Convolutional Autoencoder with gamma correction module (GAE) for automatic image colorization

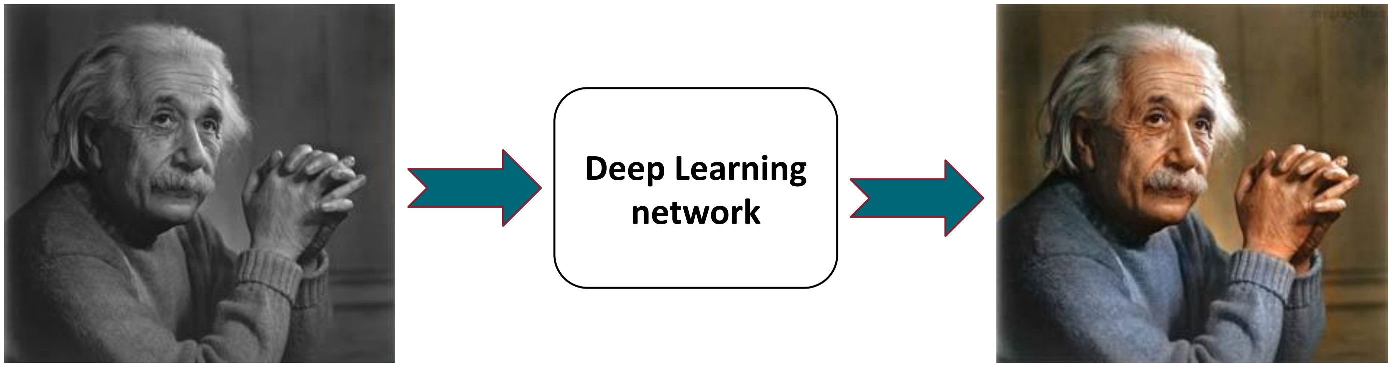
SAPIENZA
UNIVERSITÀ DI ROMA

course: Computer Vision 2023/24
student: Angelo Gianfelice 1851260
PROJECT PRESENTATION

Outline:

- Introduction
- Related Works
- Proposed methods
- Datasets and Metrics
- Implementation details
- Experimental results
- Conclusion and Future Works

Deep Learning Image Colorization



- **Improving many computer vision tasks by augmenting feature extraction, contrast, and interpretability.**

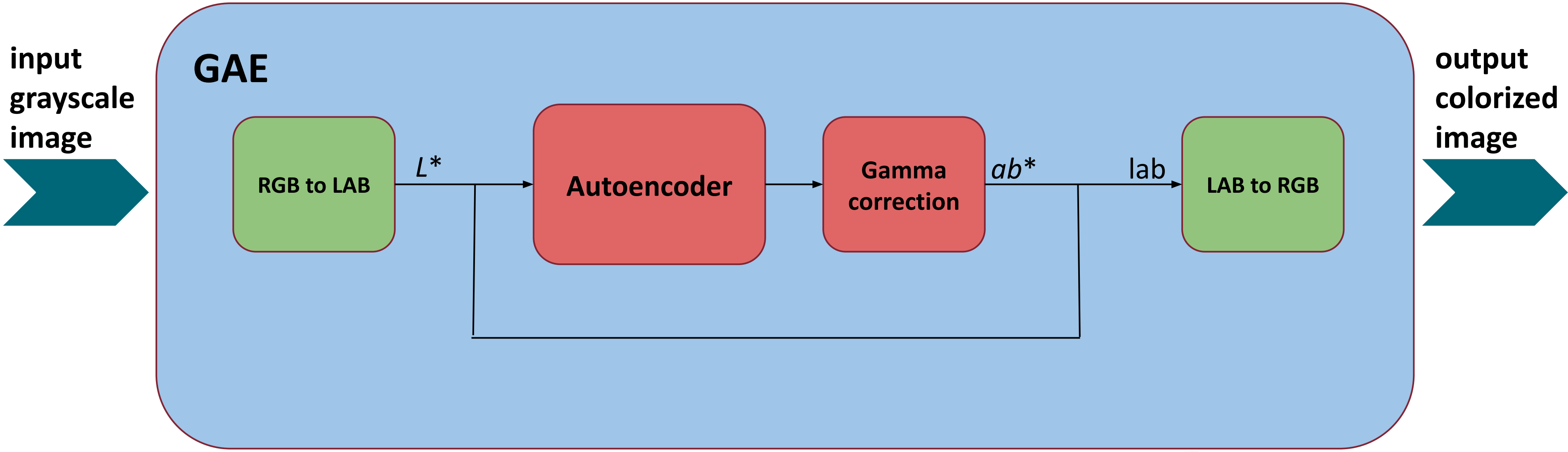
1. **Historical and cultural restoration**
2. **Enhancing medical imaging**
3. **Object Detection & Scene Understanding**

