

From Gray to Vivid: Methods and Metrics for Image Colorization

Beaula Mahima V
MA21BTECH11002

Riya Ann Easow
EE21BTECH11044

Prasham Walvekar
CS21BTECH11047

Kallu Rithika
AI22BTECH11010

Abstract

Image colorization is a fundamental task in computer vision that aims to restore plausible and aesthetically pleasing colors to grayscale images. Recent advances in deep learning have led to the development of generative adversarial networks (GANs), convolutional neural networks (CNNs), and transformer-based approaches for colorization. Despite significant progress, evaluating colorization remains a challenge because of the inherent difficulties in gauging the perceptual and natural qualities of image colorings.

1. Introduction

Color plays a crucial role in computer vision, influencing tasks such as object detection, tracking, and recognition. Image colorization is the process of assigning realistic colors to grayscale images, a task that requires understanding object semantics, lighting conditions, and natural color distributions. Important real-world applications include restoring historical photographs, enhancing medical imaging, improving autonomous driving perception, and aiding creative content generation.

The problem is severely ill-posed as two out of the three image dimensions are missing. While humans intuitively assign colors based on their understanding of the world, automatic image colorization remains a challenging problem due to its inherent ambiguity—many objects can have multiple plausible colorizations. Early colorization methods relied on reference images or user inputs, but recent advances in deep learning have enabled fully automatic approaches.

Deep learning approaches, particularly CNNs, GANs, VAEs, and Transformers, have significantly advanced this field by leveraging large datasets and learning complex image features. However, despite these improvements, challenges remain in balancing realism, diversity, computational efficiency, and handling a wide range of object categories and scenes. Additionally, evaluating the quality of colorized images is difficult, as traditional metrics like PSNR (Peak Signal-to-Noise Ratio) and SSIM (Structural Similarity Index) fail to capture perceptual realism, while

newer metrics like FID (Fréchet Inception Distance) and LPIPS (Learned Perceptual Image Patch Similarity) still have limitations.

2. Literature Review

The automatic image colorization techniques can be broadly divided into 3 categories: Early Architectures, Diverse Colorization Networks, Multi-path Networks. [1]

2.1. Early Architectures

Early colorization models relied on simple convolutional architectures with minimal skip connections, leading to slower learning and higher data requirements. Recent advancements leverage CNNs and deep learning techniques to enhance automation, semantic understanding, and performance across different image modalities.

Cheng et al. [6] introduced Deep Colorization, the first CNN-based model for automatic colorization, eliminating manual input by mapping grayscale features to chrominance values using a large-scale dataset. A joint bilateral filtering step [20] was employed to reduce artifacts. Zhang et al. [26] proposed Colorful Image Colorization, a fully automatic CNN-based approach incorporating class rebalancing to enhance rare colors and prevent desaturated outputs. It also demonstrated colorization as an effective pretext task for self-supervised feature learning, achieving state-of-the-art representation learning performance.

Carlucci et al. [5] introduced Deep Depth Colorization for depth image colorization, enhancing object recognition by learning depth-to-RGB mappings. The $(DE)^2CO$ algorithm outperformed traditional handcrafted approaches, achieving significant performance gains. Hu et al. [12] developed a U-Net-based grayscale image colorization model operating in the Lab color space, where the L component predicts the a and b chrominance values. This method provides a foundation for further advancements in automatic colorization using deep architectures.

2.2. Diverse Colorization Networks

Deep learning based techniques for image colorization primarily employ GANs and VAEs, with methods generating

either multiple diverse colorizations or a single realistic output based on semantic information.

Cao et al. [4] introduced unsupervised diverse colorization using conditional GANs, incorporating noise channels to ensure multiple plausible outputs. Nazeri et al. [18] proposed ICGAN, a fully convolutional network-inspired model that predicts a single realistic colorization using a modified generator loss for improved quality.

Frans [10] developed Tandem Adversarial Networks for line art colorization, utilizing two adversarial networks for color prediction and shading, enhanced by skip connections and L2 or adversarial losses. Deshpande et al. [9] employed a VAE-Mixture Density Network (MDN) to generate diverse colorizations, leveraging specificity, colorfulness, and gradient-based losses.

ChromaGAN [23] utilized self-supervised learning with PatchGAN to predict a single realistic colorization, integrating chrominance prediction and class distribution modeling. MemoPainter [24] incorporated memory networks with GAN-based colorization, using threshold triplet loss to enhance rare object color retrieval.

