HIGH SCHOOL EDITION
Journal of Student Research
Volume 13 Issue 2 (2024)

Deep Learning Based Anime Character Sketch Quality Recognition

Ishita Yanamadala[1], Atul Dubey[#] and Divya Rajagiri[#]

[1]Dougherty Valley High School, USA
[#]Advisor

ABSTRACT

The explosion of the global anime market is poised to redefine the entertainment industry, boasting a valuation of USD 31.23 billion as of 2023 and projected to climb at a CAGR of 9.8% until 2030. At the heart of this growth is a vibrant community of artists and enthusiasts who navigate challenges such as artist scarcity and the lack of advanced tools for qualitative feedback and effective content promotion. Our study targets the core of anime creation sketching by evaluating the essential drawing elements and assessing their implementation in professional-quality anime portraits. We introduce an automated approach to predicting the anime quality using advances in deep learning. Utilizing transfer learning, we enhance three pre-trained models—MobileNetV2, ResNet50, and VGG16—with a customized dense layer, refining their capabilities for the binary classification of anime character sketches. The dataset employed comprises 155 images, categorized as 'Good' and 'Bad' to reflect the quality of sketches. A balanced split into training, validation, and testing subsets ensures a quantitative evaluation of data not seen previously by the model. The models, pre-trained on ImageNet and fine-tuned with our dataset, demonstrate varied sensitivity to hyperparameters, with the MobileNetV2 and ResNet50 model attaining a peak validation accuracy of 94% and the highest test accuracy of 79% indicating its potential as a robust tool for quality assessment in the anime industry. An interesting outcome of the research is that a lightweight MobileNetV2 model with much fewer parameters compared to other models resulted in the highest test accuracy.

Introduction

The global anime market has witnessed a remarkable surge in popularity, with an estimated value of USD 31.23 billion in 2023[1,2]. As projected by Grand View Research, this dynamic industry is poised for a robust growth trajectory, with a compound annual growth rate (CAGR) of 9.8% from 2024 to 2030 [1]. This market is fueled by a growing number of anime enthusiasts and a vibrant community of artists dedicated to bringing imaginative worlds to life. However, the anime industry faces significant challenges that could potentially hinder its growth. A critical issue is the scarcity of anime artists, attributed to demanding work schedules and inadequate compensation [3]. Furthermore, there is a notable gap in the availability of tools that can provide artists with constructive feedback on their work and assist them in marketing their creations more effectively. Addressing these challenges is crucial for sustaining the growth and vibrancy of the anime market.

Mastering the art of drawing and sketching is a foundational skill for any artist, especially in the realm of anime creation. The intricacies of this craft lie in understanding the various elements and components that contribute to the overall aesthetic of the artwork. A single piece of art can incorporate multiple fundamentals, depending on the style and medium used. These fundamentals play a crucial role in determining the quality and accuracy of the artwork, allowing artists to bring their sketches to life with greater realism and depth. Among the ten fundamentals for drawing and sketching, forms, anatomy, lighting, values, perspective, lines, texture,

composition, space, and color theory are critical. Each of these elements adds a unique dimension to the artwork, enhancing its visual appeal and emotional resonance [4].

Recognizing the style of anime is a complex task that extends beyond basic image recognition. Unlike traditional biometric recognition, such as face or iris recognition, anime style recognition (ASR) deals with a larger semantic gap and has received relatively less attention in the research community. Haotang Li et al. proposed a challenging ASR benchmark, which includes a large-scale dataset (LSASRD) containing 20,937 images from 190 anime works [5]. This dataset presents various challenges, such as complex illuminations, diverse poses, and exaggerated compositions, making ASR a more challenging computer vision task. The goal of ASR is to learn the abstract painting style of anime works and determine whether different images belong to the same work, thereby pushing the boundaries of semantic understanding in artificial intelligence.

Artificial intelligence (AI) is revolutionizing the creation of professional anime portraits by transforming incomplete rough sketches into high-quality images. Despite advancements in Generative Adversarial Networks (GANs), generating detailed images from partial sketches remains challenging, particularly due to the abstract nature of anime lines. Recent research has demonstrated the potential of AI in this domain, particularly through the exploration of StyleGAN's latent space and a two-stage training approach, which together can produce refined results from incomplete sketches [6].

