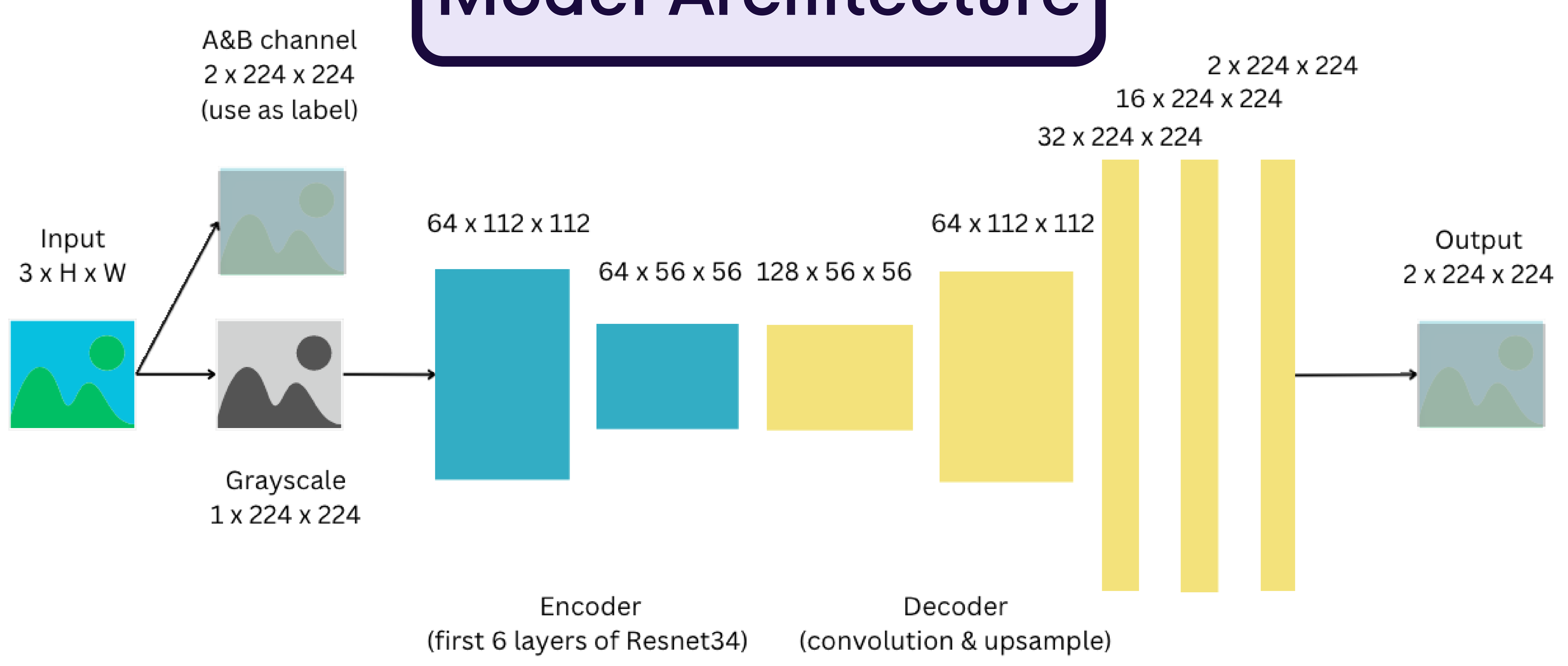
Image Colorization with CNN

This project focuses on developing a deep learning model for colorizing grayscale images using convolutional neural networks (CNNs). The goal is to create a model that can generate realistic and visually-appealing colorized images that accurately represent the original content. The potential applications of this technology are vast, from enhancing old black and white photographs to aiding medical diagnoses and environmental studies.

