LEARNING TO SEE IN THE DARK

Image data - img2img

Edoardo Fava 851665, Alberto Porrini 826306,
Davide Prati 845926

Foundation of Deep Learning

Università degli Studi di Milano-Bicocca
Year 2022/2023

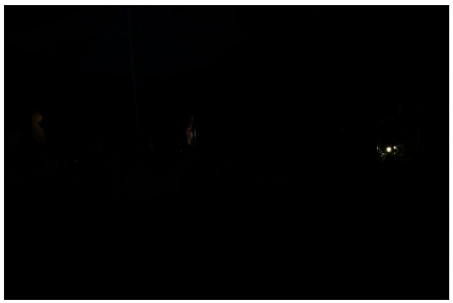
27/06/2023

Summary

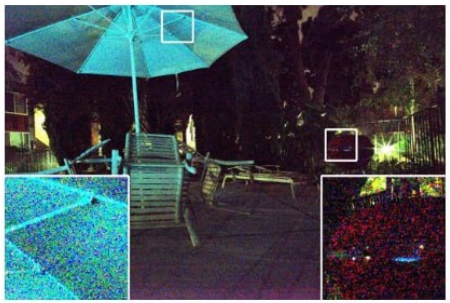
UNIVERSITÀ DEGLI STUDI DI MILANO BICOCCA

Problem introduction

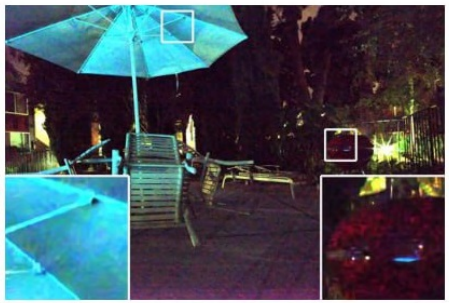
The goal of this project is to become able to 'see in the dark', intended as an extraction which starts from an image, taken in conditions of absence of light, that obtains another picture which instead manages to show the original subject of the frame.



(a) Original dark picture



(b) Traditional pipeline



(c) ...followed by BM3D denoising

(d) Original bright picture

UNIVERSITÀ DEGLI STUDI DI MILANO BICOCCA

Analysis of the data - 1

Our dataset consists in 2697 short exposure images, both indoor and outdoor, each with a corresponding long-exposure image (231).

They are taken with:

- **Camera**: Sony α7S II
- **Illuminance**:
  - Outdoor: between 0.2 and 5 lux
  - Indoor: between 0.03 and 0.3 lux
- **Objective**:

	Focal length	Aperture
First one	21 mm	2.8
Second one	24 - 240 mm	3.5 - 6.3

UNIVERSITÀ DEGLI STUDI DI MILANO BICOCCA

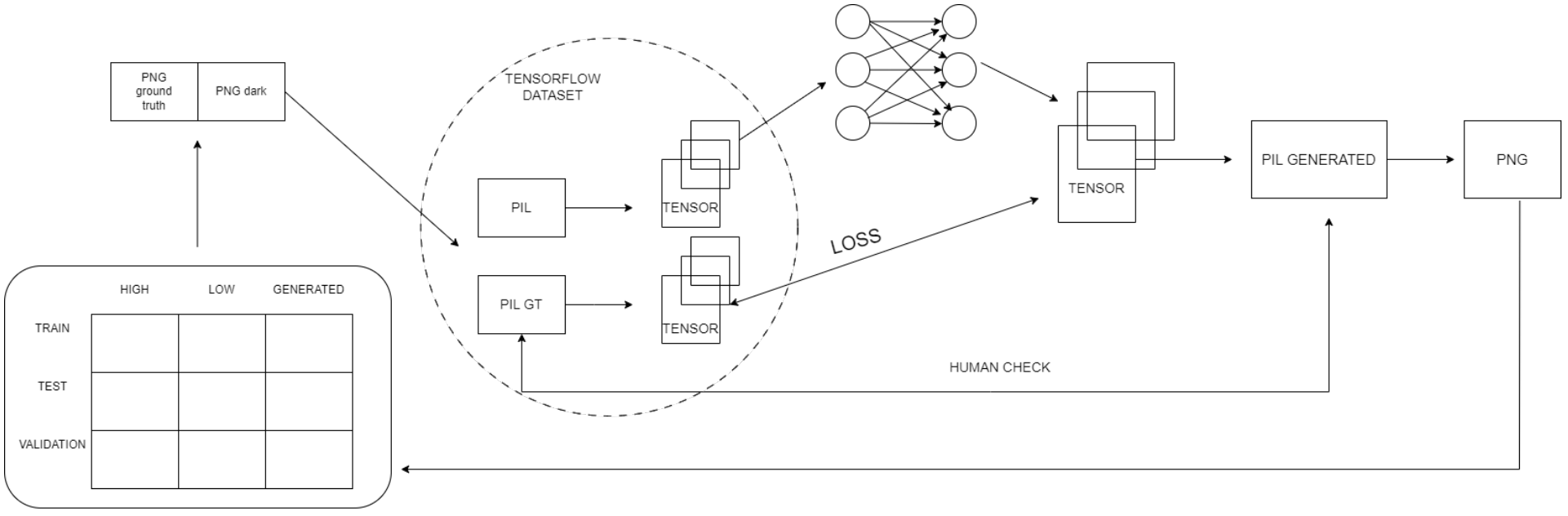
Analysis of the data - 2

The images characteristics are:

