# Image reproduction with flags

**TNM097 - Image Reproduction and Image Quality**

Viktor Sjögren David Robín Karlsson

March 17, 2021

**Abstract**

In this report, two image reproduction methods using flags as primitives are presented. An input image is divided into cells which are replaced by the perceptually most similar flag. The flag is found using two methods, lowest mean colour difference in the CIELAB colour space ($\Delta E_{ab}$) and its extension, S-CIELAB. The flag database contains 218 flags. The reproduction's efficiency was increased using two different optimisations of the database. A general optimisation which reduced the number of flags without reducing the colour span and an image specific optimisation where flags unimportant to the reproduction were removed from the database. The resulting reproductions were visually assessed and measured using the objective measurements signal-to-noise ratio, structural similarity, mean and max full-reference S-SCIELAB differences. The two methods proved to be capable of producing acceptable reproductions, however suffers from inconsistencies that was largely based on the limitations of the flags.

# Contents

# 1 Introduction

Image reproduction can take form in different ways, commonly with the goal of resemblance to the original image when viewing at a distance but dissimilar at close inspection. Image reproduction can be used to cut costs in print by reducing the necessary amount of ink through halftoning. There is also an artistic angle, such as the use of mosaic where individual pieces of an image are switched with another similar looking part not present in the image.

This report outlines an image reproduction method based on the mosaic principle, wherein an image is reproduced using a subset of the national flags of the world. The aim is to create an image reproduction method where the reproduced image is similar to the original at a distance while the flags are easily discernible at close inspection.

## 1.1 The flags

The key to a good mosaic reproduction is a vast database of pieces to choose from. Luckily there are a lot of countries on this planet. The flag database was compiled using [1] which contains 250 different flags, although a lot of the flags were culled. E.g, as a result of British imperialism, a lot of flags are almost identical, such as Anguilla, Cayman Islands and Cook Islands. Some of the flags also belonged to autonomous regions, not nations per say, an example being Åland. Since this project focused on national flags, flags such as these were culled from the database, which ended up being cut down to 218 images.

The database consists of PNG images with the dimensions $32 \times 16$ pixels. A weakness inherent in these dimensions comes from the fact that not all flags have the aspect ratio $2:1$. The author of [1] resolved this by making redundant pixels transparent, which becomes problematic when using the database in image reproduction. Since this mosaic method needs to be able to modify every single pixel of the image without any blind spots or overlap, the dimensions of the flags needs to be the same. This was resolved with a script filling out the images horizontally using the rightmost nontransparent column of colour. This worked well for flags with a largely horizontal layout such as Sweden or Norway, but worked less well for centred designs such as Switzerland or South Korea. This solution was chosen since the time investment necessary to manually extend the flags correctly would was deemed to large.

Finally, the size of all flags were halved to $16 \times 8$ pixels since the original size was deemed to large for a proper image reproduction as described in section 1. The culled and modified

**Figure 1.1:** *All flags in the culled and modified database.*

flag database can be seen in figure 1.1.

# 2 Method

Two different, yet similar methods for reproduction were used in this project. They only differ in the process of finding which flag to substitute into the image.

## 2.1 Flag database usage

In order to reproduce an input image using the flag database, a predicate determining which flag to substitute into the reproduced image based on the flag's characteristics is necessary. In this project, the characteristic is the mean colour of a flag. The mean colour is calculated in the CIELAB colour space since differences in this colour space quantify how human eyes perceive colour differences. The mean colour of every flag can be seen in figure 2.1.

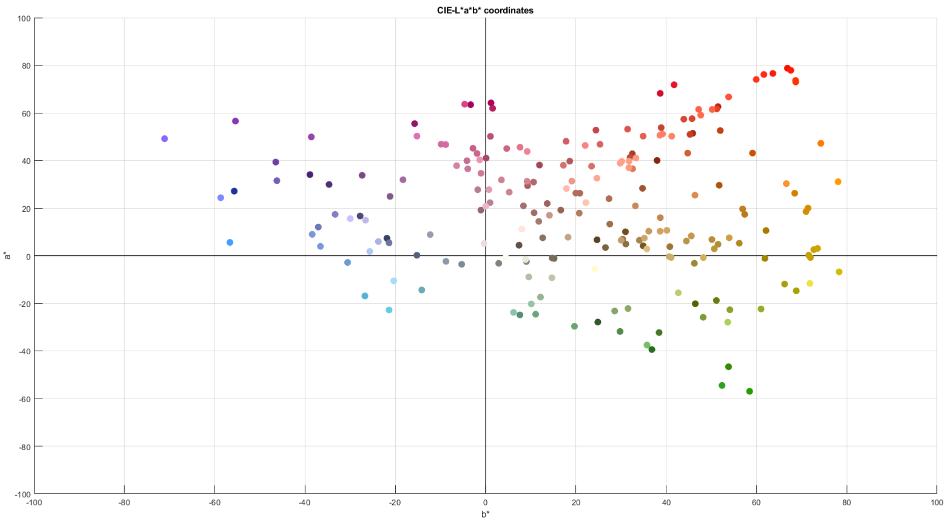



**Figure 2.1:** *The mean CIELAB colours of all flags in the database.*

## 2.2 Initial image processing

The image to be reproduced is subjected to some processing before the reproduction begins. Firstly, if the image is too small or large, the image size is adjusted. If the image did not satisfy the heuristically obtained constraints

$$500 < \text{height} < 2000$$
$$1000 < \text{width} < 2500$$

the image was resized, with bicubic interpolation, to the width = 1920 and the height being adjusted to keep the aspect ratio of the image intact.

Secondly, to ensure that a flag can be substituted into every part of the image without missing any pixels, the height and width of the image was lightly adjusted. If the image did not satisfy the constraints

$$(\text{height}) \mod 8 = 0$$
$$(\text{width}) \mod 16 = 0$$

the rightmost column or bottom row was deleted until the constraints were met.

The image was then divided into $16 \times 8$ pixels large cells wherein the mean CIELAB colour is calculated as well. The cells' positions and mean colour values were stored in a list which was used in the image reproduction.

## 2.3 Image reproduction

The list of the input image's cells, described in section 2.2 above, were compared to the mean colours of all 218 flags in an attempt to find which flag is the most perceptually similar in colour. This comparison was made using two methods, conventional CIELAB colour difference $\Delta E_{ab}$ and full-reference S-CIELAB[1] difference. No consideration was taken to neighbouring cells in the image, similar to a nearest-neighbour type algorithm.

Since every cell in the image is compared to every flag in the database, the image reproduction algorithm will always have a quadratic time complexity $O(mn)$ where $m$ is the number of cells and $n$ is the number of flags.

