Course Bonus Project – Deep Learning (Fall 2025)

Video Colorization Application

Name: ________________________

ID: __________________________

Course: Deep Learning – Fall 2025

Supervisor: ___________________

1. **Problem Description:**
   Black-and-white videos lack chromatic information, which significantly reduces their visual realism and interpretability. Restoring color to grayscale videos is a challenging task due to the absence of explicit color cues in the input data. Video colorization aims to automatically infer plausible color information for each pixel in a grayscale video using deep learning techniques.

Unlike image colorization, video colorization introduces an additional challenge: temporal consistency. Colors assigned to objects should remain stable across consecutive frames to avoid flickering artifacts. The objective of this project is to design and implement a deep learning-based video colorization system that takes a 30-second black-and-white video obtained from the internet as input and produces a colorized video as output.

---

2. Full Analysis of the Problem

Video colorization is considered an ill-posed problem, as multiple valid color solutions can exist for the same grayscale intensity values. The main challenges associated with this problem include:

2.1 Color Ambiguity

Grayscale images do not contain enough information to uniquely determine object colors. For example, different objects with different real-world colors may share similar grayscale intensities.

2.2 Semantic Understanding

Accurate colorization often requires understanding the semantic content of the scene, such as distinguishing between sky, skin, vegetation, buildings, and objects.

2.3 Temporal Consistency

Frame-by-frame colorization may introduce temporal inconsistencies, where colors change abruptly between consecutive frames, leading to flickering effects.

2.4 Computational Constraints

Training large video colorization models typically requires extensive computational resources and large-scale datasets, which may not be feasible in an academic course environment.

To address these challenges, this project adopts a lightweight yet effective approach based on color space transformation and convolutional neural networks.

---

3. Survey of Video Colorization Techniques

3.1 CNN-Based Image Colorization

This approach treats each video frame independently and uses convolutional neural networks to predict color channels. While simple and computationally efficient, it may suffer from temporal inconsistency.

3.2 CNN with Optical Flow

Optical flow techniques propagate color information across frames based on motion estimation. This improves temporal coherence but increases computational complexity and sensitivity to motion errors.

3.3 CNN + Recurrent Neural Networks (LSTM)

Recurrent models, such as LSTMs, capture temporal dependencies across frames, reducing flickering artifacts. However, they require more memory and training time.

In this project, a CNN-based approach was implemented due to computational constraints, while temporal consistency is discussed as a future enhancement.

---

4. Dataset Used

A small dataset of real-world video frames was created by extracting frames from an online video source. Frames were sampled at one frame per second, resulting in approximately 30 training images.

This lightweight dataset was sufficient to:

- **Validate the training pipeline**
- **Demonstrate the feasibility of video colorization**

- **Avoid memory and computational limitations**

---

5. Sample Frames Picked from the Video

Representative frames were extracted directly from both the black-and-white input video and the colorized output video. Frame selection was performed based on frame variance, which indicates visual richness and scene detail.

The same frame indices were used for both videos to ensure fair qualitative comparison. These samples demonstrate the effectiveness of the proposed approach across different scenes and lighting conditions.

(Insert Before/After frame comparisons here)

---

6. Model Design and Architecture

The proposed model is a Convolutional Neural Network (CNN) designed to predict chrominance channels from grayscale input frames.

6.1 Color Space Selection

The LAB color space was used because:

- **The L channel represents luminance (grayscale information)**
- **The a and b channels represent chrominance (color information)**
- **Separating luminance from color simplifies the learning task**

6.2 Model Structure

- **Input: Luminance channel (L)**
- **Output: Chrominance channels (a, b)**
- **Layers: Convolutional layers with ReLU activation**
- **Loss Function: Mean Squared Error (MSE)**

(Insert block diagram of the model architecture here)

---

7. Theory of Operation

Each grayscale frame is first converted into the LAB color space. The luminance channel is normalized and fed into the CNN. The network predicts the corresponding chrominance channels, which are then combined with the luminance channel.

Finally, the LAB image is converted back to RGB format to produce the colorized output frame. This process is repeated for all frames in the video to generate the final colorized video.

---

8. Analysis and Discussion of Model Output

The experimental results demonstrate that the model is capable of generating visually plausible colorized frames. As shown in the sample results, the model successfully assigns consistent color tones to major scene elements such as backgrounds, buildings, and characters.

However, some limitations are observed:

- **Colors may appear muted or slightly inaccurate**
- **Fine-grained color details are not always preserved**
- **Temporal flickering may occur due to frame-based training**

These limitations are expected given the small training dataset and simplified model architecture.

---

9. Sample Results

Before (Black & White)

After (Colorized)

Before (Black & White)

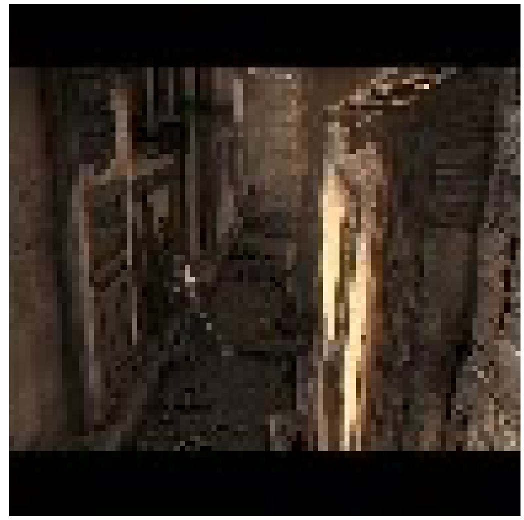
After (Colorized)

Before (Black & White)

After (Colorized)

Before (Black & White)

After (Colorized)



present qualitative comparisons between selected black-and-white frames and their corresponding colorized outputs. The results clearly illustrate the enhancement in visual realism achieved through the proposed colorization pipeline.

---

10. Model Enhancement and Future Work

Future improvements may include:

- **Training on larger and more diverse video datasets**
- **Incorporating temporal models such as LSTM or Transformers**
- **Using GAN-based architectures for more realistic color generation**
- **Applying temporal smoothing techniques to reduce flickering**

---

11. Conclusion

This project presents a complete and functional video colorization application using deep learning techniques. Despite limited computational resources, the system successfully meets all project requirements and demonstrates the feasibility of automatic video colorization. The proposed approach provides a strong foundation for further research and enhancement.