- **Resolution**: 4240×2832 pixels
- **Exposure**:
  - Input: between $\frac{1}{30}$ and $\frac{1}{10}$ seconds
  - Ground truth: between 10 and 30 seconds
- **ISO**: set according to ground truth image
- **F-stop**: set according to ground truth image

UNIVERSITÀ DEGLI STUDI DI MILANO BICOCCA

Pipeline

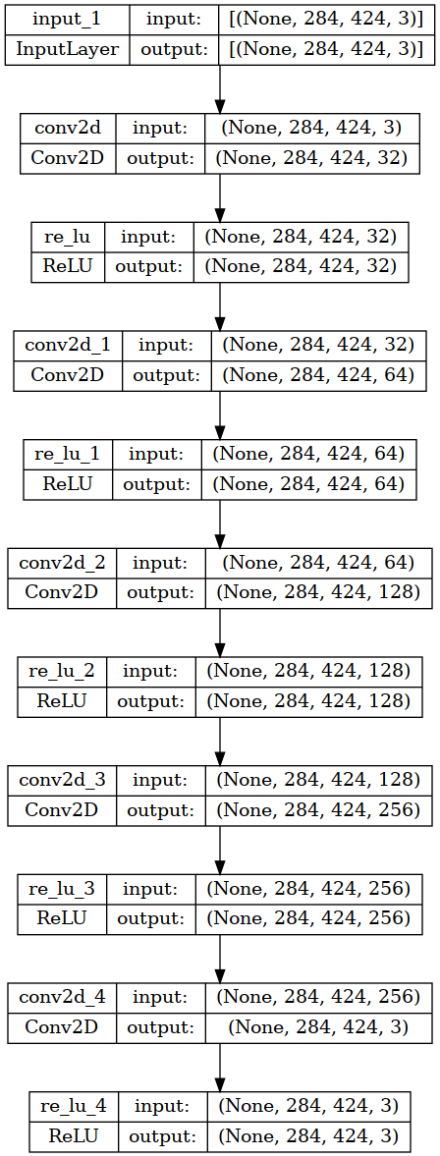


To accelerate the whole process, we reduced images dimensions of $\frac{1}{10}$. We also converted them to png from raw format, to support a more common format.

The tensorflow dataset allows to load images in memory only when necessary.

