Image Synthesis from image-to-image translation of coarsely quantized images

Akshay Dodwadmath[1] and Manuela Ceron Viveros[2]

[1] Saarland University, Saarbrücken, Germany
akdo00001@stud-uni-saarland.de
[2] Saarland University, Saarbrücken, Germany
mace00001@stud-uni-saarland.de

Abstract. We propose a method for generation of dequantized or finely-quantized images. We cast this as an image-to-image translation task using conditional adversarial networks. Compared to standard image generation problems, which often need all the information in an image, our generation task is dependent on only the color features in image, which assists the model to perform finer color quantization as well as reduces its complexity. To the best of our knowledge, this is the first work to focus solely on dequantization of images. Using images from a custom dataset with limited data, we demonstrate that our approach can generate finely quantized images, while outperforming the baseline. We further show that our method can be extended to applications such as video regeneration.

Keywords: Quantization · Conditional GAN · LAB color space.

1 Introduction

Image quantization is the process of discretizing the codomain of an image by mapping a range of color or grey values to a single value. A *coarser* quantization is when the image has a reduced color/grayscale palette, while a *finer* quantization or dequantization is when the image has an enhanced color/grayscale palette.

The motivation for our task is two-fold: (i) Extensive time and expert effort are currently required to create high-quality, finely quantized graphics assets, so it would be valuable to have a trainable model that can synthesize such assets from their coarser counterparts with minimal human supervision; and (ii) Image quantization is a commonly used lossy compression technique, and recovery of the original image from the quantized version is an ill-posed problem, since quantization induces a many-to-one mapping of pixel intensities. If a model could closely reconstruct the source image from its coarser quantized version, this could reduce the loss incurred due to image compression.

In this paper, we show that the task of finer quantization can be performed in a supervised manner by modelling it as an image-to-image translation task.

Ground Truth Input Output Ground Truth Input Output

Fig. 1. Here we show some results of the proposed model. In each case we show the original image, its coarsely-quantized version and the generated image that is finely-quantized.

We use a conditional adversarial network to perform the input-output mapping based on the pix2pix[5] network. We further separate out the color information in an image. This allows us to better constrain the color de-quantization process. At inference time, we generate finer quantized images without seeing the original images. We evaluate the quality of our synthesized images both quantitatively and qualitative visualization. Our experiments show that the images generated by our method are more colorful and comparable to the original set of images. We further show that a sequence of generated images can be used for retro video game regeneration.

2 Related work

Multiple approaches have been proposed to perform coarse color quantization in images. [7] translates images to LAB space and reduce color content by growing regions with local color average, and reducing number of colors between regions. We also make use of the LAB space, but only make use of the color channels. [6]reduces the number of colors to a small number by using octree quantization. In contrast to methods such as [7] and [6] we are increasing the color content in images rather than reducing it, which we feel is a harder task as additional information needs to be learnt.

Using conditional adversarial networks as general purpose solution for image-to-image-translation was proposed by [5]. They propose a network to learn mapping from input to output and perform image translation for different applications. We extend its use for the task of finer color quantization by integrating it into our network. [9] also performs image to image translation using conditional GAN but makes use of unpaired data.

A closely related task to ours is the Colorization of images which has been researched using different methods. [8] obtains feature vector for each pixel in grayscale image and use SVMs to obtain the geometric margin of the feature vector for each possible color. [5] also performs colorization of images using conditional GANs. Compared to colorization, our task is a less ill-posed problem as we have a general idea of the color content needed in the image.

3 Proposed method explained

Let

$$x, x' \in X = R^{\mathrm{HxWxC}}$$

be the RGB images of an object: coarsely quantized image and finely quantized image. We call x the source image and *x'* the target image. We transform to the LAB color space and separate out the luma and color information(channels) from both images, and name them as l and c for source image, *l'* and *c'* for target image.