(and much more …)

Related Works

- **Early Approaches**
  - *Levin et al*. (2004): Scribble-based optimization
  - *Irony et al*. (2005): Image retrieval for color transfer
- **CNN-Based Methods**
  - *Zhang et al*. (2016): End-to-end CNN with classification loss
  - *Iizuka et al*. (2016): Global-local context model
- **GAN-Based Colorization**
  - *Nazeri et al.* (2018): Adversarial learning for vibrant colors
  - *Cao et al*. (2017): User-guided GAN for interactive colorization
- **Transformer & Diffusion Models**
  - *Kumar et al*. (2021): Colorization Transformer
  - *Liu et al.* (2023): Diffusion models for semantic colorization

Proposed Architecture (GAE)



SAPIENZA
UNIVERSITÀ DI ROMA

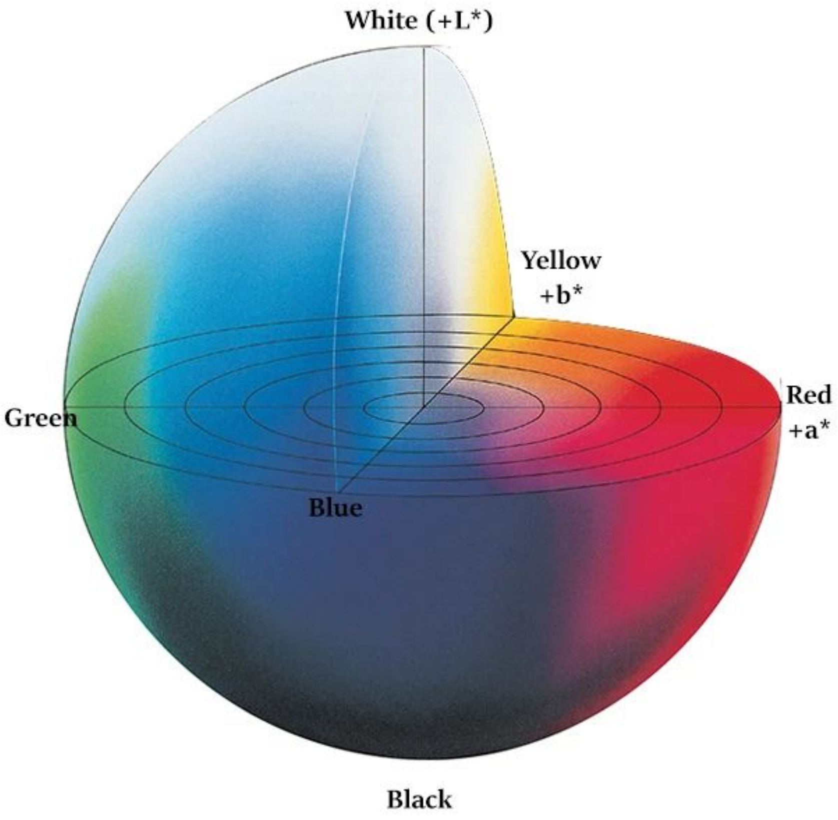
LAB Color Space

- designed to match **human vision** closer than rgb
- LAB color has 3 components:

  L* → Lightness (0 = black, 100 = white)

  A* → Green–Red axis (-128 = green, 127 = red)