## 2.4 Quality measures

The quality of each reproduced image was gauged through visual assessment and the objective measurements signal-to-noise ratio and structural similarity, mean and max colour differences measured in full-reference S-CIELAB evaluations. Although the S-CIELAB values were the main focus in quality assessment, as they are more representative of human perception, the strictly objective signal-to-noise ratio and structural similarity values were used as contrast. This would allow analysis of which reproductions were favoured by which quality measures. The viewing distances used in the S-CIELAB evaluations are 20

[1] Calculated by assuming a 0.5 meter distance to a 91.79 PPI monitor.

cm, 50 cm and 100 cm representing close, normal and far viewing distances. The screen resolutions used are 91.79 PPI and 183 PPI, representing two 24 inch monitors with full HD and 4K resolution.

## 2.5 Optimisations

Two types of optimisations were implemented, removing flags without largely reducing the general colour span and removing flags deemed unnecessary for a particular image.

With the flag dataset, the colour span that the reproduction algorithm uses can be seen in Figure 2.1. This revealed that some of the flags were perceived as almost identical in colour. Therefore, the dataset was optimised where each of the flags within a certain distance ($\Delta E_{ab}$) from each other were removed.

As seen in figure 2.1, there are no true black colours in the dataset, the closest match would rather be a dark blue. This created problems when recreating largely dark images. This was remedied when optimising for particular images. In the process of removing unnecessary flags, the jolly roger flag was added into the database as it's a largely black flag. This optimisation was performed on the paintings Girl with a Pearl Earring and Mona Lisa.

Since the painting Girl with a Pearl Earring contains mostly yellow, blue and black, the optimisation focused on removing flags with a mostly red mean colour.

Since the painting Mona Lisa contains mostly red, yellow and dark parts, the optimisation focused on removing flags with mostly blue and green mean colours.

# 3 Results

Image reproductions, visual assessments and quality measurements are presented in this chapter. The original images and the reproductions are displayed next to each other for reference. The mean and max S-CIELAB differences are presented in a table below the reproduction.

The two reproductions of Bernie Sanders, presented in figure 3.1 and 3.2, display large differences from each other on the clothing and the white background. The authors agree that the S-CIELAB based reproduction is the better one since it displays more details around Bernie. This coincides with the S-CIELAB quality measurements but contradicts SNR and structural similarity, which favour the CIELAB reproduction.

The two reproductions of trees, presented in figure 3.3 and 3.4, display large differences on the undersides of the tree canopies. The authors agree that the S-CIELAB based reproduction is the better one since it displays more details, a better reproduction of the elaborate sprig system. Once again, this coincides with the S-CIELAB quality measures but also SNR. Structural similarity once again slightly favours the regular CIELAB reproduction.

The two reproductions of the iconic wave, presented in figure 3.5 and 3.6, display an approximately same amount of detail while differing in colour in the background and certain parts of the wave. Once again, the authors favour the S-CIELAB reproduction since the background colour is more reminiscent of the original. This is contrary to the S-CIELAB quality measures and SNR which are just slightly favoured to the CIELAB reproduction. The structural similarity is identical between the reproductions, which the authors agree with considering the two reproduction mostly differ in colour.

The two reproductions of Mona Lisa, presented in figure 3.7 and 3.8, is a test of how large image are reproduced. They differ mostly in the background hues and the shadows on the clothing. The authors prefer the CIELAB reproduction since it more accurately represents the background and shadows, the S-CIELAB reproduction barely discerns any of the shadows at all. This contradicts the SNR but agrees with the structural similarity which slightly favours the CIELAB reproduction. This contradicts the S-CIELAB quality measures that slightly favour the S-CIELAB reproduction.

The resulting flags and colour span of optimising colours for Mona Lisa can be seen in figure 3.9 and the resulting reproductions in 3.10. In the optimised CIELAB reproduction, the addition of the mostly black Jolly Roger flag improves the clothing and shadowed parts immensely. The removal of blue and green flags significantly improved the background as well, barring some missing green tones in the up parts of the painting. The optimised

S-CIELAB reproduction completely butchered the original image, filling out almost all dark regions, clothes and hair with red flags and the background with a yellow flag.

The resulting flags and colour span of optimising colours for Girl with a Pearl Earring can be seen in figure 3.11 and the resulting reproductions in 3.12. In the optimised CIELAB reproduction the background is once again greatly improved with the addition of the Jolly Roger flag. The yellow and blue clothing is left mostly intact while the skin tones have worsened significantly, giving the girl a sickly skin tone. The optimised S-CIELAB reproduction once again butchered the original image, filling out the background with a mostly purple flag, destroying detail and skin tones in the face. Only the yellow clothing and blue headband bares some resemblance to the original image.

The resulting flags and colour span of the general optimisation can be seen in figure 3.13 and the resulting reproduction in figure 3.14. The original image has been butchered. This is a consequence of the flags not having a uniform colour span distribution which makes a general colour optimisation highly unreliable. All quality measures reinforces that this is bad reproduction.

| **Mean** | Close | Normal | Far |
|---|---|---|---|
| Full HD | 2.67 | 2.15 | 1.61 |
| 4K | 2.31 | 1.61 | 1.13 |

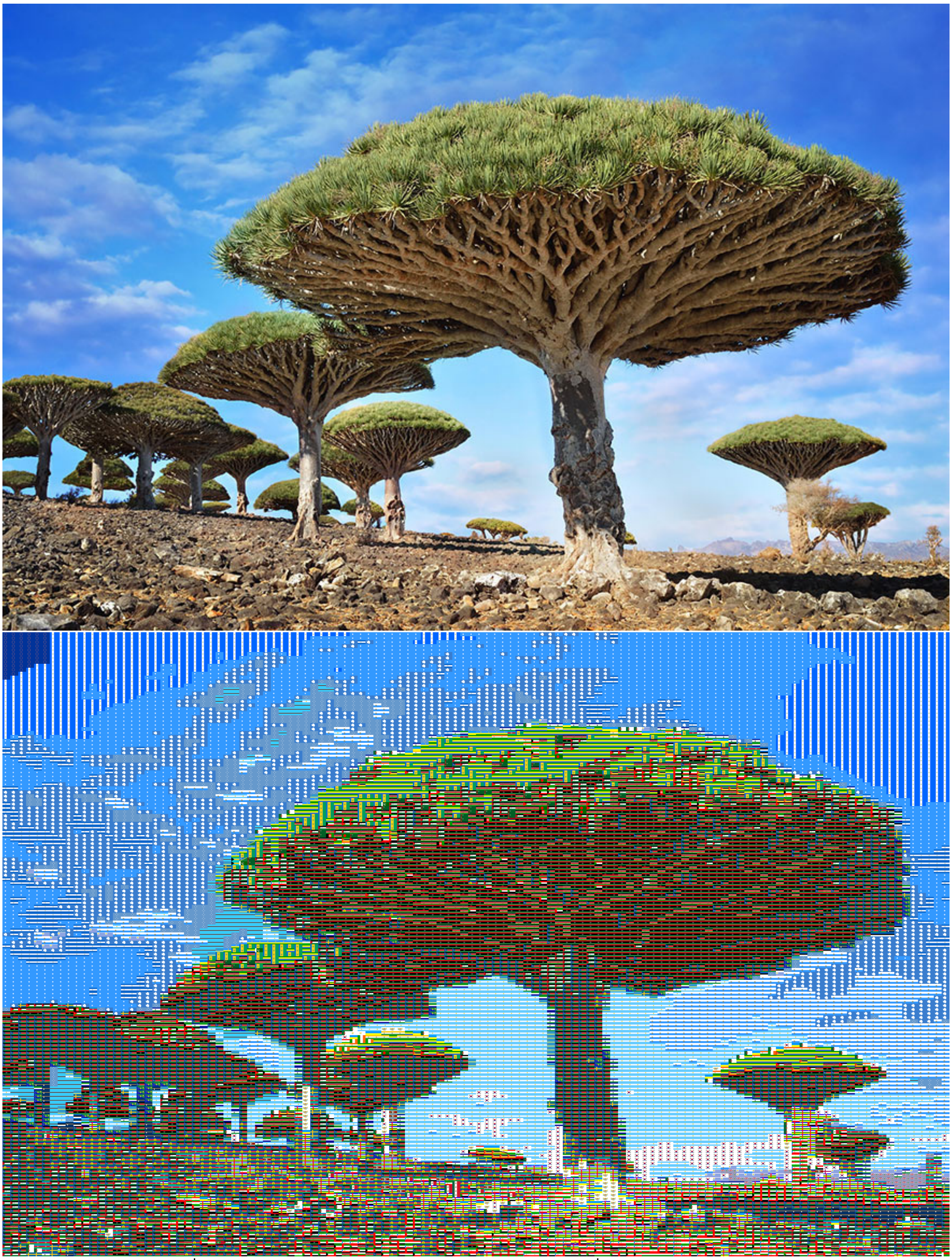
| **Max** | Close | Normal | Far |
|---|---|---|---|
| Full HD | 10.67 | 9.21 | 6.95 |
| 4K | 9.46 | 6.97 | 5.70 |

