Ethnicity-Aware Auto-colorization of Grayscale Human Portraits Using CNN

M.Sc. Project Final Defense

Presented By:
Rupesh Poudel (THA079MSISE012)

Supervised By:
Er. Umesh Kanta Ghimire

Department of Electronics and Computer Engineering
Institute of Engineering, Thapathali Campus

August 21, 2024

Presentation Outlines

- Motivation
- Background
- Problem Statement
- Objectives of Project
- Scope of Project
- Originality of Project
- Potential Applications
- Literature Review
- Methodology
- Results
- Discussion and Analysis
- Future Enhancements
- Conclusion
- Project Schedule
- References

Motivation

- Grayscale images lack the richness of color information which can be vital for interpretation.
- Exploring how well neural networks can interpret and reconstruct color information.

Background

- Grayscale images have been a fundamental component of visual representation in the past
- The absence of color limits the ability to convey the full richness and depth of visual content
- Existing colorization methods often apply uniform color schemes, failing to accurately represent diverse ethnic groups
- Incorporating ethnicity detection with colorization enhances realism, emotional impact and provide aesthetic appeal

Problem Statement

- Existing colorization methods often apply uniform color schemes that do not account for ethnic diversity which leads to unrealistic results.
- Developing system capable of accurately identifying ethnic characteristics and applying corresponding color palettes.
- Maintaining consistent and natural coloring across visible regions such as the neck and hands regions

Objectives of Project

- Develop a CNN to predict realistic colors for grayscale images, trained on diverse human portrait datasets with ethnic labels.
- Integrate a colorization algorithm using ethnic information to apply culturally accurate colors, ensuring consistent colorization of faces, necks, and hands.

Scope of Project

- Detect the ethnicity of individuals in grayscale images and apply realistic colors.
- Accurately colorize facial features, necks, and hands, ensuring realistic and appropriate representation for each ethnicity.
- Struggle with images containing multiple ethnicities, varying lighting conditions
- Performance is dependent on the diversity and quality of the training dataset, and computational constraints limit the dataset size.

Originality of Project

- Proposing a novel method that integrates ethnicity detection into image colorization to ensure accurate representation of diverse ethnic backgrounds.
- Enhancing colorization by not only including facial skin color but also considering eyes, hair, beards, and consistently coloring neck and hands using diverse datasets.

Potential Applications

- Transform old photographs by restoring color to engage modern audiences.
- Use in educational materials for realistic, culturally accurate visuals in textbooks and documentaries.
- Assisting law enforcement and forensic experts in enhancing grayscale surveillance footage and crime scene photos.
- Revive old family photos for genealogy to enhance connections with past members.

Literature Review [1]

Paper	Classification of Ethnicity Using…[1]	Human Face Image Colorization…[2]	Automatic Gray Image Coloring…[3]	Fine-grained semantic ethnic costume…[4]	Colorization of black-and-white…[5]
Authors	Abdulwahid Al Abdulwahid, 2023	Ben Wan et al., 2022	Jiayi Fan et al., 2022	Di Wu et al., 2021	David Futschik, 2018
Focus	Ethnicity classification using CNN	Colorizing human face	Automatic image coloring focusing on color collocation and quality improvement	High-resolution image colorization focusing on ethnic costumes	Colorizing cartoon images from video sequences using deep neural networks
Dataset	MORPH, FERET datasets	CelebA-HQ human face dataset	Custom dataset	Custom ethnic costume dataset	Rumcajs Dataset containing cartoon images hand-colorized by scribble methods

Literature Review [2]

Paper	Classification of Ethnicity Using…[1]	Human Face Image Colorization…[2]	Automatic Gray Image Coloring…[3]	Fine-grained semantic ethnic costume…[4]	Colorization of black-and-white…[5]
Method	- Efficient CNN models - Holdout approach for training and testing	Dual-Scale Attention U-Net	CNN-based automatic gray image coloring	Conditional GAN with Pix2PixHD backbone	Plain CNN and residual CNN models inspired by ResNet
Results	**Model A:** 85% accuracy for gender classification. **Model B:** 86% accuracy for ethnicity classification.	- **PSNR**: 28.639 dB, - **SSIM**: 96.1%; - Reduced boundary leakage and detail loss,	Achieves good coloring effects with advantages in network size and coloring quality over VGG and differential networks.	- **PSNR**: 26.45 dB - **SSIM**: 91.4% - Superior visual performance in preserving details and color distribution of ethnic costumes	- Plain CNNs excelling in backgrounds and residual CNNs handling smaller objects better.

Literature Review [3]

Paper	Classification of Ethnicity Using…[1]	Human Face Image Colorization…[2]	Automatic Gray Image Coloring…[3]	Fine-grained semantic ethnic costume…[4]	Colorization of black-and-white…[5]
Strength	-High accuracy -Efficient use of central region of the face.	-Excellent deep feature extraction with effective dual-scale and attention mechanisms	- Hybrid foreground & background coloring approach. - Effective CNN implementation. -Vivid, realistic results with detailed lighting/shading	- Uses fine-grained semantic information for ethnic costumes - Superior performance in preserving local details and color distribution	- Predict per-pixel color histograms using classification - Use ensemble and segmentation post-processing for consistent colorizations.
Weakness	- Lack of detailed analysis of model architectures and feature extraction techniques.	-Long training time and unstable colorization of background regions	- Reliance on manual selection of reference images. - Fusion between foreground and background could be improved.	- Relies on manually annotated fine-grained semantic masks - Limited experiments on a custom dataset of four Chinese ethnic minority costumes	- Trained and evaluated in limited dataset - Shows models struggle with large uniform areas but excel on smaller objects and characters

Methodology [1]

- ## Dataset Collection

  Includes images of human portraits from various ethnic backgrounds, which is crucial for the ethnicity detection and colorization task.

Dataset	Description
FairFace Dataset	Contains human portraits of 6 different ethnic groups: White, Black, Indian, Asian, Middle Eastern and Latino Hispanic
Kaggle Human Face dataset	Contains collection of images with mix of all common races, age groups, genders and lighting conditions along with some GAN generated images.
Manual Curation	Created a few thousands images manually from open source image libraries like Unsplash featuring human portraits with multiple ethnic group in single frame. And also including nepali people images.

Methodology [2]

- **Dataset Contents**
  - **Images**

    The dataset contained coloured images of human portraits having diverse ethnic groups with mix of genders and various ages. Image folder contains image named systematically (eg. image001.jpg, image 002.jpg, etc).

  - **Labels**

    CSV file will includes detailed labels for each image, specifying the ethnic group. It includes the following columns:

    - **imageName:** The file name of image (eg. image001.jpg)
    - **ethnicity:** The ethnicity label of individual in image (eg. Asian, White, Indian, Middle Eastern, Hispanic and Black.)

Methodology [3]

- The final structure of the dataset is segregated like mentioned below

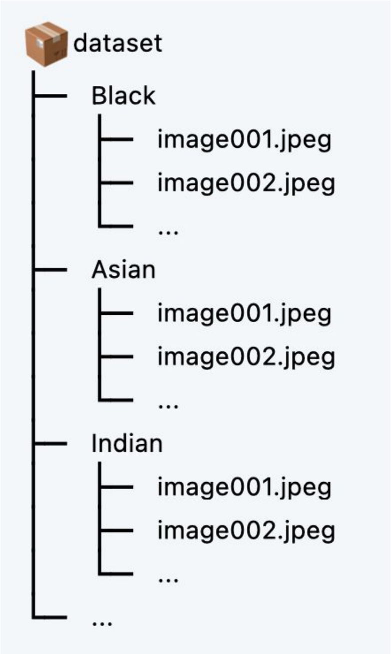


Fig: Dataset Folder Structure

Methodology [4]

- **Dataset Preprocessing**
  - **Dataset Cleaning:** Manual removal of outliers
  - **Image resizing:** All the images resized to a consistent 256x256 pixels.
  - **Data Augmentation**: Applied techniques like rotation, flipping, zooming and shearing for a more robust model. Also brightness shift as a photometric augmentation
  - **Color Space Conversion**: RGB images converted to Lab color space to separate lightness ('L') from color channels ('a' and 'b').

