

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

BIOMEDICAL IMAGING GROUP

Image deconvolution using Deep-Learning

Clémentine AGUET

Supervised by
Emmanuel SOUBIES
Kyong JIN
Professor Michael UNSER

April 2, 2017



ÉCOLE POLYTECHNIQUE
FÉDÉRALE DE LAUSANNE

1 Introduction

Recovering degradation, such as blurring and noise addition, occurring during microscopy image acquisition, is typically classified as inverse problem. They are classically solved by Wiener filter and deconvolution by minimizing regularized Least-Squares. Those methods have the inconveniences of difficult hyperparameter optimization and to require some prior knowledges about the degradation model (PSF, noise, constraint, etc.). Deep learning is an emerging trend, particularly used and powerful for image classification or segmentation. The possibilities of such network have also started to be considered for questions related to inverse problems and image restoration. In this project, the potential of convolutional neural network (CNN) was explored for the purpose of reconstructing high quality microscopy images.

2 Methods and models

2.1 Degradation model

During the process of microscopy image acquisition, the original image is unfortunately degraded resulting in a blurry and noisy one. It mainly comes from the resolution of the devices, small shifts between equipment and sample or acquisition conditions (exposure time, illumination, temperature) [10].

The principal factor limiting optical microscopy resolution and restraining the ability to observe fine details of biological molecules or processes comes from light diffraction. The maximal resolution can be described by Abbe's diffraction limit [3]. This value is characterized in term of the light wavelength and the numerical aperture (NA) (1).

$$d = \frac{\lambda}{2 \cdot NA} \quad (1)$$

Looking at the reciprocal space, the microscopy resolution is here represented as a circle, commonly named observable region, where low resolution information is near the center and frequencies higher than the cutoff frequency are not recorded by the device.

Image degradation can be mathematically modeled as a combination of two operations in a linear process (2). The original image is first convolved with a Point Spread Function (PSF), followed by Gaussian noise addition $n(x, y)$.

$$b(x, y) = PSF(x, y) * s(x, y) + n(x, y) \quad (2)$$

The PSF describes the transformation occurring during the image acquisition process. It represents the light diffraction pattern. The convolution operation with the PSF has an impact on and reduces the high frequency components. It smoothens the edges, lowering the contrast.

The PSF of a microscope with a circular aperture corresponds to an airy disk pattern (centered disk with surrounding rings) [12]. The diameter of the first disk is defined by the numerical aperture of the optical system (NA) and takes part of the theoretical calculation of the resolution. The cutoff frequency is related to the numerical aperture by the following equation: $f_c = \frac{2 \cdot NA}{\lambda}$ with λ the illumination wavelength [9]. As the cutoff frequency decreases, the diameter of the airy disk grows, producing a blurrier image.

For simpler and faster computation, equation (2) can be written in the frequency domain (3), where the convolution operator becomes a multiplication.

$$B(u, v) = OTF(u, v) \cdot S(u, v) + N(u, v) \quad (3)$$

The Optical Transfer Function (OTF) corresponds to the Fourier transform of the PSF. Equation (4) defines the airy disk pattern in Fourier space [2][14].

$$OTF(f) = \begin{cases} \frac{1}{\pi} \left[2 \cos^{-1} \left(\frac{f}{f_c} \right) - \sin \left(2 \cos^{-1} \left(\frac{f}{f_c} \right) \right) \right] & \text{if } f < f_c \\ 0 & \text{if } f > f_c \end{cases} \quad (4)$$

The noise introduced during the imaging process can be modeled as a Gaussian noise. It is generated by assuming independence at a pixel level and is based on its mean and standard deviation.

$$n(x, y) = \sigma_{noise} \cdot randn(imagesize) + \mu_{noise} \quad (5)$$

The *randn* function in equation (5) returns normally distributed random numbers. The noise variance can be computed from the Signal-to-Noise ratio (SNR) by isolating it from equation (6). Smaller the SNR, larger the noise impact.

$$SNR = 10 \cdot \log_{10} \left(\frac{\sigma_{image}^2}{\sigma_{noise}^2} \right) \quad (6)$$

2.2 Inverse problem

From the degraded images, the aim is to produce a high-resolution image as closed as possible to the original one, which can be assessed using restoration algorithms. Denoising and deconvolution are such methods. They are characterized as inverse problems.

Most of the commonly used restoration algorithms are so called non-blind methods. They require assumptions and prior knowledges regarding the PSF, the type and level of noise added or non-negativity constraint [10]. It is the case of inverse filters such as the classical direct Wiener deconvolution or an iterative least-square deconvolution.