We are interested to learn a function ϕ(c) that maps the color information from the source image to color image. In order to learn the map ϕ, we consider the problem of conditional image generation. Namely, we wish to learn a generator function

$$\phi(c) : (c; c') \rightarrow c''$$

such that *c"* is an approximation of *c'*

$$(c'' \approx c')$$

We finally combine the generated color channel with the luma information from the source image and convert back to RGB color space to synthesize an image *x"* which will be an approximation of target image
i.e.

$$l + c'' = x''$$

such that

$$x'' \approx x'$$

3.1 Model development

In Figure 2 is diagrammed a high-level architecture of our model Pix2Pix-FQ. It consists of a pre-processing color disentanglement module, a conditional-GAN model(pix2pix) and a color entanglement module as a post-processor. We also add an an optional video generator module which can be used to convert a sequence of generated frames into a finely-quantized video.

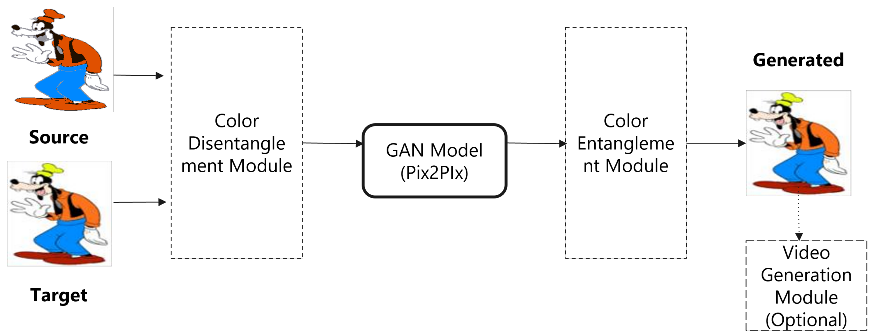

Fig. 2. Pix2Pix-FQ model architecture.

Color Disentanglement Module The aim of this module is to transform the input (source and target RGB images x and x') into LAB color space, and separate brightness information of Luma channel l and l' from color data of AB channels c and c'. c and c' are then passed on as input to the GAN model, whereas l and l' are stored in a buffer and are not considered by the training process (See Fig. 3).

GAN Model The GAN model is trained to learn the mapping between c and c' and synthesize a 2-channel image c''. We have considered the conditional-GAN model pix2pix for our task, but it could be replaced with other conditional-GAN models which can perform image to image translation task.

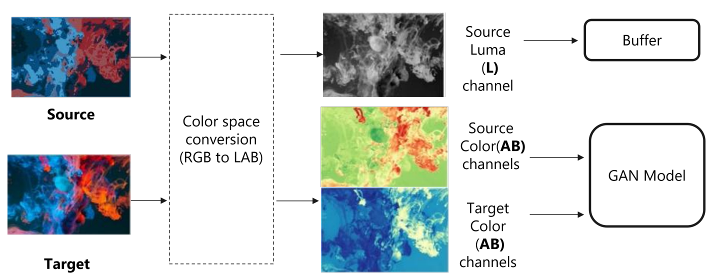

Fig. 3. Color Disentanglement Module.

Color Entanglement Module The GAN model generates fake images consisting of only color information(c''). Further, this generated 2-channel image is joined with the luma channel of source image (l), converted back to RGB color space to reproduce the desired finely-quantized image (See Fig. 4).

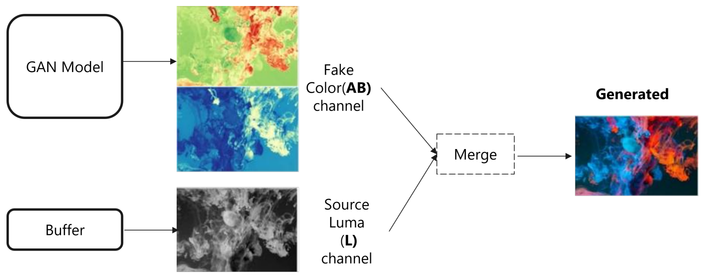

Fig. 4. Color Entanglement Module.