Methodology [5]

- **Overall Workflow**
  - **Training Stage [Two Models]**
    - Ethnicity Classification CNN Model Training
    - Colorization CNN Model Training
  - **Testing Stage**
    - Ethnicity Classification
    - Image Colorization Testing

Methodology [6]

System Block Diagram

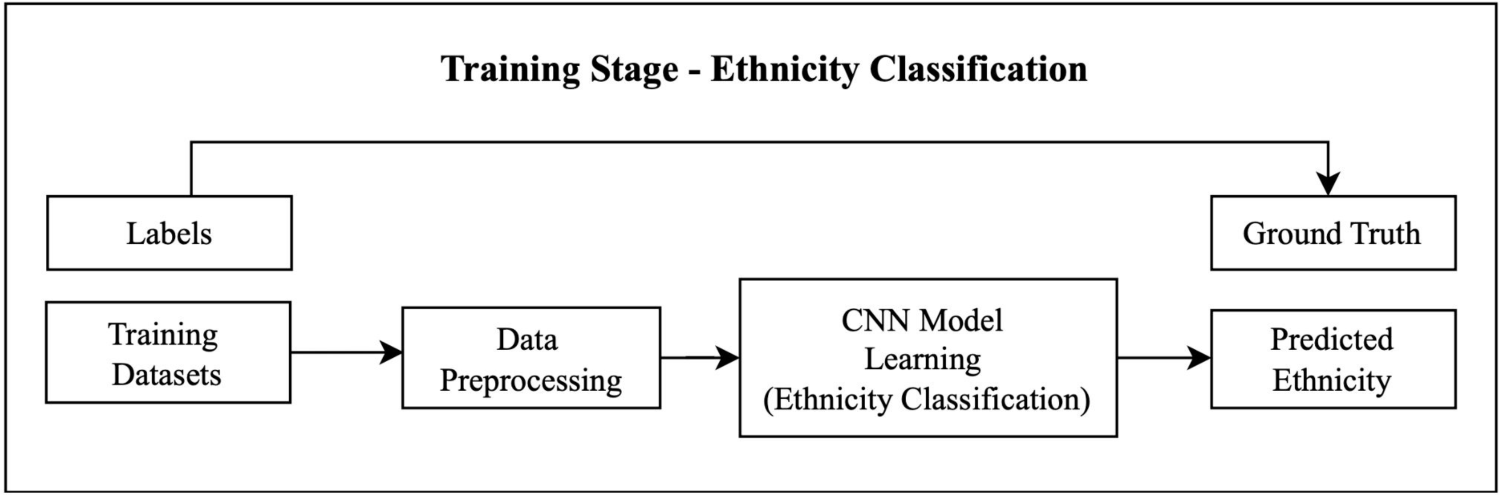


Fig: System block diagram - Training Stage of Ethnicity Classification Model

Methodology [7]

System Block Diagram

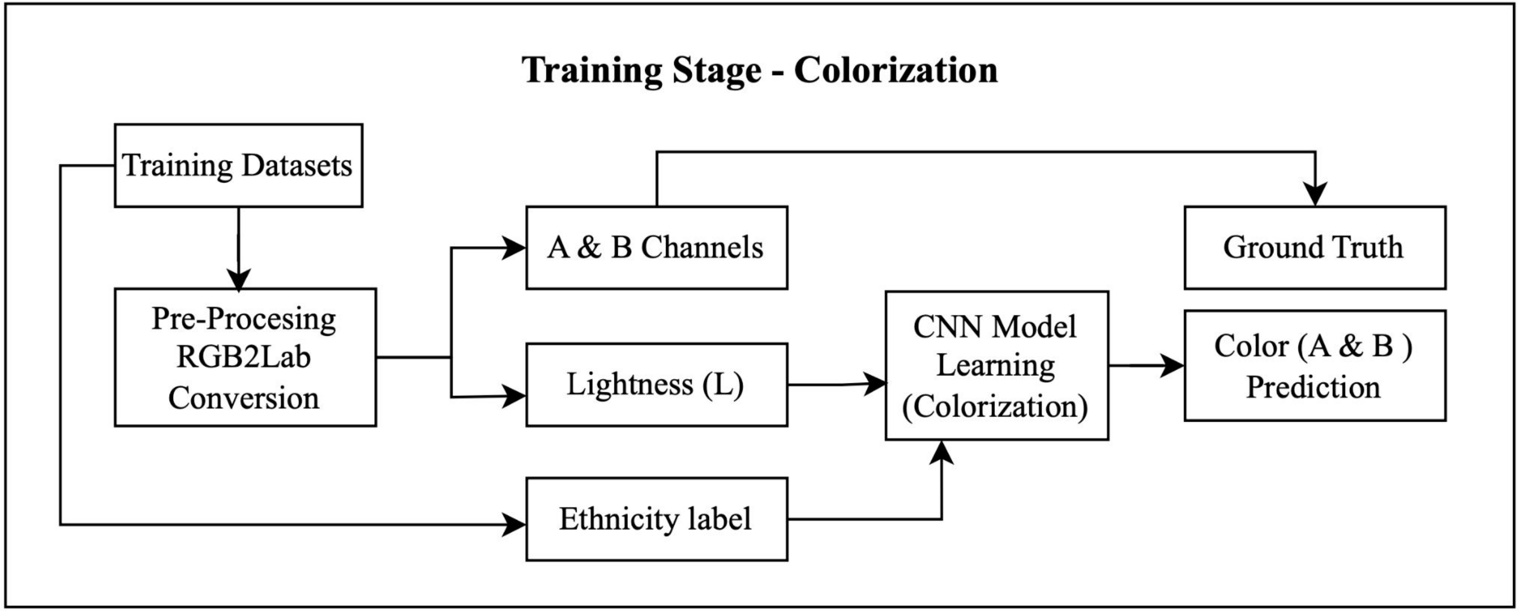


Fig: System block diagram - Training Stage of Colorization Model

Methodology [8]

