



e-ISSN: 2319-8753 | p-ISSN: 2347-6710

IJRSET

International Journal of Innovative Research in
SCIENCE | ENGINEERING | TECHNOLOGY



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

Volume 13, Issue 4, April 2024

ISSN INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA

Impact Factor: 8.423

9940 572 462

6381 907 438

ijirset@gmail.com

www.ijirset.com

Image-Based Content Classification for Filtering Sensitive Images

Dr. Chethan L.S¹, Sathvik Patil², Sathwik G Hegde³, Siddharth K M⁴, Skanda⁵

Associate Professor, Department of CSE, PES Institute of Technology and Management, Shivamogga, Karnataka, India¹

U.G Student, Department of CSE, PES Institute of Technology and Management, Shivamogga, Karnataka, India²⁻⁵

ABSTRACT: The program addresses concerns that caregivers may not share their phones with children or visitors due to harmful images. To solve this problem, we developed an application that scans the photos on your phone and hides unnecessary content. We use a model that uses the Inception model to detect inappropriate images. Once the model is created, we seamlessly integrate it with the application created for the end user. The app has an intuitive user interface that allows users to easily filter unwanted images. This integration simplifies the process and allows users to easily switch between hiding and showing content based on standard image validation.

KEYWORDS: Inception model, Standard image validation.

I. INTRODUCTION

In today's digital environment, children need to interact with their parents or guardians online. The mobile phone is the source of education, entertainment and social thought. But the different roles that adults play in their work and personal lives highlight the need for children's parents or guardians to use their mobile phones. Not participating in inappropriate images poses a serious risk to children and the mentally ill. Parents who are professionals such as doctors or police often have valid reasons for holding images that may be considered inappropriate for teenagers. For example, doctors can save surgical or anatomical images as an important reference for treatment. Similarly, police officers may take photographs of crimes or accidents as part of their duties. Although these visuals are important tools for professionals, children's resistance to these elements can lead to negative consequences. Medical or legal imaging can be difficult for teens to understand and may cause anxiety or stress. Therefore, it is very important to strike a balance between appropriate and safe digital environment for children. Research, including a recent study by the American Academy of Paediatrics, shows that long-term exposure to violence or sexual abuse can lead to increased anxiety, depersonalization, and behavioural problems in children. Additionally, for emotional people, graphic images can cause emotional stress, increase anxiety, and have the potential to affect mental health. The effect is not limited to temporary discomfort, as research in the Journal of Child Depression and Psychiatry shows a strong association between early exposure to inappropriate content and chronic mental health problems. Children and people with high levels of anxiety may experience problems in emotional regulation, cognitive development, and sensitivity to stress. Our project offers the most advanced filtering capabilities to solve the problem of accidental exposure of sensitive content. This feature carefully filters images containing violence, blood and brutality and is important for the health of users. It provides fast and accurate detection of vision problems. This feature is designed for children who have access to a parent or guardian's phone, creating a safe digital space while watching different adults. Our services effectively reduce the risks associated with sensitive images, support the reconciliation of digital needs, and support the safety and health of users of all ages.

II. LITERATURE SURVEY

[1] Quan Minh Vo, Nhan Thi Cao and An Hoa Ton-That published paper "Unsafe image classification using convolutional neural network for brand safety" in the year 2020 under the Asia-Pacific Conference on Computer Science and Data Engineering. It addresses brand safety in contextual advertising, focusing on image classification to filter inappropriate websites. It introduces a contextual advertising framework involving publishers, advertisers, advertising networks, and end users. The target is to avoid unsafe websites during the advertising examination process. Recognizing the effectiveness of images in capturing user attention, the paper proposes an image multi-class

classification method using deep residual learning and network pruning. A dataset of 56,501 images in six categories supports the evaluation, revealing significant performance improvements. The paper's structure includes sections on related work, deep residual learning, network pruning, experiments, and conclusions. Noteworthy advantages are robust filtering, improved performance, and a comprehensive dataset for diverse brand safety needs. Challenges include category-specific variations and potential computational overhead.