Li et al. [16] proposed SS-CycleGAN, employing dual generators and discriminators with self-attention and cascaded dilated convolutions for improved structural consistency. Shafiq and Lee [22] introduced a Transformer-GAN hybrid, where a VGG-based encoder captured global semantics, and a Swin Transformer processed color-specific features, refining outputs with perceptual, adversarial, and color losses.

These methods demonstrate diverse architectural strategies to enhance image colorization, optimizing for realism, diversity, and semantic coherence.

2.3. Multi-path Networks

Multi-path networks facilitate learning diverse feature representations but incur higher computational costs.

Iizuka et al. [13] proposed 'Let There Be Color,' a CNN-based model integrating global and local features for resolution-independent image colorization via a joint classification-colorization loss. Larsson et al. [15] utilized hypercolumns from VGG16 to predict per-pixel hue and chroma distributions, employing KL divergence loss for improved semantic coherence.

PixColor [11] introduced a two-stage approach combining a conditional PixelCNN for low-resolution chroma prediction with a refinement CNN for high-resolution colorization. ColorCapsNet [19] enhanced CapsNet with VGG-19 feature extraction, batch normalization, and capsule reduction, learning color distributions in the CIE Lab space.

Pixelated [27] employed a dual-branch network for color embedding and semantic segmentation, utilizing a conditional PixelCNN and multi-scale atrous spatial pyramid pooling. Mohammad et al. [2] proposed a tree-structured

network generating multiple color hypotheses per pixel, leveraging a shared convolutional trunk for efficient color reconstruction.

These methodologies exemplify advancements in deep learning-based colorization, leveraging diverse architectures to enhance quality, efficiency, and semantic accuracy.

2.4. Evaluation Metrics

Evaluating image colorization remains an open challenge, as traditional metrics like PSNR (Peak Signal-to-Noise Ratio) and SSIM (Structural Similarity Index) often fail to capture the perceptual quality of colorized images. Consequently, a number of perceptual and naturalness evaluation metrics have been explored.

However, a number of image colorization models in the literature have been assessed through human evaluation, which remains the gold standard to this day. For example, in [26], a "Colorization Turing Test" was conducted, where human observers were asked to distinguish between real and colorized images. Similar subjective methods were used to evaluate model performance in [8, 13]. However, human evaluation is subjective, time-consuming, expensive, and not scalable, making it impractical for large-scale assessments. Therefore, there is a pressing need for more efficient and scalable metrics tailored specifically to the colorization task.

3. Datasets

Choosing the right dataset is crucial for training an image colorization model that can handle diverse cases effectively.

1. **COCO-Stuff** [3] contains 164k images across 172 categories, covering complex real-world scenes.
2. **ImageNet** [7] consists of 1.2M images spanning 1000 categories, featuring varied objects, lighting conditions, and real-world diversity. This dataset enhances the model's ability to colorize a wide range of objects accurately and generalize well to new images.
3. **Places205** [28] provides 2.5M images across 205 scene categories, focusing on indoor and outdoor environments rather than individual objects. This dataset is valuable for learning scene-based colorization, ensuring the model captures natural color distributions in different locations.

Other datasets were considered but found less suitable. CIFAR-100's low resolution (32×32) is a limitation, while Pascal VOC's 11k images focus on segmentation-based colorization with limited scale and variety. Instead, we choose COCO-Stuff, Places205, and ImageNet for their complementary strengths: Places205 enhances scene understanding, while COCO-Stuff and ImageNet provide diverse, large-scale data for better real-world adaptability. Further, since we do not require any labels for our task, we could

make use of the large amounts of unlabeled data available by web scraping, to model the natural color distribution more accurately.

4. Project Proposal

In this project, we aim to analyze state-of-the-art architectures and explore ways to enhance their performance. The following are some of the ways we plan to explore to achieve the same.

4.1. Monotonic Curriculum

One possible approach to improve the state of the art is to incorporate a monotonic curriculum strategy as done in [17] in the field of deepfake detection. The overview of the pipeline is shown in Fig. 1.

Previous colorization approaches have trained models by simply mixing different datasets together. Instead, inspired by curriculum learning, we propose a gradual data transition strategy that adjusts the proportion of samples from each dataset over the course of training. For this experiment, we use two datasets: ImageNet-100 and COCO-Stuff. Our key idea is based on the nature of these datasets:

- ImageNet-100 primarily focuses on a single object per image, making it well-suited for learning object-specific colors.
- COCO-Stuff, being an object detection dataset, captures more complex scenes with a broader view, better representing natural color distributions.

