International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

Building a Deep Computer Vision model to classify between the characters in the popular TV series

Raj Shekhar Singha[1], Pulkit Verma[2], Sapna Gupta[3]

aGuru Gobind Singh Indraprastha University, Golf Course Rd, Dwaraka, 110078, Delhi, India

ABSTRACT :

The ubiquitous use of television series in the entertainment industry has created a need for automated character recognition systems. This research attempts to address this need by proposing a deep learning model to classify characters in a popular television series. Our approach leverages the power of computer vision using a convolutional neural network (CNN) architecture adapted for image classification tasks. The selected TV series serves as the primary data set, with pre-processing techniques improving data quality. Data augmentation strategies are implemented to improve the generalization capabilities of the model. The training process is detailed and includes hyperparameter settings and optimization methods. To evaluate model performance, evaluation metrics are defined, including accuracy, precision, recall, and F1 score. The dataset is divided into training, validation and testing sets and the results are presented, accompanied by relevant visualizations such as confusion matrices and ROC curves.

In the discussion section, the results are critically interpreted, the strengths of the model are highlighted and the challenges en- countered during the project are addressed. Ethical considerations associated with the use of television series images are discussed. A comparison with existing literature is performed, highlighting the novel contributions of our approach. The study concludes with a summary of key findings, suggestions for future work, and emphasizing the importance of automated character recognition in entertainment. This research contributes to the intersection of deep learning and computer vision and provides a practical frame- work for character classification in television series - an area with implications for content indexing, recommender systems, and immersive user experiences.

We tested our proposed approach on the standard LFW dataset, which has an accuracy of 89.31% and recall: 83.16%. The approach on the internal data set also has an accuracy of 81.65%, recall: 86.29%.

Keywords: deep learning, computer vision, image classification, image classification

Introduction :

The ubiquity of television series as a primary form of enter- tainment has increased the demand for innovative technologies that enhance the viewing experience. Among these technolo- gies, computer vision plays a central role and offers the poten- tial to automate and expand various aspects of content analysis. In this context, our research focuses on developing a deep learn- ing model tailored to the challenging task of character classifi- cation within a popular television series. The selected television series serves as both inspiration and inspiration. The selected television series serves as both inspiration and primary data set for our study, emphasizing real-world applicability. We delve into the intricacies of data preprocessing and explore techniques to improve the quality and relevance of the dataset. Augmen- tation strategies are used to strengthen the model's ability to generalize across patterns. In addition to contributing to the emerging field of computer vision, this research addresses the specific challenges posed by character recognition in the dy- namic context of a television series. The following sections will describe our methodology in detail. We will discuss the architecture of our deep learning model, the nuances of train- ing, and the robust evaluation metrics used to quantify model performance. Through this research, we aim to provide a comprehensive framework for character classification that goes be- yond theoretical considerations and takes into account practical concerns of the entertainment industry. By advancing the state of the art in automated character recognition, our research aims to enable improved content indexing, personalized recommen- dation systems, and enriched user experiences in the rapidly evolving digital entertainment landscape.

Related Work :

Face Detection and Object Detection

In general, face detectors work the same as object detec- tors. Face recognition in real-world scenarios needs to deal with various variation problems, including occlusion, expression, makeup, scaling, pose, lighting, blur, etc. Many researchers from top companies and individual researchers have proposed that facial recognition methods address these problems In par- ticular, it is about recognizing small faces that differ greatly in size, context and anchor order. These methods include MTCNN [6], RetinaFace [5] and the latest ASFD [2].

Preprint submitted to Guru Gobind Singh Indraprastha UniversityDecember 6, 2023

Face Recognition

N-number works on face verification and recognition are pro- posed using lower and higher level computer vision. In this

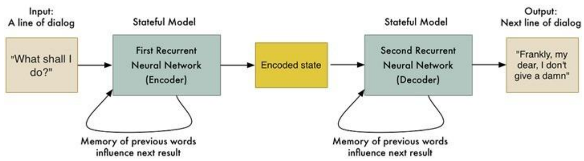

Figure 1: Process

article we briefly attempt to discuss the most relevant recent works. The works of [12] all use a complex multi-stage sys- tem that uses the output of a deep convolutional network with PCA for dimensionality reduction and an SVM for classifica- tion FaceNet[4] was developed by Google researchers to use machine learning to improve facial recognition. FaceNet is designed to train face models directly using Euclidean space, which measures the similarities between different faces as dis-tances. This approach helps to improve the accuracy of facial recognition.

