Privacy-Preserving AI-based Age Verification using Low Quality Facial Images

Damon Wargo, Jyh-haw Yeh

July 2024

1 Abstract

As younger generations gain easier access to the internet, the need for effective online age verification becomes increasingly critical to ensure age-restricted activities remain accessible only to those who meet legal requirements. Modern online age verification often suffers from either intrusive personal data collection or ineffective blockage. Accurate age verification requires government ID or sensitive information to be exposed, while methods that don't collect any information from users aren't suitable for preventing access. This research aims to find an online age verification method that is both accurate and respectful of user privacy using AI and machine learning. The most commonly used metric for AI-based age estimation is facial images, which have proven relatively accurate for determining exact age. This method holds promise for balancing privacy with age gates as it does not require a user's name, address, or other sensitive information—something especially important for minors. To further protect user privacy, we explore age estimation using low-quality facial images, which provide less detail for analysis by both people and computers. If results with low-quality images are shown to retain accuracy, they could pave the way for future age estimation software.

2 Introduction

In the age of the internet, the widespread availability of easily-accessible restricted or adult content has prompted the need for online age verification software. The internet has the benefit of being able to be used by people of all ages, but this comes with the risk of underage users being exposed to inappropriate content. Social media and online gaming platforms generally have ratings and content unsuitable for children of an elementary age. Similarly, websites containing adult content, such as tobacco/alcohol sales, pornography, and gambling, require users to be of a particular legal age before they are allowed access. Although age gates are often implemented into these websites, they rarely serve as an effective barrier preventing usage from ineligible users.

Further complicating the issue is the interest of privacy and security of online users. While physical settings allow for straightforward age verification methods like visual analysis, ID checks, or parental consent, online websites have fewer ways to determine information about a user without requiring them to divulge personal details such as name, location, or ID details. Given the sensitive nature of online data, users are often wary of providing information for fear of it being misused or finding itself in the wrong hands. Bearing in mind the need for verification and user safety, a balance of accuracy and user privacy is required.

This research explores the development of a machine learning method for effective, non-intrusive age verification on an online platform using low-quality facial images. Extensive work has been done on facial age estimation, and this research aims to further protect user privacy and comfort by investigating the effectiveness of the age estimation when given blurry or poor resolution images. The research trains models using facial images of varying quality to observe how model accuracy will change with respect to decreasing accuracy. In aims to provide a thorough analysis, different thresholds for verification will be calculated to account for model error and assess whether low-quality facial images are a realistic strategy for age verification. Different thresholds for verification will be calculated to account for model error and assess whether low-quality facial images are a realistic strategy for age verification.

3 Backround/Related Work

3.1 Current Age Verification

The large majority of current content restriction methods are ineffective in preventing underage audiences from accessing websites. Common verification methods include checkboxes, where a user only needs to check a box to confirm they are of a valid age, and birthdate entry, where a user enters a date, and that date is used to calculate age (Barry et al., 2021; Madison, 2023). With these methods, users can easily click a box to confirm their supposed age or enter a valid birthday via trial and error. There is no way to verify whether the user is giving legitimate information and thus access can eventually be granted to inappropriate content.

Examples of current ineffectiveness include alcohol and cannabis sales. The effectiveness of age gates in alcohol brands showed that most websites allow trial and error access, wherein 91% of the studied group had date of birth entry and 9% had checkboxes (Barry et al., 2021). Of those surveyed, only 20% blocked further attempts to access the website after an attempt was failed for the birthdate entry, allowing unlimited trials for users before granting them the rest of the website. Cannabis sales, an emerging industry with differing legality in various regions, are even less effective in preventing underage access– only 25% of 97 dispensary websites had any sort of age gate, with the rest using checkboxes or date of birth entry (Madison, 2023). These checkbox methods are found throughout the internet– whether it's part of substance sales, social media, or adult content, they aren't effective means of content restriction.

Effective age verification methods in online content are rare and often invasive. When a website wants to definitively determine a user's age, methods include scanning a user's physical ID, live face detection with ID, credit card information, and checking a user's personal information within a database (Yoti Ltd., 2024; Jumio Co., 2024). All of these methods require a user to either display government identification or provide sensitive personal/payment information– something many internet users aren't comfortable with providing for sites they are unfamiliar with (Brown, 2024).