To leverage these differences, we begin training with mostly ImageNet samples and gradually introduce more COCO-Stuff images. We follow the curriculum schedule from [17]:

$$q(t) = \sin(t/\epsilon) \quad (1)$$

where t is the current training epoch and $\epsilon = 2T/\pi$, ensuring a smooth and monotonic increase in COCO-Stuff data over time. Let the total number of images in a batch be b . The number of images from the ImageNet dataset will be given by $(1 - q(t)) * b$ and the number of images from the COCO-Stuff dataset will be given by $q(t) * b$. Thus, in the initial stages, the model will be trained on mostly ImageNet data. As the number of epochs progresses, images from the COCO-Stuff data will be introduced to the model. This approach helps the model learn the color distributions of the individual objects first and then learn a more general distribution of the color in images.

4.2. Image Aesthetic Metrics for Evaluation

Most colorization studies in the literature primarily use metrics that capture structural similarity (e.g., SSIM) or perceptual quality (e.g., PCQI, LPIPS). However, a key application of image colorization is enhancing the "aesthetic appeal"

of grayscale historical images. Therefore, we believe that incorporating aesthetic quality metrics alongside structural and perceptual evaluation measures would provide a more comprehensive assessment. We plan to use the following three metrics, inspired by the ideas presented in [25], to capture the aesthetic quality of our colorizations.

4.2.1. Colorfulness Index (CI)

The Colorfulness Index (CI) measures the vibrancy and aesthetic appeal of images by analyzing their chromatic properties. It evaluates two key aspects: **chromatic contrast**, which captures variability in color differences (red-green and yellow-blue channels), and **saturation**, which reflects the intensity of colors. Higher CI values correspond to more visually striking and vibrant images, making it an effective tool for assessing aesthetic improvements in tasks like image colorization.

The CI is calculated using the formula:

$$CI = \sqrt{\sigma_{rg}^2 + \sigma_{yb}^2} + 0.3 \times \sqrt{\mu_{rg}^2 + \mu_{yb}^2}$$

where:

- σ_{rg} and σ_{yb} : Standard deviations of red-green ($rg = |R - G|$) and yellow-blue ($yb = |0.5(R + G) - B|$) differences, representing chromatic contrast.
- μ_{rg} and μ_{yb} : Mean values of rg and yb differences, representing color saturation.
- The weighting factor 0.3 balances the contribution of saturation relative to contrast.

4.2.2. Color Harmony Metric

The Color Harmony metric is designed to evaluate the aesthetic quality of an image based on the distribution of hues in its color palette. It operates under the principle that harmonious images tend to have a more cohesive and balanced arrangement of colors, which can be quantified by analyzing the variance in their hue values. The metric converts the image to HSV (Hue, Saturation, Value) color space and calculates the variance of the hue channel. A lower hue variance indicates a more harmonious color scheme, while higher variance suggests greater diversity or imbalance in the hues.

The harmony score is mathematically defined as:

$$\text{Harmony Score} = \frac{1}{1 + \sigma_h^2} \quad (2)$$

Where σ_h^2 is the variance of the hue values in the image. The score is inversely proportional to hue variance, ensuring that images with lower hue variability receive higher scores. This metric provides a simple yet effective way to assess color harmony, with scores typically ranging between 0 and 1, where values closer to 1 indicate stronger harmony.