Model Architecture



- **Encoder - Modified ResNet34**: The first convolutional layer has been adjusted to accept a single channel input by summing over the input channel dimension and unsqueezing the result to add a new singleton dimension. This first 6 layers of the modified ResNet-34 model is used to generate the feature map on the input grayscale image.
- **Decoder - Convolution & Upsampling**: The feature map is then fed into the decoder network, which consists of several convolutional layers with batch normalization and ReLU activation functions. The upsampling layers in the decoder network are used to increase the spatial resolution of the feature maps. The final layer of the decoder produces a two-channel image corresponding to the 'a' and 'b' channels of LAB color space.
- **Justification of Choice**: This architecture is suitable because it leverages the power of pre-trained models while providing a flexible and customizable decoder. The use of the LAB color space, which separates the luminance and chrominance components of the image, provides a better framework for colorizing grayscale images.
- **Improvements**: The Resnet 18 model used in the original implementation was replaced with a larger model. More layers of the pre-trained model were unfreezed during training. The batch size was increased from 64 to 128 to improve training efficiency.

Dataset

ImageNet data　　Processed input data

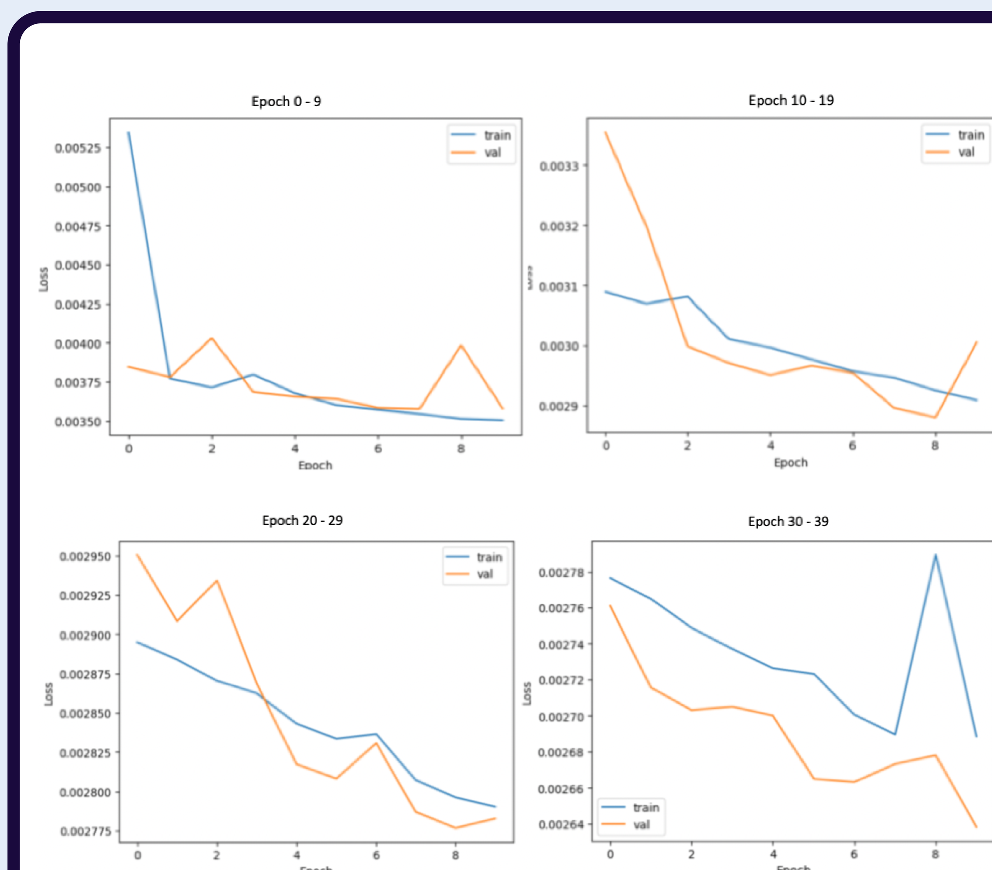
- **Dataset**: ImageNet
  - Contains 14 million images in 1,000 object classes.
  - The images are of high quality and resolution, with a sufficient amount of detail for CNN to learn from.
  - A wide range of color variations, including different shades and hues.
  - Collected using a combination of web crawling and crowdsourcing and cover a wide range of object categories and scenes.
- **Pre-processing & Augmentation**:
  - converted to RGB when opened because the dataset contains grayscale images
  - resized to 224 x 224 to fit the ResNet34 model
  - horizontally flipped with a probability of 0.5 for training to increase dataset size
  - converted from RGB to LAB color space
  - The LAB color values are normalized to 0-1 by adding 128 and dividing by 255
  - Target: The "ab" channels are extracted from the LAB image and converted to a PyTorch tensor, which is used as the target of the algorithm
  - Input: The image is converted to grayscale using the rgb2gray, then converted to a PyTorch tensor and a batch dimension is added