3.2 Machine Learning and Age Estimation

Methods of machine learning driven age estimation have been studied for both their accuracy and usage in fields such as forensics and social media. With these methods, a machine learning model is trained on a set of data and used to estimate the age of an unknown individual.

When estimating a person's age, models are split into two categories– age ranges and actual age (Angulu et al., 2018). Age range models categorize new data into subjective categories (e.g., child, teen, adult) or numerical ranges (e.g., 0-10, 11-20, 21-30) and are then tested for effectiveness in sorting new data into the ranges. Age range models can be useful for general classification accuracy but suffer from non-specificity and inaccuracy on edge case ages. Actual age models assign a single numerical value to new data and are tested for how close the estimation was to the actual age associated with the data. Due to the much lower margin for error, actual age models have much fewer correct guesses but don't have the same issues with edge cases. For our purposes, we would need to use an actual age model for usage in an age gate and ability to determine the exact number of years a model is incorrect.

An issue associated with nearly all models is the difficulty in amassing a comprehensive dataset for training. An ideal dataset would have points of data for each person across their lifetime to observe changes over time, but this takes time to collect and requires consistent data collection of a massive amount of individuals during a lifetime. Furthermore, to be comprehensive, the data would need to encompass a large swath of ages, ethnicities, cultures, and backgrounds to account for the variety of factors affecting age and how it is presented across the world's population (Angulu et al., 2018). Some of these datasets do exist, but the difficulty nonetheless remains a limitation depending on the application and model used.

3.3 Methods considered for research

3.3.1 Facial Images

The most studied method of age estimation using machine learning is facial age estimation. These models are trained on datasets of pictures of human faces, often of varying sizes and quality, and then given new images to classify. Facial models using age ranges have reached a remarkable level of accuracy, both in subjective terms and actual numerical values. A study in 2023 achieved 98% classification accuracy using a dataset with the age groups labeled young, child, teenage, adult, and senior (Khalid & Al-Jibory, 2023). Numerically, a joint photo and real-time image study achieved 96.54% accuracy on a dataset with photos labeled with ten-year age ranges, continuing the trend of accurate range estimates (Kumar & Misra, 2024).

Actual age prediction with facial analysis has also proven to be quite accurate, to the point of being used commercially in select circumstances. Using a dataset of all ages, Yoti Ltd achieved a mean absolute error for predicted age of 2.7 in a range of ages from 6-70, and an impressive 1.4 for age ranges 6-12 and 13-18 (Yoti Ltd., 2023). The model was deemed accurate enough to be used for the Age Check Certification Scheme of the United Kingdom for providing certain age certifications namely 18, 21, and 25 when using a threshold for error (Yoti Ltd., 2024). The thresholds account for any errors that may occur with the model. For example, to confidently determine whether someone is at least 18 years old, the model would only let photos it estimates to be 21 or older pass.

Facial recognition provides an ideal balance between accuracy and intrusiveness. While it won't be as accurate as facial matching– determining whether a face matches an ID, for example– it requires far less information from the user and can be done with just a webcam picture. A picture is far less information to provide compared to a picture associated with a name, or a name along with an address. That said, some users might be uncomfortable with the idea of providing a high-quality picture to be used for age verification. Lower quality images can potentially solve this problem.

3.3.2 Speech Analysis

Speech analysis is the method of determining a person's age from a sample of that person's talking– often a short phrase or utterance. Studies using this method often used smaller ranges for prediction due to the difficulty of finding comprehensive datasets One study used phone conversation to train a model to identify the age ranges of 7-14, 15-24, 25-54, and 55-80 reached 72.29% accuracy (Piel & Alumäe, 2018). Narrowing the ranges to children in the grades K-3, 4-7, and 8-10 and using a combination of read and spontaneous speech in another study yielded 85.8% accuracy (Safavi et al., 2018).