**Figure 3.1:** *Lowest CIELAB difference reproduction. SNR = 3.20, SSIM = 0.05.*

| Mean | Close | Normal | Far |
|---|---|---|---|
| Full HD | 2.52 | 2.02 | 1.44 |
| 4K | 2.18 | 1.44 | 0.87 |

| Max | Close | Normal | Far |
|---|---|---|---|
| Full HD | 12.40 | 10.34 | 7.67 |
| 4K | 11.18 | 7.68 | 4.66 |

**Figure 3.2:** *Lowest mean S-CIELAB difference reproduction. SNR = 2.83, SSIM = 0.04.*



| **Mean** | Close | Normal | Far |
|---|---|---|---|
| Full HD | 3.04 | 2.48 | 1.91 |
| 4K | 2.65 | 1.91 | 1.38 |

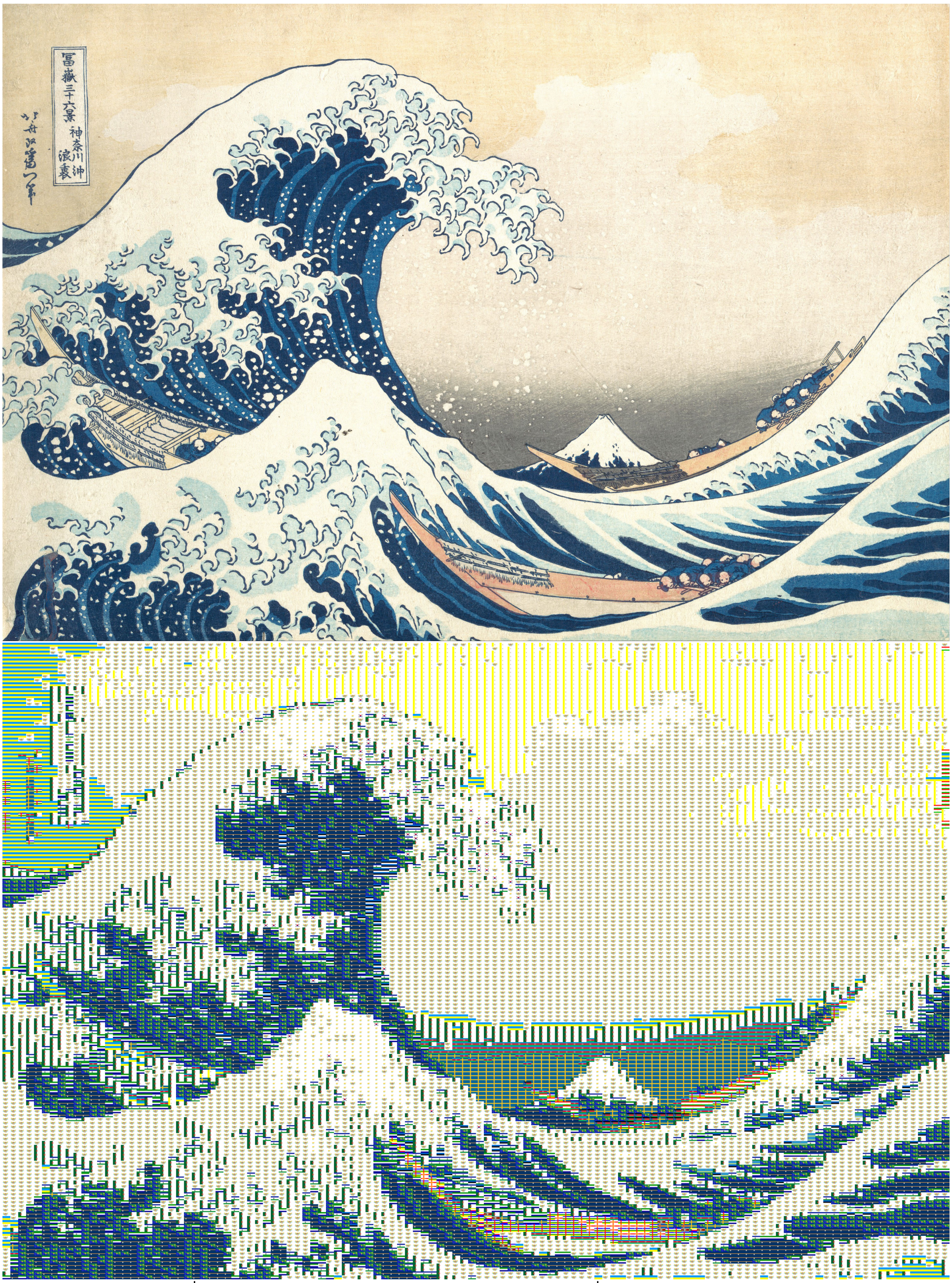
| **Max** | Close | Normal | Far |
|---|---|---|---|
| Full HD | 14.03 | 10.98 | 9.17 |
| 4K | 11.77 | 9.18 | 6.07 |