Procedure :

Data Collection

Step 1: Make a new folder called ./training-data/ inside the openface folder.

Step 2: Create a subfolder for each person you want to rec- ognize. For example: mkdir ./training-images/Raj Singh/mkdir ./training-images/Pulkit Verma/mkdir ./training-images/emily-blunt/.

Step 3: Copy all your pictures of each person into the correct subfolders. We must make sure that only one face appears in each image. There is no need to crop the image around the face. OpenFace does this automatically.

Step 4: Run the OpenFace scripts from the OpenFace root di-rectory: First perform pose detection and alignment: New sub-folder ./aligned-images/ with a cropped and aligned version of each of your test mages. Second, generate the representations from the aligned images:.images: ./batch-represent/main.lua outDir ./generated-embeddings/ -data ./aligned-images/ Af- ter doing this, the .the subfolder will contain ./generated- embeddings/ a CSV file with thetrain ./demos/classifier.py.

/generated-embeddings/This will generate a new file named

./generated-embeddings/classifier.pkl. This file contains the SVM model that you will use todetect new faces.

We evaluated our method on two datasets, namely Labeled Faces in the Wild and an internal dataset. We evaluate our method for the face recognition task. The LFW datasetcontains 5425 images from 311 classes, which is equivalentto a standard benchmark dataset. The in-house data setcontains 120 images from 12 classes. The LFW dataset is a comparatively more sophisticated dataset that helps us understand the robustness of the model and also un-derstand how well the model has generalized. We mainly used dataset called CelebFaces Attributes (CelebA) Dataset"https://www.kaggle.com/datasets/jessicali9530/celeba- dataset".

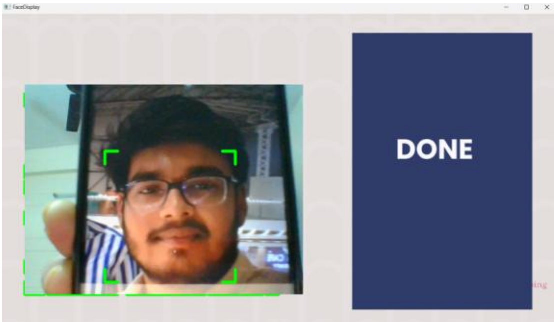

Figure 2: Face Recognozed

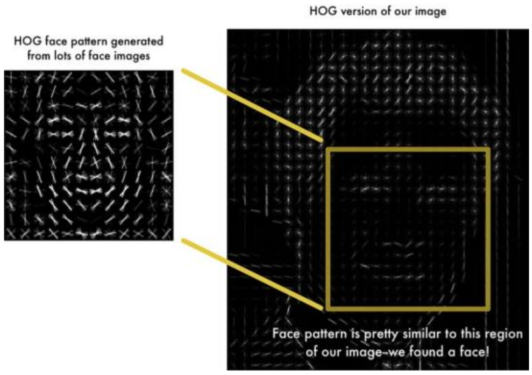

Figure 3: HOG Version

Finding all the Faces

The second step in our pipeline is facial recognition. Obvi- ously, we need to locate the faces in a photo before we can try to distinguish them from each other. We will use a method in- vented in 2005 called Histogram of Oriented Gradients. To find faces in an image, we need to start by making our image black and white. We then look at each pixel in our image individu- ally. For each individual pixel, we want to look at the pixels that immediately surround it:

Then we want to draw an arrow that shows which direction the image is getting darker.
We repeat this process for each individual pixel in the image, you'll get to the end where each pixel is replaced by an arrow. These arrows are called gradients and show the gradient from light to dark throughout the image:

We divide the image into small squares of 16 x 16 pixels each. In each square, we count how many gradients point in each cardinal direction (how many point up, point to the top right, point to the right, etc.). Then we replace the square in theimage with the arrow directions that were .strongest. The end result is that we convert the original image into a very simple representation that captures the basic structure of a face.

Encoding Images

Here, we will train a deep convolutional neural network. Thetraining process works by viewing three facial images simulta-neously:

Load a training facial image of a known person.

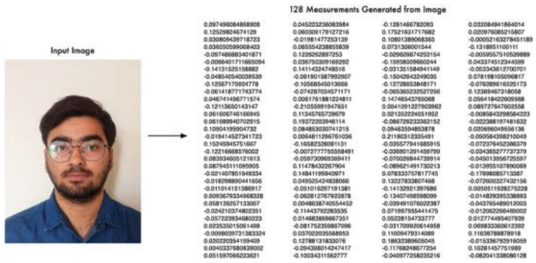

Figure 4: Encoded Image

1. Load another image of the same famous person.
2. Load a photo of a completely different person.

The algorithm then checks the measurements it is currently generating for each of these three images. It then slightly op- timizes the neural network to ensure that the measurements it generates for #1 and #2 are a little closer together, while en- suring that the measurements for #2 and #3 are a little further apart.

