

CSC2516-Assignment II

Sergio E. Betancourt (998548585)

2019-02-28

Part A

1. Below is a description of the conv layers in the model `RegressionCNN`. In the given setting, `kernel=3` and `num_filters=32`.

Conv Layer	Filter Size	No. of Filters
1st	kernel	num_filters
2nd	kernel	2×num_filters
3rd	kernel	2×num_filters
4th	kernel	num_filters
5th	kernel	3
6th	kernel	3

2. 25 epochs in the given setting.

3. I re-trained the given model with 10 and 90 epochs. The original model (25 epochs) achieved a training loss of .008.

For the 10 epoch run, this model achieved a training loss of .01. When compared to the 25 and the 90-epoch runs, the output images here seem to have a slightly greener (monochromatic) appearance and the details are less clear.

For the 90 epoch run, the training loss is lower (.006) than that of the 25-epoch. The output images show less vibrant colors when compared to the 25-epoch run; however, the resulting images appear sharper.

4. Consider the Euclidian distance between two colors in the RGB (Red-Blue-Green) color space,

$$D(\text{col}_i, \text{col}_j)^2 = (R_i - R_j)^2 + (G_i - G_j)^2 + (B_i - B_j)^2.$$

Let $\text{col}_1 = \text{"Red"} = (R=\text{max}, B=0, G=0)$, $\text{col}_2 = \text{"Green"} = (R=0, B=0, G=\text{max})$, and $\text{col}_3 = \text{"Black"} = (R=0, B=0, G=0)$. Then, $D(\text{col}_1, \text{col}_3) = D(\text{col}_2, \text{col}_3)$, and these two colors are indistinguishable from each other when compared to black, in terms of distance. However, these are all three very distinct colors to the human eye. This translates to binary comparison, as the distance between two colors in this color space does not indicate **how** they are different, only that they are.

5. Framing the colorization problem as a regression problem will tend to average out predicted colors (minimizing the error) so that the colorized output does not display sharp colorization. In contrast, framing this as a classification problem should alleviate this issue by allowing the model to choose from a bucket of pre-defined, distinct colors, prioritizing the probability of choosing the right color, and thus limiting ambiguity.

Part B

2. Running this new CNN with the same number of epochs as before (25), yields images with higher contrast; however, the resulting colors are not terribly close to the target. Note how in some of the CNN images 3rd and 7th), the colorization has added green to places that neither had it in the original image, nor in the image resulting from the model in **RegressionCNN**.

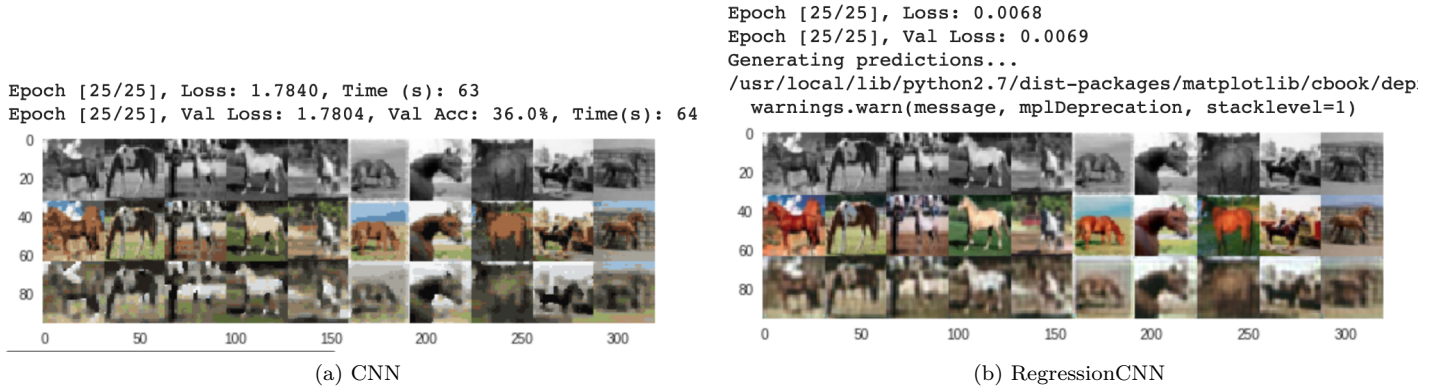


Figure 1: Output images produced by the CNN and the Regression CNN models

Part C

2. The skip connections added in UNet improved the training loss ($1.74 \rightarrow 1.66$), validation loss ($1.74 \rightarrow 1.65$), and validation accuracy ($36\% \rightarrow 40\%$) for the same number of epochs, when compared to the vanilla CNN. Qualitatively, the skip connections improved the output's details very slightly, most noticeably in the 1st, 4th, and 8th images. Nonetheless, it is still imperfect.

