and fellow researchers. In summation, this paper encapsulates two pivotal narratives. Primarily, it accentuates the Jetson Nano's prowess in edge computing, exemplifying its resilience and efficacy within constrained contexts [19]. Secondly, and of profound significance, it accentuates the imperative of confronting and mitigating gender bias in AI-enabled image captioning. By seamlessly intertwining the domains of ethical AI discourse with edge computing realities, this work resonates as a clarion call, envisioning a horizon characterized by both equity and operational excellence. As the narrative draws to a close, a spirited invitation echoes, calling upon emerging enthusiasts and established researchers within the vast domains of machine learning and deep learning. With the Jetson Nano in hand, the aspiration to democratize AI emerges, imagining a future where artificial intelligence transcends elitism to become an omnipresent catalyst, championing inclusivity, equity, and groundbreaking innovation. With this shared vision, the journey ahead is one towards a radiant horizon brimming with possibilities and aspirations.

References

[1]. Yusuke Hirota, Yuta Nakashima, Noa Garcia, "Model-Agnostic Gender Debiased Image Captioning", 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp.15191-15200, 2023.

[2]. Valladares, S., Toscano, M., Tufiño, R., Morillo, P., Vallejo-Huanga, D., “Performance Evaluation of the Nvidia Jetson Nano Through a Real-Time Machine Learning Application”, Intelligent Human Systems Integration (IHSI 2021), Springer, vol 1322, 2021.

[3]. Zhenxing Zhou, Yisiang Neo, King-Shan Lui, Vincent W.L. Tam, Edmund Y. Lam, and Ngai Wong, “A Portable Hong Kong Sign Language Translation Platform with Deep Learning and Jetson Nano”, In Proceedings of the 22nd International ACM SIGACCESS Conference on Computers and Accessibility (ASSETS '20), Association for Computing Machinery, New York, NY, USA, Article 89, 1–4, 2020.

[4]. Cass, Stephen, “Nvidia makes it easy to embed AI:

The Jetson nano packs a lot of machine-learning power into DIY projects - [Hands on]", IEEE Spectrum, 57, 14-16, 2020.

[5]. Xinlei Chen, Hao Fang, Tsung-Yi Lin, Ramakrishna Vedantam, Saurabh Gupta, Piotr Dollár, et al., "Microsoft COCO captions: Data collection and evaluation server", arXiv preprint, 2015.

[6]. Dora Zhao, Angelina Wang, and Olga Russakovsky, "Understanding and evaluating racial biases in image captioning", ICCV, 2021.

[7]. Mihaela Dobreva, Tea Rukavina, Vivian Stamou, Anastasia Nefeli Vidaki, Lida Zacharopoulou, "A Multimodal Installation Exploring Gender Bias in Artificial Intelligence", Universal Access in Human-Computer Interaction, vol.14020, pp.27, 2023.

[8]. Shengyu Jia, Tao Meng, Jieyu Zhao, and Kai-Wei Chang, "Mitigating gender bias amplification in distribution by posterior regularization", ACL, 2020.

[9]. Ruixiang Tang, Mengnan Du, Yuening Li, Zirui Liu, Na Zou and Xia Hu, "Mitigating gender bias in captioning systems", WWW, 2021.

[10].Tianlu Wang, Jieyu Zhao, Mark Yatskar, Kai-Wei Chang, and Vicente Ordonez, "Balanced datasets are not enough: Estimating and mitigating gender bias in deep image representations", ICCV, 2019.

[11].Jieyu Zhao, Tianlu Wang, Mark Yatskar, Vicente Ordonez, and Kai-Wei Chang, "Men also like shopping: Reducing gender bias amplification using corpus-level constraints", EMNLP, 2017.

[12].Rishabh Bhardwaj, Navonil Majumder, and Soujanya Poria, "Investigating gender bias in BERT", Cognitive Computation, 2021.

[13].Joy Buolamwini and Timnit Gebru, "Gender shades: Inter-sectional accuracy disparities in commercial gender classification", In ACM FAccT, 2018.

[14].Quanzeng You, Hailin Jin, Zhaowen Wang, Chen Fang, and Jiebo Luo, "Image captioning with semantic attention", CVPR, 2016.

[15].Peter Anderson, Xiaodong He, Chris Buehler, Damien Teney, Mark Johnson, Stephen Gould, et al., "Bottom-up and top-down attention for image captioning and visual question answering", CVPR, 2018.

[16].Pal, Kaushik, et al. "Influence of carbon blacks on butadiene rubber/high styrene rubber/natural rubber with nano-silica: morphology and wear." Materials & Design 31.3 (2010): 1156-1164.

[17].Nayak, Ganesh Ch, et al. "Novel approach for the selective dispersion of MWCNTs in the Nylon/SAN blend system." Composites Part A: Applied Science and Manufacturing 43.8 (2012): 1242-1251.

[18].Nayak, Ganesh Ch, R. Rajasekar, and Chapal Kumar Das. "Effect of SiC coated MWCNTs on the thermal and mechanical properties of PEI/LCP blend." Composites Part A: Applied Science and Manufacturing 41.11 (2010): 1662-1667.

[19].Mukherjee, M., et al. "Improvement of the properties of PC/LCP blends in the presence of carbon nanotubes." Composites Part A: Applied Science and Manufacturing 40.8 (2009): 1291-1298.