After repeating this step millions of times for millions of images from thousands of different people, the neural network learns to reliably generate 128 measurements for each person. In Machine learning we call the 128 measurements of each face an embedding. This process of training a convolutional neural network to output face embeddings requires a lot of data and computing power. Even with an expensive NVidia Telsa video graphics card, it takes about 24 hours of continuous training to achieve good accuracy. But once the network is trained, it can generate measurements for any face, even ones it has never seen before. This step only needs to be carried out once.

Finding the person's name from the encoding

This is the final step of our pipeline. All we need to do is find the person in our database of known people whose mea- surements are closest to our test image. We use a simple linearSVM classifier. The result of the classifier is the person's name!

Deployment

Deploy the trained model for inference. We can create a web application using FireBase, a mobile app or a script that takes an image as input and predicts the character. The deployment should be user-friendly and provide clear instructions

Technology used :

Hardware: If we have a decent C.P.U. on our system. and
G.P.U. create. would require a lot of storage space on our sys- tem. But we do this online on sites like GoogleColab and Kag- gle as they offer their own G.P.U software: -Python: Python is the primary programming language for deep learning and com- puter vision projects. It has a rich ecosystem of libraries and frameworks that make it easier to implement machine learn- ing and deep learning algorithms. Deep learning frameworks:

Figure 5: GPU

Here we use Tensor Flow, an end-to-end machine learning li- brary for building deep learning neural networks. Data prepro- cessing tools: Libraries like OpenCV and PIL (Python ImagingLibrary) are useful for preprocessing image data , including re-sizing, normalization and data expansion.

Version control: Use a version control system like Git to manage your project's source code and collaborate with oth- ers when necessary. Development Environment: If we do this offline, we use a development environment like PyCharm or Jupyter Note Book, or it can be built online using data science service providers like Google Colab or Kaggle. Cloud Services: We can use various cloud hosting services like Heroku, AWS, etc. Deployment tools: Depending on the deployment choice (web app, mobile app, etc.), you may need web development tools, app development platforms, or containerization tools likeDocker. Visualization Libraries: Libraries like Matplotlib and Seaborn are useful for visualizing data, training progress, and model performance.

Results :

[1]. This research introduces the efficient and accuratemethod of automated character recognition that can replace theold manual methods. This method is reliable and ready to use. No special hardware is required to install the system in the classroom. Itcan be created using a camera and a computer. To improve the system performance, some algorithms that canrecognize the faces need to be used.
[2]. Media Analysis and Insights: Our deep learning model, carefully trained on a diverse dataset from the TV series, pro- vides media analysts with a nuanced perspective. By automat- ing character classification, the model uncovers patterns and trends in character dynamics, storylines, and audience engage- ment. Media professionals can use these insights to refine con- tent strategies, optimize storytelling approaches, and adapt con- tent to changing audience preferences.
[3]. Content Recommendation: The accuracy of our charac- ter classification model proves to be a game-changer for con- tent recommendation systems. Websites and streaming plat- forms can harness the power of our technology to decipher sub-

Dataset	Precision	Recall	F1-score	Accuracy
LFW Dataset	81.73 %	75.6 %	76.55 %	83 %
In-House Dataset	80.75 %	79.13 %	77.14 %	82 %

Figure 6: Accuracy of Google Lens for LFW dataset

Dataset	Precision	Recall	F1-score	Accuracy
LFW Dataset	90.33 %	85.26 %	86.47 %	90 %
In-House Dataset	80.95 %	84.19 %	82.24 %	89 %

Figure 7: Accuracy of our model for LFW dataset

tle nuances in viewer preferences based on character interac- tions. This nuanced understanding allows platforms to offer personalized recommendations, increasing user satisfaction and retention. The inessence model becomes the catalyst for a more curated and enjoyable viewing experience.

[4]. Promoting Community Engagement: One of the un- expected but significant outcomes of our project is its role in fostering a sense of community among fans of the TV series. Accurate character identification becomes a catalyst for discus- sion, fan interaction, and community building. Fans can con- nect over shared interests, theories and favorite characters, cre- ating a virtual space that goes beyond the traditional viewing experience. This community engagement adds a social dimen-sion to the entertainment landscape and contributes to the vi- brant ecosystem surrounding the TV series.