**Figure 3.3:** *Lowest CIELAB difference reproduction. SNR = 5.79, SSIM = 0.34.*

| **Mean** | Close | Normal | Far |
|---|---|---|---|
| Full HD | 2.73 | 2.24 | 1.72 |
| 4K | 2.38 | 1.72 | 1.20 |

| **Max** | Close | Normal | Far |
|---|---|---|---|
| Full HD | 16.36 | 13.41 | 8.88 |
| 4K | 14.53 | 8.89 | 5.90 |

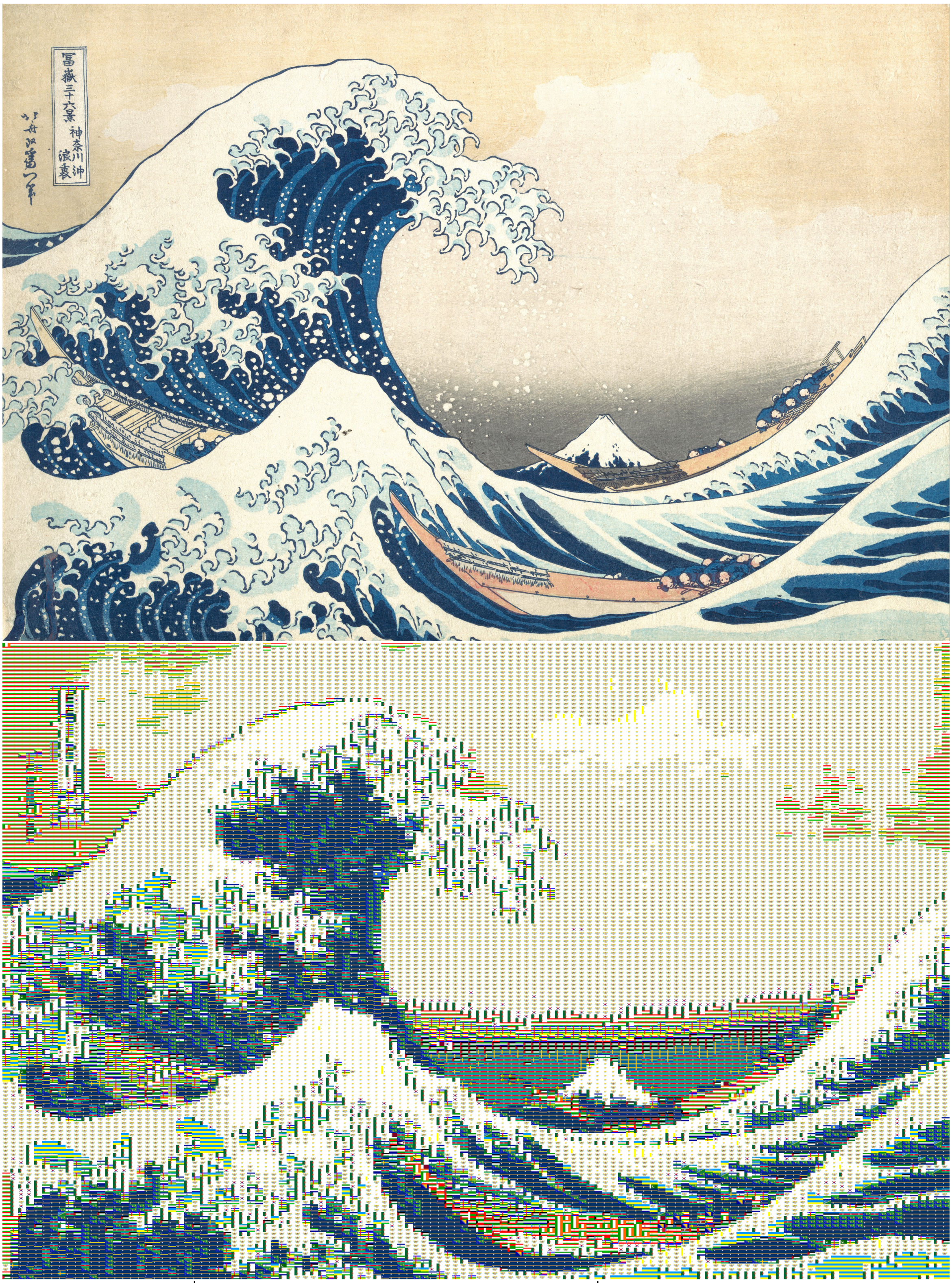
**Figure 3.4:** *Lowest mean S-CIELAB difference reproduction. SNR = 5.97, SSIM = 0.32.*




| Mean | Close | Normal | Far |
|---|---|---|---|
| Full HD | 3.16 | 2.62 | 2.12 |
| 4K | 2.77 | 2.12 | 1.60 |

| Max | Close | Normal | Far |
|---|---|---|---|
| Full HD | 13.99 | 11.42 | 9.39 |
| 4K | 12.15 | 9.40 | 7.28 |

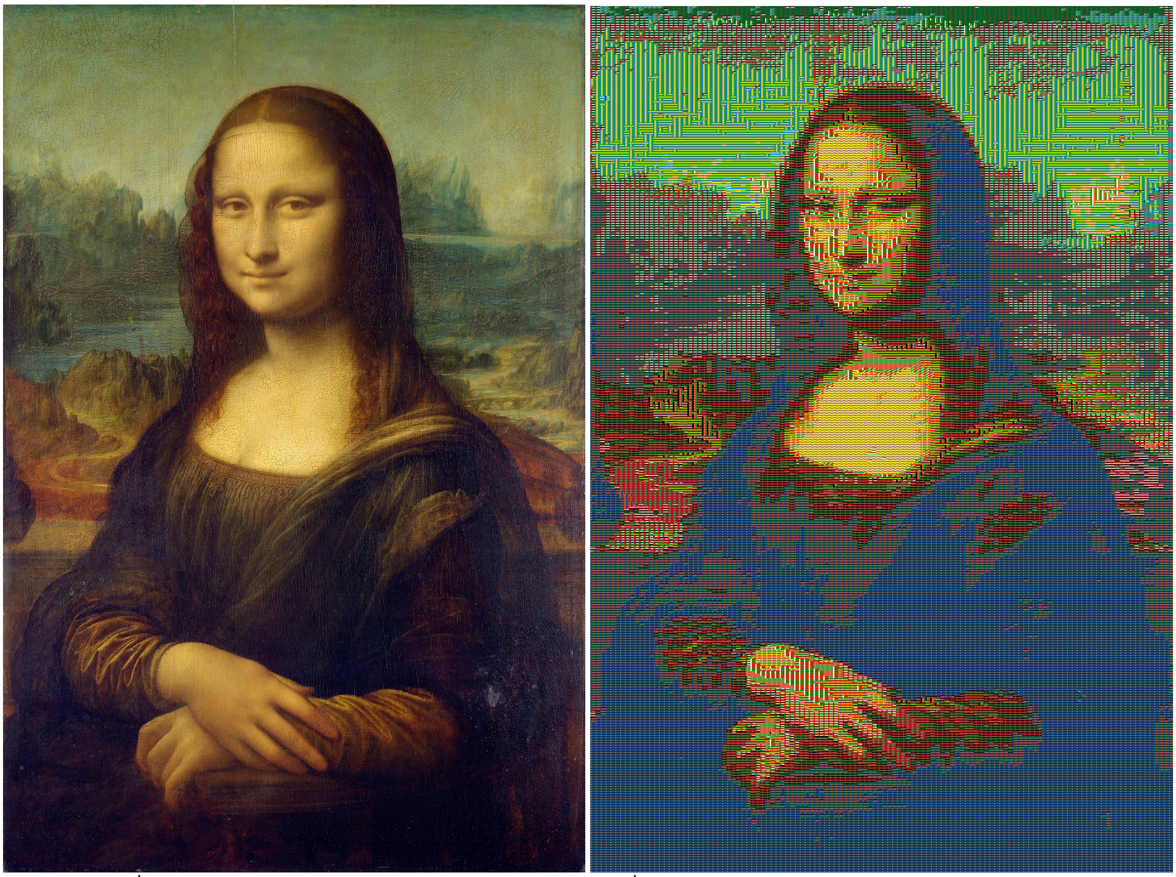
**Figure 3.5:** *Lowest CIELAB difference reproduction. SNR = 8.87, SSIM = 0.15.*