More recent studies have used modern datasets in attempts to narrow results to actual age estimations and use the model to determine an age cutoff. 87.97% success rate was achieved when using 18 as a boundary for a model trained using the 2021 Common Voice Dataset, and a study using curated vowel sounds from children ages 4-15 had an RMSE for age prediction of 1.29 (Abdulmohsin et al., 2022; Novotny et al., 2023). While these results show promise for an non intrusive method, however, speech verification greatly struggles from a current lack of data. Widely available datasets are either very small or insufficiently labeled, resulting in a lack of diversity or inability to be used for exact age respectively. Future research with exactly labeled and diverse data would enable future research into uses for age verification, but it's unrealistic to be developed today.

3.3.3 Author Profiling

Author profiling is the method by which a machine learning algorithm will take a text sample and attempt to determine information about the author based on the way that the text is written. One of the pieces of information that can potentially be determined is age. Results have been produced from studies investigating social media profiles on Twitter, yielding a 66.90% accuracy of age range classification in the ranges 18-24, 25-34, 35-49, 50-64, and 65+ (Ashraf et al., 2020). When looking at individual posts, an accuracy of 75.3% was achieved.

Author profiling presents a lucrative method for age identification as users would merely need to answer questions with text, providing no personal information in the process. That said, some challenges likely disqualify it from a viable age verification method. Given that it already shows lower accuracy than other methods for identification in ranges, it's unlikely to be accurate for actual age estimation. Accurate author profiling also requires that the text analyzed be the same genre as the text that a model was trained on. When analyzing from one domain to another, e.g. from one social media site to another, the frequency of words and phrases changes and results in a model not properly trained for analysis in the foreign domain (Neto & Paraboni, 2022). A verification software would require a large amount of data collection for all ages in a specific context in order to be viable for age estimation.

4 Problem

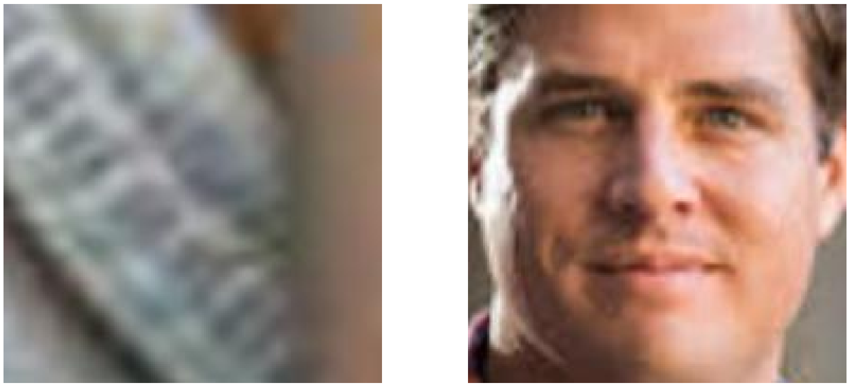
The prevalence of age-restricted on the internet and the ineffectiveness of age gates on websites, due to their reliance on easily bypassed methods like checkboxes or birthdate entries, allow for children and adolescents to easily access inappropriate content while online. While accurate age verification methods exist, they intrude on the privacy of internet users by requiring sensitive personal information or physical identification. Thus, new methods must be developed to accurately, but non intrusively, verify the age of internet users. The goal of this research is to investigate how the quality of images used for training a machine learning model affects the model's accuracy in age estimation.

5 Methods

The literature review indicated that age estimation based on facial data is the most promising method for age verification. Extensive research has demonstrated the potential for high accuracy models using facial data, and it is supported by the availability of extensive datasets suitable for training robust models.

5.1 Data

The dataset used for this research was the UTKFace dataset. The UTKFace dataset consists of approximately 23,708 cropped and adjusted facial images labeled with age, race, and gender, covering ages from 0 to 116 years (Zhang et al., 2017). Before using the dataset, a manual review was conducted to remove 120 noise images that were either incorrectly labeled or did not contain a discernible face. Examples of removed images are shown in Figure 1.