UNIVERSITÀ DEGLI STUDI DI MILANO BICOCCA

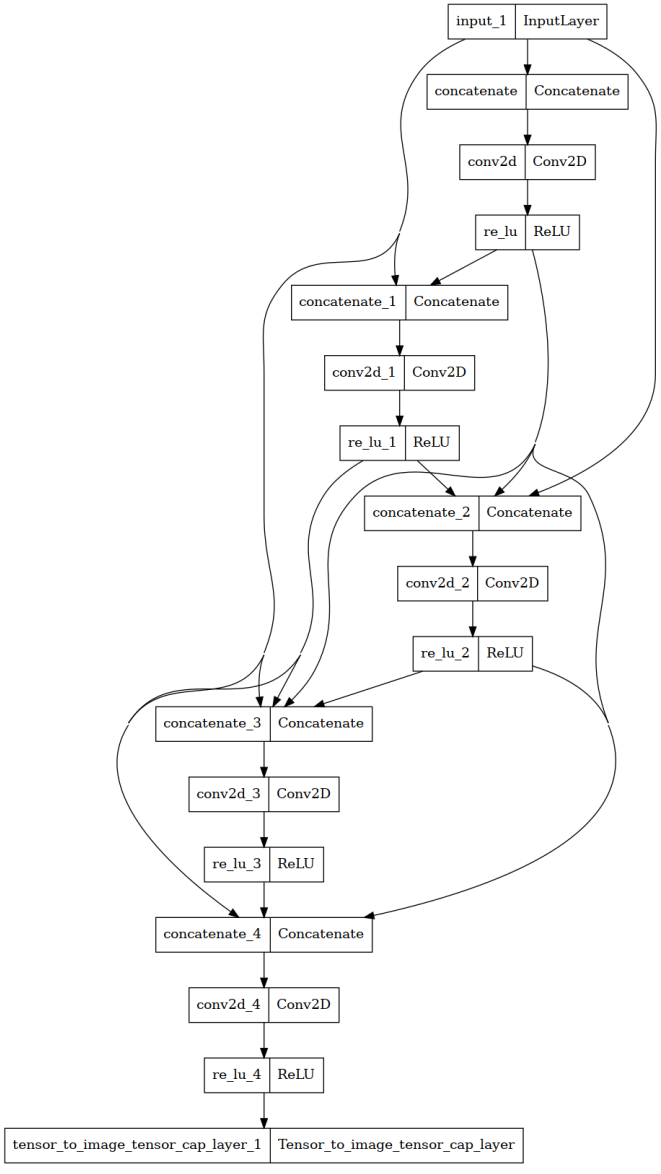
Models - Convolutional Neural Network



- Width and height dimensions (424 and 284) do not change in the whole network (padding = same).
- The filters number doubles at each level starting from an initial value of 32.
- The last layer has a 3 size filter, to respect the number of the original RGB channels.

UNIVERSITÀ DEGLI STUDI DI MILANO BICOCCA

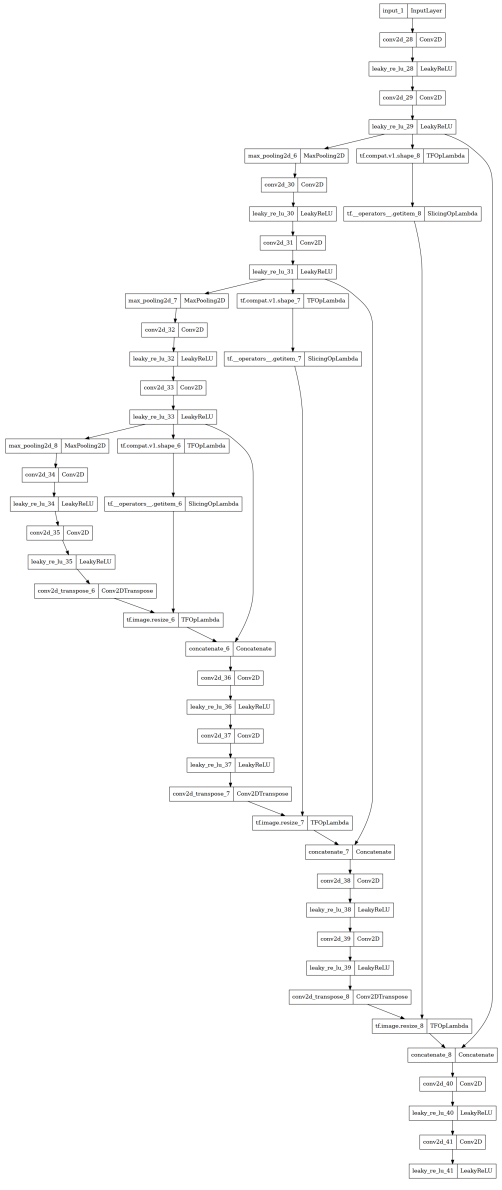
Models - Dense CNN



- After each block of layers, dense networks concatenate the output of all the previous blocks. This allows the loss to propagate (almost) directly to every layer block.
- We built a function that creates dense convolutional networks referring to the standard convolutional networks in the previous slide.

UNIVERSITÀ DEGLI STUDI DI MILANO BICOCCA

Models - U-Net



- The filters number quadruples at each level starting from an initial value of 16.
- The depth is equal to 3, i.e. there are 3 encoder and 3 decoder blocks.
- Due to one unit width and/or height difference between the skip tensor and the decoded tensor, we added a resize operation.

UNIVERSITÀ DEGLI STUDI DI MILANO BICOCCA

Setup - Data augmentation

Considering that we have to avoid the data augmentation options that modify image data values, we make data augmentation only by flipping our image, both vertically and horizontally.
Setting the probabilities of the two flipping both equal to 1/2 we have a 1/4 probability of obtaining each of the following:

Original

Mirrored

Upside down

Rotated of 180 degrees

Setup - Losses to test

We considered three kinds of losses while training:

- L1: Mean Absolute Error (MAE)
- L2: Mean Squared Error (MSE)
- SSIM, discarded because of NaN results