2.2.1 Wiener deconvolution

Wiener deconvolution is one of the most widely used inverse filter. This method attempts to reduce the noise effect while simultaneously removing the blur.

An approximation of the original image can be defined by inverting the problem of equation (2):

$$\hat{s}(x, y) = g(x, y) * b(x, y) \quad (7)$$

This estimation is performed in the frequency domain [10]

$$\hat{S}(u, v) = G(u, v) \cdot B(u, v) \quad (8)$$

$$G(u, v) = \frac{H(u, v) \cdot W(u, v)}{|H(u, v)|^2 \cdot W(u, v) + N(u, v)} \quad (9)$$

2.2.2 Deconvolution by minimizing regularized Least-Squares (LSTV)

The aim of this approach is to minimize the mean squares error with total variation (TV) regularization and positivity constraint. This ill-posed inverse problem can be mathematically expressed as follow:

$$\arg \min_{\mathbf{f}} \frac{1}{2} \|\mathbf{H}\mathbf{f} - \mathbf{y}\|^2 + \lambda \|\mathbf{f}\|_{TV} + i_{\geq 0}(\mathbf{f}) \quad (10)$$

The vector \mathbf{y} represents the measurements, \mathbf{H} corresponds to the PSF matrix of the system and \mathbf{f} is the reconstructed image. λ is a non-negative hyperparameter that should be defined. The $\|\mathbf{f}\|_{TV}$ term denotes the regularizer, while the last element of this minimisation problem induces

non-negativity constraint. By incorporating appropriate regularizers, images of great quality can be reconstructed. The regularizer considered here is total variation (TV) defined by

$$\|\mathbf{f}\|_{TV} = \sum_n \sqrt{(\nabla_x \mathbf{f})_n^2 + (\nabla_y \mathbf{f})_n^2} \quad (11)$$

As it is not possible to derive a closed form solution from equation (10), iterative algorithms (ISTA, ADMM, FISTA or SALSA) are the most common approach to solve ill-posed inverse problems.

There are multiple optimization algorithms to solve inverse problems. The alternating direction method of multipliers (ADMM) is an efficient one. It takes advantage of variable splitting, well suited when dealing with regularizers such as total-variance and conditioning like positivity [13].

The functions used for this deconvolution are the one implemented in the inverse problem library developed in the BIG laboratory.

2.2.3 Deep Convolutional Neural Network

The main drawbacks of iterative algorithms are related to hyperparameter optimization as well as prior model knowledge. An emerging trend considering those points and capable of solving ill-posed inverse problems is deep convolutional neural network. This approach has already proved its worth regarding image classification and segmentation. Applying deep learning for the purpose of denoising [17] or deconvolution [18] has also started to be investigated.

The potential of this emerging method to solve inverse problems is addressed in this project. By giving as inputs to a Convolutional Neural Network a degraded image or the result of a first deconvolution, would it be possible to obtain as output of a high quality image, as close as possible to the original one? This approach is challenged in this project to evaluate its power regarding restoration of microscopy images.

For this purpose, CNN is preferred to regular neural network as it is well adapted to image input and deals with less parameters that must be learned. Moreover, the spatial structure is a point considered in CNN [16], where distant or adjacent pixels are handled in the different way. Another important aspect of CNN is parameters sharing. Assuming that a feature relevant at a certain position should also be relevant at other, weights and biases are shared in a feature map.

The CNN architecture used in this project is the one developed by K.H. Jin et al.[15], which is itself based on U-Net model [19]. It was first implemented for the purpose of image segmentation but contains some characteristics having analogs in iterative methods.

The first one is multi-level decomposition based on max-pooling, which ensures that a layer is surrounded by layers having smaller effective filter size. Max-pooling represents a downsampling operation, which reduces the spatial dimension and thus lowers the number of parameters and computations [20]. This decomposition operation can be compared to the multi-resolution wavelets approach in iterative methods [15]. The second interesting property is multichannel filtering. In order to increase the neural network capacity, several feature maps can be placed within each layer. For a similar approach in iterative algorithms such as ADMM, the split variables can be considered as channels.

The residual learning is an important concept that is incorporated to enhance the U-net. At a layer node, the inputs of previous layer are accessible and it is the variation between the input and output that is learned by the network. It facilitates the optimization and reduces the vanishing

gradient problem [21]. Zero-padding is added to the U-net architecture. It has the potential to preserve spatial size of output volume after each convolution step [20]. The last implementation corresponds to the use of a convolutional layer for the last one in order to obtain a unique output image.