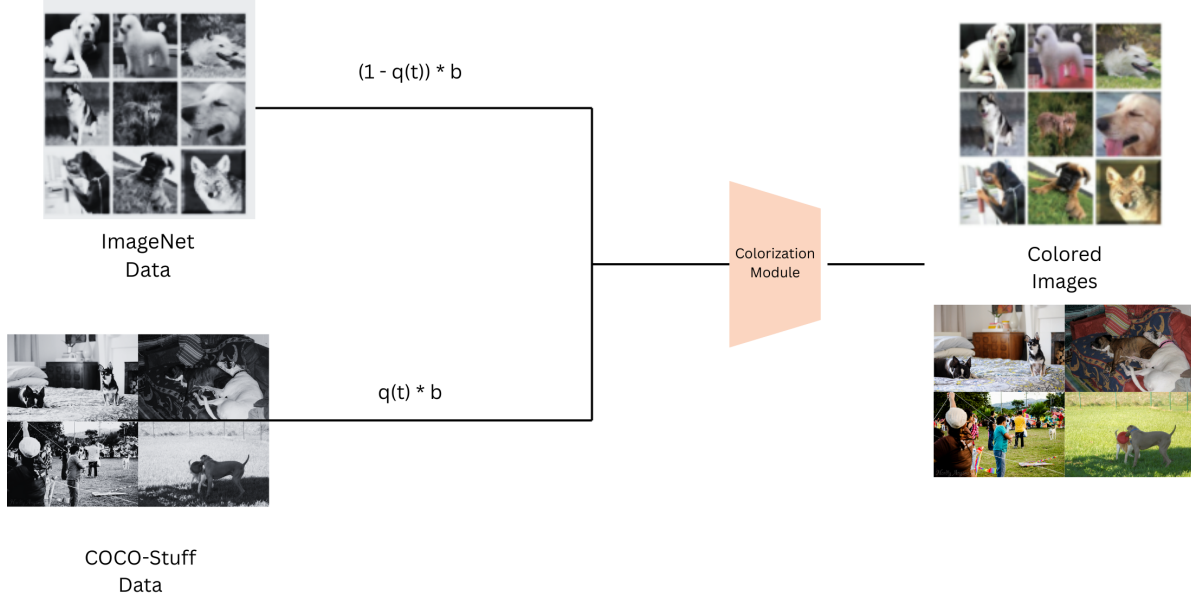


Figure 1. Pipeline for Monotonic Curriculum Strategy for Image Colorization implemented for training the Colorization model. Here, b is the batch size used during training. Finally, from the colored images, loss is computed to update the model.

4.2.3. Color Distribution Balance Metric

The Color Distribution Balance metric evaluates the uniformity of color distribution in an image by calculating the entropy of the hue histogram in the HSV color space. A higher entropy indicates a more balanced and diverse color distribution.

4.3. Ensemble with a Feature Extraction Module

Various state-of-the-art (SOTA) models have been developed, each excelling in different types of images (e.g. natural scenes, portraits, or historical images). However, a single model often struggles to generalize across all types of inputs. A potential solution is to employ an ensemble-based approach that combines multiple SOTA colorization models with a feature extraction module, which dynamically assigns weights to each model’s output to optimize the final colorized image.

Extracting features from the input grayscale image using a lightweight CNN or Vision Transformer (ViT) along with a weighting module (e.g., a small neural network or attention mechanism) that assigns different weights to the outputs of the ensemble models based on extracted features could yield better final colorized images.

4.4. VGG-based Perceptual Loss Metric

In image colorization, relying solely on pixel-wise losses like MSE often leads to blurry or desaturated results, as these losses cannot capture high-level semantic information. To address this, we could incorporate a VGG-based perceptual loss as a regularization term in the overall loss

function. [21]

The perceptual loss compares deep features extracted from a pre-trained VGG network for both the predicted and ground truth color images. These features, taken from intermediate layers, capture textures and object-level details. The loss is defined as:

$$\mathcal{L}_{\text{perceptual}} = \sum_l \|\phi_l(I_{\text{pred}}) - \phi_l(I_{\text{gt}})\|_2^2$$

The total loss used for training is:

$$\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{pixel}} + \alpha \cdot \mathcal{L}_{\text{perceptual}}$$

where α balances the contribution of the perceptual term. This regularization encourages the model to produce colorizations that are not just pixel-accurate but also perceptually realistic and semantically consistent.

4.5. Pretraining with Perceptual Loss (for GANs)

Most recent architectures in colorization literature are GAN-based, but a key challenge is the instability of GAN training, where the generator initially produces low-quality outputs and the discriminator provides weak feedback. A solution to this, inspired by SRGAN (Ledig et al., 2017), is to pretrain the generator with L1 / L2 loss before adversarial training. This allows the generator to first learn a stable grayscale-to-color mapping, preserving structural details before adversarial loss refines the realism of the generated images. This approach enhances training stability, prevents mode collapse, and improves colorization quality.

Additionally, initializing the generator’s encoder with a pre-trained model like ResNet further strengthens performance by leveraging learned feature representations, ensuring a more robust starting point.

5. Preliminary Experiments

The first step towards progressing with the above proposed approaches is to get a better understanding of the STOA models in literature. For the same, we have tested out the following two models for the image colorisation task.

5.1. Colorful Image Colorisation (Model 1)

This is a CNN-based colorization model [26] that maps grayscale input to a distribution over quantized color outputs. Instead of using L2 loss (which tends to produce desaturated results due to color ambiguity), the model uses a classification approach, treating color prediction as multinomial classification over 313 quantized color bins in the CIE Lab color space.