Training Procedure

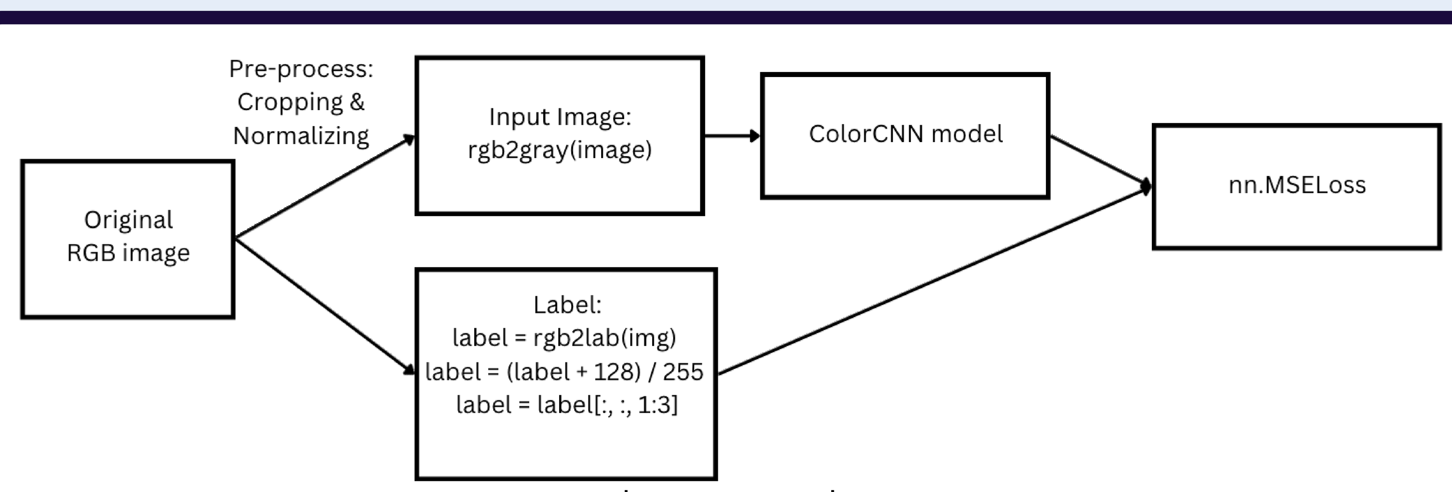
- **Training Process**:
  - Platform: Google Colaboratory, a cloud-based Jupyter notebook environment
  - Hardware: GPU provided by Google Colab (NVIDIA Tesla T4 with 16 GB of memory)
  - Framework: PyTorch
- **Hyperparameters**:
  - Batch size: 128
  - Optimizer: Adam
  - Learning rate: 0.001
  - Loss function: mean squared error (MSE)
- **Techniques**:
  - **Mini-batch training**: 40% of images is randomly selected from the dataset for every 10 epochs. Repeated 4 times for a total of 40 epochs.
  - **Data augmentation**: randomly flipping the image horizontally
  - **Early stopping**: Early stopping was used to stop training if the validation loss did not improve for five consecutive epochs.