Skip connections may enhance the performance of our vanilla CNN by controlling for output deviations (in shape or texture) from the input image. Moreover, they assist network training by back-propagating the gradient to lower layers.

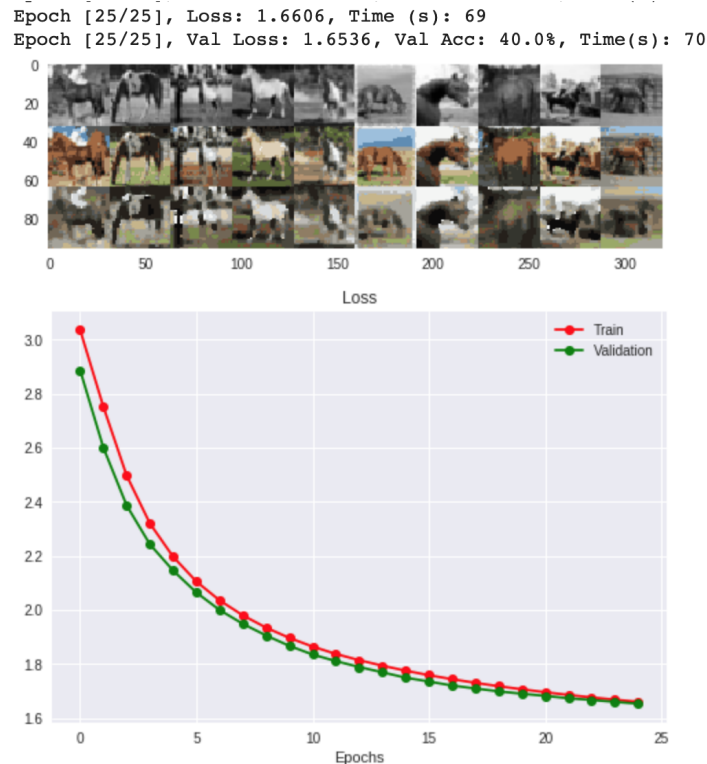


Figure 2: Output images produced by the UNet model with different mini-batch of 100

3. Below are the relevant outputs from runs with different mini-batches. Note the differences between their training and validation loss curves. Moreover, these different mini-batch sizes produced different images for the output—the TA Xuchan confirmed that this is normal.

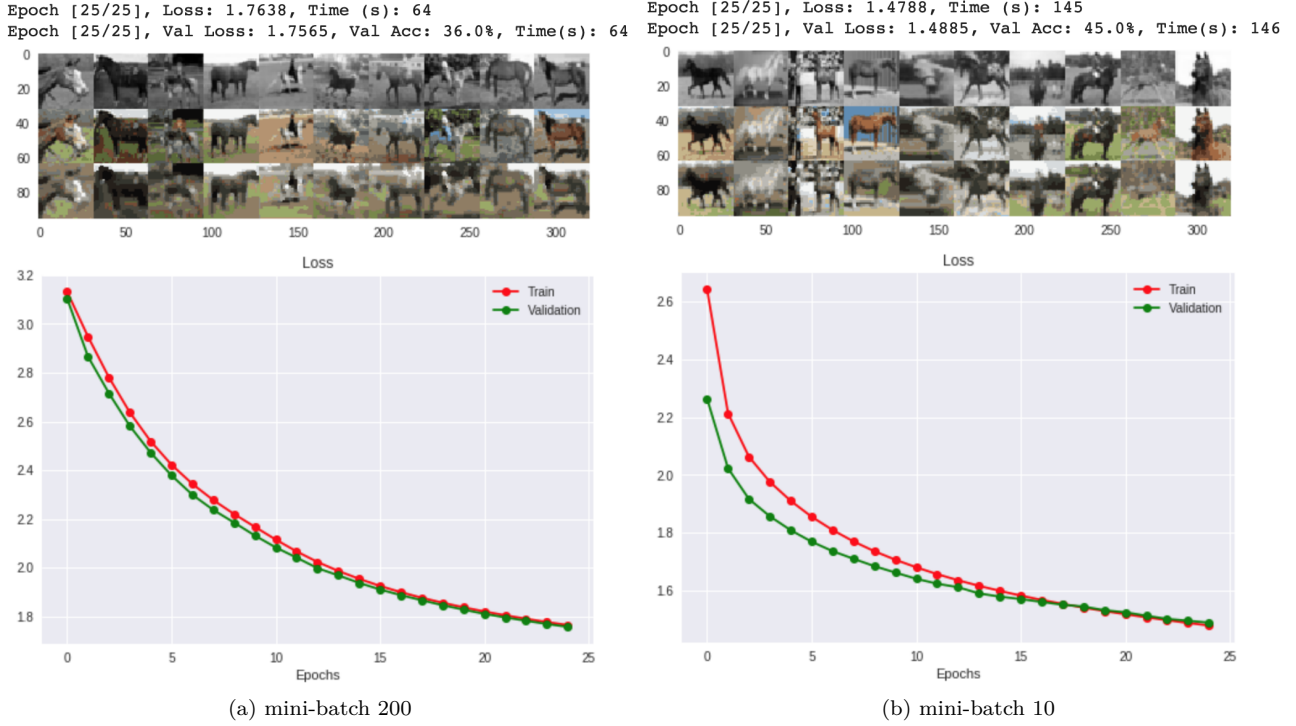


Figure 3: Output images produced by the UNet model with different mini-batch sizes