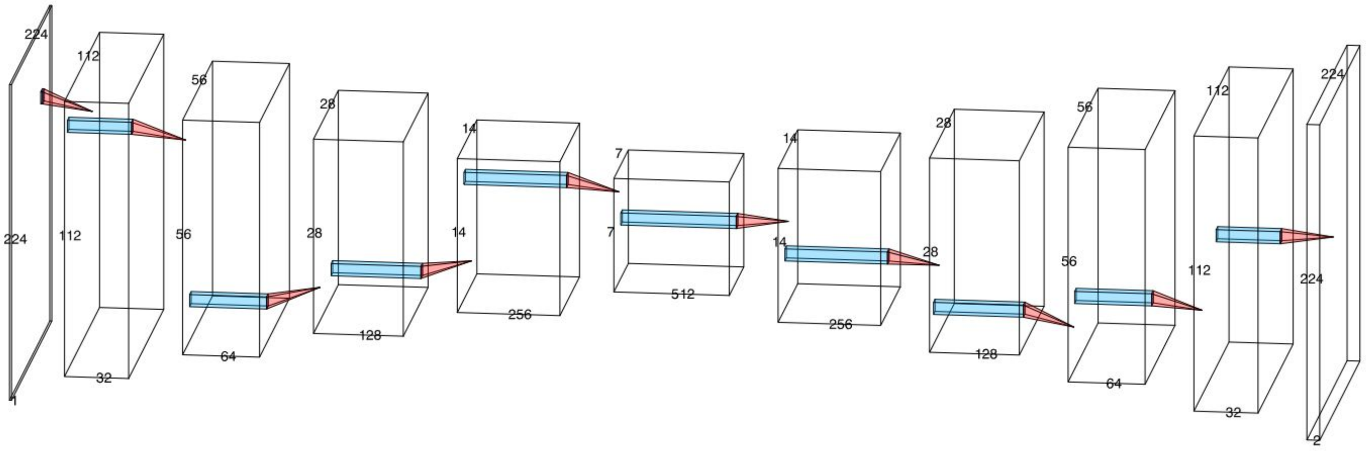
  B* → Blue–Yellow axis (-128 = blue, 127 = yellow)
- color information stored in only 2 channels(ab)
- widely used in **colorization tasks**



SAPIENZA
UNIVERSITÀ DI ROMA

Autoencoder architecture

- **symmetric** 10-layer convolutional autoencoder
- **batch normalization** after every encoder layer
- upsampling through **deconvolution** in decoder
- **Tanh** final activation function to predict a*b* channels (normalized in [-1,1])



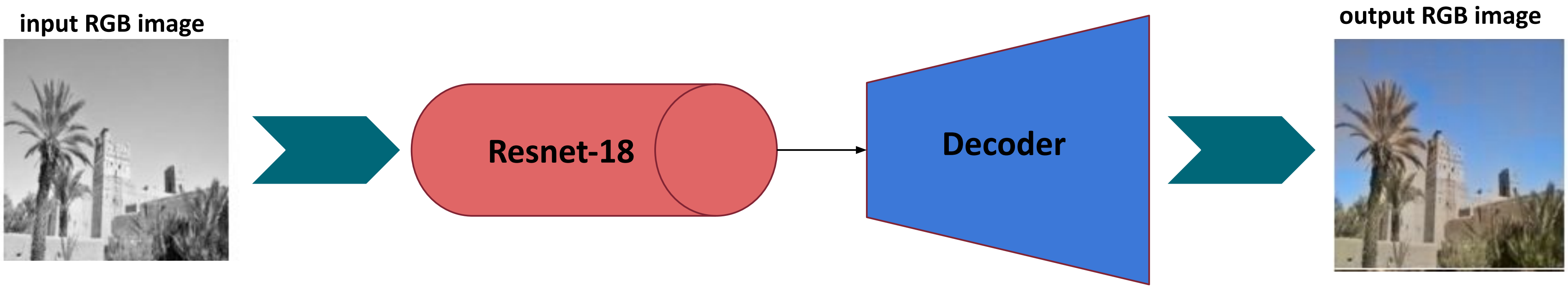
Gamma Correction

- empirically I noticed that colorized images often are very **dull** and **unsaturated,** as the model does very conservative predictions in order to minimize loss

$$X_{out} = X_{in}^{\gamma}$$

- non linear operator to enhance colorization, augmenting **saturation** and **contrast**
- **learnable** γ (initialized to 1)

Baseline



- *Resnet-18* model as a **fixed feature extractor**
- Feature fed directly to **decoder-only** architecture
- faster training, but worst results as only the decoder part is trained for the colorization task

Dataset

- small subset of the **Places dataset,** which contains 10 million images from diverse environments, from natural landscapes to urban settings.

- **30K** image dataset divided in a **70-15-15** split
- resized images to **224x224**
- **data augmentation:** random rotation, flipping and cropping

Evaluation Metrics

- **pixel-wise** metrics:
  - *MSE*: average squared difference between corresponding pixels.

    $$MSE = \frac{\sum_{i=1}^{n}(x_i - \widehat{x}_i)^2}{n}$$

  - *PSNR* (Peak Signal-to-Noise Ratio): Measures image quality in decibels (dB)

    $$PSNR = 10 \cdot \log_{10}\left(\frac{MAX_I^2}{MSE}\right)$$

- **perceptual** metrics:
  - *SSIM*: measures luminance, contrast, and structure similarity.