Evaluation Procedure

- **Test set**: randoml 10,000 images from the dataset that is not used in train or validation.
- **Metric**: MSE between the predicted and actual a and b values in LAB color space.
- **Qualitative Analysis**: applied to a variety of images that are not from the ImageNet dataset
  - A study is conducted where 5 participants are asked to identify the images colored by AI out of a total of 5 original images and 5 colored images. None of the participants succeeded in identifying all AI-colored images, with an average of 57% accuracy.
  - Fails to capture small details when the background is complex or blurred.
- **Quantitative Analysis**: tested on three distinct image datasets - ImageNet (0.0027), CVPR09 (0.0047), and CelebA (0.0058)
  - Performs well with animals and scenery but poorly with people and in-door scenes

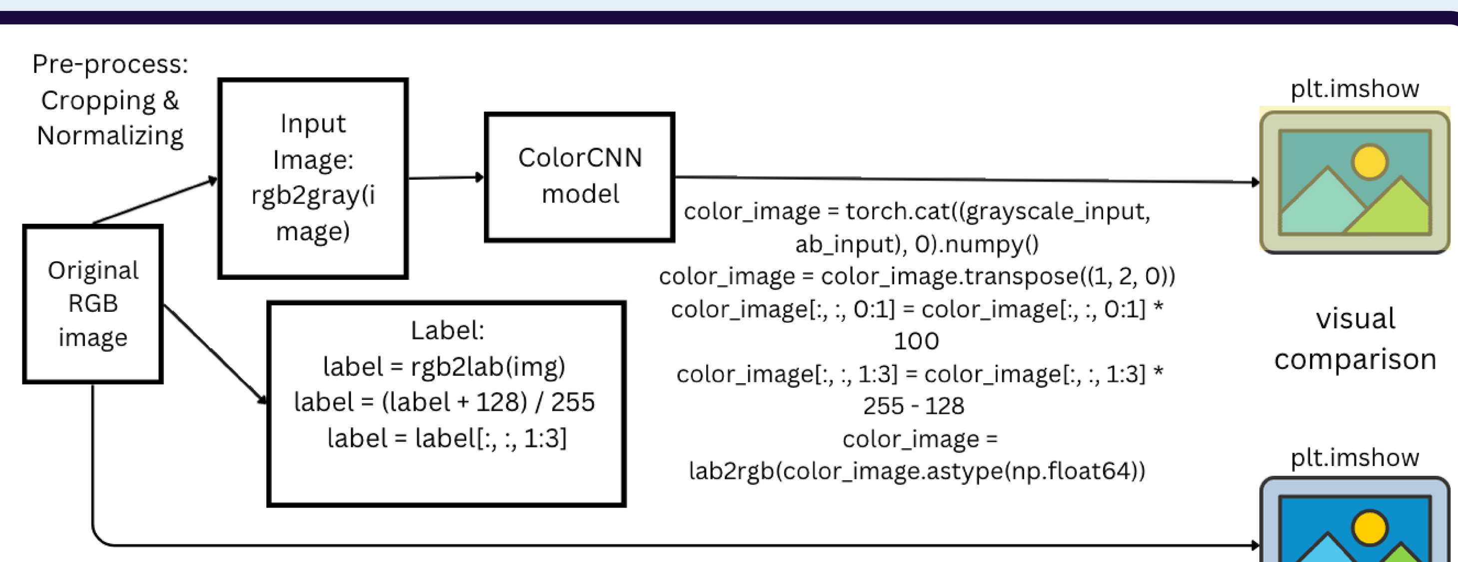


Train & Validation Loss Curve



Evaluation Pipeline

1. Images converted to grayscale for input and to LAB (only A and B) for targets.
2. A data loader for the test dataset.
3. MSE: difference between the output A&B channels and the target.