2.2.4 LSTVConvNet

K.H. Jin et al.[15] focused their research on biomedical images reconstruction, such as CT (Computed Tomography), MRI (Magnetic resonance imaging) and DT (diffraction tomography). For this purpose, they developed a model combining a fast approximation with a CNN. FBPCo-vNet algorithm performs filtered back projection (FBP) of the inputs and use the FBP outcomes to train the CNN. This approach facilitates the learning and results in improved reconstruction performance compared or iterative methods.

They also demonstrated that by applying FBPCo-vNet, efficient sparse-view reconstruction can be achieved [15]. This reduction of view is quite relevant for human medical imaging as it is highly proportional to radiation received.

This project is not dealing with CT images but the main objective seems similar. By using an equivalent CNN, would it be possible to restore degraded microscopy images?

2.3 Structured Illumination Microscopy

The performance of a deep learning network is highly influenced by the data used during the training phase, and the system needs a gold standard as output. An idea would be to use super-resolution SIM images as so and by the same occasion to improve resolution starting from a wide-field.

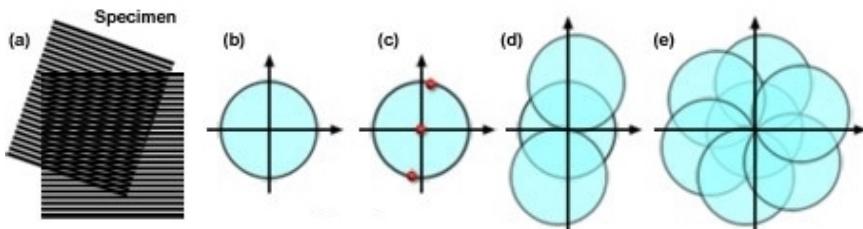


Figure 1: Resolution enhancement with Structured Illumination Microscopy [7]

Super-resolution microscopy techniques were developed to bypass this diffraction limit and collect information out of the observable region. Structured Illumination Microscopy (SIM) is one of them, capable of doubling the spatial resolution (Figure 1.e) [1]. The basic principle underlying SIM is an illumination of the object with spatially structured light. It takes advantage of the Moiré effect [3], where the superposition of two periodic patterns leads to the apparition of Moiré fringe (Figure 1.a). The observation region itself is not enlarged, but high frequency previously beyond the resolution limit are shifted into the observable region. After acquiring images with the pattern of varying orientations and phases, a super-resolution image is computationally reconstructed.

As the technique adapts easily to varying fluorescent probes and does not require high illumination power, it is a huge potential for live-cell imaging [3], which is essential in biological research. However, the temporal resolution of such system is mainly restricted by the large number of acquisition required for the computational reconstruction of super-resolution. This condition lowers the process speed and frequently induces photobleaching [8]. Sample illumination with

uncontrolled random speckled [4] instead of known patterns is of great interest for SIM as it facilitates the pattern generation process, but the question regarding the number of acquisitions for the reconstruction is even more important. Furthermore than using SIM super-resolution images as output to the CNN, the idea would be to take advantage of deep learning to reduce the number of acquisitions and to estimate the illumination patterns. Would it be possible to process uniquely the wide-field or the reconstruction of fewer acquisitions through a convolutional neural network and still obtained a super-resolution image? This question reveals how this project could be further developed.

2.4 Metrics

The original image and the restored version of the blurry noisy one obtained with either the Neural Network or the regularized non-negative Least-Squares approach (LSTV) can be compared in order to get an evaluation of the restoration performance. Human visual system can be defined as the absolute referee of image quality. The idea is to find a measurement metric approaching it as much as possible [22]. A constrain is often the requirement of a noiseless reference signal. Most of the commonly used quality measurement computations are based on the Mean Squared Error (MSE) (12), with the advantage to be parameter free with a low computation complexity.

$$\text{MSE}(X, Y) = \frac{1}{m \cdot m} \sum_{j=1}^m \sum_{i=1}^n (X_{ij} - Y_{ij})^2 \quad (12)$$

This quantitative value gives information regarding the degree of distortion between the prediction and the actual observation. The image restoration quality increases as the error gets smaller. The first error estimation computed in this project is the Root Mean Squared Error (RMSE) defined below (13).