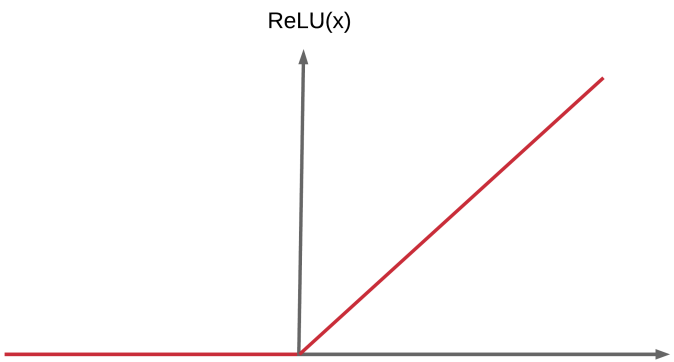
$$\text{MAE} = \frac{1}{b}\sum_{i=1}^{b}|y_i - \hat{y}_i| \qquad \text{MSE} = \frac{1}{b}\sum_{i=1}^{b}(y_i - \hat{y}_i)^2$$

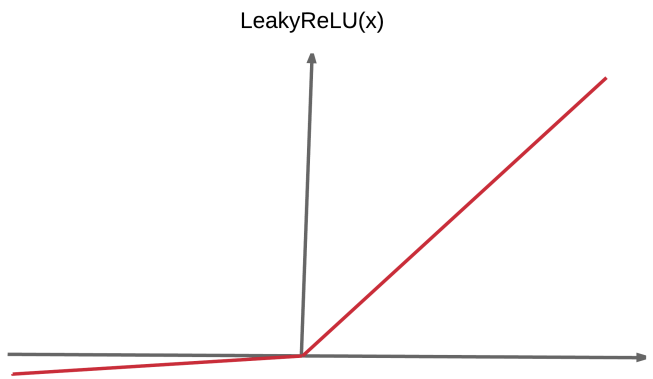
UNIVERSITÀ DEGLI STUDI DI MILANO BICOCCA

Setup - Activation functions to test

Regarding activation functions, we tested:

- ReLU
- Leaky ReLU with slope $\alpha = 0.2$





UNIVERSITÀ DEGLI STUDI DI MILANO BICOCCA

We implemented three callbacks which for each epoch:

- simply return the number of the current epoch
- save the model weights in checkpoints, which are files that are used to restart the training from the weights saved, without losing memory of the epochs already done
- force an early stopping in case there is a given number of consecutive epochs (5) without improvements on a given parameter (loss or validation loss values).

Setup - Output conversion

Problem: while the values of an RGB image go from 0 to 255, the output values of the neural networks are float values, that include negative numbers, numbers over 255, infinities, and non-integer numbers.

Reasoning: We think that it is correct to consider non-integer numbers during the loss computation, but we are aware that values below 0 and over 255 have no meaning.

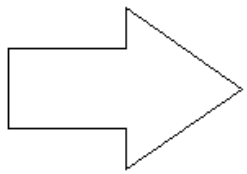
Solution: So, in the last layer we set all negative values equal to 0, and all values above 255 equal to 255 itself.

```
[[[[-1.00e+00 -1.00e+00 -1.00e+00]
   [ 2.00e+00  2.00e+00  2.00e+00]]

  [[ 2.55e+02  2.56e+02       inf]
   [ 4.00e+00  4.00e+00  4.00e+00]]]


 [[[-1.00e+00 -1.00e+00 -1.00e+03]
   [ 2.00e+00  2.00e+00  2.00e+00]]

  [[ 1.00e+36  1.00e+37  1.00e+38]
   [ 4.00e+00  4.00e+00  4.00e+00]]]]
```



```
[[[[  0.,   0.,   0.],
   [  2.,   2.,   2.]],

  [[255., 255., 255.],
   [  4.,   4.,   4.]]],


 [[[  0.,   0.,   0.],
   [  2.,   2.,   2.]],

  [[255., 255., 255.],
   [  4.,   4.,   4.]]]]
```

UNIVERSITÀ DEGLI STUDI DI MILANO BICOCCA

Setup - Human check

We implemented a function that allows to make a direct comparison between the output of one model and the ground truth image.

By default, we decided that the function always takes the same input image, in a way that makes possible to also compare two or more different models.

Output example of our human check function

UNIVERSITÀ DEGLI STUDI DI MILANO BICOCCA

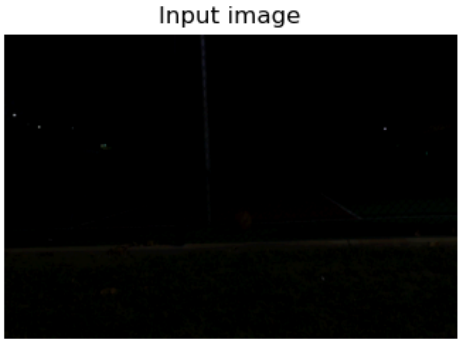
For all our training we set as fixed values the following:

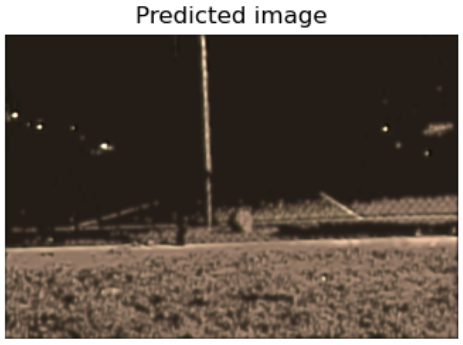
- **number of epochs**: 20
- **batch size**: 10
- **optimizer**: Adam with learning rate = 0.001
- **early stopping callback**: monitoring loss and patience equal to 5

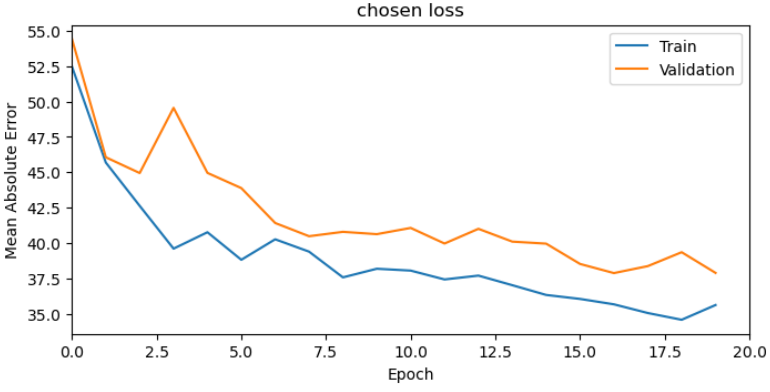
while we let vary:

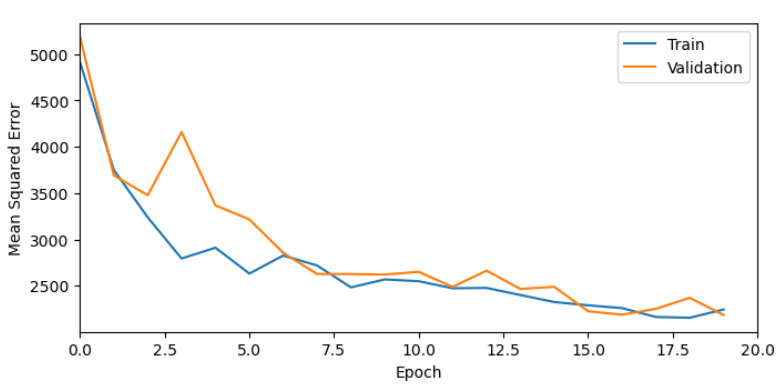
- **the model**: CNN, dense CNN or U-Net
- **data augmentation**: yes or not
- **activation function after the convolutions**: ReLU or LReLU
- **loss function**: MAE or MSE.

CNN - ReLU, MAE and augmentation





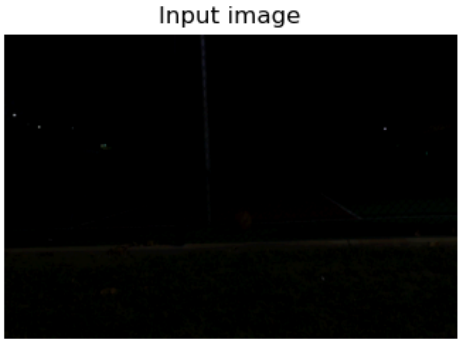


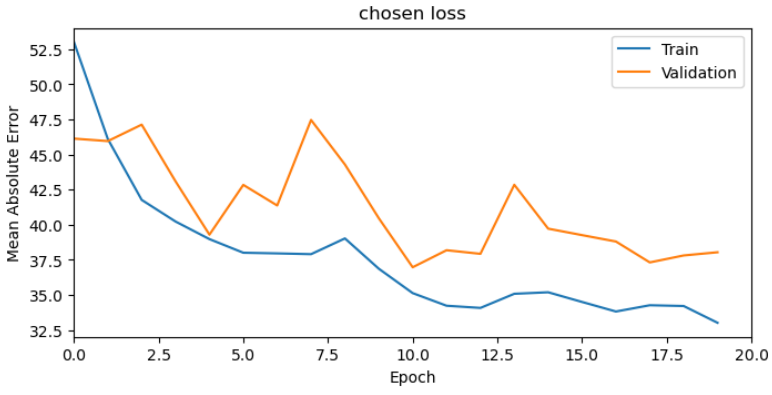


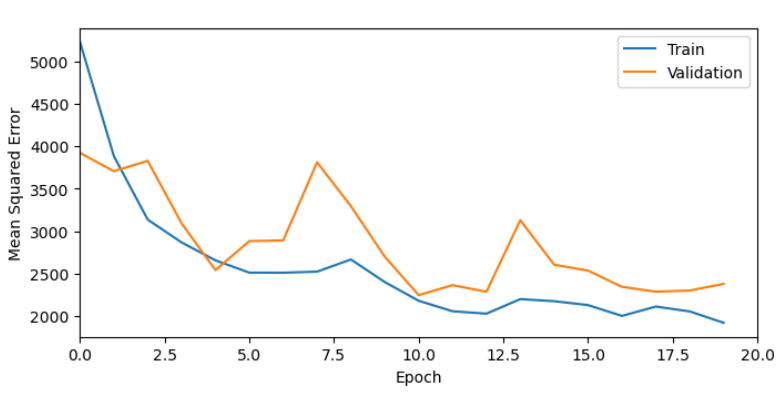
Testing results: MAE = 34.754, MSE = 2057.030.