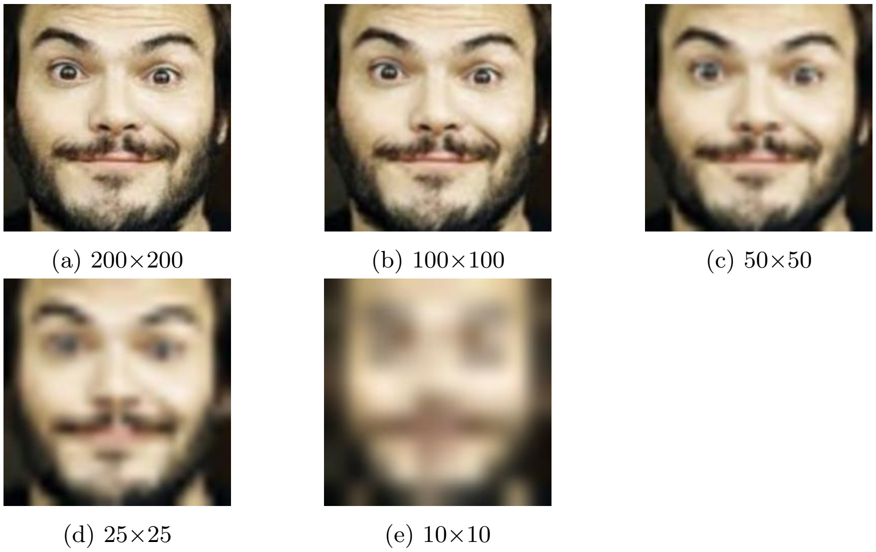


(a) Non-facial image

(b) Image labeled as two-year old

Figure 1: Examples of noise images removed

For testing purposes, only images within the age range of 1-80 were used to create a realistic baseline for the model. The final dataset comprised 23,052 images, encompassing a variety of photo conditions and including all races and genders present in the dataset. Images in the dataset had a baseline resolution of 200×200 pixels. The quality of the images was artificially lowered via downscaling and upscaling. Specifically, images were downscaled to dimensions of 100×100, 50×50, 25×25, and 10×10 pixels for each test before being upscaled back to 200×200 pixels.



(a) 200×200

(b) 100×100

(c) 50×50

(d) 25×25

(e) 10×10

Figure 2: Examples of images after downscaling and upscaling, labeled with downscaled dimensions

5.2 Models

Of the models reviewed in the literature, deep learning methods, specifically convolutional neural networks (CNNs), have been shown to produce the best results for exact age estimation (Angulu et al., 2018). Given their ease of use and customizability, this study opted to adapt a pre-trained model for image classification into a regression model for predicting age.

The model chosen for this study was EfficientNet from Keras. EfficientNet has demonstrated higher performance with fewer parameters, making it ideal for training on a limited dataset such as facial images (Tan & Le, 2020). Additionally, EfficientNet performs comparably to traditional CNNs when used for age estimation (Christopher et al., 2023). With its high performance and improved training times, EfficientNet was the ideal candidate for testing.

The code used in this experiment was adapted from an example regression model using EfficientNet to predict housing prices, authored by Markus Rosenfelder (Rosenfelder, 2020). Once adapted for use with facial images, EfficientNetB0 was used as the base model. An additional dense layer was added to increase model complexity, followed by a final regression layer to output the model's age predictions.

5.3 Training

Before training, data augmentation techniques such as rotation, flipping, and zooming were applied to the training set to increase the model's robustness and prevent overfitting. The images were resized to a consistent input size of 224x224 pixels as per EfficientNetB0 recommendations.