System Block Diagram

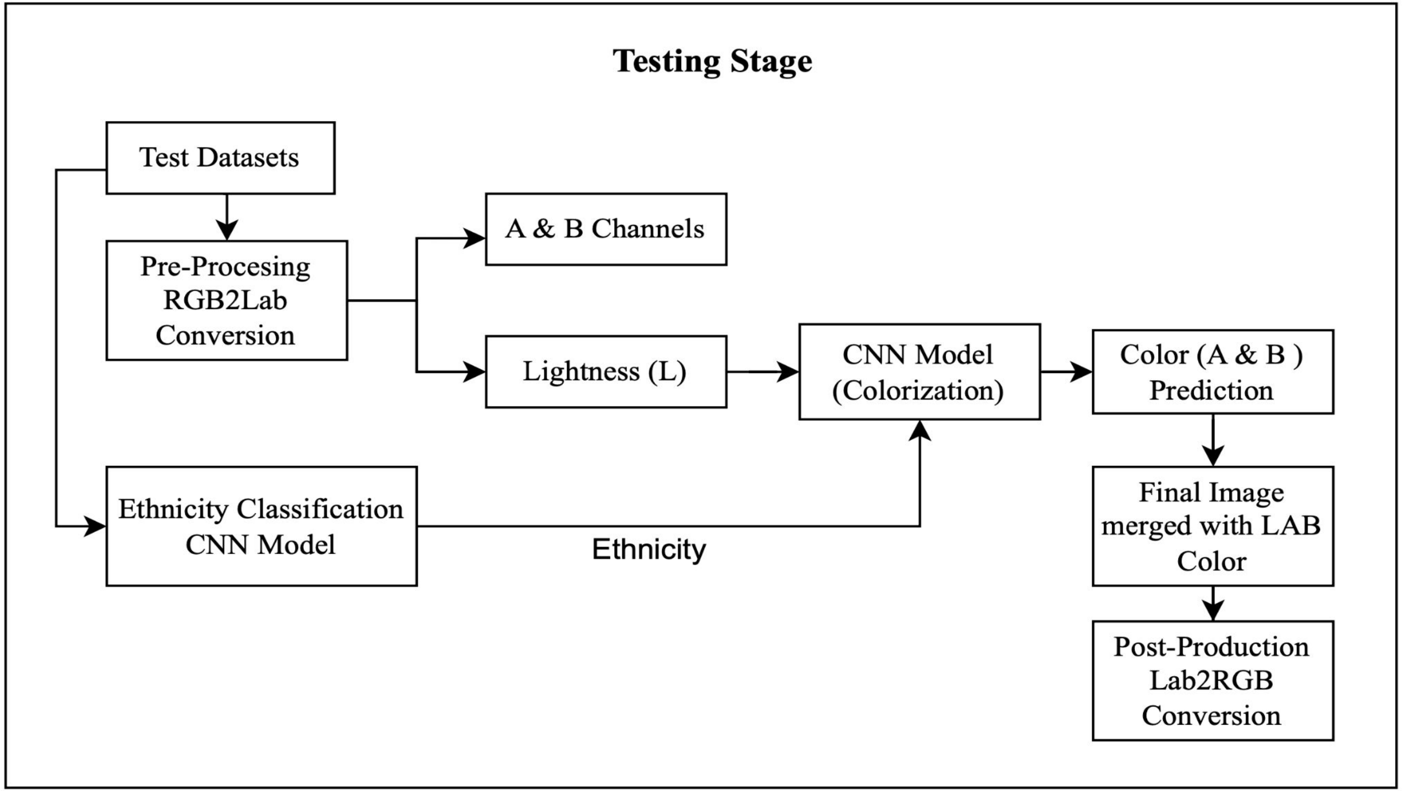


Fig: System block diagram - Testing Stage

Methodology [9]

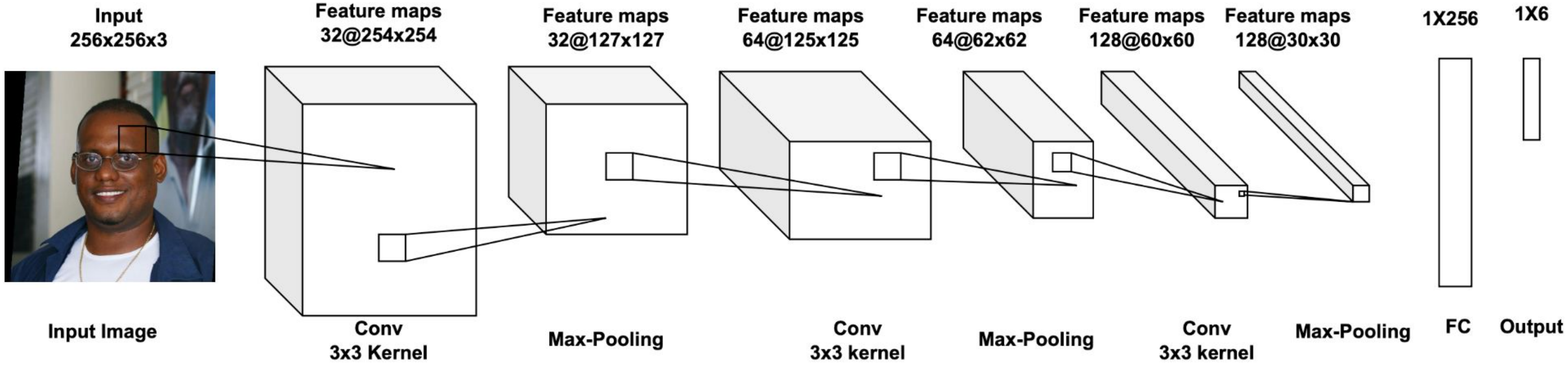


Fig: Model Architecture of Ethnicity Detection Model

Methodology [10]

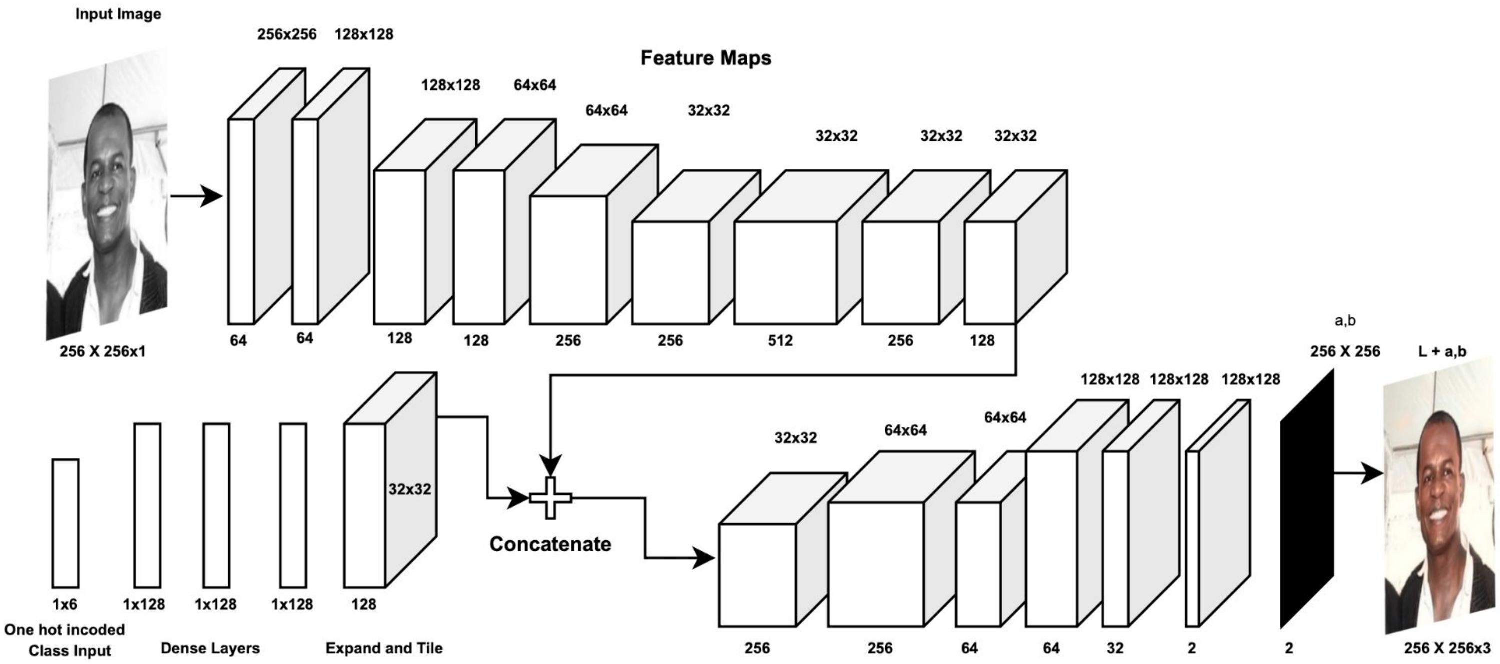


Fig: Model Architecture of Colorization Model

Methodology [11]

- **Validation Techniques**
  - We evaluate the colorization model's performance using metrics such as :
    - *Mean Squared Error (MSE)*
    - *Peak Signal-to-Noise Ratio (PSNR)*
    - *Structural Similarity Index Metric (SSIM)*
    - *Learned Perceptual Image Patch Similarity (LPIPS)*
  - We also conduct qualitative evaluations by visually comparing the colorized images with the ground truth images

