

A Comparative Analysis: Dual-Task CNN vs. Single-Task CNNs for Gender and Age Prediction in Facial Images

Paolo Speziali
paolo.speziali@studenti.unipg.it

Abstract—Lorem ipsum dolor sit amet, consectetur adipiscing elit. Suspendisse non eleifend elit. Lorem ipsum dolor sit amet, consectetur adipiscing elit. Ut ligula nulla, placerat ut porta vitae, efficitur ut ipsum. Aenean sodales lacus et mauris faucibus, at congue turpis consectetur. Mauris volutpat a velit ac commodo. In dui urna, pulvinar interdum varius at, volutpat non ipsum. Integer hendrerit convallis laoreet. Nunc in mattis diam. Donec at hendrerit risus, vel pellentesque tortor. In hendrerit malesuada elementum. Suspendisse nibh dolor, condimentum non nisi id, laoreet tincidunt mi.

I. INTRODUCTION

In the realm of computer vision and image processing, the ability to accurately predict gender and age from facial images holds significant importance across various applications that benefit from the demographic data of their users.

Biometric information can, in fact, be used in a plethora of ways, ranging from targeted commercial use [1] to intelligent non-profit campaigns [2] and even extending to Orwellian credit scoring systems [3].

A heated debate continues to unfold regarding the application and potential abuse of Machine Learning and Computer Vision technologies in the daily lives of citizens, particularly heightened since the advent of Convolutional Neural Networks (CNN). The transformative capabilities of CNNs lie in their ability to process vast amounts of data and generate remarkably accurate predictions. Notably, these networks eliminate the necessity for manual feature engineering tasks, such as feature extraction, thereby rendering the implementation and utilization of these technologies more convenient than ever before [4].

This evolution raises significant questions about privacy, ethics, and the broader societal impact of seamlessly integrating advanced algorithms into various aspects of our lives [5]. Notably, entities such as the European Union Commission and Parliament have actively addressed these concerns by formulating the AI Act [6]. This legislative initiative seeks to classify the risks associated with ML and CV technologies, especially concerning the citizens of the confederation. The primary objective is to safeguard individuals from the inappropriate use of their biometric data, acknowledging the critical need to establish regulatory frameworks that balance the advancement of technology with the protection of individual rights and privacy [7].

Given the circumstances, it is crucial to understand the design techniques and architectures upon which these mod-

els are based to utilize and implement them with increased awareness and consideration for the effects and consequences on end-users. In particular, we will undertake a comparison between a multi-task CNN and two single-task CNNs for age and perceived gender detection to assess their results and differences in task execution.

II. RELATED WORK

Citation example [?]

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Suspendisse non eleifend elit. Lorem ipsum dolor sit amet, consectetur adipiscing elit. Ut ligula nulla, placerat ut porta vitae, efficitur ut ipsum. Aenean sodales lacus et mauris faucibus, at congue turpis consectetur. Mauris volutpat a velit ac commodo. In dui urna, pulvinar interdum varius at, volutpat non ipsum. Integer hendrerit convallis laoreet. Nunc in mattis diam. Donec at hendrerit risus, vel pellentesque tortor. In hendrerit malesuada elementum. Suspendisse nibh dolor, condimentum non nisi id, laoreet tincidunt mi.

III. PROPOSED APPROACH

A. Dataset and Preprocessing

The dataset we will be utilizing is the UTKFace dataset, which consists of 23,708 aligned and cropped facial images, annotated with age, gender, and ethnicity labels. This dataset was created with the intention of covering a wide range of variations, including pose, facial expression, illumination, occlusion, resolution, and more. In our analysis, we will specifically concentrate on the first two attributes within the dataset: age and gender.

In our analysis, we will exclusively consider 70% of the dataset for our training set, with the remaining 30% designated for testing purposes. This division allows us to train our models on a substantial portion of the data while maintaining a separate, untouched set for rigorous evaluation.

The training set exhibits an age distribution depicted in the histogram shown in Fig. 1, along with a balanced gender distribution, as illustrated in Fig. 2.

We further split the training set into a 70% training set and a 30% validation set. The validation set plays a critical role in refining and making our model more robust, preventing overfitting through techniques such as early stopping.

Despite this subdivision, we have taken measures to maintain the balance in the dataset: the accompanying histogram