[2] Akashkumar Gupta, Anubha Gajargaonkar, Krutika Rajput and Prof. Vidya Bharde published a paper "Real time explicit image filtering using convolutional neural networks" in the year 2022 journal. It introduces a Chrome extension employing Convolutional Neural Networks (CNN) for explicit image detection and blurring on the internet. TensorFlow and Keras are utilized for model development, integrating face and nude body detection APIs. The CNN classifier, trained on a dataset of 16,000 images, distinguishes between nude and non-nude categories. The extension, ensuring child-safe browsing, automatically blurs explicit images and disables clickable links. Advantages include user-friendly model development, seamless integration with TensorFlow and Keras, and potential for future improvements. Challenges include dependence on training data quality, hardware limitations for non-NVIDIA users, and sensitivity to evolving internet content. The research addresses a critical problem in computer vision, proposing a CNN-based model for accurate and robust detection of explicit images

III. PROBLEM STATEMENT

Safeguarding young users, our application conceals inappropriate images on mobile phones, ensuring a secure digital space for children and targeted audiences

IV. PROBLEM DESCRIPTION

Acknowledging the potential harm inflicted upon the targeted audience by exposure to inappropriate images, our project takes a proactive approach. We meticulously analyze all the images present on the user's mobile phone and discreetly hide those that could be distressing. This initiative is rooted in the commitment to creating a safer digital space, shielding users from content that may have adverse effects, particularly on minors. By automating the process of identifying and concealing such images, our project not only establishes a secure environment but also significantly saves the user's time. The automated nature of this system eliminates the need for manual intervention, providing a hassle-free experience for guardians, parents, or phone owners. Ultimately, our goal is to enhance user convenience, contribute to a protected digital landscape, and ensure the well-being of the intended audience in the face of potential harmful content.

V. OBJECTIVES

- S Creating a model using CNN (Inception model) to flag images
- Using the model to identify and filter the images
- Easy user interface

VI. EXISTING SYSTEM

Various models, including VGG11, Slim-VGG11, ResNet34, L1-ResNet34, ResNet50, and Regression - ResNet50, have been employed to train inappropriate images. Each model possesses distinct advantages and drawbacks. The variation is mainly based on speed and accuracy. Here are the outcomes for each model when trained on individual categories.

Inference model: Average inference speed over 100 runs are reported in Table I for batch size equal to 1, 32, 64, 128, and 256. Inference speed is measured in frames per second (FPS).

Table 1: Inference Speed of Different Model

Model	1	32	64	128	256
VGG11	190	486	489	491	510
Slim-VGG11	283	496	512	512	493
ResNet34	257	768	783	796	800
L1-ResNet34	253	916	931	1004	1020
ResNet50	189	389	404	408	413
Regression-ResNet50	132	414	432	439	444

It is found that the existing system of VGG19 is some existing Nudity Detection Methods that use skin filters based on color space combined with other features such as color histogram, texture analysis, and shape measures. The system attained 60% of precision on a test set composed of 138 nude images and 1401 assorted control images. The Old proposed method obtains a precision of 79.3% on a test set composed of 1,200 nude images and 1,200 assorted non-nude images. Some of the non-nude sample are images of people with no nudity. In comparison with the traditional methods, this combination can achieve 78.9% of accuracy on a test set of 400 nude images and 400 assorted images.

VII. SYSTEM DESIGN

The primary goal of the project is to train a model capable of accurately identifying inappropriate images. The Inception method is employed for the model training due to its efficient feature extraction and computational prowess. The choice of the Inception model is rooted in its streamlined feature extraction and computational efficiency. To facilitate model training, we opt for the TensorFlow model, known for its user-friendly deployment nature. A crucial prerequisite is a diverse dataset, sourced from Kaggle, encompassing various image categories such as accidents, fire, documents, and gore. Each category is meticulously collected from the Kaggle platform. To train the model the images has to have same properties for efficient training, So, we resize the image by 224x224 pixels.