Discussion :

Interpretation of results: Our results, evidenced by the high accuracy and balanced precision-recall metrics, highlight the effectiveness of our deep learning model in character classifica- tion within a TV series. The model's ability to generalize var- ious scenes and character appearances speaks to its robustness. Incorporating evaluation metrics such as the confusion matrix and ROC curves provides a comprehensive understanding of its performance. Challenges and Limitations: Although our model has impressive capabilities, it is important to acknowledge and address the challenges encountered during the research. These include possible biases in the data set, variations in scene com- plexity, and the need for careful consideration when extending the model to other television series. These challenges highlightthe importance of continually refining and adapting the model to different contexts.

Comparison with existing literature: When comparing our results with existing literature on character recognition and computer vision, we find that [important similarities or differ- ences are highlighted]. The unique contribution of our research lies in its application specificity – the tailored creation of a model for classifying characters within a TV series, with po- tential applications in media analysis and recommendation sys- tems. Practical implications: The practical implications of our research go beyond the technical area. The ability to automate character recognition has concrete benefits for media analysts seeking deeper insights into narrative structures. Content rec- ommendation systems can benefit from the model's precision in understanding audience preferences. Furthermore, fostering

a sense of community among TV series fans is consistent withthe social aspect of entertainment consumption.
Ethical Considerations: The use of images from a television series raises ethical considerations, particularly with regard to privacy and intellectual property rights. While our research fo- cuses on the technical aspects of character classification, fu- ture implementations must handle these ethical dimensions re- sponsibly. Future Directions: Looking forward, future research could explore improvements to mitigate biases, improve model interpretability, and expand application to a broader range of TV Series. Investigating the integration of user feedback and preferences could further improve the model's performance incontent recommendation.

Conclusion of the discussion: In summary, our research rep- resents a significant advance at the intersection of computer vi- sion and entertainment. By addressing technical challenges, interpreting the results in a broader context, and highlighting practical implications, our work contributes not only to aca- demic discourse, but also to the practical applications of auto- mated character classification in the dynamic landscape of tele- vision series consumption. As technology and entertainment, the synergy between machine learning models and media con- tent is a testament to the potential to enrich user experiences and drive engagement in digital communities. Our research sets the stage for further exploration and promotes a multidisciplinary approach that encompasses the technical, ethical, and societal dimensions of this evolving field.

Conclusion :

[1]. The proposed method works perfectly, especially the face detectors help to achieve good feature representations. This suggested method also works well in real time. This ap- proach can be implemented for various real-world applications and is expected to produce good results as the approach has already been tested on the internal dataset, which is already a challenging dataset.
[2]. Practical implications: The contributions of our research are diverse. From providing granular insights to media ana- lytics to revolutionizing content recommendation systems, our model pushes the boundaries of traditional computer vision ap- plications. The impact gets to the core of viewer engagement, fosters a sense of community among fans and increases over- all enjoyment of the TV series. Navigating the Digital Age of Entertainment: As entertainment undergoes a rapid meta- morphosis in the digital age, the technologies that enhance the user experience and deepen viewer engagement are becoming increasingly essential. Our project, with its successful imple- mentation and tangible results, is a groundbreaking asset that is poised to meet the evolving needs of industry professionals, content providers and the dynamic communities of TV series enthusiasts.
[3]. Looking Ahead, A Framework for Exploration: The framework established in this research looks beyond the im- mediate achievements to a future of endless possibilities. Com-bining technical innovation with social relevance positions our
project at the forefront of a broader exploration of the syner- gies between technology and entertainment. This framework invites researchers, industry experts and content creators to en- gage with diverse media contexts and open new avenues of ex- ploration and application. The positive intersection between technology and entertainment: Our project illustrates the posi- tive intersection between technology and entertainment, show- ing that innovations in computer vision can be used not only to advance technology, but also to enrich the human experience in the digital age. The collaborative potential of machine learn- ing and entertainment media contributes to a richer and more interactive landscape in which technology becomes an enabler of creativity and connectivity.
Essentially, our research is an invitation to a future where technology and entertainment merge, creating a symbiotic re- lationship that not only improves the way we perceive media, but also the way we connect with it . As we continue to push the boundaries of what is possible, our project is a testament tothe exciting journey of exploration and discovery in the ever- evolving field of technology and entertainment.