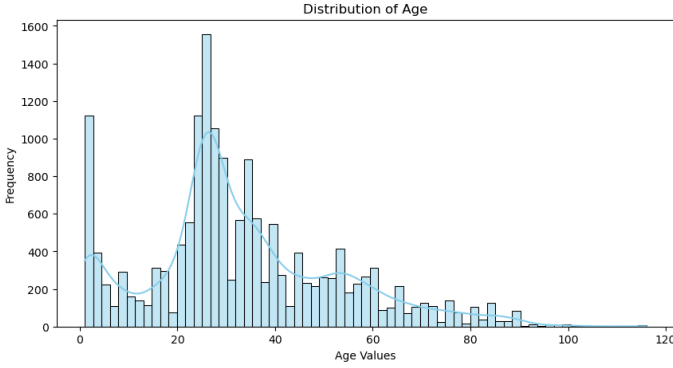


Fig. 1: Histogram of age distribution in the training set

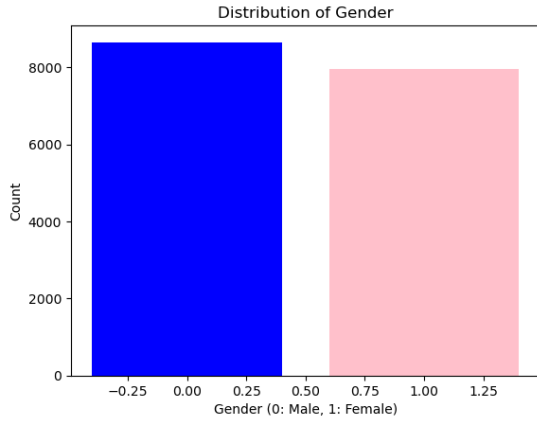


Fig. 2: Histogram of gender distribution in the training set

in illustrates that, concerning gender, the distribution remains similar in both datasets. As for age, a continuous value that poses challenges for balance assessment, we will employ the non-parametric Kolmogorov-Smirnov test to scrutinize the absence of statistically significant differences in the distributions of its values between the two datasets. This rigorous approach ensures the reliability and fairness of our evaluation process.



Fig. 3: Histogram of gender distribution in the new training set and the validation set

The outcome of the aforementioned test yields a p -value of approximately 0.38. This value is reasonably high, leading us to consider it sufficient evidence to accept the null hypothesis. Thus, we conclude that the two distributions of age in the two datasets do not exhibit statistically significant differences.

The concluding step in our dataset preprocessing involves the application of data augmentation techniques. Specifically, we expand the dataset by incorporating horizontally mirrored images and introducing random rotations of up to 10 degrees. Additionally, we employ random adjustments to the brightness, contrast, saturation, and hue of the images. This augmentation strategy has been employed based on empirical evidence suggesting that enlarging the dataset enhances the model's performance. By introducing these variations in orientation and color, we aim to expose the model to a more diverse set of examples, ultimately improving its ability to generalize and make accurate predictions on unseen data.

B. Model Architecture

As previously mentioned, in this project, we will compare two CNN architectures, one comprising a single multi-task network and another consisting of two single-task networks. The overarching goal is common between them.

Let's delve into the details of the first architecture.

1) Multi-task Network:

IV. EXPERIMENTS

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Suspendisse non eleifend elit. Lorem ipsum dolor sit amet, consectetur adipiscing elit. Ut ligula nulla, placerat ut porta vitae, efficitur ut ipsum. Aenean sodales lacus et mauris faucibus, at congue turpis consectetur. Mauris volutpat a velit ac commodo. In dui urna, pulvinar interdum varius at, volutpat non ipsum. Integer hendrerit convallis laoreet. Nunc in mattis diam. Donec at hendrerit risus, vel pellentesque tortor. In hendrerit malesuada elementum. Suspendisse nibh dolor, condimentum non nisi id, laoreet tincidunt mi.

V. CONCLUSION

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Suspendisse non eleifend elit. Lorem ipsum dolor sit amet, consectetur adipiscing elit. Ut ligula nulla, placerat ut porta vitae, efficitur ut ipsum. Aenean sodales lacus et mauris faucibus, at congue turpis consectetur. Mauris volutpat a velit ac commodo. In dui urna, pulvinar interdum varius at, volutpat non ipsum. Integer hendrerit convallis laoreet. Nunc in mattis diam. Donec at hendrerit risus, vel pellentesque tortor. In hendrerit malesuada elementum. Suspendisse nibh dolor, condimentum non nisi id, laoreet tincidunt mi.

REFERENCES

- [1] M. Fitzpatrick, "Advertising billboards use facial recognition to target shoppers," Sep 2010, accessed: 23/01/2024. [Online]. Available: <https://www.theguardian.com/media/pda/2010/sep/27/advertising-billboards-facial-recognition-japan>
- [2] L. Chen, "Gender discriminating ads," Feb 2012, accessed: 23/01/2024. [Online]. Available: <https://www.trendhunter.com/trends/facial-recognition-billboard>
- [3] N. Kobie, "The complicated truth about China's Social Credit System," Jun 2019, accessed: 23/01/2024. [Online]. Available: <https://www.wired.co.uk/article/china-social-credit-system-explained>
- [4] Datagen, "Convolutional Neural Network: Benefits, Types, and Applications," May 2023, accessed: 23/01/2024. [Online]. Available: <https://datagen.tech/guides/computer-vision/cnn-convolutional-neural-network/#>

- [5] N. Gladstone, "How facial recognition technology permeated everyday life," Sep 2018, accessed: 23/01/2024. [Online]. Available: <https://www.cigionline.org/articles/how-facial-recognition-technology-permeated-everyday-life/>
- [6] European Union, "EU Artificial Intelligence Act," accessed: 23/01/2024. [Online]. Available: <https://artificialintelligenceact.eu/the-act/>
- [7] L. Zorloni, "Abbiamo letto l'ultima versione dell'AI Act, la legge europea sull'intelligenza artificiale," Jan 2024, accessed: 23/01/2024. [Online]. Available: <https://www.wired.it/article/ai-act-testo-ultima-versione-gennaio-divieti-riconoscimento-facciale>