Our training methodology involves initially training the model for individual categories and progressively incorporating diverse image categories to enhance its proficiency. Introducing a flatten layer and a dense layer adds adaptability to the InceptionV3 architecture, enabling it to learn task-specific patterns. The output size of the dense layer corresponds to the number of classes in the dataset, facilitating accurate predictions. Model compilation involves specifying categorical cross entropy as the loss function, optimizing with Adam, and evaluating performance based on accuracy. This configuration establishes a comprehensive framework for model training and assessment. Following the training phase, the model undergoes testing to determine its accuracy. The resultant model is seamlessly integrated into the upcoming application, functioning in the background and necessitating access to the user's image folder. Upon execution, the application autonomously evaluates each device image individually without user supervision. Every image undergoes assessment through the trained model, flagged as either appropriate or inappropriate.

Inappropriate images are discreetly hidden in a newly created folder. The hiding and unhiding process is facilitated by a toggle button within the application. Activating the toggle conceals images, while deactivation restores hidden images to their original folders. This user-friendly process underscores the effectiveness of our innovative solution.

VIII. METHODOLOGY

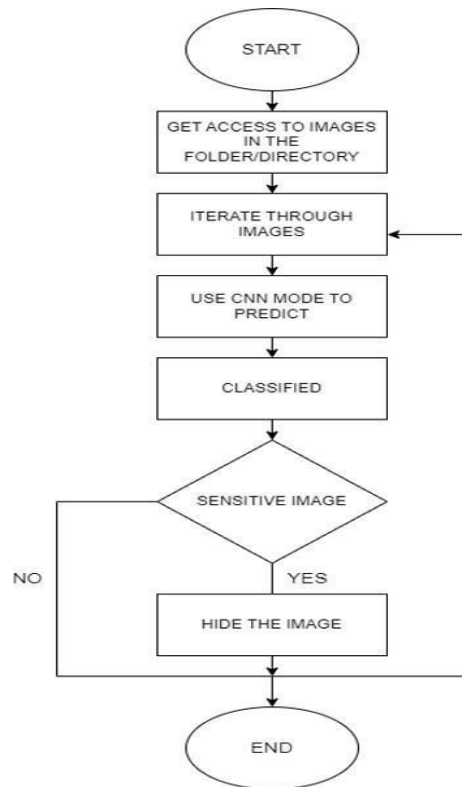


Fig 1: Methodology

Inception Model

Regarding model training, we exclusively employ the Inception model. The model undergoes training for the accident and gore image categories, ensuring a nuanced and efficient approach to image classification.

In summary, our project encompasses meticulous data collection, preprocessing, Inception model training, and user-centric design. The simplicity of the application, coupled with the toggle functionality, reflects our commitment to creating an efficient and accessible solution for managing and filtering accident and gore images.

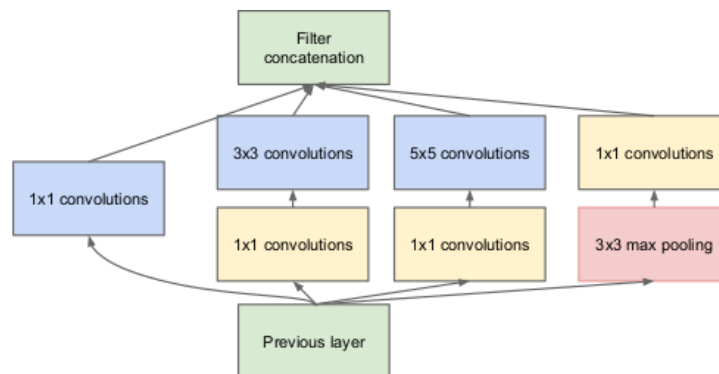


Fig 2: Inception Model V1 [11]



IX. SYSTEM IMPLEMENTATION AND RESULTS

Essential libraries like TensorFlow and Keras form the foundation for constructing and training convolutional neural networks. Random and numpy libraries offer additional functionalities, while matplotlib facilitates the visualization of training metrics. Standardizing the image size to 224x224 pixels ensures uniform input dimensions for the model, promoting consistency in feature extraction. Specifying paths for training and validation datasets streamlines subsequent data loading and preprocessing steps. Initializing the InceptionV3 model with pre-trained ImageNet weights harnesses a robust feature extractor, leveraging knowledge acquired from a diverse set of images. The exclusion of the top classification layer tailors the model for a specific classification task.