$$\text{RMSE}(X, Y) = \sqrt{\text{MSE}(X, Y)} \quad (13)$$

Another frequent quality indicator in image processing is the Peak Signal to Noise Ration (PSNR) (14), which is also a measurement based on MSE. Higher PSNR value corresponds to better image quality.

$$\text{PSNR}(X, Y) = 10 \cdot \log_{10} \left(\frac{\max(X)^2}{\text{MSE}(X, Y)} \right) \quad (14)$$

2.5 Data simulation

Before testing on real SIM images the approach of deep learning to reduce the number of acquisitions, the feasibility of this method as well as the training set selection are performed on synthetic data. There exist multiple possibilities to generate such images. As the purpose of this project is related to microscopy imaging, the data simulated should be close to biological structures commonly observable and coherent with SIM real data. Multiple segments crossing each others is a schema typically present in micro-scale organisms (microtubules, microfilaments, etc.). Therefore, structures containing segments of random size, orientation and location are generated and serve as reference images for the following part of the project.

The number of segments present are divided into two categories, low density with a value randomly chosen between 10 and 100 or high density with a number between 110 or 200. Same kind of choice for the size. Segments characterized as small have a length between 5 and 30, while they are considered large if the value is between 35 and 60. Each segment has an orientation randomly chosen between 0 and 180 degree. Finally, the segments are arranged in a particular distribution chosen between five possibilities, uniform, Gaussian, Gaussian multivariate, Gaussian mixture and Gaussian ring. In total twenty different configuration types were created and are represented on Figure 2.

Larger the number of data available to train and test the model, more accurate the estimation of its performance. This concept makes it relevant to generate a diverse dataset.