Furthermore, the work by Tracy Hammond and colleagues at Texas A&M University and the Georgia Institute of Technology has made significant contributions to sketch-based learning and AI-assisted art creation. Their research focuses on modeling expert sketching ability, identifying key metrics that differentiate expert and novice sketches. By leveraging the SketchTivity intelligent tutoring system, they aim to provide personalized feedback to enhance students' sketching skills. This research not only advances our understanding of sketch assessment but also suggests new directions for AI-driven feedback in art education [7].

Anime production houses require good anime characters to portray a character in their story. However, Anime studios have to spend hours trying to find the right artist to draw the characters. This can be made easier by building a deep learning based software tool to access the quality of work of the artists and filter accordingly. Also, this tool can be helpful for publishing platforms in quality management by highlighting the good work and banning inappropriate content. The software that we are building will save time for anime studios, artists, and publishing platforms.

In this paper, we have presented a deep learning-based solution for the quality assessment of Anime character drawings. We have used the transfer learning technique with 3 pretrained models - MobileNetv2, RESNET50, VGG16. We have added a dense custom layer with 100 neurons to these pre-trained models to train the model for our problem.

Methods

Dataset

The dataset used for the experimentation contains 155 images. It is organized into 2 folders (Good and Bad) representing the 2 different categories to classify the quality of an Anime character sketch. Subsequently, the dataset was partitioned into training, testing, and validation sets. The training set comprises 74 'Good' and 81 'Bad' images, the testing set contains 9 'Good' and 10 'Bad' images, and the validation set consists of 8 'Good' and 9 'Bad' images. This distribution was strategically chosen to maintain a balanced representation of both classes across each subset, thereby enabling the model to learn from a diverse set of examples and ensuring its robust evaluation on unseen data.

Deep Learning Algorithms

For our experimentation, we selected three deep learning models based on their performance with the ImageNet dataset and their architectural characteristics. The deep learning models used for experimentation are described in the following paragraphs. All the architectures are pre-trained on the ImageNet database containing more than a million images.

VGG16 is a 16-layer deep convoluted neural network model. The pre-trained network is capable of classifying images into 1000 object categories, such as keyboard, mouse, pencil, and many animals. ResNet50 is a 50-layer deep convoluted neural network model. It can also classify 1000 objects present in the Imagenet dataset. MobileNetV2 is a 53-layer deep neural network model. It is based on an inverted residual structure where the residual connections are between the bottleneck layers. It is popular for its compactness and ability to run even on mobile devices.

We chose the above 3 models for experimentation based on their performance with the ImageNet dataset and their sizes as per the number of parameters and layers.

The flowchart shown in Figure 1 outlines the pipeline for the development and deployment of a deep learning model developed in this study. The models employed in this study were pre-trained on the ImageNet challenge dataset. To adapt these models for our specific task, all layers were frozen, and the prediction layer was removed. A single dense layer with 100 neurons was added to tailor the models for the binary classification of Anime character sketches.

During training, the train subset of the dataset was utilized with varying hyperparameter values, including learning rates ranging from 0.05 to 0.000001 and epochs between 10 and 50. The models in each case were evaluated on the validation dataset.