Freezing pre-trained layers retains previously learned hierarchical features, preventing them from being altered during the fine-tuning process. This approach ensures that the model builds upon existing knowledge rather than overfitting to the new dataset. Introducing a flatten layer and a dense layer adds adaptability to the InceptionV3 architecture, enabling it to learn task-specific patterns. The output size of the dense layer corresponds to the number of classes in the dataset, facilitating accurate predictions. Model compilation involves specifying categorical cross entropy as the loss function, optimizing with Adam, and evaluating performance based on accuracy. This configuration establishes a comprehensive framework for model training and assessment.

Incorporating image data augmentation into the training set introduces variability through shearing, zooming, and horizontal flipping. This diversification enhances the model's ability to generalize to unseen data, contributing to improved overall performance. The training and testing data generators play a crucial role in efficiently handling large datasets, ensuring a continuous flow of augmented images to the model during training and evaluation. The batch size and target size parameters balance computational efficiency and model accuracy. Model training occurs through multiple epochs, allowing it to learn patterns and adjust weights iteratively. The steps per epoch and validation steps parameters dictate the frequency of weight updates and validation assessments during the training process. Plotting training and validation metrics provides a visual representation of the model's learning progress. These plots serve as valuable insights into the model's convergence and potential areas for improvement. Saving the trained model to 'imfilter.h5' preserves the model's architecture and learned weights, facilitating easy deployment and further analysis without the need for retraining. The generated predictions are valuable for assessing the model's performance on the test set. Extracting the predicted classes allows for a detailed examination of the model's accuracy and potential areas of misclassification.

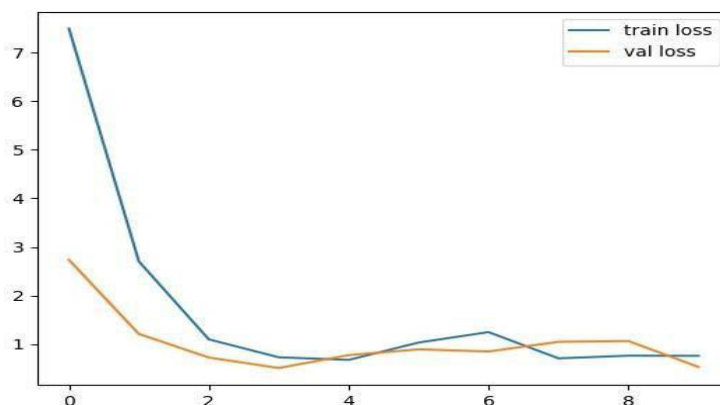


Fig 3: Loss Graph

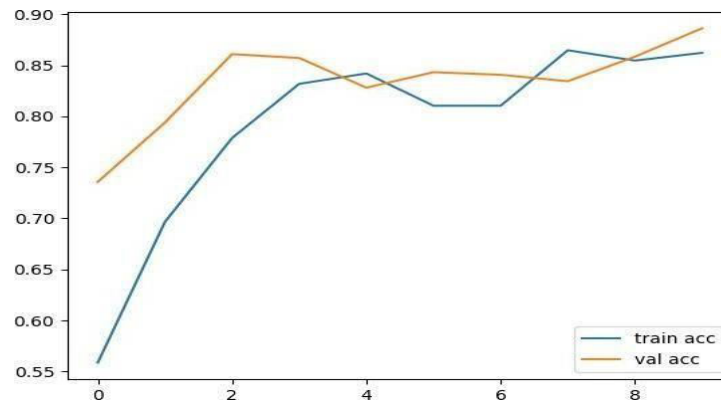


Fig 4: Accuracy Graph

X. CONCLUSION

In summary, the introduction of image filtering in the state has successfully resolved the issues related to children accessing their parent's or guardian's phone. The program reflects the importance of suitability and design to ensure a safe digital environment for young people by monitoring users health and carefully filtering out ambiguous content such as violence, blood and gore. Recognizing the different roles played by adults, particularly in industries such as medicine and law, this feature provides a safe place for children while also taking into account the need for a certain image. The way to reduce undesirable consequences is to follow research showing the negative impact of graphic content on children's health. Finally, the program supports the integration of digital needs and promotes the safety and well-being of users of all ages in today's digital environment.