Results [1]

Ethnicity Detection Model

Experiments

- Experiment 1: Total 9000 images with unbalanced images per class, Batch size of 16, learning rate of 0.0001 and 60 epochs.
- Experiment 2: Total 9000 images with 1500 images per class, Batch size of 32, leaning rate of 0.0001 and 120 epochs.
- Experiment 3: Total 9000 images with 1500 images per class, Batch size of 32, learning rate of 0.00001 with 134 epochs with early stopping

Results [2]

- **Loss Plot Ethnicity detection model:**

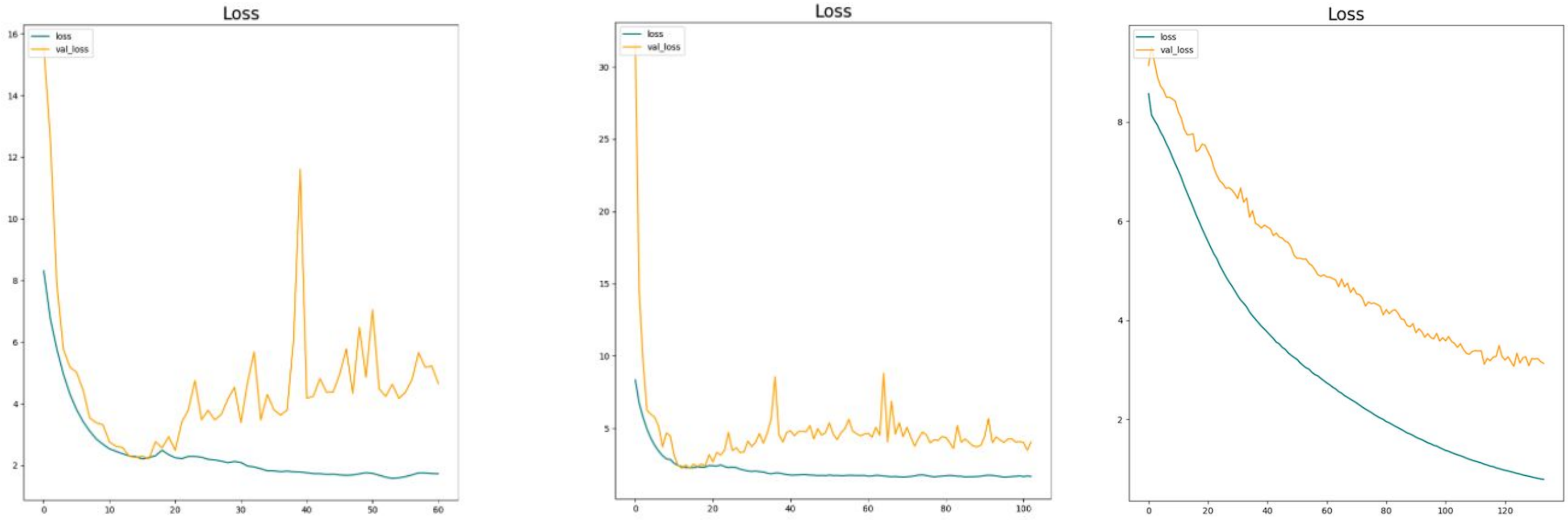


Fig: Loss Plot of Experiment 1,2 and 3 respectively

Results [3]

- **Confusion Matrix Ethnicity detection model:**

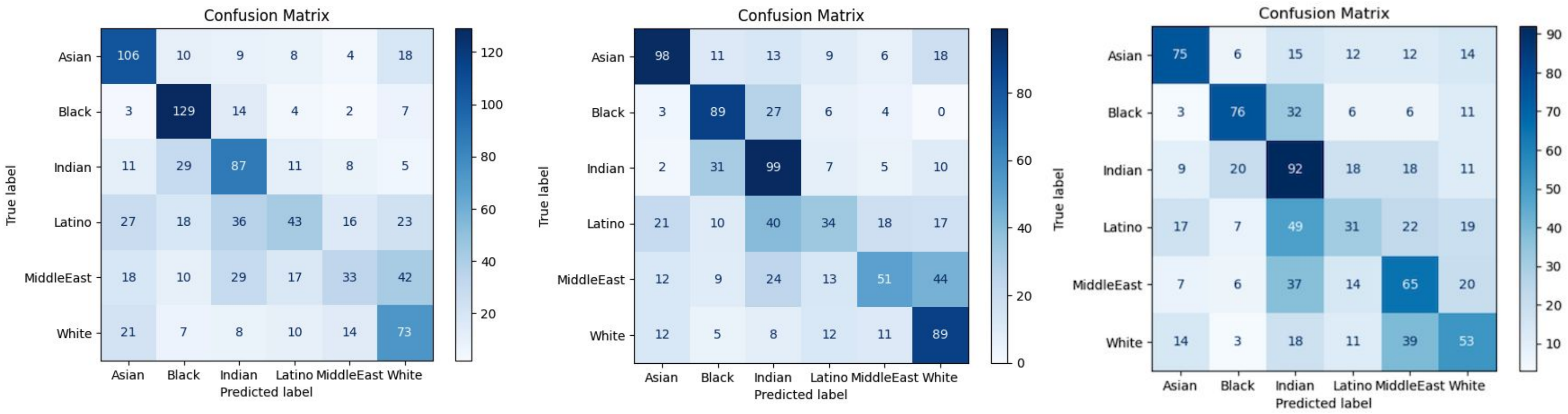


Fig: Confusion Matrix of Experiment 1,2 and 3 respectively

Results [4]

- **ROC Curve of Ethnicity detection model:**

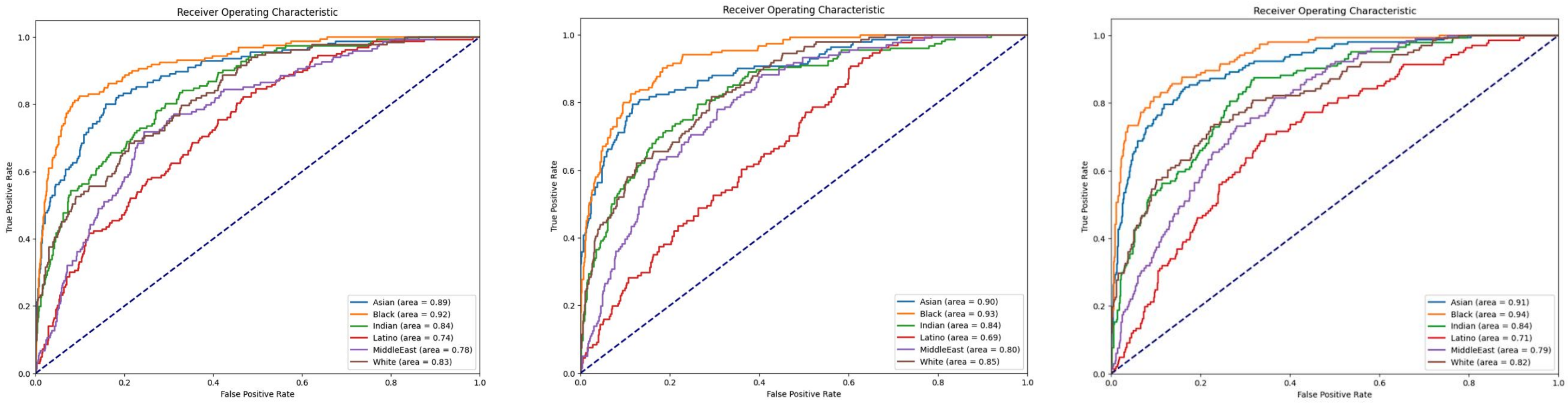


Fig: ROC Curve of Experiment 1, 2 and 3 respectively

Results [5]

Image	True Value	Predicted Value	Prediction Score
	Asian	Asian	0.9998
	Black	Black	1.0000
	White	White	0.9153
	Indian	Indian	0.9968

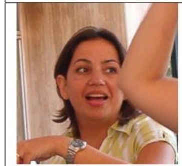
Image	True Value	Predicted Value	Prediction Score
	Latino	Indian	0.9991
	Latino	Middle East	0.6481
	Middle East	Asian	0.8202
	Indian	Black	1.0000

Fig: Best [Left] and Worst [Right] Case Scenario of Ethnicity Model

Results [6]

Initial Experiments of Colorization Model

- Experiment 1: Total 1200 images with 200 images per class, Batch size of 16, and 100 epochs.
- Experiment 2: Total 1200 images with 200 images per class, Batch size of 32, and 300 epochs.
- Experiment 3: Total 2400 images with 400 images per class, Batch size of 32, 150 epochs and removing outlier datasets.