To address the class imbalance in natural images (bias toward desaturated colors due to dominance of objects like clouds), they apply loss reweighting based on color rarity using a smoothed empirical distribution. Finally, the model maps the predicted distribution to color values using a temperature-controlled annealed-mean, balancing vibrancy (mode) and spatial consistency (mean). The weights used to train this model were obtained by training on ImageNet data.

5.2. CGAN with U-Net (Model 2)

The CGAN with U-Net approach (the link to the blog can be found [here](#)) follows a two-step strategy for image colorization using deep learning. First, the author implements a U-Net-based conditional GAN (cGAN), following the approach from an existing research paper [14]. The generator is a U-Net, which consists of an encoder-decoder structure with skip connections, gradually downsampling and upsampling the input. This model takes grayscale images and predicts the two missing color channels. The discriminator is a PatchGAN, which classifies 70x70 patches of the generated image as real or fake rather than evaluating the entire image. The training process alternates between optimizing the discriminator and the generator, where the generator minimizes a combination of adversarial loss (to fool the discriminator) and L1 loss (to ensure pixel-level accuracy). Despite producing reasonable results after 100 epochs, the baseline model struggled with rare objects, color spillovers, and unnatural color distributions.

So, in order to enhance the model’s performance, the author introduces a super-resolution-inspired pretraining strategy to improve the generator. Instead of training the generator from scratch, a ResNet18 model pretrained on ImageNet is used as the encoder (downsampling path) of the

U-Net. The generator is then pretrained with only L1 loss on the colorization task before integrating it into the GAN framework. This pretraining stabilizes the GAN training by providing the generator with a more informed initialization, mitigating the “blind leading the blind” problem where both the generator and discriminator start with no knowledge. The pretrained generator is then fine-tuned within the adversarial framework, leading to significantly improved results. This approach reduces the training time (from 100 epochs to around 20) and enhances the model’s ability to predict realistic colors, even for less common objects. The weights used to train this model were obtained by training on COCO dataset.

5.3. Experiment Results

The evaluation of both the models was done on 1000 images of the COCO dataset, using pretrained weights.

Table 1. Performance Comparison.

Metrics	Model 1	Model 2
SSIM \uparrow	0.86 ± 0.060	0.80 ± 0.069
PCQI \uparrow	1.44 ± 0.37	1.72 ± 0.44
Colorfulness \uparrow	115.61 ± 46.64	135.87 ± 36.43
Color Harmony \downarrow	0.004 ± 0.009	0.007 ± 0.070
Color Distribution Balance \uparrow	1.53 ± 0.13	1.65 ± 0.18

The table reports mean \pm standard deviation for each metric. Model 1 corresponds to **Colorful Image Colorization**, and Model 2 to **cGAN**.

5.4. Code References

- For evaluating the models as part of our initial experiments, we have written a custom python script (which is a part of our submission).
- **Model-1: Colorful Image Colorisation:** The code can be found [here](#)
- **Model 2: CGAN with U-Net:** The code can be found [here](#)

References

- [1] Saeed Anwar, Muhammad Tahir, Chongyi Li, Ajmal Mian, Fahad Shahbaz Khan, and Abdul Wahab Muzaffar. Image colorization: A survey and dataset. *Information Fusion*, 114: 102720, 2025. 1
- [2] Mohammad Haris Baig and Lorenzo Torresani. Multiple hypothesis colorization and its application to image compression. *Computer Vision and Image Understanding*, 164:111–123, 2017. 2
- [3] Holger Caesar, Jasper Uijlings, and Vittorio Ferrari. Coco-stuff: Thing and stuff classes in context. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1209–1218, 2018. 2