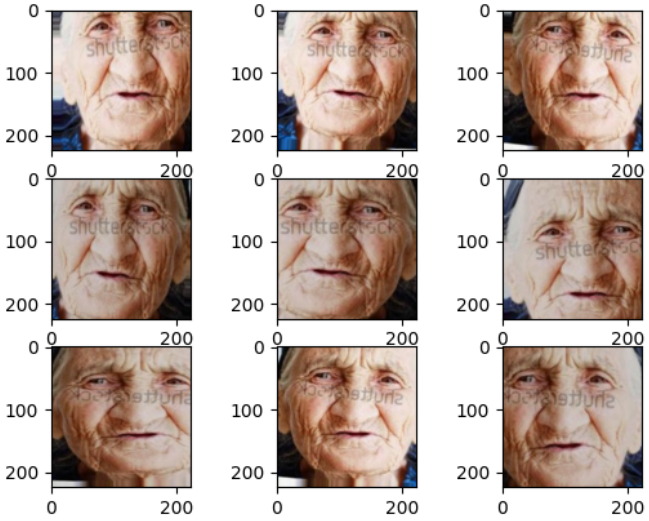


Figure 3: Example adjustments applied to images

For training and testing the model, the data was split into 80% for training, 10% for validation, and 10% for testing. The data was normalized to ensure consistent input for the model. Using MAE as the loss function, models were trained until no MAE improvement greater than 0.01 years was observed after 10 epochs. MAE and Root mean squared error (RMSE) were monitored throughout training to evaluate model performance.

5.4 Limitations

Due to the limited time available for testing, only the baseline of the model could be evaluated. Specific data subsets, such as particular age ranges or isolated genders, were not tested to observe changes in accuracy with reduced quality. Runtime constraints also greatly limited the number of tests that could be conducted. As a result, the least complex version of EfficientNet was used, leading to a low-complexity model that might not produce ideal baseline results. These limitations will be addressed in future research, as described in the respective section later in this report.

6 Results

Without using rescaling, the model achieved an MAE of 13.98 on the dataset containing the ages 1-80 after 36 epochs with an associated RMSE of 18.41. Further tests with more dramatic rescaling showed decreasing accuracy as picture detail decreased. The full results can be seen below.

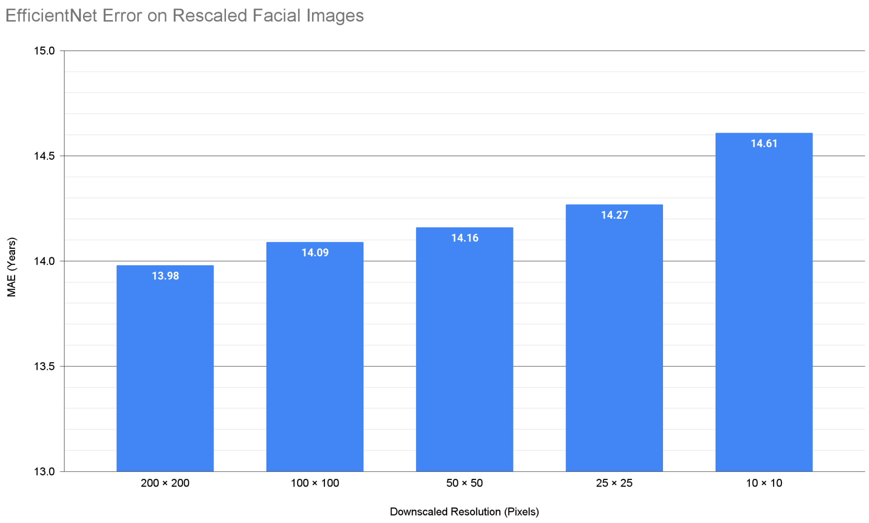


Figure 4: Model performance with different levels of image rescaling

Image Resolution	**MAE**	**RMSE**
200x200	13.98	18.41
100x100	14.09	19.07
50x50	14.18	18.88
25x25	14.27	19.06
10x10	14.61	18.98

Table 1: Results of model performance with different image resolutions. MAE and RMSE values are shown for each resolution.

The baseline results for testing were poor, with an MAE of 13.98 indicating significant errors in age estimation. This likely boils down to a model complexity issue. Given that EfficientNetB0 is the least complex version of EfficientNet, the model has a drastically reduced ability to accurately estimate facial age.

Despite the poor baseline performance, the results showed only a small increase in error as image detail decreased. This suggests that the quality of the images has little impact on model accuracy. The baseline being so poor allows for little insight into how accurate the model could be, but nonetheless provides some hope for low-quality facial images given the very small losses in accuracy as image detail worsened. If this trend continues with more complex models, low-quality images still hold strong promise for usage in age estimation.