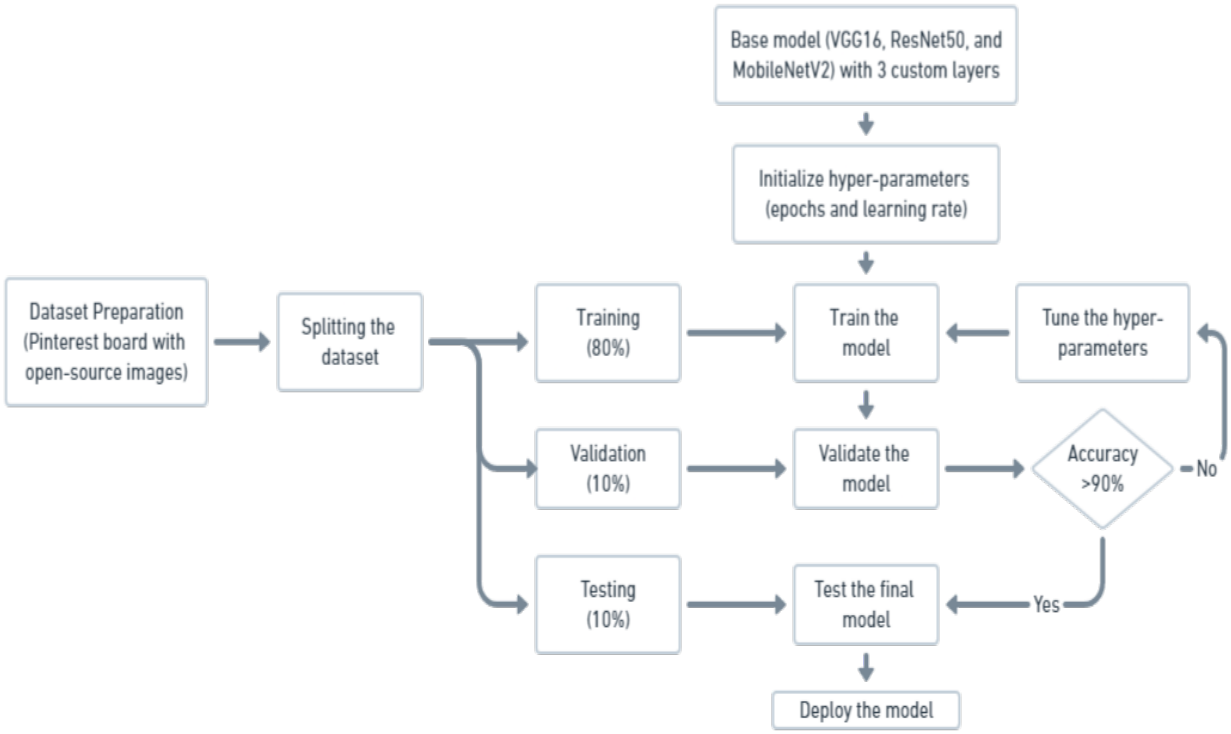


Figure 1. Deep Learning pipeline.

The final models were tested using the test dataset to evaluate their performance in classifying the quality of unseen Anime character sketches. The testing phase provided insights into the models' effectiveness and their generalization capabilities on new data.

Results

In this section, we detail the experimental results for the three neural network architectures utilized in our study. The models were subjected to hyper-parameter tuning, where learning rates were adjusted within the range of 0.000001 to 0.05 and epoch numbers varied from 10 to 50.

The validation accuracies recorded during this process are illustrated in Figures 2, 3, and 4. The peak validation accuracy reached 94% for both MobileNetV2 and ResNet50, while VGG16 demonstrated a maximum of 82% accuracy.

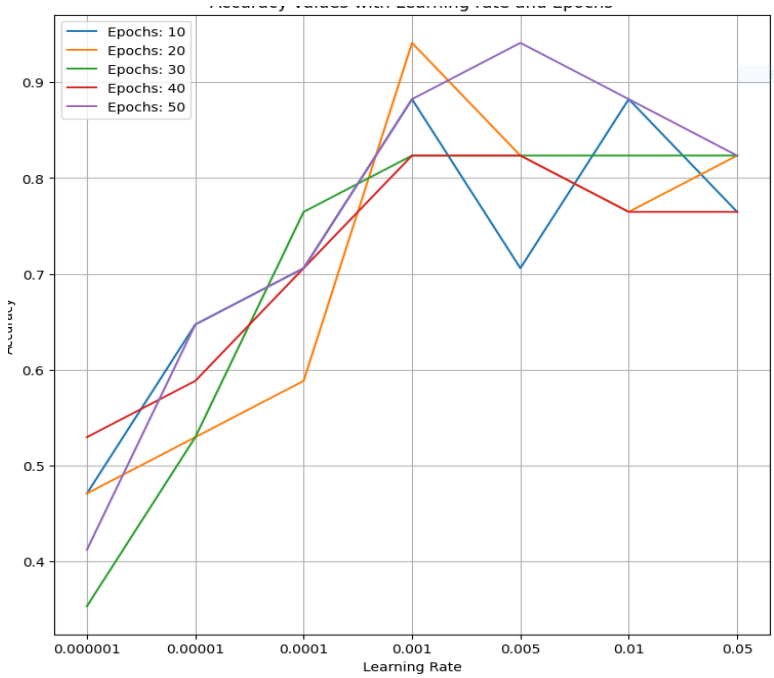


Figure 2. Accuracy outcomes of hyper-parameter optimization for the MobileNetV2 architecture. The x-axis represents the learning rate, the y-axis denotes model accuracy, and each line, as detailed in the legend, corresponds to a distinct epoch value.