| Mean | Close | Normal | Far |
|---|---|---|---|
| Full HD | 3.16 | 2.64 | 2.09 |
| 4K | 2.80 | 2.10 | 1.52 |

| Max | Close | Normal | Far |
|---|---|---|---|
| Full HD | 15.38 | 12.37 | 9.94 |
| 4K | 13.23 | 9.95 | 7.13 |

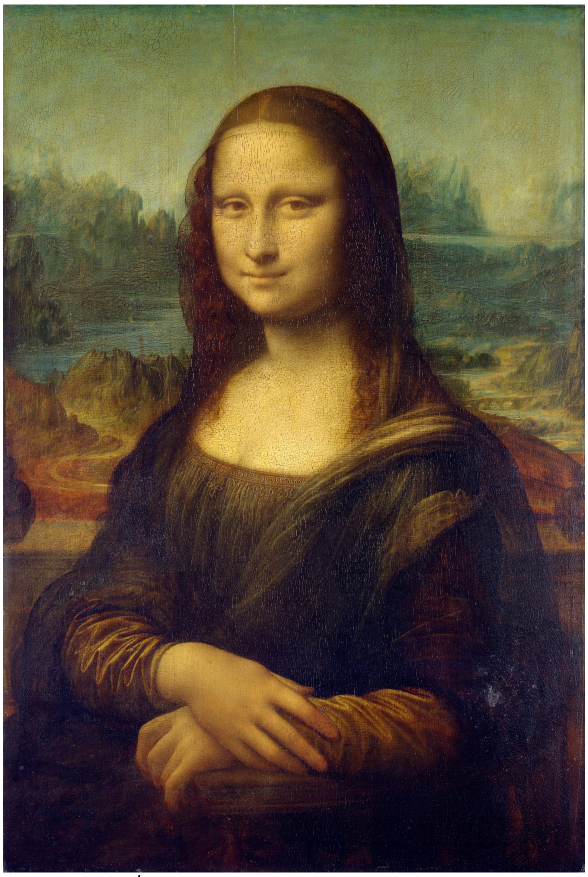
**Figure 3.6:** *Lowest mean S-CIELAB difference reproduction. SNR = 8.41, SSIM = 0.15.*

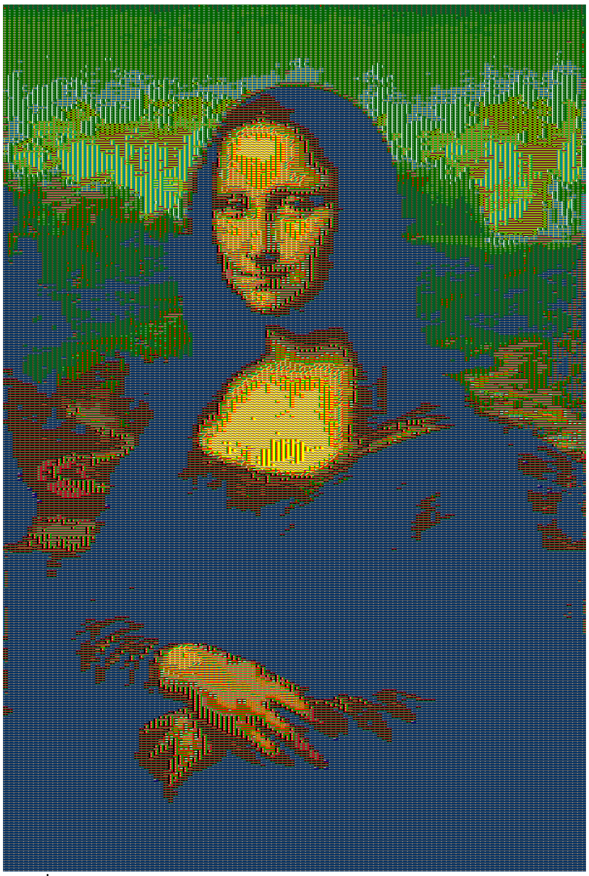


| **Mean** | Close | Normal | Far |
|---|---|---|---|
| Full HD | 2.34 | 1.88 | 1.44 |
| 4K | 2.01 | 1.44 | 1.10 |

| **Max** | Close | Normal | Far |
|---|---|---|---|
| Full HD | 12.14 | 9.51 | 6.95 |
| 4K | 10.37 | 6.96 | 4.58 |

**Figure 3.7:** *Lowest CIELAB difference reproduction. SNR = 1.02, SSIM = 0.11.*





| **Mean** | Close | Normal | Far |
|---|---|---|---|
| Full HD | 1.81 | 1.47 | 1.14 |
| 4K | 1.57 | 1.15 | 0.86 |

| **Max** | Close | Normal | Far |
|---|---|---|---|
| Full HD | 11.58 | 9.19 | 6.88 |
| 4K | 10.06 | 6.90 | 4.27 |