    $$\mathrm{SSIM}(x,y) = \frac{(2\mu_x\mu_y + c_1)(2\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)}$$

  - *ΔE(76)*: measures perceptual color difference

    $$\Delta E_{ab}^* = \sqrt{(L_2^* - L_1^*)^2 + (a_2^* - a_1^*)^2 + (b_2^* - b_1^*)^2}$$

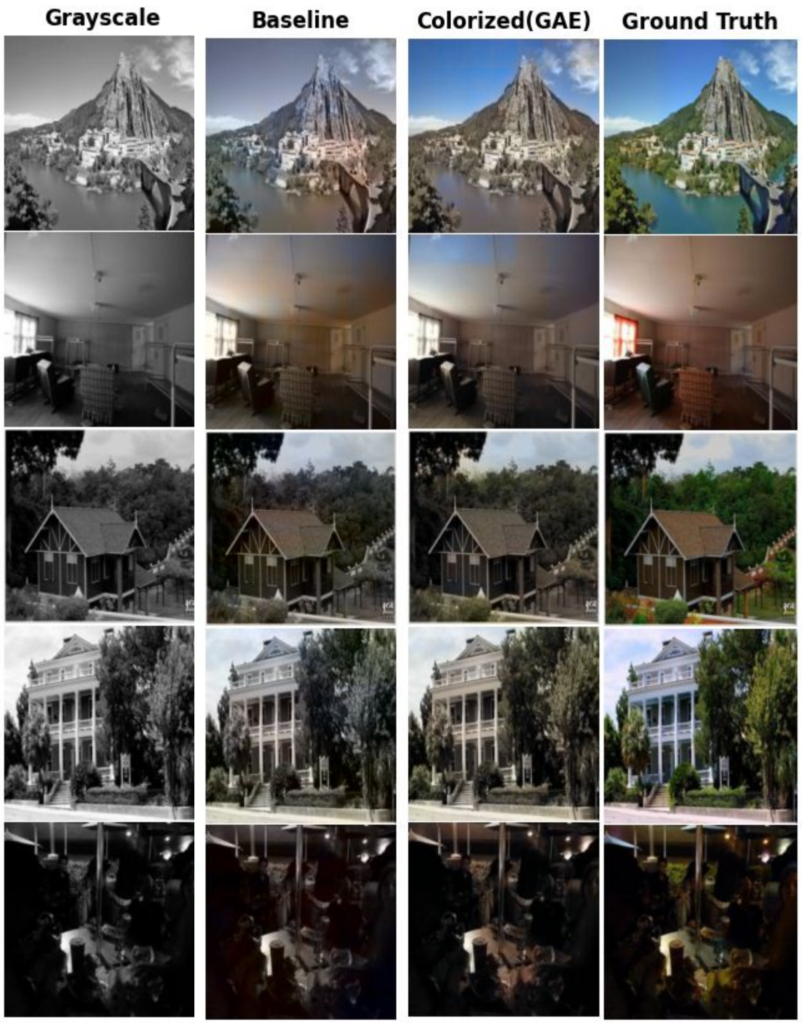
Results: Baseline Comparison

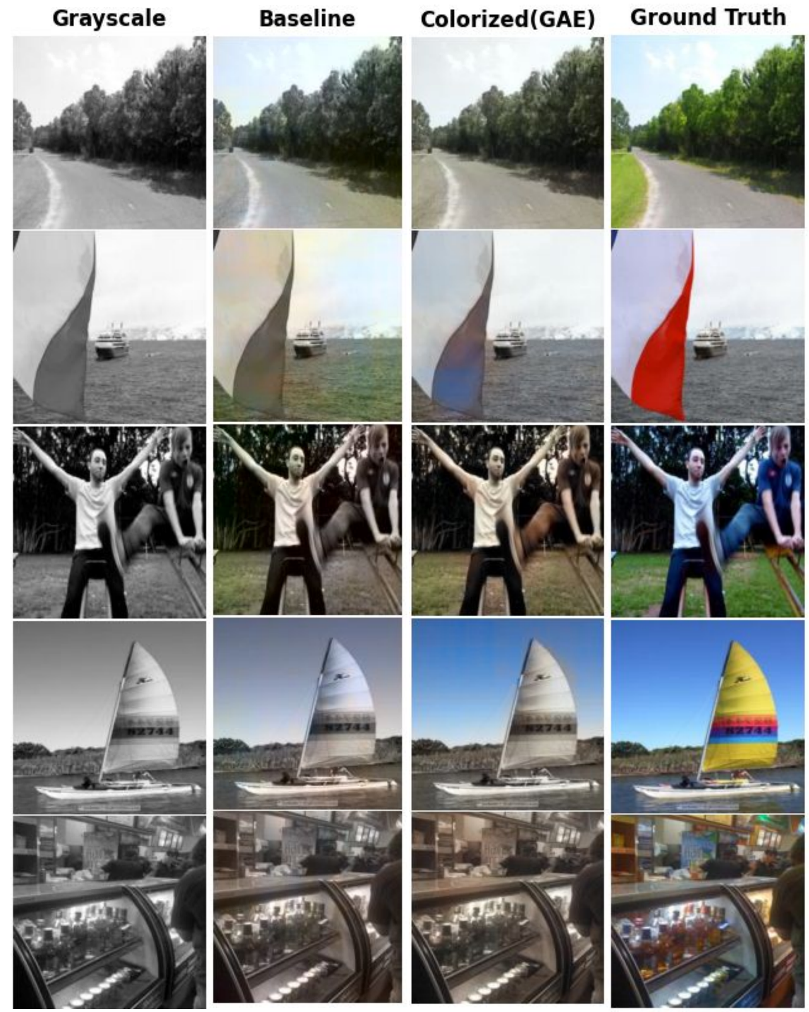
- model tested on **test set** (4500 images) and compared with baseline