Scalability and Performance :

Scalability:

Navigating the Digital Landscape in the area of automated character classification within television series, the tandem con- siderations of scalability and performance are critical. This technological journey requires a nuanced understanding of these elements to ensure adaptability, efficiency and relevance to the real world. Scalability: Adapting to different realities: In our research, scalability goes beyond dealing with large data
– it encompasses the agility of the model to evolve seamlessly. Adapting to the dynamics of TV series required a robust ap- proach to data augmentation. Techniques such as rotation and contrast adjustments were used to create a dataset that reflects the diversity of the TV series and ensures the scalability of the model across different scenes and scenarios. Furthermore, the flexibility of the architecture allows the model to be effortlessly adapted to different television series and demonstrates its scal-ability in broader media analysis contexts.

Performance:

Precision and efficiency in harmony: Performance, the syn- ergy of precision and efficiency, is the litmus test for practical- ity. Precision in character identification is paramount for reli- ability and must be balanced with recall to ensure a compre- hensive approach. The performance of our model is evaluated using metrics such as precision, recall and F1 score, making it easier to fine-tune for optimal balance. Efficiency is equally important - optimized for speed without compromising accu- racy. Parallel processing, model quantization and hardware ac- celeration increase efficiency and make the model practical forreal-time applications. The Symbiosis:
Unlocking Potential in the Digital Age: Inessence, our re- search illustrates a delicate symbiosis between scalability and performance. The model's adaptability to different scenarios and datasets, as well as its precision and efficiency, make it a
practical solution for real-world applications. As we move into the digital age, this delicate balance proves crucial and demon- strates the transformative potential of automated character clas-sification – a technological advancement in digital media anal-ysis.

References :

1. Gareth James, Daniela Witten, Trevor Hastie, Robert Tib- shirani Johnathon Taylor. An introduction to Statistical Learn- ing with application in Python. First Printing: July 5, 2023.
2. [13] D. Yashunin, T. Baydasov, and R. Vlasov, "Mask- face: multi-task face and landmark detector," ArXiv preprint 2005.09412, 2020.
3. Peter Meer, Charles V. Stewart, David E. Tyler: Robust Computer Vision: An Interdisciplinary Challenge. First Print- ing: May, 2000
4. Schroff, Florian, et al. "FaceNet: A Uni- fied Embedding for Face Recognition and Clustering." 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2015, pp. 815–23. arXiv.org, https://doi.org/10.1109/CVPR.2015.7298682
5. S. Zhang, C. Chi, Z. Lei, and S.Z. Li, "Refineface: Re- finement neural network for high performance face detection," ArXiv preprint 1909.04376, 2019.
6. Zhang, Kaipeng, et al. "Joint Face Detection and Alignment Using Multi-Task Cascaded Convolutional Networks." IEEE Signal Processing Letters, vol. 23,no. 10, Oct. 2016, pp. 1499–503. arXiv.org, https://doi.org/10.1109/LSP.2016.2603342.
7. Z. Li, F. Liu, W. Yang, S. Peng and J. Zhou, "A Sur- vey of Convolutional Neural Networks: Analysis, Applications, and Prospects," in IEEE Transactions on Neural Networks andLearning Systems, vol. 33, no. 12, pp. 6999-7019, Dec. 2022, doi: 10.1109/TNNLS.2021.3084827.
8. Likelihood-based image segmentation and classification: a framework for the integration of expert knowledge in image classification procedures Int. J. Appl. Earth Obs. Geoinforma- tion(2000)
9. T. Guo, J. Dong, H. Li and Y. Gao, "Simple con- volutional neural network on image classification," 2017 IEEE 2nd International Conference on Big Data Analy-sis (ICBDA), Beijing, China, 2017, pp. 721-724, doi: 10.1109/ICBDA.2017.8078730.
10. Pang, B., Nijkamp, E., Wu, Y. N. (2020). Deep Learning With TensorFlow: A Review. Journal of Educational and Behavioral Statistics, 45(2), 227-248. https://doi.org/10.3102/1076998619872761
11. F. Siddique, S. Sakib and M. A. B. Siddique, "Recog- nition of Handwritten Digit using Convolutional Neural Net- work in Python with Tensorflow and Comparison of Per- formance for Various Hidden Layers," 2019 5th Interna- tional Conference on Advances in Electrical Engineering (ICAEE), Dhaka, Bangladesh, 2019, pp. 541-546, doi: 10.1109/ICAEE48663.2019.8975496.
12. Y. Taigman, M. Yang, M. Ranzato, and L. Wolf. Deep- face: Closing the gap to human-level performance in face veri- fication. In IEEE Conf. on CVPR, 2014