4 Experimental results

4.1 Dataset generation

Due to lack of available datasets suitable for our task, we developed custom made datasets. We used original images from low resolution datasets Linnaeus 5[2] and Akvelon Cartoon Dataset[1] and downsampled images from high resolution datasets DIV2K[3] and Holopix50[4]. A fine-to-coarse quantization module was added to generate the coarsely-quantized counterpart images. The original images are of RGB format consisting of 24 channels; the module generates quantized images with the quantization level set in a random manner ranging from 10% to 50%(12 to 22channels). The final resolution of both the original and quantizated images used for training was 256x256 .

4.2 Baseline

As baseline, we setup the Pix2Pix model without disentanglement of color information. It is trained on the original RGB images of coarsely quantized-finely quantized pair. We set all the training parameters same for both the baseline and our model.

4.3 Qualitative results

In Figure 5 we show some results from our model. The left side of each pair represents the coarsely-quantized image from the Linnaeus 5 dataset, whereas the right side shows the generated image. The model is able to generate images with increased color content in different parts of the object. For instance, the generated bird image recovers more colors in bird feet, breast and in the right most background.

We also compare generated images from baseline and Pix2Pix-FQ. We found that in some cases the baseline is not able to generate complete color information, as can be seen in Figure 6, The baseline does not retain some textures or edges, and even creates fake colors in some parts. On the other hand, our model generates more colorful images and also with more detail, for instance in

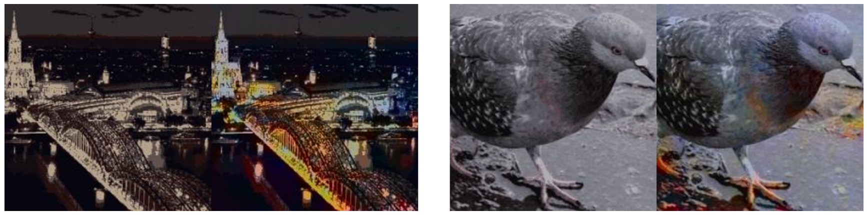

Fig. 5. In each pair of images: Coarsely-quantized (Left) and Finely-quantized image generated by Pix2Pix-FQ (Right).

the forehead, eyes, and legs of the dog where the edges and details are better defined.

Fig. 6. From left to right: (a) Coarsely-quantized, (b) Original finely-quantized, (c) Finely-quantized image generated by Pix2Pix and (d) by Pix2Pix-FQ model.

4.4 Quantitative results

In Figure 7, we compare generator and discriminator training losses between the baseline and our model Pix2Pix-FQ. We get a discriminator error range between 0 and 1. Our model exhibits more stability with respect to the loss functions, which can be seen in the loss curves: the generator loss keeps increasing for the baseline, whereas it fluctuates within a range for Pix2Pix-FQ. The L1-loss is also reduced for our model.

We used different parameters to evaluate the performance of the models. Frechet Inception Distance (FID) is a popular metric for evaluating the quality of generated images and specifically developed to evaluate the performance of generative adversarial networks. We also used the Peak Signal-to-Noise Ratio (PSNR) and Mean Square Error(MSE) that measure quality of image reconstruction. Table 1 shows performance differences between Pix2Pix and Pix2Pix-FQ. Both models were trained on custom Linnaeus training dataset, consisting of 1500 images for 200 epochs. Our model is able to outperform the baseline in all metrics, getting lower FID and higher PSNR and MSE values.