Results [7]

- **Accuracy:** It shows limitations of using accuracy as a metric for colorization tasks.

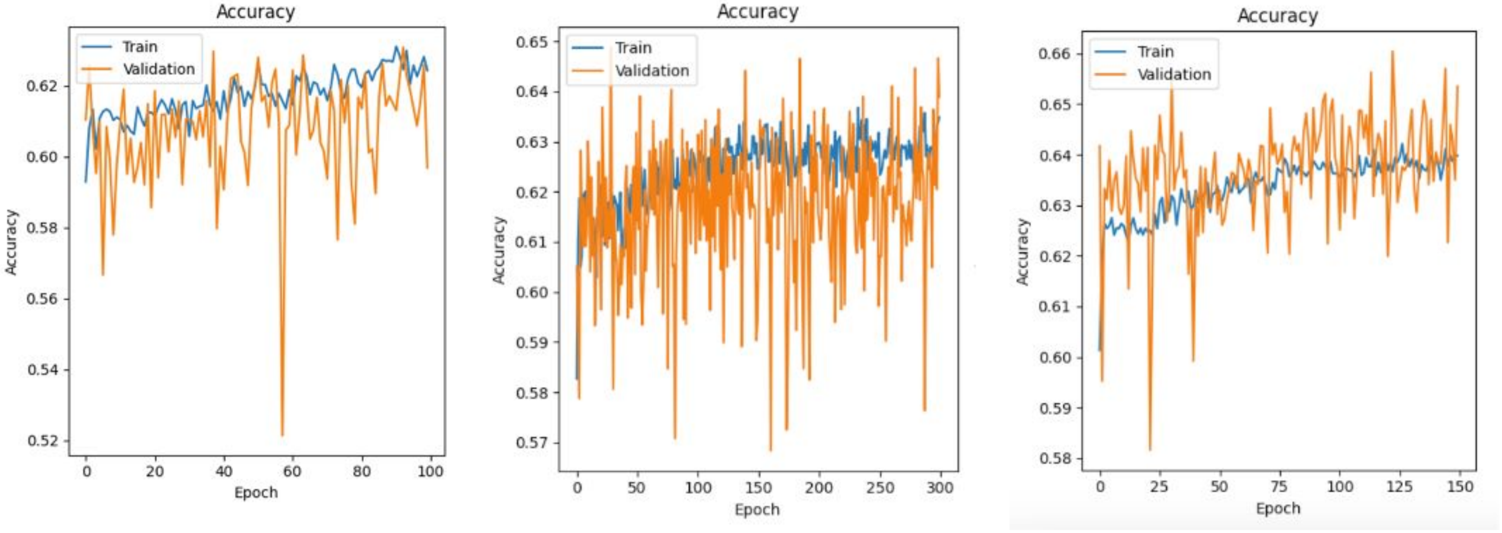


Fig: Accuracy Plot of Experiment 1, 2 and 3 respectively

Results [8]

- **Model Loss:** Experiment 3, with 2400 images, showed the best overall performance with lower MSE, indicating better model generalization and colorization quality.

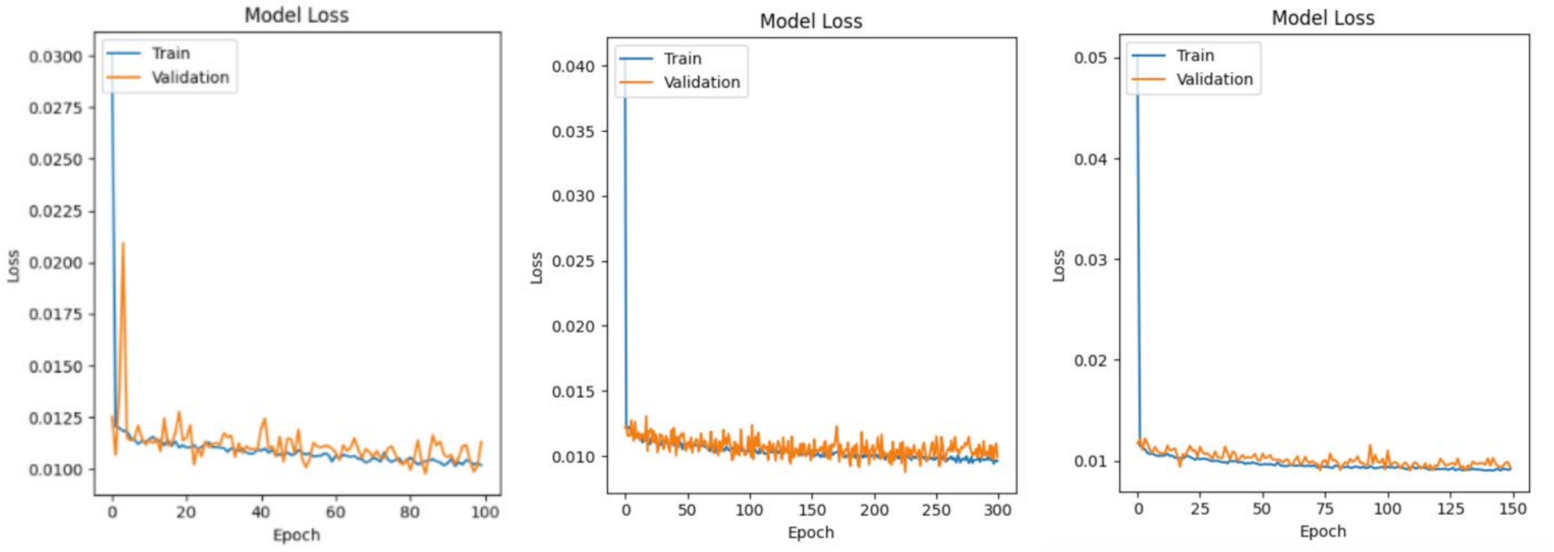


Fig: Loss Plot of Experiment 1, 2 and 3 respectively

Results [9]

- **PSNR:** Experiment 1 and 2 have more fluctuating validation plot of PSNR but the experiment 3 have a bit less fluctuation after removing outliers from dataset and increasing dataset number.