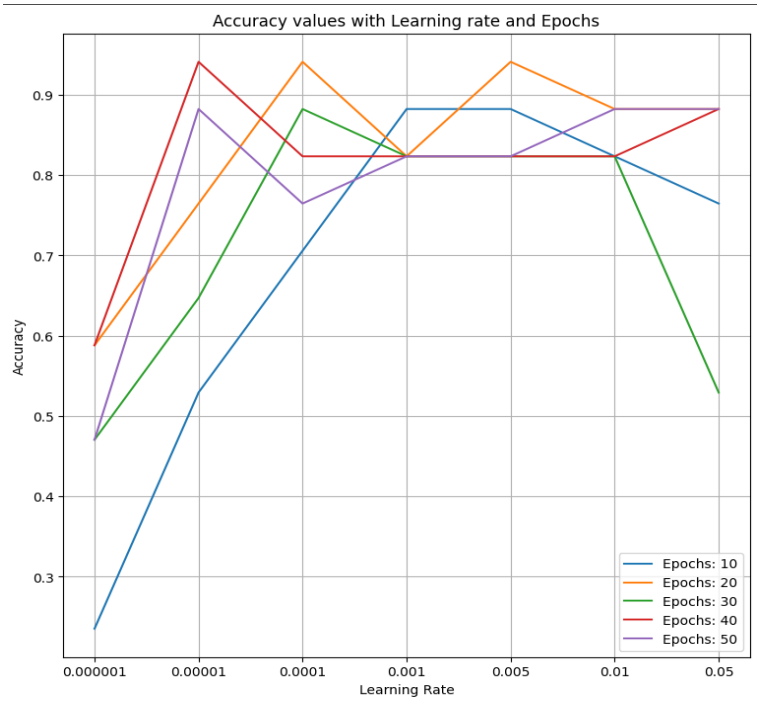


Figure 3. Accuracy outcomes of hyper-parameter optimization for the ResNet50 architecture. The x-axis represents the learning rate, the y-axis denotes model accuracy, and each line, as detailed in the legend, corresponds to a distinct epoch value.

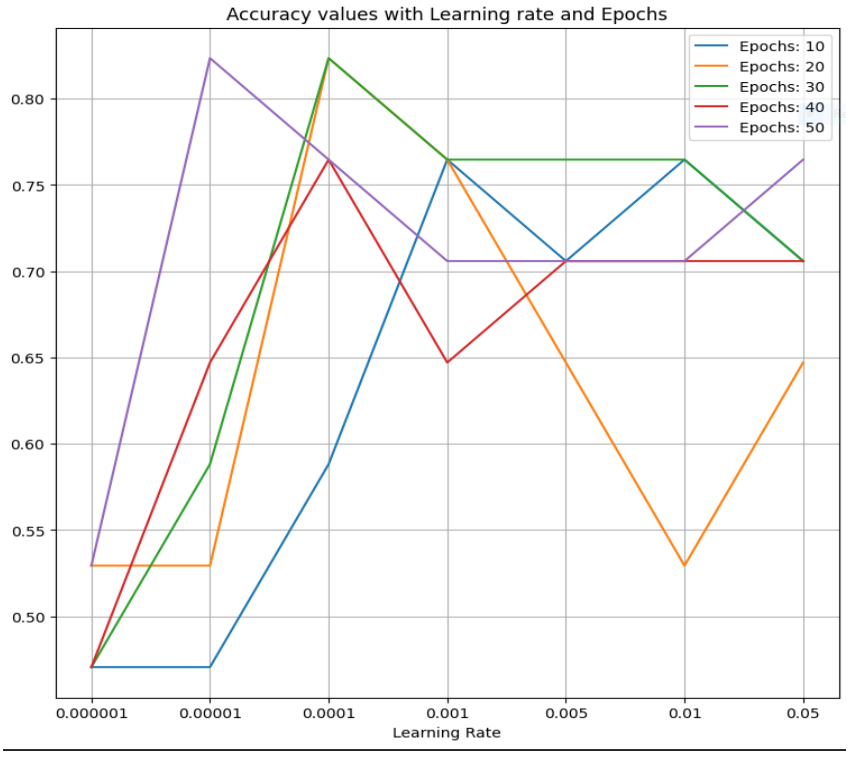
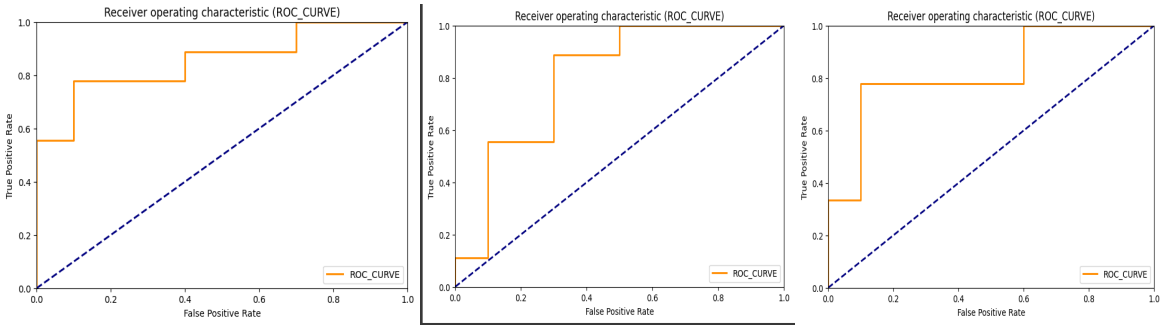


