1. Each color value has 3 components, so when shade values are integers between 0 and 255, there are 256\*256\*256=16777216 color values. Each gray has only one component, so there are 256 possible grays. Each 33 gray pixel patch has 9 grays, so there are 256\*9=2304 possible 33 gray pixel patches. A 33 gray pixel patch is not enough to guess the center color. A 33 pixel patch is a very small area in the picture. It can’t be very specific. It’s highly possible that there’s little different between pixels in the patch, and they are all in the same texture. So 33 gray pixel patch provide such little information to guess the center color. When the picture is small and monochrome, it may be enough to try to associate every shade of gray with a specific color. Classical color to gray conversion formula is. Then same gray can be converted from different. Actually, the corresponding relationship between gray and color is like normal distribution. So normally it’s hard to associate shade of gray, whose values is near the middle of range, with a specific color. But if in a picture, two color values of every pixel are fixed or slightly changed, and the left color changes. Then in this picture, we can absolutely associate every shade of gray with a specific color.
2. By reducing the color-space and the patch-space, capacity of training data can effectively reduce that can improve training efficient without reducing the effect of colorizing. For example, if there’s color value is [120, 50, 80] and another color value is [130, 45, 75]. These two color values have slight different, but people can’t clearly differentiate them. So it’s fair to lump them together. And all values in the range between them can also be lumped together, this will obviously reduce number of possible color values and lower data capacity. Also, if two 33 gray pixel patches are slightly different, they maybe belong to same object in the picture and represent similar color in the color image. So it is fair to lump them together and reduce the patch-space.

When it was fairly confident a pixel should be green or red, but certainly not blue. It may remove patch-color pairs whose color is blue in the training data and retrain the model. If we remove patch-color pairs whose color is blue in the training data, the training model will focus on distinguishing red and green instead of distinguishing red, green and blue. Then this model can distinguish red and green more correctly and maybe can confident this pixel should be red or confident this pixel should be green.

8. As we discussed in the question 1, the corresponding relationship between gray and the number of color is like normal distribution, one gray can be convert from different. After training algorithm on pictures of palm trees, it makes model more likely to convert gray to colors of palm trees, like brown and green. Then if we use this model to color a grayscale of a tiger, it’s more likely to covert gray to colors of palm trees instead of colors of tigers. It will get an image of a tiger with weird colors.