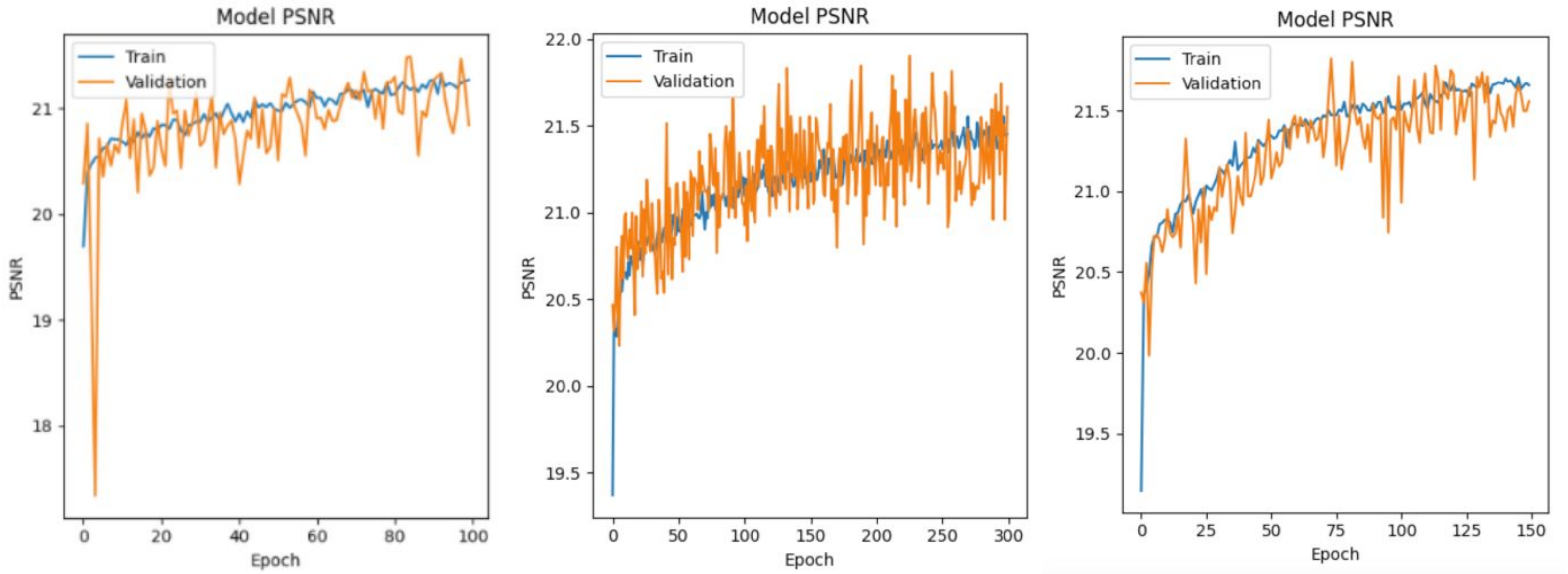


Fig: PSNR Plot of Experiment 1, 2 and 3 respectively

Results [10]

- **SSIM:** Experiment 1 and 2 have more fluctuating validation plot of SSIM but the experiment 3 have a bit less fluctuation after removing outliers from dataset and increasing dataset number.

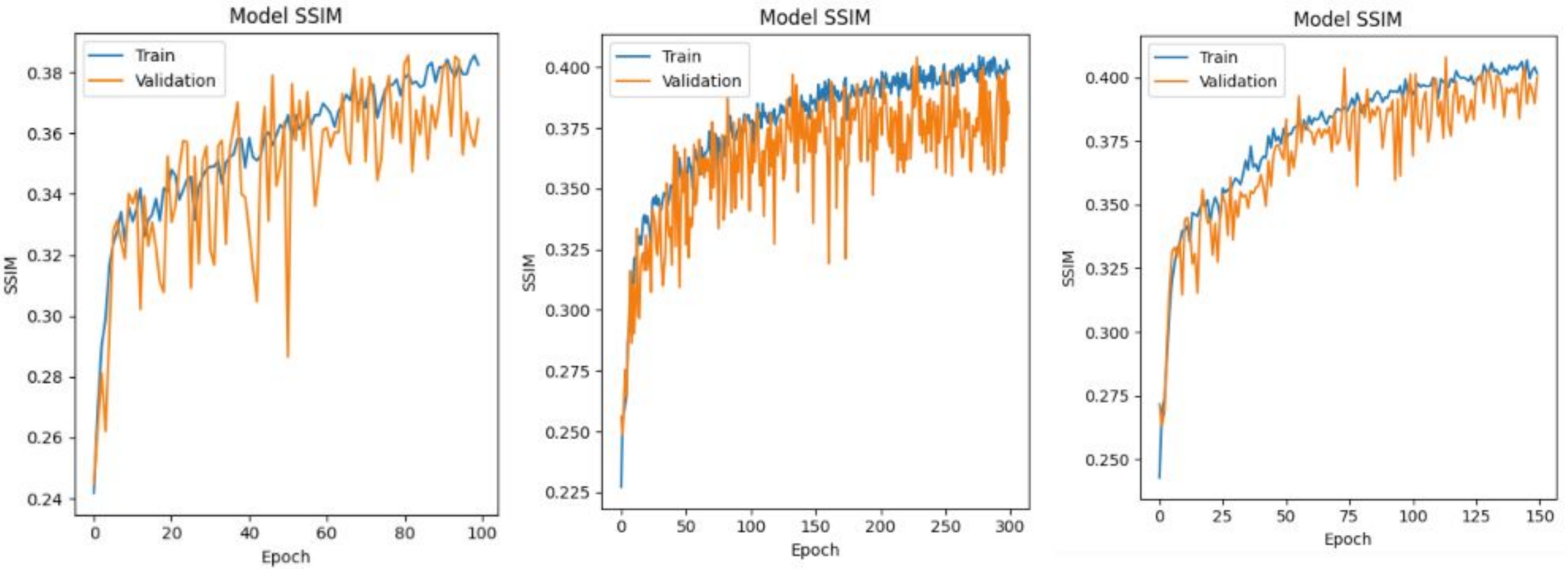


Fig: SSIM Plot of Experiment 1, 2 and 3 respectively

Results [11]

Large Scale Training of Colorization Model

- Experiment 4:
    - Total Images: 9000 (1500 per class)
    - Batch Size: 16
    - Initial Training: 100 epochs
    - Fine-tuning:
        - Lower learning rate: 1e-5
        - Early stopping to prevent overfitting
        - Learning rate reduction if validation loss plateaus
        - Additional 25 epochs

Results [12]

- **PSNR:** During initial training, PSNR values improved but plateaued. Fine-tuning with a reduced learning rate and adaptive adjustments led to higher PSNR values, ensuring better colorization quality.

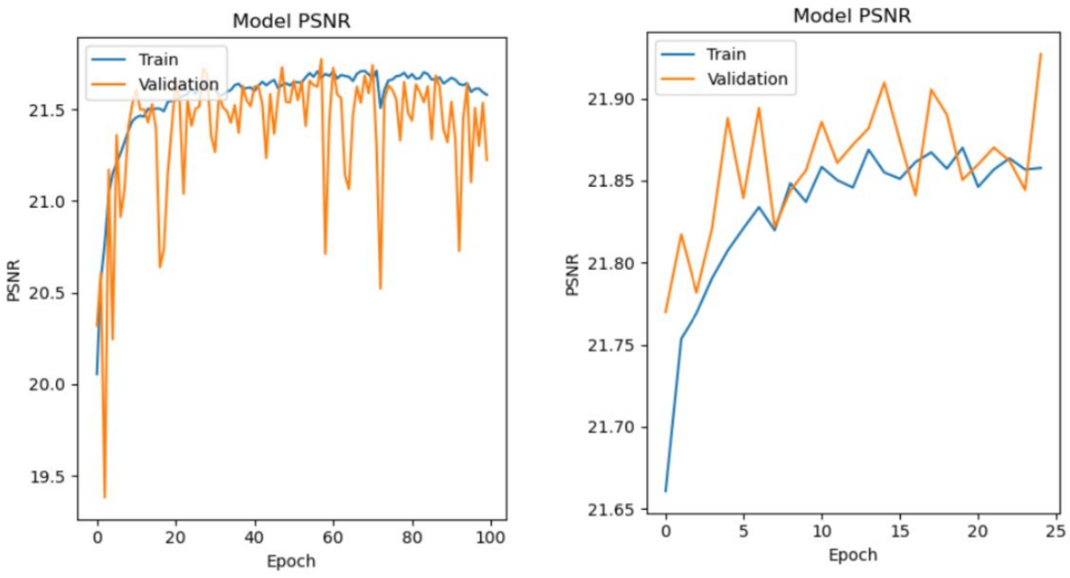


Fig: PSNR plot of experiment 4 initial training and fine tuned model respectively

Results [13]

- **SSIM: I**nitial training showed steady SSIM increase with jitteriness. Fine-tuning improved SSIM further, enhancing structural detail and preservation.