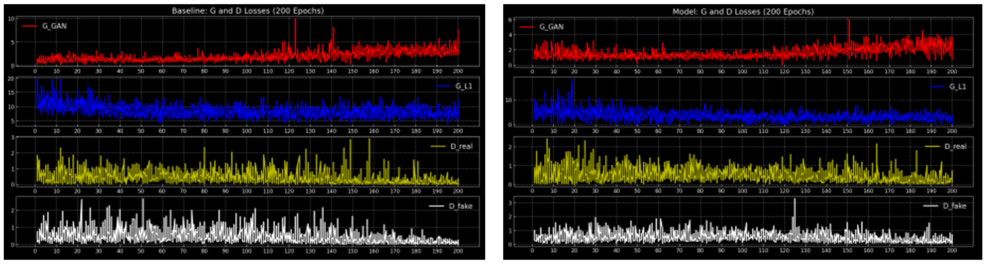

Fig. 7. Pix2Pix (Left) and Pix2Pix-FQ (Right) Training Losses.

Table 1. Performance comparison between Pix2Pix and Pix2Pix-FQ models. For FID, lower the better. For PSNR and MSE, higher the better

Dataset	Model	FID	PSNR	MSE
Custom Linnaeus (2200 test images)	Pix2Pix	42.35	24.04	284.10
	Pix2Pix-FQ	**36.05**	**25.18**	**229.56**
Akvelon Cartoon (600 test images)	Pix2Pix	94.96	19.37	786.77
	Pix2Pix-FQ	**85.82**	**21.60**	**503.62**

4.5 Ablation Studies

We performed many ablation studies. Similar to the baseline, we found that using L1 loss along with the adversarial loss results in a lower FID value, which represents a better performance of the model. Besides, we explored different number of epochs for training our model. We found that increasing it beyond 200 epochs for the custom Linnaeus dataset does not improve the performance.

Table 2. FID values got with different set up of losses and epochs in training phase of Custom Linnaeus dataset.

Adversarial	L1+Adversarial
59.66	36.05
200 epochs	**400 epochs**
36.05	36.84

4.6 Limitations

Our model is able to perform very well for quantization levels upto 30% but it does not perform as well for heavily quantized images. We feel this can be improved if the dataset size is increased along with the training time.Further, our model sometimes fails to recover information such as texture and depth in images.This can be seen in Figure 8, where some of the texture information on the bag is lost. This is to be expected as we are making the model to focus more on the color content in the image. Adding an additional module which disentangles the other needed information could improve the result in such cases.

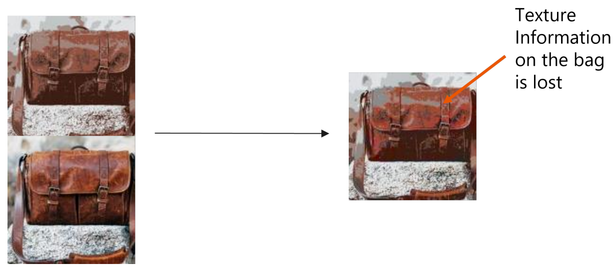

Fig. 8. Example of Pix2Pix-FQ limitations.

4.7 Retro Video Game Regeneration

One interesting application our model can be used for is video regeneration. Retro Video games usually contain poor color content and show a pixelated effect due to the low resolution. Our model can be extended to include a Video generation module, which takes in a sequence of generated images as input and generates a video which is finely quantized in color. We show one such example in Figure 9. Our model is able to regenerate a Mario game with richer color content and reduced pixelated effect.

5 Conclusions and Future work

In this paper, we have shown that a conditional adversarial network combined with disentanglement of color content can be utilized to perform color dequantization. The disentanglement of color information makes the model to generate more colorful content as well as reduces its complexity, as the number of features and parameters get reduced in the network. This is especially useful for small datasets. Even though we used limited data, we feel with the availability of more data, the model performance will be better. Future work can consider the disentanglement of latent space to have more control over generated output. Multimodal generation can also be explored to increase the color diversity in generated images and videos.