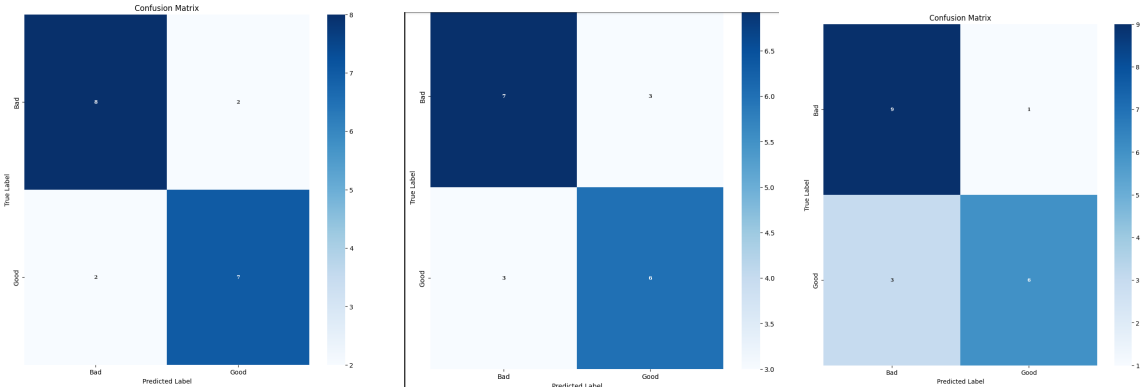
Figure 4. Accuracy outcomes of hyper-parameter optimization for the VGG16 architecture. The x-axis represents the learning rate, the y-axis denotes model accuracy, and each line, as detailed in the legend, corresponds to a distinct epoch value.

To assess the models' predictive capabilities, the best-performing models from each architecture were selected based on validation accuracy and applied to the test dataset. The Receiver Operating Characteristic (ROC) curves, which display the trade-off between sensitivity and specificity for the selected models, are presented in Figure 3.



(a) MobileNetV2 (b) ResNet50 (c)VGG16

Figure 3. ROC curve of the best models on the test subset of the data for different architectures.



(a) MobileNetV2 (b) RestNet50 (c) MobileNetV2

Figure 5. Confusion matrix showing the test results with different model architectures.

The test dataset's performance was summarized through a confusion matrix for each architecture, as shown in Figure 5. This is complemented by a comprehensive metrics report, tabulated in Table 1. Contrary to the validation phase, the test accuracies reveal a disparate trend. Despite this, all models exhibited similar efficacy levels in the testing phase. It is noteworthy that the MobileNetV2 model, which boasts the least parameters, achieved the highest test accuracy at 79%, underscoring its efficiency.

Table 1. Test results showing the accuracy, precision and recall for different model architectures

Model	Accuracy	Precision	Recall

MobileNetV2	79	79	79
RestNet50	68	68	68
VGG16	79	80	78

Discussion

Across the VGG16, ResNet50, and MobileNetV2 models, we observe distinct trends in image classification performance relative to learning rates and epochs. All three models demonstrate sensitivity to learning rate adjustments, with optimal performance achieved within specific ranges.