UNIVERSITÀ DEGLI STUDI DI MILANO BICOCCA

CNN - LReLU, MAE and augmentation



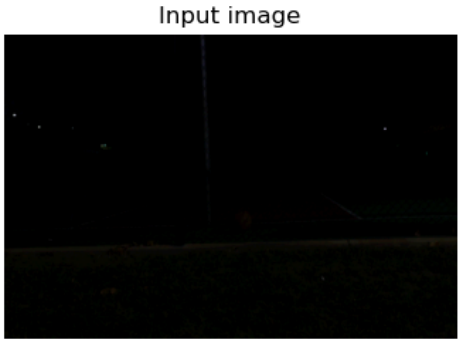


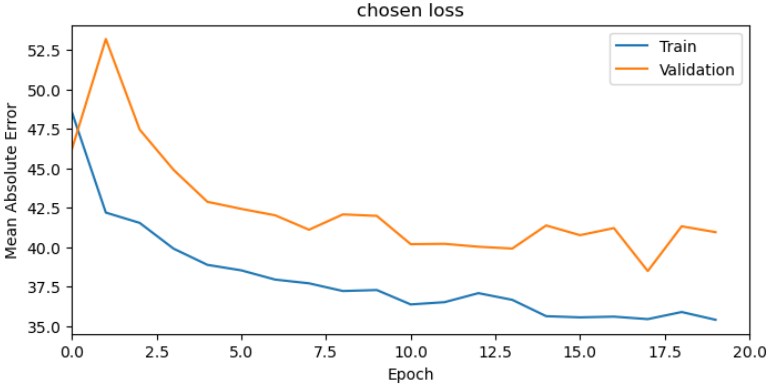


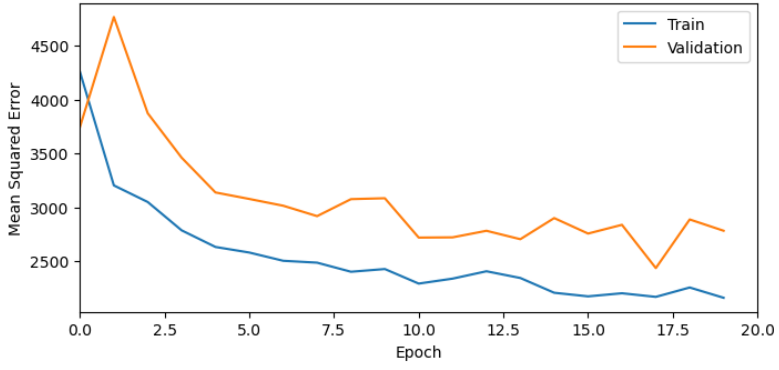
Testing results: MAE = 34.882, MSE = 2071.059.

