



# IMAGE COLORIZATION

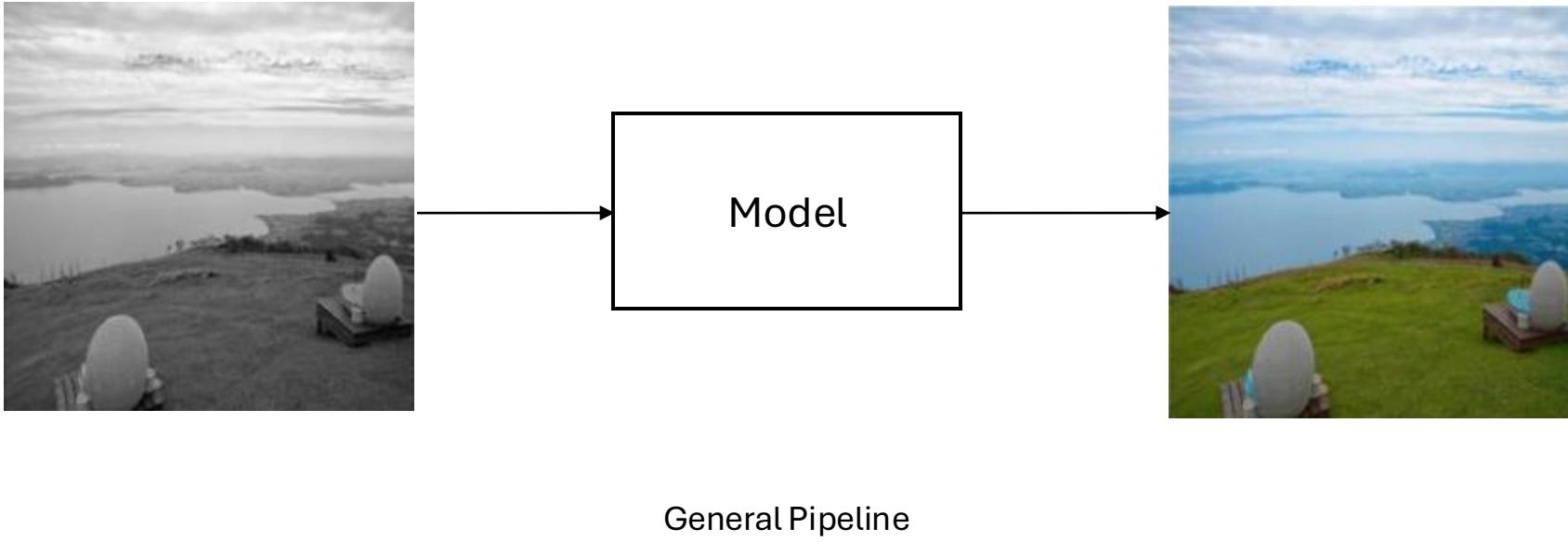
Jayanth Shreekumar

ECE 60131 – Learning and Inference in Generative Models

1st December 2025



# PRELIMINARIES



- 1) Classical methods
- 2) Deep Learning

1) Anwar, Saeed, et al. "Image colorization: A survey and dataset." *arXiv preprint arXiv:2008.10774* (2020).

2) Zhang, Richard, Phillip Isola, and Alexei A. Efros. "Colorful image colorization." *European conference on computer vision*. Cham: Springer International Publishing, 2016.

# PROJECT GOALS AND DATASET

- Experience building several deep learning models.
- Compare behavior across models.
- Gain experience with training, losses, and metrics.



Sample images from the dataset

Dataset used is a custom one scraped off the internet – has 3600 training images, 618 validation images, 64 test images for a total of 4282 color images.

Dataloader converts all images to size 256x256, creates the grayscale version as input, performs normalization, and feeds out both grayscale and color images.