7 Future Work

The results of this research remain inconclusive. Although the results for low-quality images show similar accuracies for age estimation, the low baseline performance indicates that the models trained are currently ineffective for reliable age verification. Future research will address this issue by employing more complex models to establish a more accurate baseline before examining how accuracy changes with different image resolutions.

Further exploration will include testing with additional datasets. The current study used only one dataset, and incorporating multiple datasets will provide a broader range of baselines and observations, which is crucial for robust model evaluation. Additionally, testing data subsets—such as estimating age for specific genders or races individually—will help determine whether accuracy variations are consistent when dataset conditions change.

8 Conclusion

This study explored the potential use of low-quality facial images for age verification with the goal of balancing accuracy and user privacy. The research showed that low-quality images hold promise for use in age verification due to small increases in error as the quality of images worsened. However, a low baseline accuracy attributed to an insufficiently complex model necessitates further testing to determine whether results can be considered conclusive. Future research involving higher complexity models and comprehensive testing of datasets will determine if low-quality images remain a possibility for privacy-preserving age verification.

References

Abdulmohsin, H. A., Stephan, J. J., Al-Khateeb, B., & Hasan, S. S. (2022). Speech age estimation using a ranking convolutional neural network. *Proceedings of International Conference on Computing and Communication Networks. Lecture Notes in Networks and Systems*, *394*, 123-130.

Angulu, R., Tapomo, J. R., & Adewumi, A. O. (2018). Age estimation via face images: A survey. *EURASIP Journal on Image and Video Processing*, *2018*(42).

Ashraf, M. A., Nawab, R. M. A., Nie, F., Pinto, D., Singh, V., & Perez, F. (2020). A study of deep learning methods for same-genre and cross-genre author profiling. *Journal of Intelligent & Fuzzy Systems*, *39*(2), 2353-2363, 2379-2389.

Barry, A. E., Primm, K., Russell, H., & Russell, A. M. (2021). Characteristics and effectiveness of alcohol website age gates preventing underage user access. *Alcohol & Alcoholism, 56*(1), 82-88.

Brown, E. N. (2024). Carding people to watch porn. *Reason*, *56*(1), 13.

Christopher, M. V., Wahid, A., Nabiilah, G. Z., & Rojali. (2023). Comparing age estimation with cnn and efficientnetv2b1. *Procedia Computer Science*, *227*, 415-421.

Jumio Co. (2024). *Online age verification system: Simple and secure.*

Khalid, Q., & Al-Jibory, F. (2023). Predicting age and gender using alexnet. *TEM Journal*, *12*, 512-518.

Kumar, B. A., & Misra, N. K. (2024). Masked face age and gender identification using caffe-modified mobilenetv2 on photo and real-time video images by transfer learning and deep learning techniques. *Expert Systems with Applications*, *246*.

Madison, M. J. (2023). You shall not pass? the design of age gates in an emerging cannabis market. *Journal of Technical Writing & Communication*, *53*(3), 240-276.

Neto, J. P. D., & Paraboni, I. (2022). Multi-source bert stack ensemble for cross-domain author profiling. *Expert Systems*, *39*, 1-19.

Novotny, M., Cmejla, R., & Tykalova, T. (2023). Automated prediction of children's age from voice acoustics. *Biomedical Signal Processing and Control*, *81*.

Piel, L. K., & Alumäe, T. (2018). Speech-based identification of children's gender and age with neural networks. *Human Language Technologies–The Baltic Perspective*, 104-111.

Rosenfelder, M. (2020). *Transfer learning with efficientnet for image regression in keras - using custom data in keras.* rosenfelder.ai.

Safavi, S., Russel, M., & Jančovič, P. (2018). Automatic speaker, age-group and gender identification from children's speech. *Computer Speech & Language*, *50*, 141-156.

Tan, M., & Le, Q. V. (2020). *Efficientnet: Rethinking model scaling for convolutional neural networks.*

Yoti Ltd. (2023). *Yoti facial age estimation - december 2023.*

Yoti Ltd. (2024). *Age verification tools for online customers and custom-built apps.*

Zhang, Zhifei, Song, Yang, & Qi, H. (2017). Age progression/regression by conditional adversarial autoencoder. In *Ieee conference on computer vision and pattern recognition (cvpr).*