UNIVERSITÀ DEGLI STUDI DI MILANO BICOCCA

Dense CNN - LReLU, MAE and augmentation



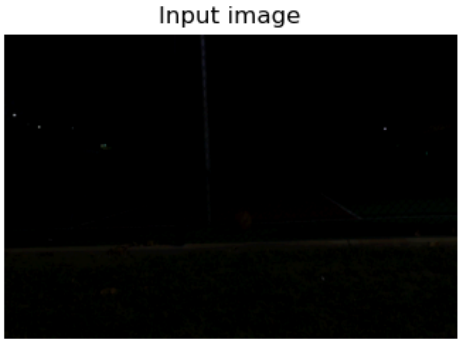


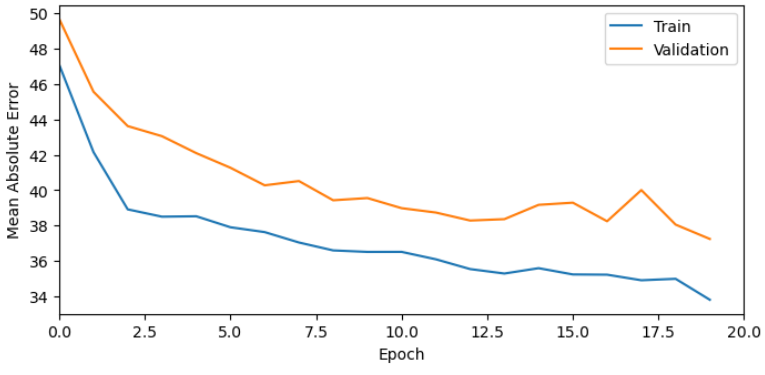


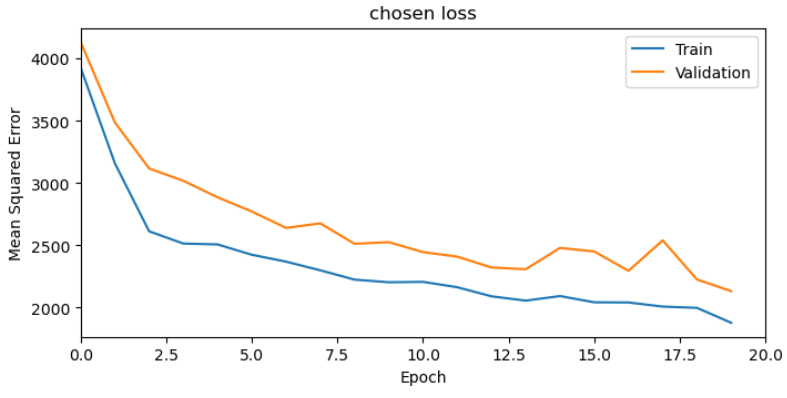
Testing results: MAE = 35.540, MSE = 2189.077.

UNIVERSITÀ DEGLI STUDI DI MILANO BICOCCA

Dense CNN - LReLU, MSE and augmentation



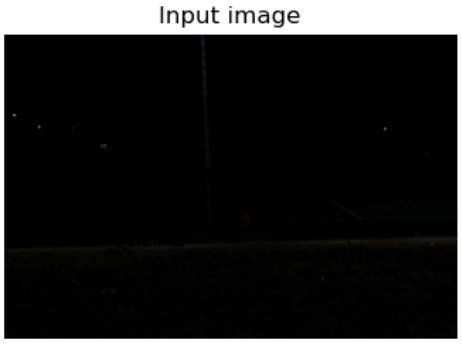


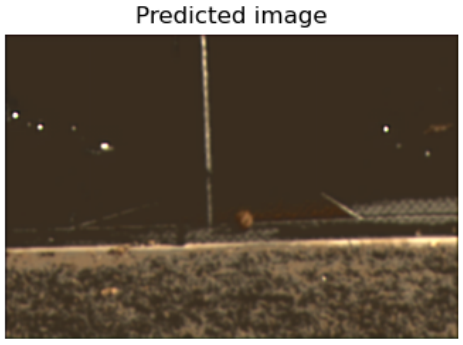


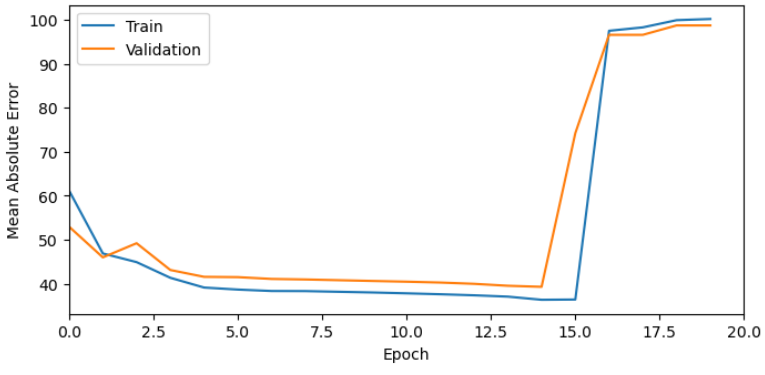
Testing results: MAE = 34.337, MSE = 1921.527.