The MobileNetV2 model's performance in image classification varies significantly with learning rates and epochs. At lower learning rates (0.000001), the model demonstrates inconsistent performance, peaking at 0.4706 in the early epochs but declining thereafter. The model achieves optimal performance at a learning rate of 0.001, reaching a peak accuracy of 0.94 by the 20th epoch. Notably, MobileNetV2 shows stable performance at higher learning rates (0.005, 0.01, and 0.05) beyond 30 epochs, indicating robustness against overfitting. At moderate learning rates (0.00001 and 0.0001), the model exhibits gradual improvement, achieving peak accuracies of 0.65 and 0.76, respectively, by the 50th epoch. Overall, a learning rate of 0.001 is most effective for this dataset, with the model displaying stability at higher learning rates.

The ResNet50 model shows gradual improvement in accuracy, reaching a peak of 0.5882 by 20 epoch at a learning rate of 0.000001. The model achieves its best performance at learning rates of 0.0001 and 0.005, with a peak accuracy of 0.94 by 20 epochs. At a learning rate of 0.05, the model's performance fluctuates significantly across epochs, indicating potential overfitting issues. The model shows stable performance at learning rates of 0.00001, 0.001, and 0.01 beyond 30 epochs, with minor fluctuations in accuracy. In summary, the ResNet50 model's performance is highly dependent on the learning rate, with 0.0001 and 0.005 being the most effective for this dataset. Higher learning rates may lead to instability, while lower rates result in slower learning.

We observed that the VGG16 model's performance in image classification is significantly influenced by the learning rate. Lower learning rates (0.000001 and 0.00001) resulted in consistently lower accuracies across all epochs, suggesting insufficient learning. Conversely, the model achieved its best performance at a learning rate of 0.0001, reaching an accuracy of 0.82 by the 20th epoch and maintaining high performance thereafter. This indicates an optimal learning rate range for this particular dataset and model architecture. At higher learning rates (0.005, 0.01, and 0.05), we noticed more fluctuations in the model's accuracy across epochs, hinting at potential overfitting issues. Specifically, at a learning rate of 0.01, the accuracy notably dropped from 0.7647 at 10 epochs to 0.5294 at 20 epochs, before stabilizing again. This suggests that while higher learning rates may accelerate initial learning, they can also lead to instability and overfitting. Furthermore, the model exhibited early convergence, particularly at learning rates of 0.0001 and 0.001, with peak accuracies reached within the first 20 epochs. Beyond this point, improvements in accuracy were marginal, indicating that longer training periods might not yield significant benefits at these learning rates.

While all three models exhibit sensitivity to learning rates, they each have distinct optimal ranges that yield the best performance. The VGG16 and ResNet50 models show early convergence and sensitivity to higher learning rates, whereas MobileNetV2 demonstrates stability at higher rates, indicating a potential advantage in training efficiency and robustness against overfitting.

In the evaluation of test results, as shown in Table 1, the MobileNetV2 and VGG16 models both achieved an accuracy of 79%, with VGG16 slightly outperforming in precision. Interestingly, both models also recorded a recall rate of 79%, indicating a balanced performance across classes. In contrast, the ResNet50 model

lagged slightly behind, with uniform scores of 68% in accuracy, precision, and recall. These results underscore the nuanced capabilities of each architecture in handling the classification task.

Adding to the discussion, it's notable that while the MobileNetV2 model demonstrates a balance of accuracy, precision, and recall, the VGG16 model's higher precision suggests it may be slightly more reliable in identifying true positives. The relatively lower performance of ResNet50 across all metrics may point to its limitations within the specific context of this dataset. Overall, the test results corroborate the observed trends in the sensitivity to learning rates and epochs, with MobileNetV2 and VGG16 showing promising generalization capabilities on unseen data.

Conclusion

In conclusion, the global anime market's exponential growth is reshaping the entertainment industry, with a valuation of USD 31.23 billion in 2023 and a projected CAGR of 9.8% until 2030. This growth is driven by a passionate community of artists and enthusiasts. However, challenges such as artist scarcity and limited tools for feedback and promotion persist. Our study focuses on enhancing anime creation through deep learning. We introduce an automated approach to predict anime quality, using transfer learning to refine three pre-trained models—MobileNetV2, ResNet50, and VGG16. The MobileNetV2 and ResNet50 models achieved peak validation accuracies of 94%, with the MobileNetV2 model demonstrating superior performance on the test set, achieving an accuracy of 79% with significantly fewer parameters. This highlights its potential as a lightweight and efficient tool for quality assessment in the anime industry.