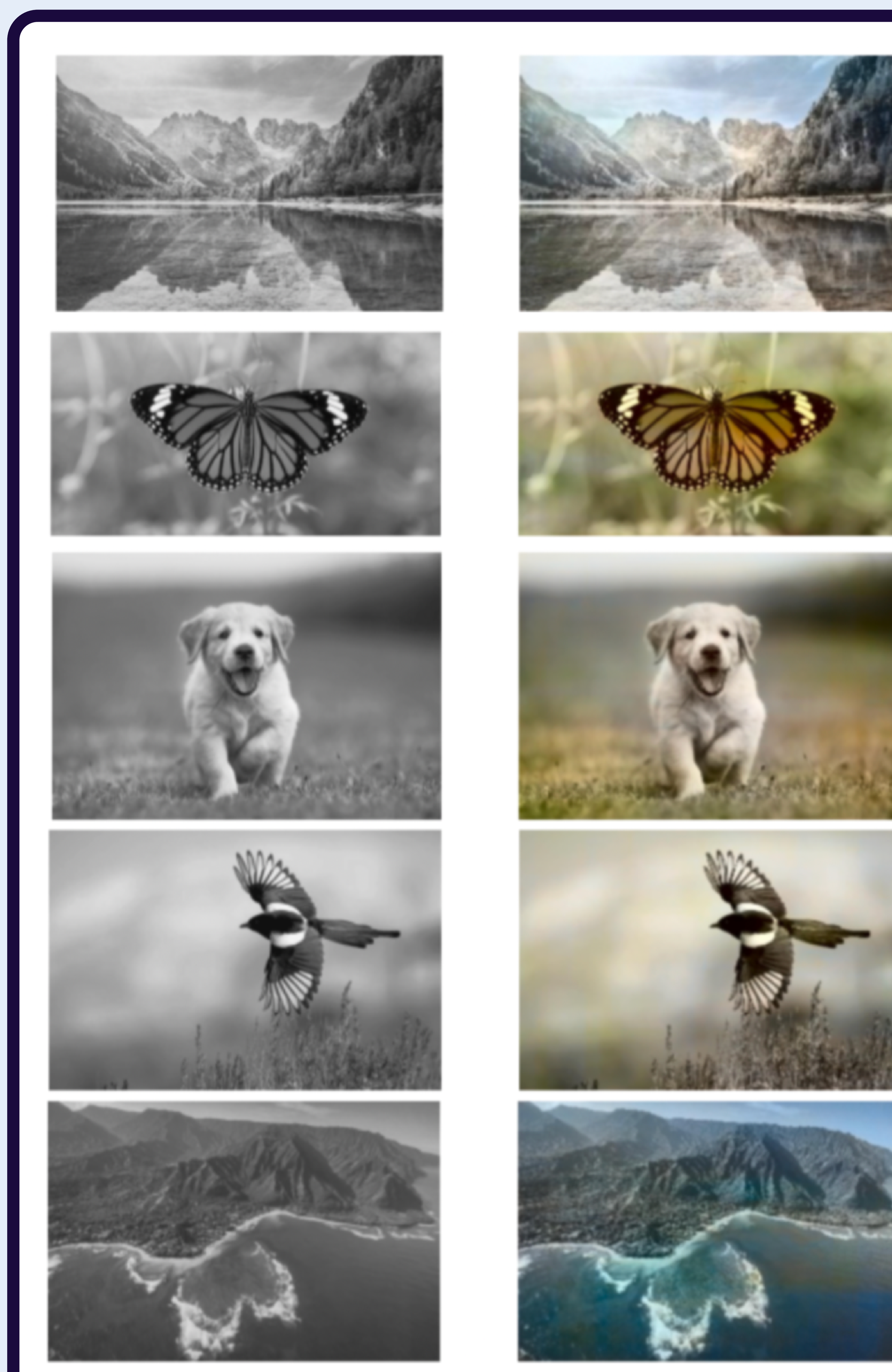


Post-Processing

concatenating the grayscale input and the model output A & B channels to get the LAB form of the image, then convert it to RGB for visualization.

Discussion

- The model produces visually appealing and realistic colorizations, capturing the color distribution in the original images.
- Qualitative analysis indicates that the model performs well with nature scenes and animals, but has limitations in capturing small details in complex or blurred backgrounds and in colorizing images of people and indoor scenes.
- The model demonstrates the potential of deep learning for image colorization tasks.
- The proposed model uses CNN and builds upon previous work in this area, producing satisfactory results with limited resources.
- Further research can be done to develop more sophisticated models that can better capture context and semantics of images and exploring other types of data, such as 3D information, to improve colorization results.
- The model could also be extended to colorize videos, useful for restoring old black-and-white films or improving the visual quality of videos by adding color.



Colorization by the model