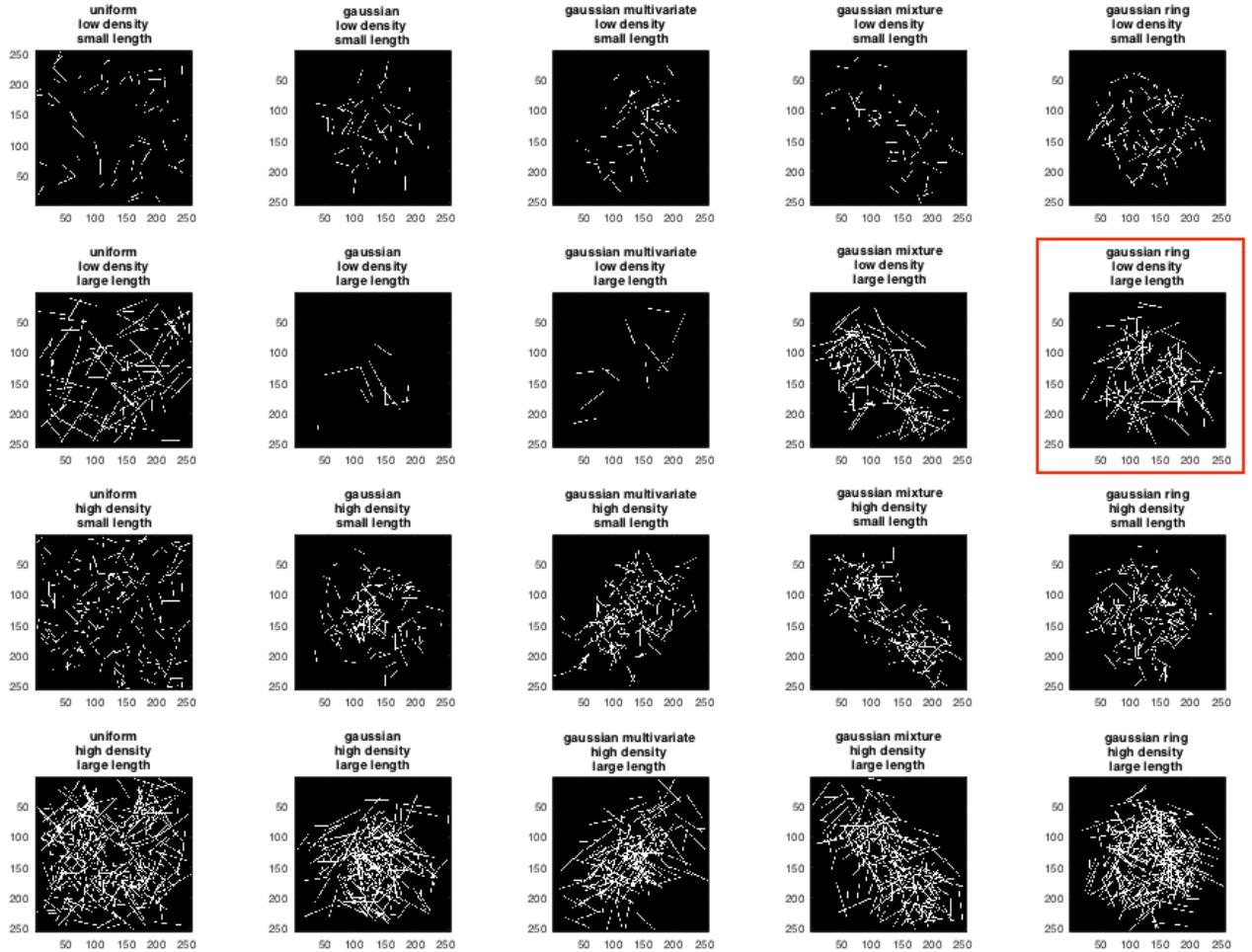


Figure 2: Various configuration type (distribution, density and segment length)

An important point to consider is that, even if only one type was considered for the training phase of the CNN, there still is a certain variability between each sample, which is well illustrated on Figure 3. The number of segments, their orientation as well as their size are chosen in quite large intervals. If it has not been the case, the model will be trained on images that look too similar to each other and will tend to overfit, which means that a high training performance will be achieved but poor results obtained when testing the model on new data.

For this project, the training phase of the model was only performed on the type where segments have a Gaussian ring distribution, are present in low density and of large length. During the selection of the best training configuration of the network, the testing set used to evaluate the system performance consists of exactly the same type. However, in order to assess the accuracy over various data configurations, the final and most efficient model is tested on all the type of data created. Depending on this result, it might be relevant to mix different types to form the training set and generate an even more generalizable model.

It is also important to point out that, as a first approach, the data generated are quite specific and not exactly similar to biological microscopy images. For instance, all the segments have the same and highest intensity, no nuance. As well, the background is completely homogenous and

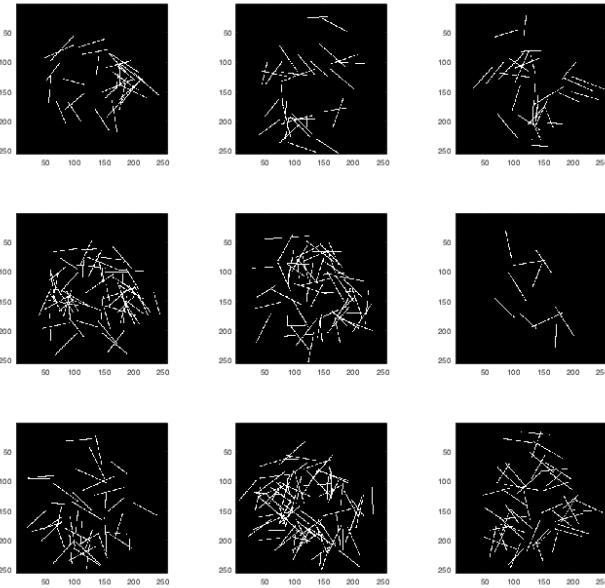


Figure 3: Data variability within the type highlighted on Figure 2

back.

A non-degraded image and a zoom on a particular region containing high frequency components are presented on Figure 4. They will serve as reference to compare the models performance during the whole project.

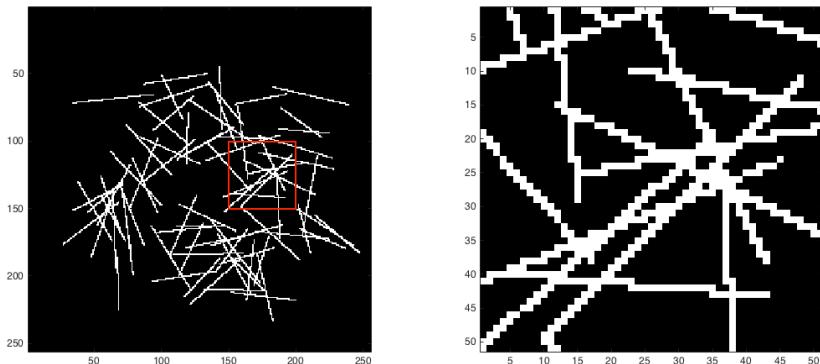


Figure 4: Reference image and a zoom on high frequency region

3 Results and discussion

3.1 Image degradation

In the project, to simulate the degradation typically occurring during image acquisition process, an airy disk PSF is created based on cutoff frequencies. This PSF is convolved with a reference image and noise is added, resulting in a blurry-noisy image used as input for restoration process. The original images are degraded with various level of blur and noise. The blurring degree is defined in term of the cutoff frequency of the OTF taken taking values between 0.1 and 0.4. With a small cutoff frequency, the segment edges are more smoothed than with a larger one. The noise rate, on the other hand, varies depending on the SNR, here between 5 and 25 dB. Lower the SNR, noisier the image.