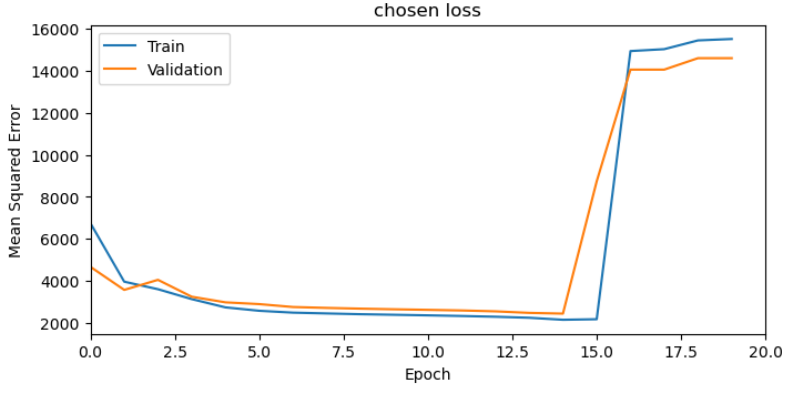
UNIVERSITÀ DEGLI STUDI DI MILANO BICOCCA

U-Net - LReLU, MSE and augmentation





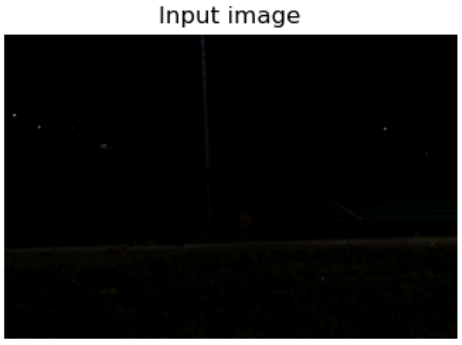


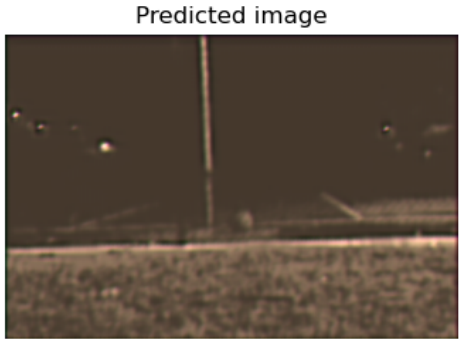


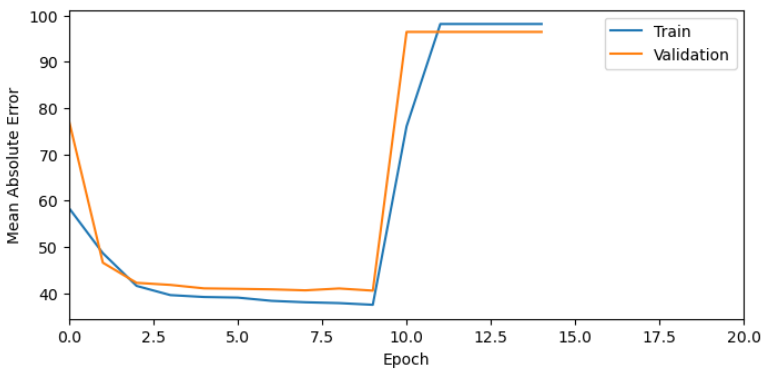
Testing results: MAE = 35.793, MSE = 2102.753.

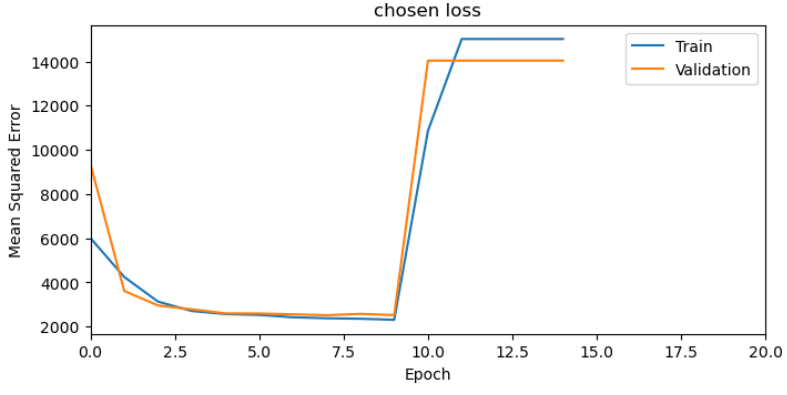
UNIVERSITÀ DEGLI STUDI DI MILANO BICOCCA

U-Net - LReLU, MSE and no augmentation









Testing results: MAE = 36.911, MSE = 2179.050.

UNIVERSITÀ DEGLI STUDI DI MILANO BICOCCA

Future developments

- Of course we have done just a subset of all the possible combinations of choices, we could develop also the others.
- We can also considerate to vary the parameters we decided to set as fixed values.
- But the ideal purpose is to have results as good as the ones of the 'Learning to see in the dark' paper, keeping in mind that our input format is png and not the raw camera data they used.

U-Net with LReLU, MAE, no augmentation, setting the epochs number to 50

UNIVERSITÀ DEGLI STUDI DI MILANO BICOCCA

Conclusions

- Firstly we have to mention that we can not make any solid conclusion since we do not have proof or statistical evidence that one training parameter (including the model) is better than its alternatives.
- From our results, the best model seems to be the dense CNN, set with LReLU, MSE and augmentation.
- We are aware that U-Net can perform at least similarly to Dense CNN, so we can conclude that a different parameters choice can make the model much better.

Thank you for your attention!

UNIVERSITA' DEGLI STUDI DI MILANO BICOCCA