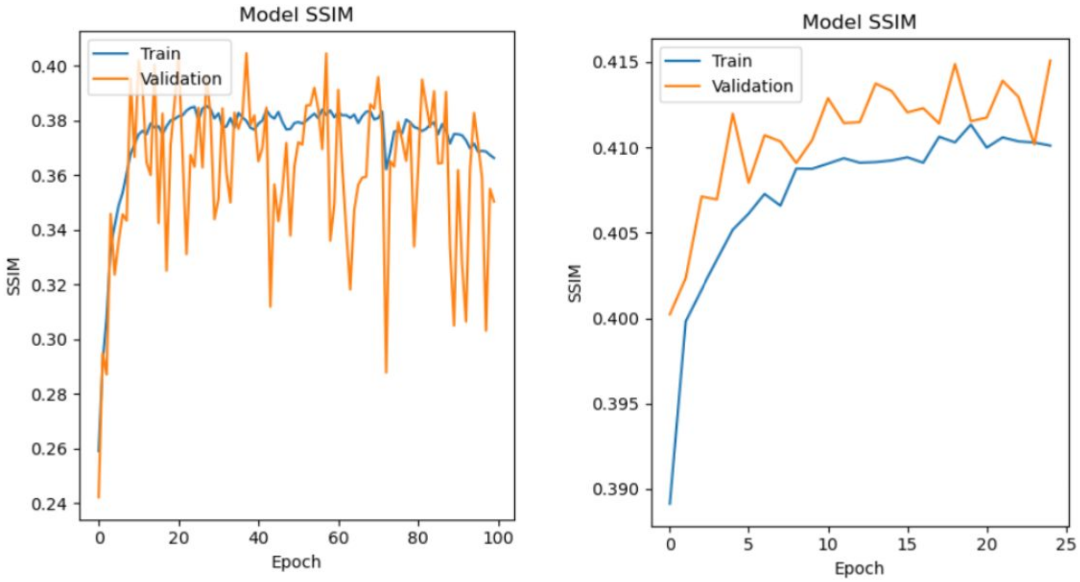


Fig: SSIM plot of experiment 4 initial training and fine tuned model respectively

Results [14]

- **Best Case Scenarios**

Grayscale Image	Ground Truth	Colorized Image	PSNR	SSIM	LPIPS
			31.56	0.97	0.0416
			25.84	0.96	0.1126

Fig: Best Case Scenario - Black Ethnic Group

Results [15]

- **Best Case Scenarios**

Grayscale Image	Ground Truth	Colorized Image	PSNR	SSIM	LPIPS
			25.6	0.94	0.0804
			23.7	0.91	0.1746

Fig: Best Case Scenario - Middle East and Indian Ethnic Group

Results [16]

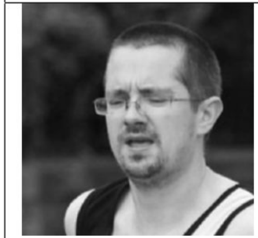
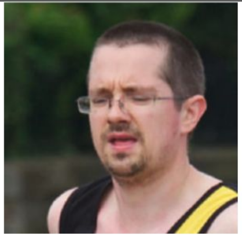
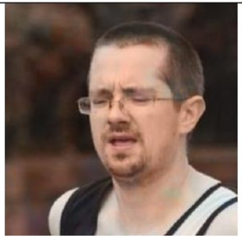
- **Best Case Scenarios**

Grayscale Image	Ground Truth	Colorized Image	PSNR	SSIM	LPIPS
			25.5	0.95	0.0784
			22.5	0.90	0.1156

Fig: Best Case Scenario - Image with multiple face in single frame

Results [17]

- **Worst Case Scenarios**

Grayscale Image	Ground Truth	Colorized Image	PSNR	SSIM	LPIPS
			24.39	0.93	0.1719

Grayscale Image	Ground Truth	Colorized Image	PSNR	SSIM	LPIPS
			22.62	0.94	0.1417

Fig: Worst Case Scenario - White Ethnic Group and Bearded Face

Results [18]

- **Worst Case Scenarios**

Grayscale Image	Ground Truth	Colorized Image	PSNR	SSIM	LPIPS
			25.0	0.96	0.0783
			24.9	0.9458	0.1580

Fig: Worst Case Scenario - Multiple people in single image

Results [19]

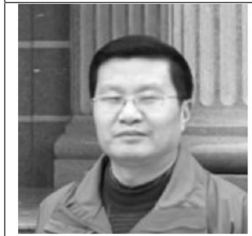
Grayscale Image	Experiment 1	Experiment 2
		
		
		

Fig: Visual comparison of initial two experiment of Colorization Model

Results [20]

Fig: Visual comparison of third experiment and fourth large scale training experiment of Colorization Model

Discussion and Analysis [1]

Theoretical Expectations vs. Simulated Outputs

- **Theoretical Predictions:**
  - Expected improvements in accuracy, PSNR, and SSIM with larger datasets and longer training durations.
  - Techniques like learning rate reduction, early stopping, and callbacks were used to fine-tune the model, prevent overfitting, and improve generalization.

Discussion and Analysis [2]

Theoretical Expectations vs. Simulated Outputs

- **Simulated Results:**
  - Confirmed improvements with performance gains in Experiment 4 (accuracy: 0.64, PSNR: 21.8, SSIM: 0.41).
  - This shows the benefits of increasing dataset size and fine tuning techniques including early stopping

Discrepancies and Analysis:

- Overfitting, Data Quality and Feature Complexity

Discussion and Analysis [3]

Sources of errors

- Data quality and diversity
- Model Architecture
- Hyperparameter settings

Error Analysis and Impact

- **Overfitting:** Evident in Experiment 2, where extended training improved performance but not as significantly as expected.
- **Underfitting:** Seen in Experiment 1, where insufficient data and epochs led to the model not fully capturing the underlying data patterns, resulting in lower performance metrics.
- **Data Quality:** The significant impact of removing outliers in Experiment 3 and the diverse dataset in Experiment 4 highlighted the critical role of clean, high-quality data.

Discussion and Analysis [4]

Comparison with SOTA

Image ID	Model	PSNR (dB)	SSIM	LPIPS
Image 1	Our Model	23.66	0.9143	0.1746
	Zhang et al.	21.96	0.9150	0.1729
Image 2	Our Model	24.38	0.9378	0.1719
	Zhang et al.	25.88	0.9587	0.1224
Image 3	Our Model	27.81	0.9574	0.0678
	Zhang et al.	25.39	0.9569	0.0796
Image 4	Our Model	29.78	0.9650	0.0697
	Zhang et al.	30.71	0.9745	0.0572
Image 5	Our Model	25.51	0.9367	0.2206
	Zhang et al.	24.88	0.9475	0.2287

Fig: Visual Comparison (Left) and Metrics Comparison (Right) with SOTA

Discussion and Analysis [5]

Comparison with SOTA

Image ID	Model	PSNR (dB)	SSIM	LPIPS
Image 6	Our Model	32.69	0.9741	0.0416
	Zhang et al.	27.36	0.9759	0.0738
Image 7	Our Model	22.71	0.9428	0.1417
	Zhang et al.	22.11	0.9474	0.1474
Image 8	Our Model	26.51	0.9344	0.1643
	Zhang et al.	25.03	0.9240	0.1854
Image 9	Our Model	25.18	0.9742	0.0808
	Zhang et al.	21.72	0.9698	0.1152
Image 10	Our Model	16.97	0.8759	0.2492
	Zhang et al.	22.18	0.9431	0.1317
Image 11	Our Model	21.59	0.9138	0.1815
	Zhang et al.	21.64	0.9362	0.1945

Fig: Visual Comparison (Left) and Metrics Comparison (Right) with SOTA

Discussion and Analysis [6]

Comparison with SOTA

Model	Average PSNR (dB)	Average SSIM	Average LPIPS
Our Model	25.1627	0.9388	0.1422
Zhang et al. Model	24.4418	0.9499	0.1372

Fig: Average Performance Metrics for Our Model and Zhang et al. Model

Discussion and Analysis [7]

Qualitative Analysis

Best Case Scenarios

- Minimum background Noise
- Front orientation of face
- Single person in a image
- Fewer obstacles in the face

Worst Case Scenarios

- Complex backgrounds
- Non-frontal orientation
- Multiple faces in image
- Obstructions in face

Future Enhancements [1]

- Improving overall results
    - Expand dataset to include more diverse ethnic groups, age ranges, and lighting conditions.
    - Utilize more powerful GPUs for efficient training.
    - Experiment with different CNN architectures or GAN to enhance colorization quality.