	MSE ↓	**PSNR ↑**	**SSIM ↑**	**ΔE ↓**
baseline	0.0063	23.7471	0.9189	13.9822
GAE	**0.0057**	**24.5263**	**0.9408**	**12.9247**

- The proposed architecture outperforms the baseline in every metric by a substantial margin

Results: Colorized Images





Results: Different Experiments

- model trained on more than 4000 **nature/landscapes** images

- model trained on a 5k subset of the **CelebA** dataset

SAPIENZA
UNIVERSITÀ DI ROMA

Historical photograph colorization

- since there exists no ground truth, the results are very hard to judge, but promising

Conclusions and future works

Even though results cannot be comparable to the state of the art, my model is successfully able to colorize images in a **natural** and **vibrant** way, while being relatively **lightweight**. Future improvement might include:

- **Increasing** number of images in **training data** for better generalization
- Use **perceptual loss** for more realistic colors
- **Finetune** model on **different datasets** for different application
- Extend model in order to colorize **videos**

THANKS FOR THE ATTENTION!!

SAPIENZA
UNIVERSITÀ DI ROMA

References

- *Places: A 10 million Image Database for Scene Recognition* B. Zhou, A. Lapedriza, A. Khosla, A. Oliva, and A. Torralba. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017*
- Levin A., Lischinski D., Weiss Y. (2004). Colorization using optimization. In *ACM SIGGRAPH 2004 Papers* (pp. 689–694).
- *Irony, R., Cohen-Or, D., & Lischinski, D. (2005). Colorization by example. Proceedings of the 16th Eurographics Conference on Rendering Techniques, 201–210. https://www.cs.tau.ac.il/~dcor/online_papers/papers/colorization05.pdf*
- *Zhang, R., Isola, P., & Efros, A. A. (2016). Colorful image colorization. European Conference on Computer Vision (ECCV), 649–666. https://doi.org/10.1007/978-3-319-46487-9_40*
- *Iizuka, S., Simo-Serra, E., & Ishikawa, H. (2016). Let there be color! Joint end-to-end learning of global and local image priors for automatic image colorization with simultaneous classification. ACM Transactions on Graphics (TOG), 35(4), 1–11. https://doi.org/10.1145/2897824.2925974*
- *Nazeri, K., Ng, E., & Ebrahimi, M. (2018). Image colorization using generative adversarial networks. arXiv preprint arXiv:1803.05400. https://arxiv.org/abs/1803.05400*
- *Zhang, R., Zhu, J. Y., Isola, P., Geng, X., Lin, A. S., Yu, T., & Efros, A. A. (2017). Real-time user-guided image colorization with learned deep priors. https://doi.org/10.48550/arXiv.1705.02999*
- *Kumar, M., Weissenborn, D., & Kalchbrenner, N. (2021). Colorization transformer. International Conference on Learning Representations (ICLR). https://arxiv.org/abs/2102.04432*
- *Liu, H., Xing, J., Xie, M., Li, C., & Wong, T.-T. (2023). Improved diffusion-based image colorization via piggybacked models. arXiv preprint arXiv:2304.11105. https://arxiv.org/abs/2304.11105*