# ARCHITECTURES IMPLEMENTED

## 1) ResNet

- Baseline Model
- Resnet downsample + standard UpSampler with Conv2D layers
- MSE loss

## 2) UNet

- Another Baseline
- Naturally lends itself to colorization
- MSE loss

## 3) VAE

- Used Conv2D layers for the encoder and ConvTranspose2D layers for the decoder
- Used the reparameterization trick to perform backpropagation
- Loss:
  - a) Reconstruction (MSE)
  - b) Regularization (KL Divergence)

$$\mathcal{L} = \underbrace{\text{MSE}(x, \hat{x})}_{\text{reconstruction}} + 0.001 \cdot \underbrace{D_{\text{KL}}(q(z|x) \| p(z))}_{\text{regularization}}$$

$$\mathcal{L} = \frac{1}{N} \sum_{i=1}^N \|\hat{x}_i - x_i\|_2^2 + 0.001 \left( -\frac{1}{2} \cdot \frac{1}{N} \sum_{i=1}^N (1 + \log \sigma_i^2 - \mu_i^2 - \sigma_i^2) \right)$$

# ARCHITECTURES IMPLEMENTED

## 4) UNetVAE

- Used UNet architecture, with skip connections between corresponding down and up layers
- Used the reparameterization trick to perform backpropagation
- Loss:
  - a) Reconstruction (MSE)
  - b) Regularization (KL Divergence)

## 5) GAN

- Used Conv2D layers for the discriminator and a combination of Conv2D and ConvTranspose2D layers for the Generator
- Binary cross entropy loss

$$\text{BCE}(y, \hat{y}) = -(y \cdot \log \hat{y} + (1 - y) \cdot \log(1 - \hat{y}))$$

## 6) UNetGAN

- Used Conv2D layers for the discriminator and UNet for the Generator
- Binary cross entropy loss

Built Diffusion models but they did not work – need more computational power

# EVALUATION METRICS

## 1) Structural Similarity Index Measure (SSIM)

- Measures **preservation of structure, luminance, and contrast** between predicted and ground truth images

## 2) Learned Perceptual Image Patch Similarity (LPIPS)

- **Measures perceptual similarity** using deep neural network features

## 3) Delta E in Lab space ( $\Delta E$ )

- Measures **color difference** between predicted and ground truth in CIE Lab space

SSIM → structure correctness

LPIPS → perceptual realism

$\Delta E$  → color fidelity

1) Wang, Zhou, et al. "Image quality assessment: from error visibility to structural similarity." IEEE transactions on image processing 13.4 (2004): 600-612.

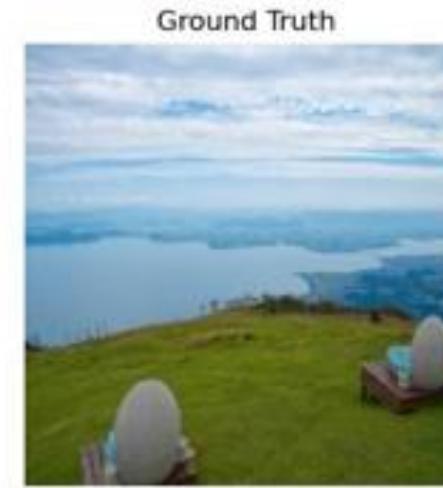
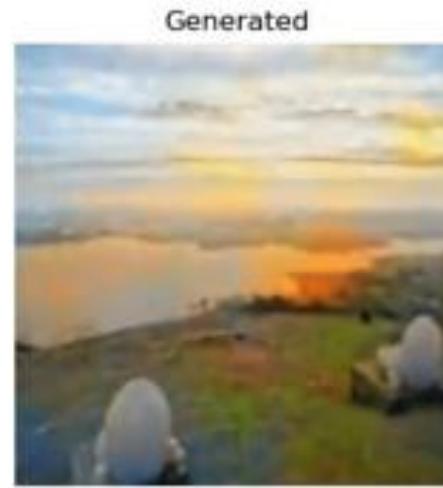
2) Zhang, Richard, et al. "The unreasonable effectiveness of deep features as a perceptual metric." Proceedings of the IEEE conference on computer vision and pattern recognition. 2018.

3) Standard, C. J. I. S. "Colorimetry-part 4: CIE 1976 L\* a\* b\* colour space." International Standard 2007 (2007).