**Figure 3.8:** *Lowest mean S-CIELAB difference reproduction. SNR = 1.56, SSIM = 0.09.*

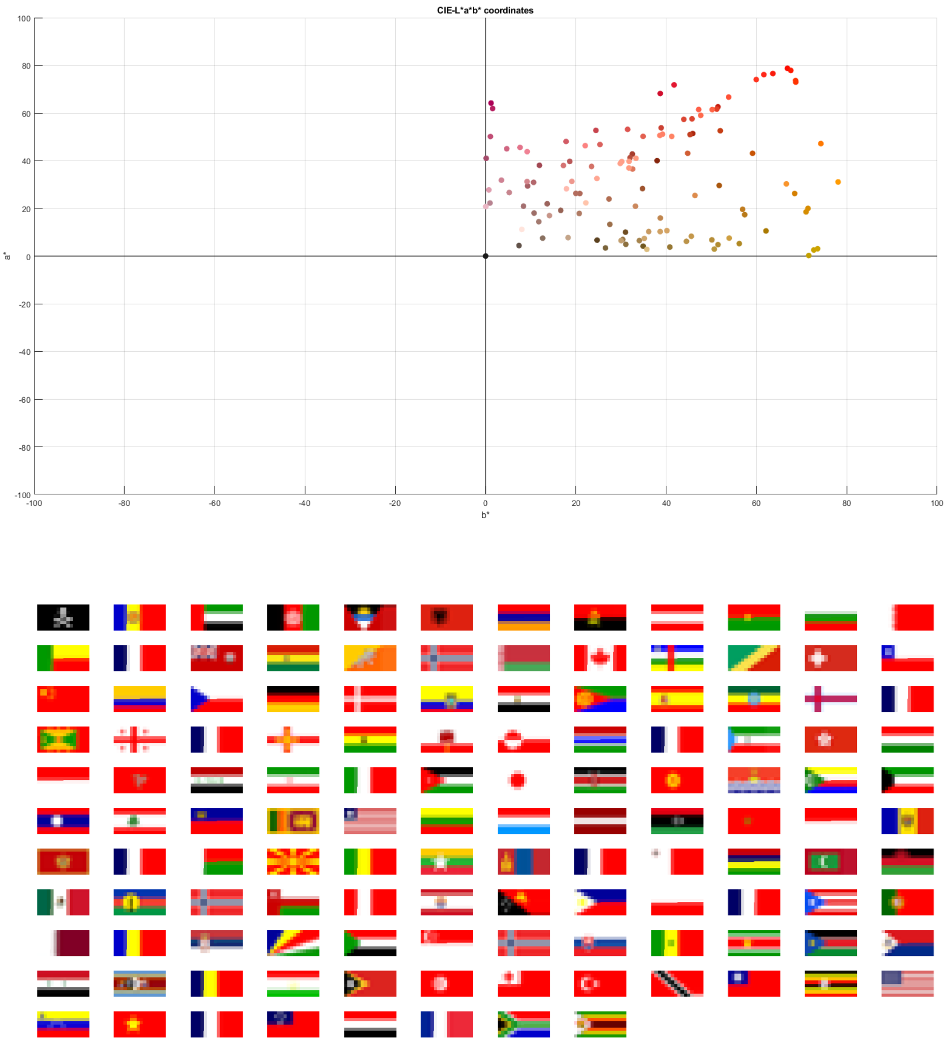



**Figure 3.9:** *Colour span and flags from the image specific optimisation of Mona Lisa. The number of flags was reduced to 128 from 218.*

**Figure 3.10:** *Reproduction of Mona Lisa using image specific optimised flag colours compared to CIELAB and S-CIELAB reproduction. Optimised reproductions in the top row, CIELAB reproductions on the left, S-CIELAB reproductions on the right.*

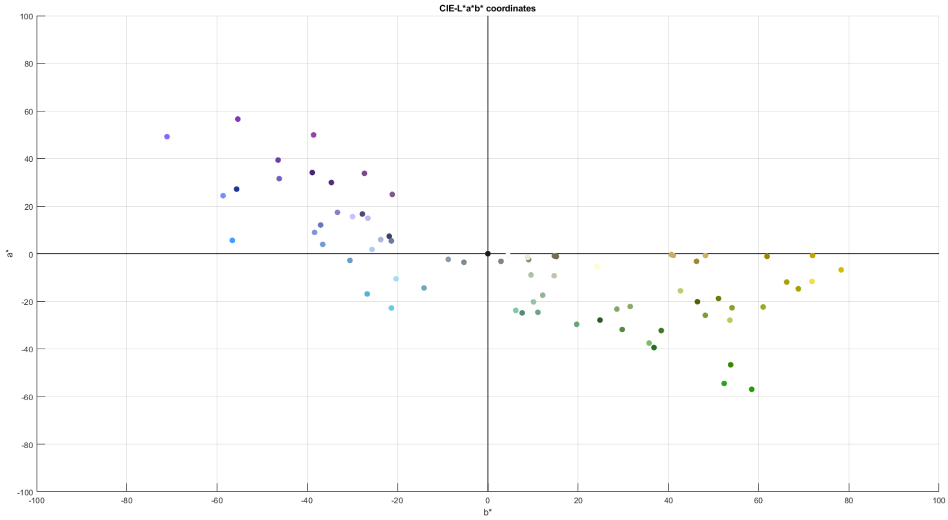



**Figure 3.11:** *Colour span and flags from the image specific optimisation of Girl with a Pearl Earring. The number of flags was reduced to 73 from 218.*