Future Enhancements [2]

- Recommendation for future researchers
    - Create a dataset that accurately represents various ethnic groups to prevent biases and ensure good generalization.
    - Develop the model iteratively, incorporating feedback and continuous testing to refine results.
    - Accelerate progress by leveraging pre-trained models.

Conclusion

- Developed a CNN-based system that accurately and culturally colorizes grayscale human portraits.
- Successfully integrated ethnicity detection into the colorization process, ensuring realistic and culturally appropriate results.
- Model performance comparable to state-of-the-art (SOTA) methods, with areas for further improvement

Tentative Timeline

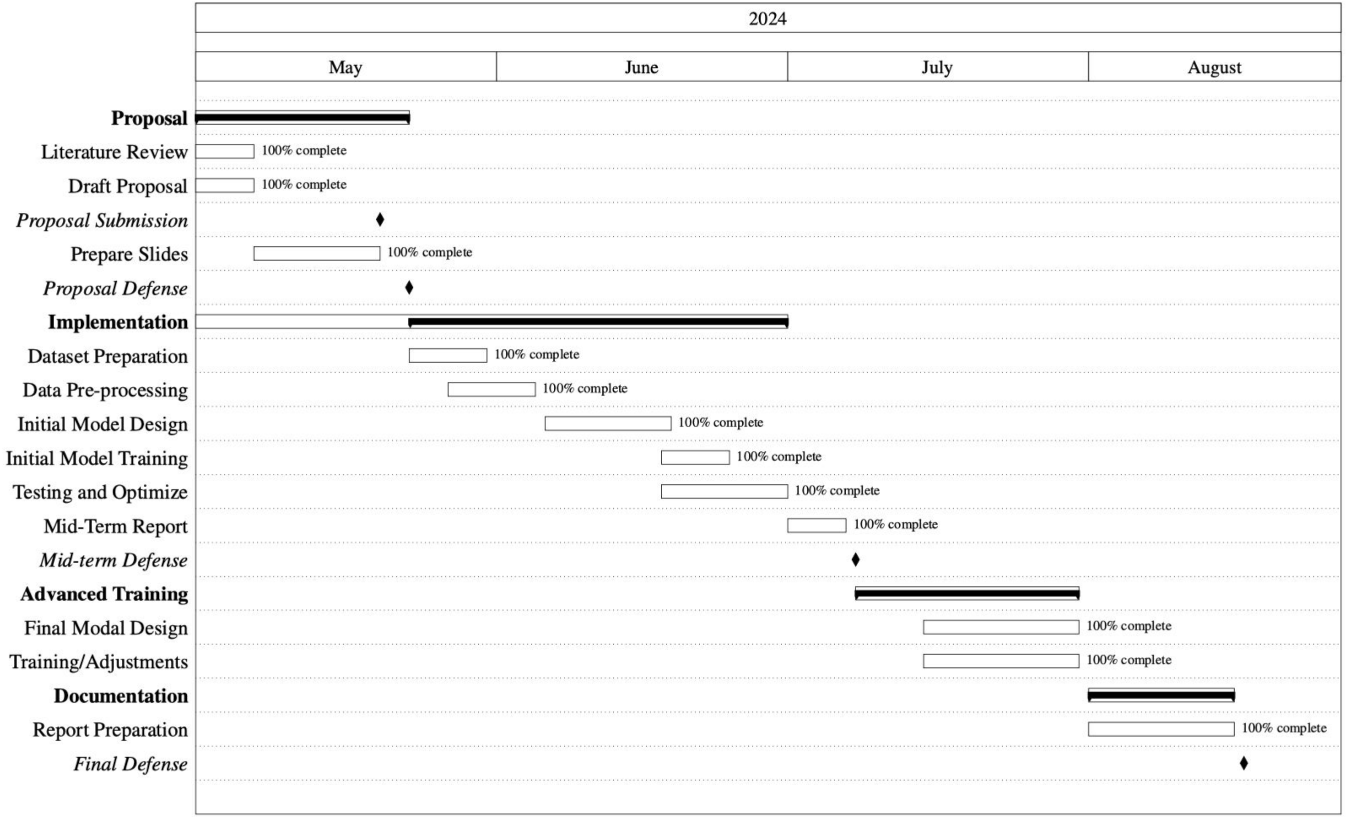


Fig: Gantt Chart showing timeline of project

References [1]

[1] Abdulwahid Al Abdulwahid. Classification of ethnicity using efficient cnn models on morph and feret datasets based on face biometrics. In *MDPI, Applied Sciences*. June 2023.

[2] Zhouxin Lu Liangqi Chen, Ben Wang. Human face image colorization with dual- scale attention u-net. In 5th International Conference on Mechatronics and Computer Technology Engineering (MCTE 2022). December 2022.

[3] Wentao Xie Jiayi Fan and Tiantian Ge. Automatic Gray Image Coloring Method Based on Convolutional Network. In *Computational Intelligence and Neuroscience*. Hindawi, China, April 2022.

[4] Jun Wang Juxiang Zhou Wei Gao Di Wu, Jianhou Gan. Fine-grained semantic ethnic costume high-resolution image colorization with conditional GAN. In *International Journal of Intelligent System*. China, November 2021.

References [2]

[5] David Futschik. Colorization of black-and-white images using deep neural networks. In Czech Technical University in Prague, Faculty of Electrical Engineering. Prague , Czech Republic, January 2018.

[6] Alexei A. Efros Richard Zhang, Phillip Isola. Colorful Image Colorization. In *CoRR*. October 2016.

[7] Hiroshi Ishikawa Satoshi Iizuka, Edgar Simo-Serra. Let there be Color! In *SIGGRAPH '16 Technical Paper*. July 2016.

[8] Alakh Sharma. Rgb to lab color space conversion: Formulas, insights, and applications. https://medium.com/@alakhsharmacs, 2024. [Accessed: May 5, 2024].

[9] Valentin Alexandru Stan. Automatic traffic-sign recognition artificial inteligence - deep learning algorithm. https://www.researchgate.net/figure/ ReLU-function-graph_fig2_346250677, 2020. [Accessed: July 1, 2024].

References [3]

[10] Valentin Alexandru Stan. Tangens hyperbolicus. https://tikz.net/tanh/. [Accessed: July 1, 2024].

[11] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. Deep Learning. MIT Press, 2016. http://www.deeplearningbook.org.

[12] Kärkkäinen et al. Fairface dataset. https://paperswithcode.com/dataset/ fairface, 2020. [Accessed: June 2, 2024].

[13] Richard Zhang. Colorful image colorization. https://richzhang.github.io/ colorization/, 2024. [Accessed: May 28, 2024].

[14] Tom Fawcett. Introduction to roc analysis. Pattern Recognition Letters, 27:861–874, 06 2006.

[17] Richard Zhang, Phillip Isola, Alexei A. Efros, Eli Shechtman, and Oliver Wang. The unreasonable effectiveness of deep features as a perceptual metric. April 2018.