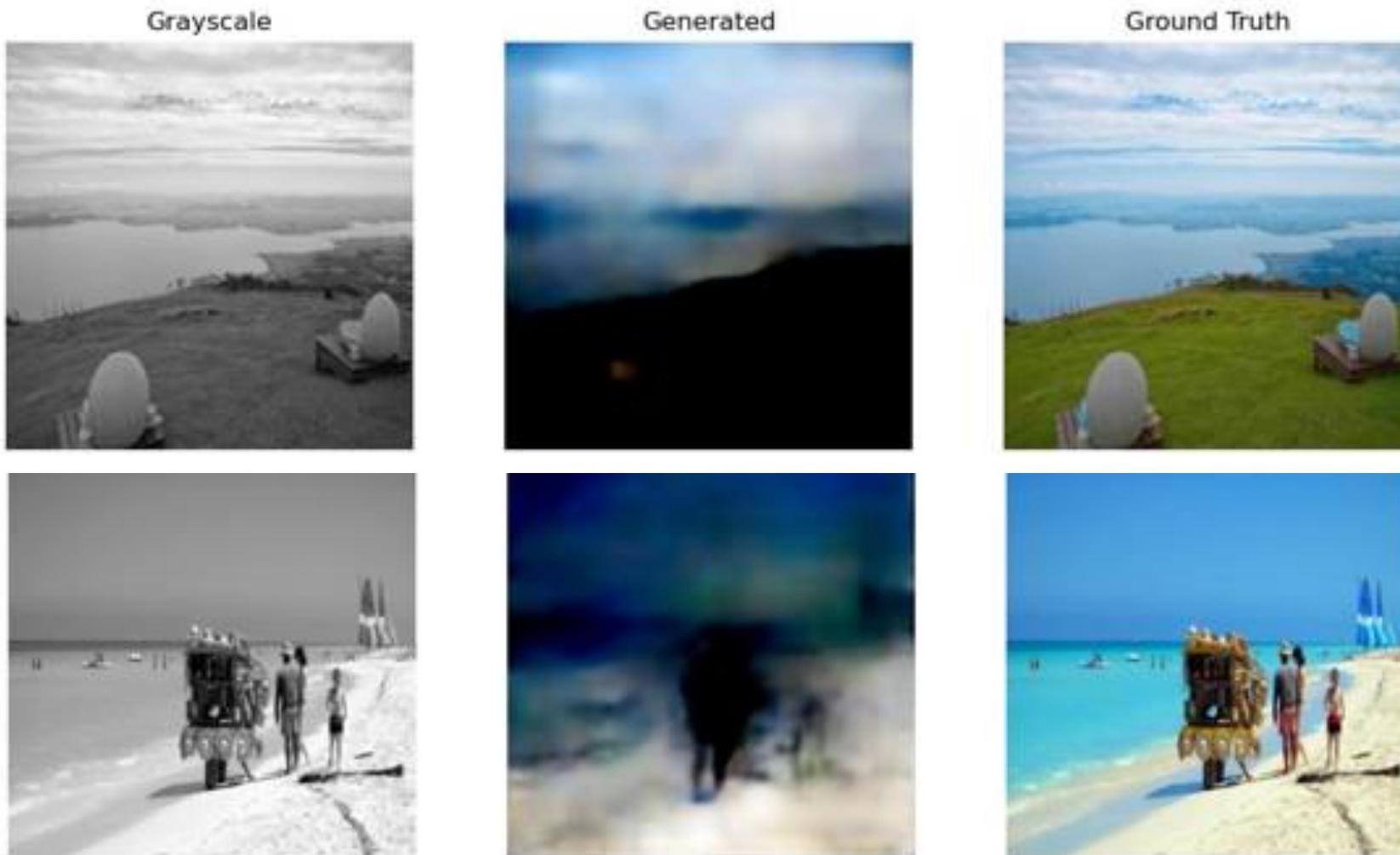
# QUANTITATIVE RESULTS

Model	Loss (no meaning)	SSIM ( $\uparrow$ )	LPIPS ( $\downarrow$ )	$\Delta E$ ( $\downarrow$ )
ResNet	0.213	<i>0.492</i>	0.428	16.533
UNet	0.181	<u>0.678</u>	<u>0.242</u>	<u>15.350</u>
GAN	13.686	0.476	0.463	<i>15.451</i>
UNetGAN	108.944	<b>0.728</b>	<b>0.239</b>	<b>14.386</b>
VAE	0.212	0.175	0.6732	21.834
UNetVAE	0.197	0.312	<i>0.427</i>	19.509

# QUALITATIVE RESULTS - RESNET



# QUALITATIVE RESULTS - VAE



# QUALITATIVE RESULTS - UNETVAE

Grayscale



Generated



Ground Truth



# QUALITATIVE RESULTS - GAN

Grayscale



Generated



Ground Truth



# QUALITATIVE RESULTS - UNET

Grayscale



Generated



Ground Truth



# QUALITATIVE RESULTS - UNETGAN



# QUESTIONS!

---