- [4] Yun Cao, Zhiming Zhou, Weinan Zhang, and Yong Yu. Unsupervised diverse colorization via generative adversarial networks. In *Machine Learning and Knowledge Discovery in Databases: European Conference, ECML PKDD 2017, Skopje, Macedonia, September 18–22, 2017, Proceedings, Part I 10*, pages 151–166. Springer, 2017. 2
- [5] Fabio Maria Carlucci, Paolo Russo, and Barbara Caputo. 2co: Deep depth colorization. *IEEE Robotics and Automation Letters*, 3(3):2386–2393, 2018. 1
- [6] Zezhou Cheng, Qingxiong Yang, and Bin Sheng. Deep colorization. In *Proceedings of the IEEE international conference on computer vision*, pages 415–423, 2015. 1
- [7] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition*, pages 248–255. Ieee, 2009. 2
- [8] Aditya Deshpande, Jason Rock, and David Forsyth. Learning large-scale automatic image colorization. In *Proceedings of the IEEE international conference on computer vision*, pages 567–575, 2015. 2
- [9] Aditya Deshpande, Jiajun Lu, Mao-Chuang Yeh, Min Jin Chong, and David Forsyth. Learning diverse image colorization. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 6837–6845, 2017. 2
- [10] Kevin Frans. Outline colorization through tandem adversarial networks. *arXiv preprint arXiv:1704.08834*, 2017. 2
- [11] Sergio Guadarrama, Ryan Dahl, David Bieber, Mohammad Norouzi, Jonathon Shlens, and Kevin Murphy. Pixel color: Pixel recursive colorization. *arXiv preprint arXiv:1705.07208*, 2017. 2
- [12] Z Hu, O Shkurat, and M Kasner. Grayscale image colorization method based on u-net network. *Int. J. Image Graph. Signal Process*, 16:70–82, 2024. 1
- [13] Satoshi Iizuka, Edgar Simo-Serra, and Hiroshi Ishikawa. 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, 2016. 2
- [14] Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A Efros. Image-to-image translation with conditional adversarial networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1125–1134, 2017. 5
- [15] Gustav Larsson, Michael Maire, and Gregory Shakhnarovich. Learning representations for automatic colorization. In *Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part IV 14*, pages 577–593. Springer, 2016. 2
- [16] Bin Li, Yi Lu, Wei Pang, and Huixin Xu. Image colorization using cyclegan with semantic and spatial rationality. *Multimedia Tools and Applications*, 82(14):21641–21655, 2023. 2
- [17] Yuzhen Lin, Wentang Song, Bin Li, Yuezun Li, Jiangqun Ni, Han Chen, and Qiushi Li. Fake it till you make it: Curricular dynamic forgery augmentations towards general deepfake detection. In *European Conference on Computer Vision*, pages 104–122. Springer, 2024. 3
- [18] Kamyar Nazeri, Eric Ng, and Mehran Ebrahimi. Image colorization using generative adversarial networks. In *Articulated Motion and Deformable Objects: 10th International Conference, AMDO 2018, Palma de Mallorca, Spain, July 12–13, 2018, Proceedings 10*, pages 85–94. Springer, 2018. 2
- [19] Gokhan Ozbulak. Image colorization by capsule networks. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, 2019. 2
- [20] Georg Petschnigg, Richard Szeliski, Maneesh Agrawala, Michael Cohen, Hugues Hoppe, and Kentaro Toyama. Digital photography with flash and no-flash image pairs. *ACM transactions on graphics (TOG)*, 23(3):664–672, 2004. 1
- [21] Rahul Sankar, Ashwin Nair, Prince Abhinav, Siva Krishna P Mothukuri, and Shashidhar G Koolagudi. Image colorization using gans and perceptual loss. In *2020 international conference on artificial intelligence and signal processing (AISP)*, pages 1–4. IEEE, 2020. 4
- [22] Hamza Shafiq and Bumshik Lee. Transforming color: A novel image colorization method. *Electronics*, 13(13):2511, 2024. 2
- [23] Patricia Vitoria, Lara Raad, and Coloma Ballester. Chromagan: Adversarial picture colorization with semantic class distribution. In *Proceedings of the IEEE/CVF winter conference on applications of computer vision*, pages 2445–2454, 2020. 2
- [24] Seungjoo Yoo, Hyojin Bahng, Sunghyo Chung, Junsoo Lee, Jaehyuk Chang, and Jaegul Choo. Coloring with limited data: Few-shot colorization via memory augmented networks. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 11283–11292, 2019. 2
- [25] Jiajing Zhang, Yongwei Miao, and Jinhui Yu. A comprehensive survey on computational aesthetic evaluation of visual art images: Metrics and challenges. *IEEE Access*, 9:77164–77187, 2021. 3
- [26] Richard Zhang, Phillip Isola, and Alexei A Efros. Colorful image colorization. In *Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part III 14*, pages 649–666. Springer, 2016. 1, 2, 5
- [27] Jiaojiao Zhao, Jungong Han, Ling Shao, and Cees GM Snoek. Pixelated semantic colorization. *International Journal of Computer Vision*, 128:818–834, 2020. 2
- [28] Bolei Zhou, Agata Lapedriza, Jianxiong Xiao, Antonio Torralba, and Aude Oliva. Learning deep features for scene recognition using places database. *Advances in neural information processing systems*, 27, 2014. 2