REFERENCES

- [1] Quan Minh Vo, Nhan Thi Cao and An Hoa Ton-That "Unsafe image classification using convolutional neural network for brand safety", In Proceeding of 2020 IEEE Asia-Pacific Conference on Computer Science and Data Engineering (CSDE),2020
- [2] Akashkumar Gupta, Anubha Gajargaonkar, Krutika Rajput and Prof.Vidya Bharde "Real-time explicit image filtering using convolutional neural networks" In International Journal of Engineering Applied Sciences and Technology, Vol. 6, 2455-2143, Pages 185-188, 2022.
- [3] Kanwal Yousaf and Tabassam Nawaz "A Deep Learning-Based Approach for Inappropriate Content Detection and Classification of Youtube Videos", IEEE ACCESS, vol. 10, 2022.
- [4] William Larocque "Gore Classification and Censoring in Images", 2021.
- [5] Donghyun Son, Kwanghee Choi, Yongsu Baek, Seungwoo Choi, Beomjun Shin "Reliable Decision from Multiple Subtasks through Threshold Optimization: Content Moderation in the Wild",In proceeding of WSDM, 2023.
- [6] Congcong Hou¹, Xiaoyu Wu, and Ge Wang "End-to-End Bloody Video Recognition by AudioVisual Feature Fusion", In proceeding of First Chinese Conference PRCV 2018 LNCS 11256, 2018, pp. 501-510.
- [7] Min-Seok Kang, Rae-Hong Park, and Hyung-Min Park "Efficient Spatio-Temporal Modeling Methods for Real-Time Violence Recognition",In proceeding of ICT Convergence Disaster/Safety Research Institute, vol. 5, 2021
- [8] Junren Chen, Gang Liang, Wenbo He, Chun Xu, Jin Yang and Ruihang Liu "A Pornographic Images Recognition Model based on Deep One-Class Classification With Visual Attention Mechanism",IEEE ACCESS, vol.8,2020
- [9] Symeon Papadopoulos, Olga Papadopoulou, Christos Koutlis, Ioannis Sarridis "Leveraging Large-Scale Multimedia Datasets to Refine Content Moderation Models", 2022
- [10] Animesh Srivastava, Anuj Dalvi, Cyrus Britto, Harshit Rai, Kavita Shekhar "Explicit Content Detection using Faster R-CNN and SSD MobileNet v2", International Research Journal of Engineering and Technology (IRJET), vol.07, pp. 5572, 2020



- [11] Saeed Iqbal, Adnan N. Qureshi, Jianqiang Li, Tariq Mahmood “On the Analyses of Medical Images Using Traditional Machine Learning Techniques and Convolutional Neural Networks” , doi: 10.1007/s11831-023-09899-9, April 2023
- [12] X. Jin, Y. Wang, and X. Tan, “Pornographic image recognition via weighted multiple instance learning,” IEEE Transactions on Cybernetics, vol. 49, no. 12, pp. 4412-4420, 2019
- [13] X. Ou, H. Ling, H. Yu, P. Li, F. Zou, and S. Liu, “Adult image and video recognition by a deep multicontext network and fine-to-coarse strategy,” ACM Trans. Intell. Syst. Technol., vol. 8, no. 5, Jul. 2017.
- [14] S.-L. Chen, C. Yang, C. Zhu, and X.-C. Yin, “Bloody image classification with global and local features,” in Pattern Recognition, T. Tan, X. Li, X. Chen, J. Zhou, J. Yang, and H. Cheng, Eds. Singapore: Springer Singapore, 2016, pp. 379-391.
- [15] Ali Qamar Bhatti, Muhammad Umer, Syed Hasan Adil, Mansoor Ebrahim, Daniyal Nawaz, and Faizan Ahmed. "Explicit content detection system: An approach towards a safe and ethical environment", Hindawi Applied Computational Intelligence and Soft Computing, 2018.
- [16] M. M. Neumann and C. Herodotou, Young children and YouTube: A global phenomenon, Childhood Educ., vol. 96, no. 4, pp. 7277, Jul. 2020, doi: 10.1080/00094056.2020.1796459.



INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

 9940 572 462  6381 907 438  ijirset@gmail.com



www.ijirset.com

Scan to save the contact details