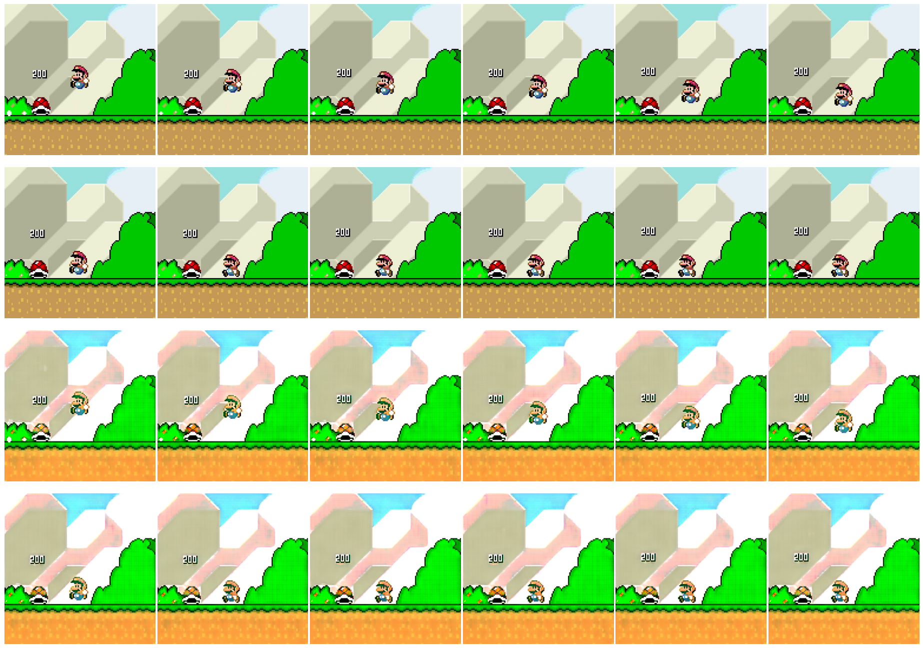

Fig. 9. Here we show a sequence of coarsely-quantized frames (Top) and its corresponding finely-quantized frames (Bottom).

References

1. Akvelon dataset, https://akvelon.com/make-pixel-art-in-seconds-with-machine-learning/
2. Linnaeus 5 dataset, http://chaladze.com/l5/
3. Agustsson, E., Timofte, R.: Ntire 2017 challenge on single image super-resolution: Dataset and study. In: The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops (July 2017)
4. Hua, Y., Kohli, P., Uplavikar, P., Ravi, A., Gunaseelan, S., Orozco, J., Li, E.: Holopix50k: A large-scale in-the-wild stereo image dataset. In: CVPR Workshop on Computer Vision for Augmented and Virtual Reality, Seattle, WA, 2020. (June 2020)
5. Isola, P., Zhu, J.Y., Zhou, T., Efros, A.A.: Image-to-image translation with conditional adversarial networks. CVPR (2017)
6. Park, H., Kim, K., Cha, E.Y.: An effective color quantization method using color importance-based self-organizing maps. Neural Network World **25**, 121–137 (2015)
7. Márquez de Silva, S., Felipe-Riverón, E., Sánchez Fernández, L.P.: A simple and effective method of color image quantization. In: Progress in Pattern Recognition, Image Analysis and Applications. pp. 749–757. Springer Berlin Heidelberg, Berlin, Heidelberg (2008)
8. Sousa, A., Kabirzadeh, R., Blaes, P.: Automatic colorization of grayscale images (2013)
9. Zhu, J.Y., Park, T., Isola, P., Efros, A.A.: Unpaired image-to-image translation using cycle-consistent adversarial networks (2020)

Report of assignments

Table 3. Report of assignments of each group member.

Task	**Akshay Dodwadmath**	**Manuela Ceron**
HT-Condor Setup	x	x
Related work research	x	x
Dataset preparation		x
Baseline Setup	x	x
Additional modules Setup	x	
Model training	x	
Model evaluation		x