For the run with a mini-batch of 200, the training and validation losses are very close to each other throughout, and they do not reach the levels of the other two models. For the run with a mini-batch of 10, not only do both curves meet sooner, and this model achieves 45% validation accuracy, but they achieve the lowest loss of these three models, albeit with twice the run-time.

Qualitatively speaking, the model with a mini-batch of 10 does the best job at preserving shapes and general contrast. However, it also suffers from green over-coloration.

Part D

1.

- Resolution of downsized input image: 8×8 ($32/2^2$)
- Resolution of output image : 32×32

2. Comparing CNN and UNet in the super-resolution context, the output of both models is very similar, albeit UNet yields colors and shapes that are closest to the input. Compared to bilinear interpolation, both of these models yield more clearly defined shapes, as well as greater contrast and variety of colors.

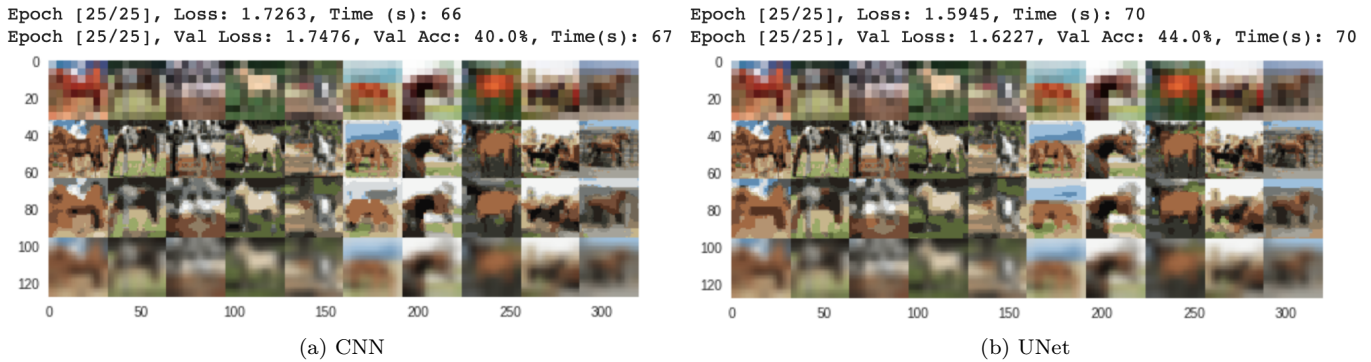


Figure 4: Output images produced by the CNN and UNet models for the super-resolution task

- Conv nets are better because they produce sharper images, i.e., more pronounced (and accurate) edges between shapes
- Conv nets express different textures and lighting better

Part E

1. With regards to the CNN, the activations in the first few layers seem to concentrate mostly on the horse's shapes and edges (which is the centerpiece of the CIFAR category in question). The later layers appear to concentrate on the surrounding features and textures, and thus look grainy.
2. Compared to the colorization CNN, the UNet activations are very similar at the lower and higher layers. However, these are much more detailed and subtle at the lower levels (the image textures and features are much more sharper) than the CNN's. The higher layers look quite grainy.
3. When compared to the activation plots of the colorization models, the activation plots for the super-res UNet appear overall less sharp in the first few layers. This is expected given how these have lower-resolution inputs. The activations are sharper in the last layers, and they somewhat retain the shapes of the lower layers, in contrast with the colorization models' plots.

Part F - Conceptual Problems

1. For the models considered in this assignment, we can tune the following hyperparameters:
 - learning rate
 - mini-batch size
 - kernel
 - num_filters
 - epochs
2. Switching the order of max-pooling and ReLU seems to yield greener pictures when colorizing via CNN. Not only are shades of brown (brown horses) in the original configuration made greener after switching the order of these layers, but green horizontal stripes are added to the colorization in areas that do not require it (according to the validation image).
3. Drawing from Johnson et al., we can improve the evaluation metrics in the models contained in this assignment by looking at perceptual loss, instead of per-pixel loss. Instead of penalizing for deviations between pixel input_{*i,j*} and output_{*i,j*}, we can instead examine differences in higher level features, such as texture, color transition, and brightness—this resembles more closely human perception.
4. To colorize test images larger than 32x32 we can modify `num_filters` to correspond to the dimension of the new, larger images. The architecture need not be changed (number of conv layers), but rather the number of filters used.