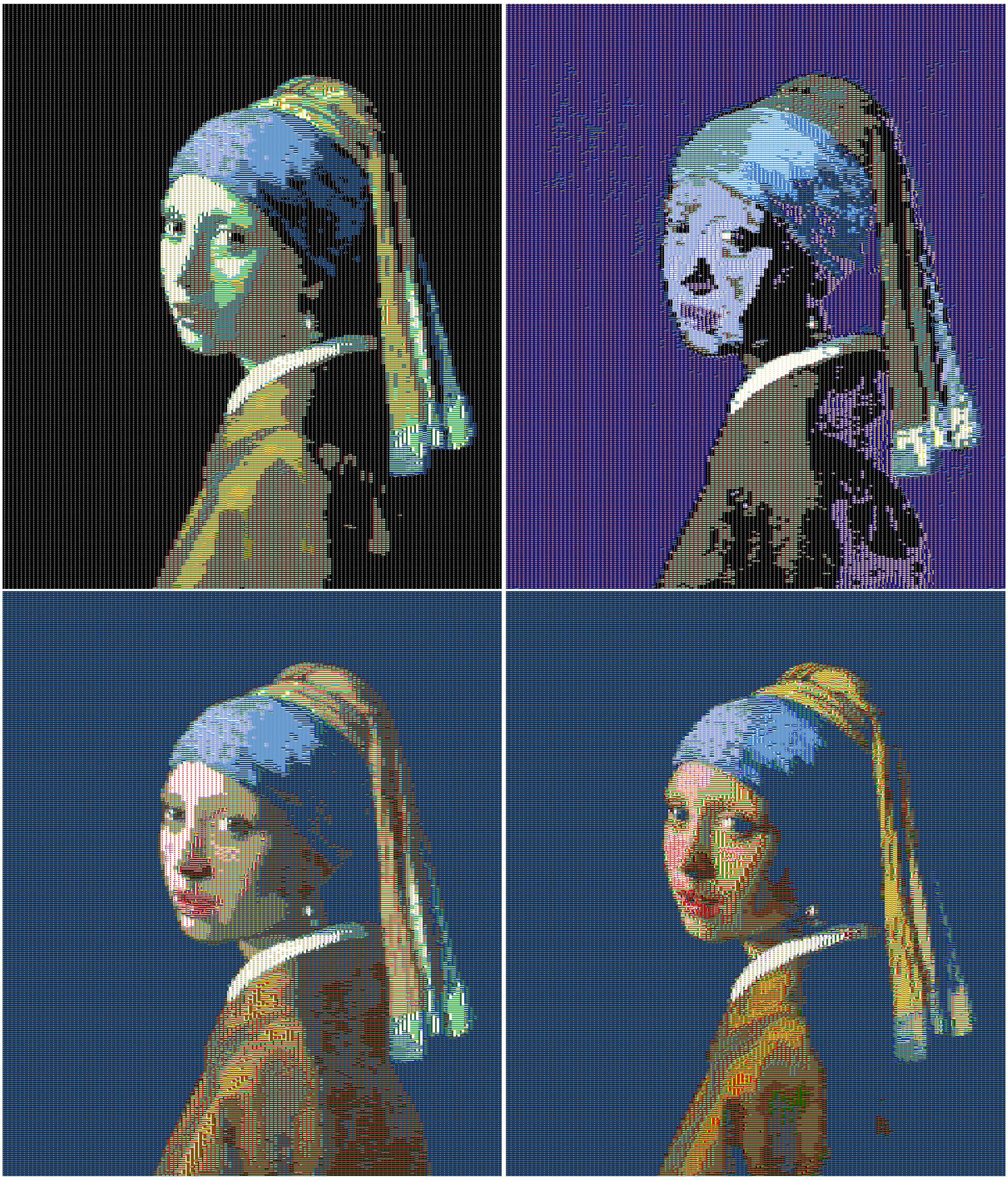


**Figure 3.12:** *Reproduction of Girl with a Pearl Earring using image specific optimised flag colours compared to CIELAB and S-CIELAB reproduction. Optimised reproductions in the top row, CIELAB reproductions on the left, S-CIELAB reproductions on the right.*

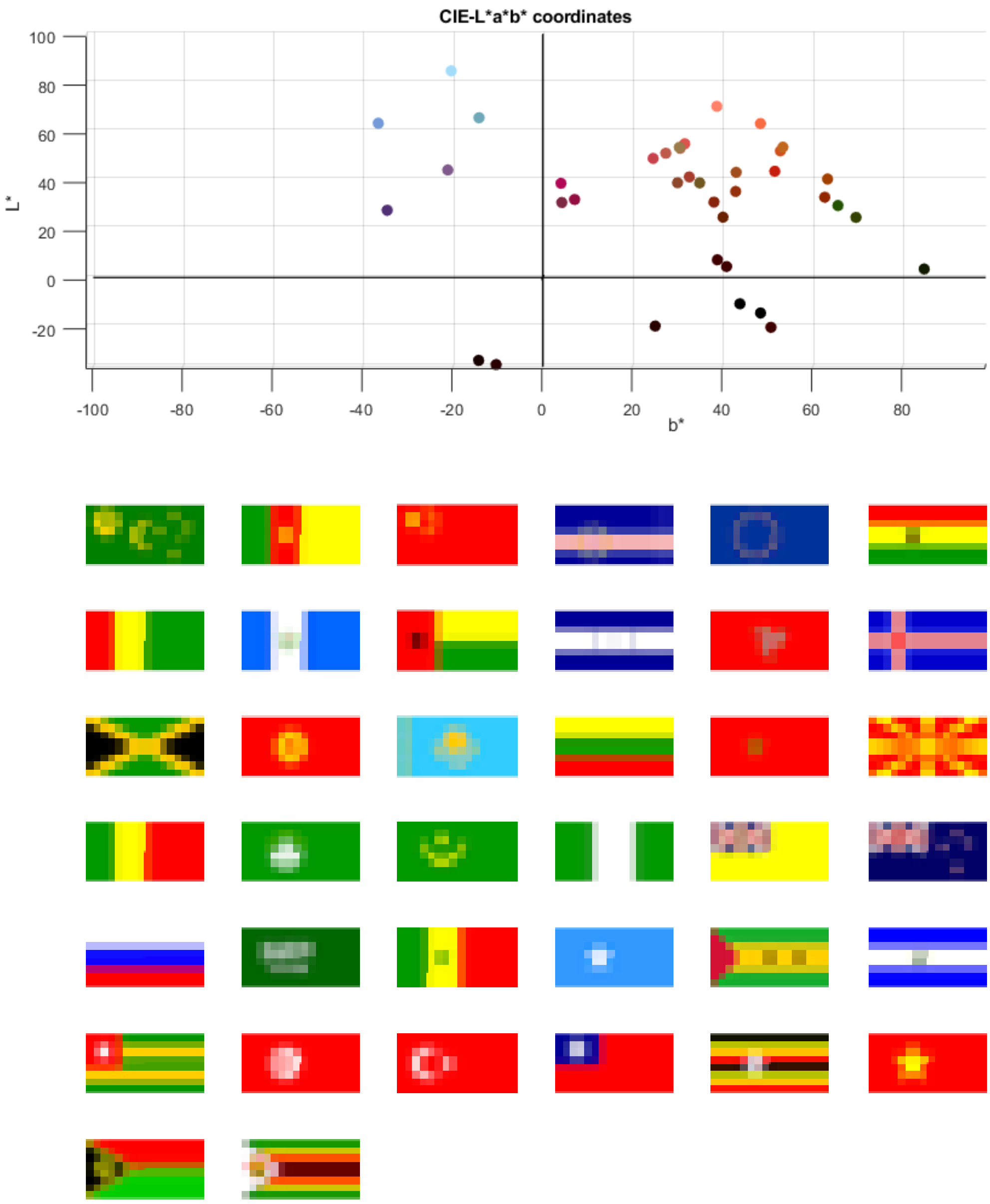



**Figure 3.13:** *Flags and colour span after general optimisation, where colours within a $\Delta E_{ab}$ distance of 130 of each other were removed..*