Limitations

Despite these advancements, our model's current limitation is its binary classification capability. Future work could involve expanding the model to recognize a broader range of quality levels. Another limitation of this work is the size of the dataset. A richer dataset will contribute to a higher performance of the deep learning models. Moreover, there are several other architectures that can be utilized for testing their performance on the dataset. Overall, our research contributes to the evolving landscape of AI in anime creation, offering new avenues for improving the quality and efficiency of anime production.

Acknowledgments

I would like to thank my advisor for the valuable insight provided to me on this topic.

References

1. Anime Market Size, Share & Trends Analysis Report By Type (T.V., Movie, Video, Internet Distribution, Merchandising, Music), By Genre (Action & Adventure, Sci-Fi & Fantasy, Romance & Drama, Sports, and Others), By Region, And Segment Forecasts, 2024 - 2030. (2023, Jan 28). Anime Market Size & Trends. https://www.grandviewresearch.com/industry-analysis/anime-market

2. Anime Market Outlook. (2023, jan 25). https://www.futuremarketinsights.com/reports/anime-market https://headphonesaddict.com/anime-statistics/

3. Pan, Y. (2018). Adapt or Die! The Social and Economic Dynamics of Japan’s Animation Industry.

https://repository.usfca.edu/capstone/762/

4. Pencil. (2023, April 6). Top 10 Fundamentals For Drawing and Sketching. Pencil Perceptions. https://www.pencilperceptions.com/10-fundamentals-for-drawing-and-sketching/

5. Li, H., Guo, S., Lyu, K., Yang, X., Chen, T., Zhu, J., & Zeng, H. (2022). A Challenging Benchmark of Anime Style Recognition. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 4721-4730). https://arxiv.org/pdf/2204.14034.pdf

6. Huang, Z., Xie, H., Fukusato, T., & Miyata, K. (2023). AniFaceDrawing: Anime Portrait Exploration during Your Sketching. arXiv preprint arXiv:2306.07476. https://dl.acm.org/doi/abs/10.1145/3588432.3591548

7. Hammond, T., Kumar, S. P. A., Runyon, M., Cherian, J., Williford, B., Keshavabhotla, S., ... & Linsey, J. (2018). It's not just about accuracy: Metrics that matter when modeling expert sketching ability. ACM Transactions on Interactive Intelligent Systems (TiiS), 8(3), 1-47. https://dl.acm.org/doi/pdf/10.1145/3181673

8. Byrne, I., Kanaoka, Y., Pollack, N. E., Rhee, H. J., & Sommers, P. M. (2019). An Analysis of Airport Delays Across the United States, 2012-2018. Journal of Student Research, 8(2). https://doi.org/10.47611/jsr.v8i2.775

9. Dong, K., Zhou, C., Ruan, Y., & Li, Y. (2020, December). MobileNetV2 model for image classification. In 2020 2nd International Conference on Information Technology and Computer Application (ITCA) (pp. 476-480). IEEE. https://ieeexplore.ieee.org/abstract/document/9422058

10. Koonce, B., & Koonce, B. (2021). ResNet 50. Convolutional Neural Networks with Swift for Tensorflow: Image Recognition and Dataset Categorization, 63-72. https://link.springer.com/chapter/10.1007/978-1-4842-6168-2_6

11. O'Shea, K., & Nash, R. (2015). An introduction to convolutional neural networks. arXiv preprint arXiv:1511.08458. https://www.researchgate.net/publication/337105858_Transfer_learning_using_VGG-16_with_Deep_Convolutional_Neural_Network_for_Classifying_Images

12. Hosna, A., Merry, E., Gyalmo, J., Alom, Z., Aung, Z., & Azim, M. A. (2022). Transfer learning: a friendly introduction. Journal of Big Data, 9(1), 102. https://journalofbigdata.springeropen.com/articles/10.1186/s40537-022-00652-w

13. Li, H., Guo, S., Lyu, K., Yang, X., Chen, T., Zhu, J., & Zeng, H. (2022). A Challenging Benchmark of Anime Style Recognition. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 4721-4730). https://openaccess.thecvf.com/content/CVPR2022W/VDU/html/Li_A_Challenging_Benchmark_of_Anime_Style_Recognition_CVPRW_2022_paper.html