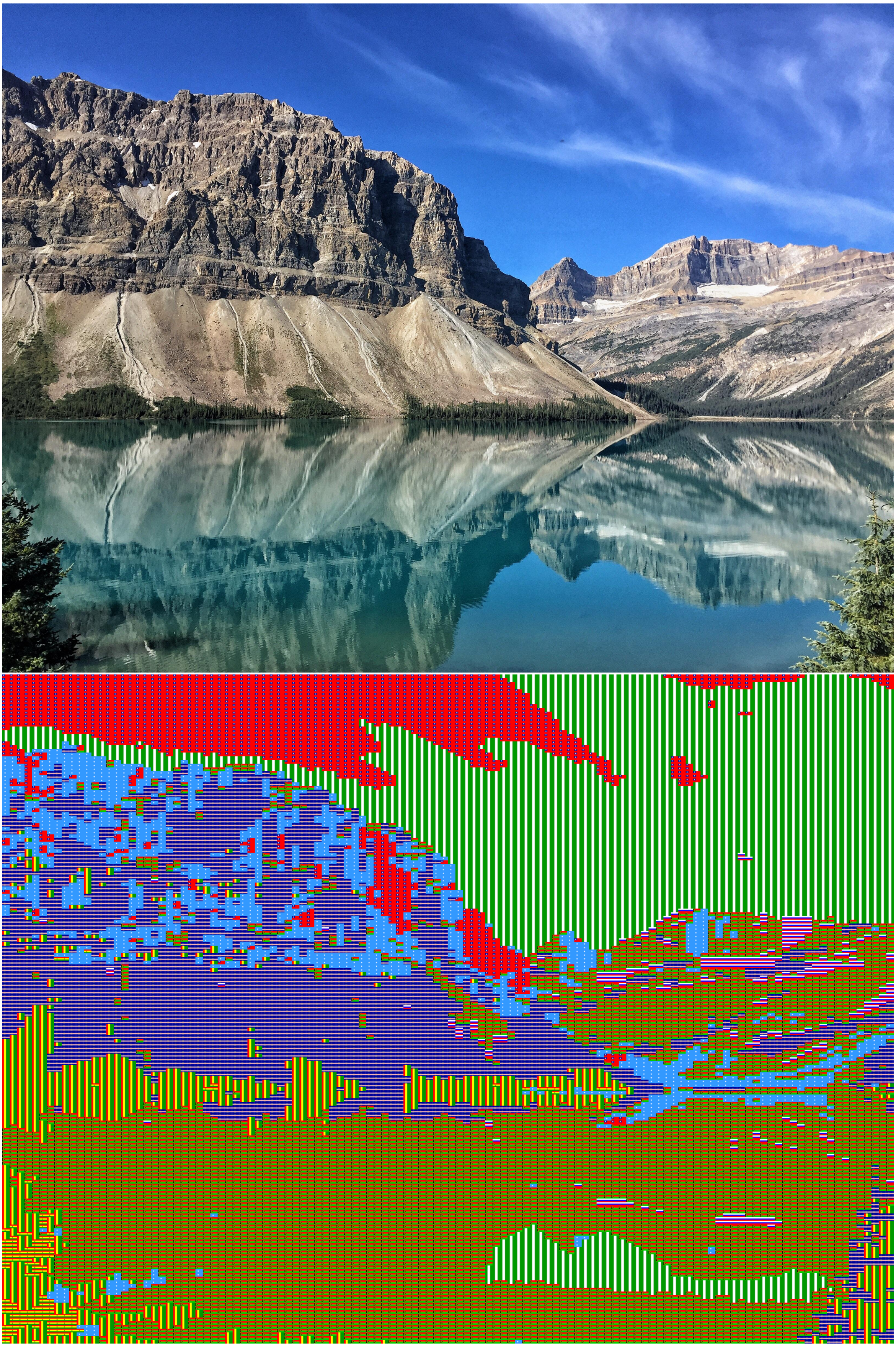


**Figure 3.14:** *CIELAB reproduction using optimised flag colours. SNR = 0.68, SSIM = -0.06.*

| **Mean** | Close | Normal | Far |
|---|---|---|---|
| Full HD | 7.05 | 6.59 | 6.13 |
| 4K | 6.74 | 6.13 | 5.73 |

| **Max** | Close | Normal | Far |
|---|---|---|---|
| Full HD | 14.51 | 13.21 | 12.68 |
| 4K | 13.64 | 12.69 | 12.44 |

# 4 Discussion

The algorithms used for this reproduction purpose proved to be sufficient for the minimum criteria of being able to reproduce images using flags. Namely that the reproduction mimics the original image at a distance but the flags are distinguishable well at a closer look. However due to a rather simplistic "best fit" approach, the algorithms had some major flaws that generated inconsistent results. This can be seen in the results where some reproduced images produced intrusive patterns or generally poor reproductions in comparison to the input image.

In an attempt to combat this problem, an alternative method was theorised. This method would instead of simplifying the flags to a singular LAB value, handle the flags differently based on their primary colours. This approach would have suited the database since almost all flags only contains a few colours and no real outliers or gradient colours. The method would have segmented the different regions of colours using the k-means clustering algorithm. The comparison between the cells in the original image and the flags would then have been based on the percentage of which colours that exist within the cell, and which flag then corresponds to that percentage the best. However, when tested for a flag such as Italy or similar flags, the segmentation rarely found all three colours. This might have been caused by imperfections in the clustering algorithm, however no real test with other clustering algorithms were made to prove this statement. Another reason could be visual artefacts produced by the downscaling of the flags as mentioned in section 1.1.

After the general optimisation of the flag dataset, the resulting reproductions quality were much lower then expected. When comparing the reproduction of the flag database in full versus the optimised 38 flags database, it was clear that the general optimisation method is flawed. This can also be seen in the colour span in figure 3.13, where many of the previously existing colours were completely removed. The could have been caused by a too extreme optimisation which heavily favoured removing flags instead of keeping them. It could also have been caused by a naive optimisation algorithm as no meaningful research was done regarding database optimisation without reducing colour span.

In contrast to the general optimisation, the image specific optimisation proved to result in higher quality reproductions. The biggest improvement factor stems from adding the Jolly Roger flag, which complements the existing colour span by adding a true black hue. Since the database was reduced around half for both optimisations, each reproduction was successfully sped up. However, the S-CIELAB based reproductions massively deteriorated the reproduction quality, especially when compared to the optimised CIELAB version and the unoptimised CIELAB and S-CIELAB versions.

As mentioned in section 2.3, the image reproduction will always have a quadratic time complexity. This can be improved by storing all flags' mean colours in a smarter data structure instead of a regular array, like a k-d tree [2]. Utilising this data structure to represent the CIELAB colour space and storing all mean flags' colours in it, searching for the perceptually closest colour can be reduced to, on average, $O(logn)$ where $n$ is the number of flags. This would give the algorithm a total time complexity of $O(mlogn)$, where $m$ is the number of cells in the image.

# 5 Conclusion

Reproduction of images using flags was an interesting idea that none of the authors truly thought would yield high-quality results. The flag primitives proved to be colourful with saturated colours, spanning a decent colour range but did lack in some areas nonetheless. In the end, majority of images that were reproduced proved to be acceptable and sometimes even good reproductions. The main exception being dark images, which luckily was improved by the image specific optimisations.

The authors perceived quality of the reproduced images and the objective measurements made it clear that the measurements were somewhat inconsistent. In some instances, the majority of the measurements strengthened the perceived opinion of the authors, while contradicting some reproductions.

# Bibliography


[1] Cristian Romanescu. (2019). "Countries," [Online]. Available: https://github.com/cristiroma/countries (visited on 03/06/2021).

[2] J. Bentley, "Multidimensional binary search trees used for associative searching," *Commun. ACM*, vol. 18, pp. 509–517, 1975.