All the possible levels of blur and noise generated are illustrated on Figure 5, with also the zoom (corresponding to Figure 4) on Figure 6.

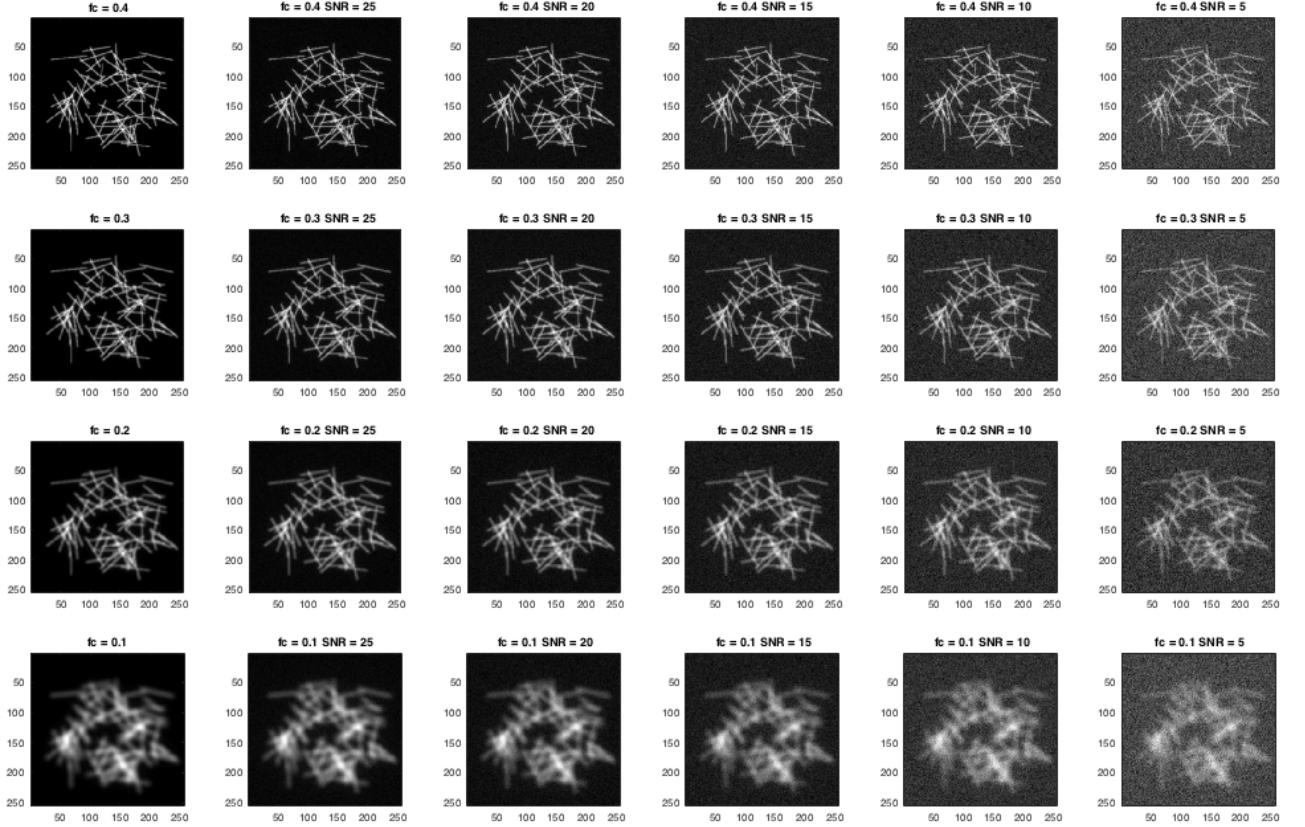


Figure 5: Various level of blur and noise

3.2 Deconvolution by minimizing regularized Least-Squares

As a first and commonly used approach to solve inverse problem, a deconvolution by minimizing TV-regularized Least-Squares (LSTV) with ADMM optimizer was performed. It serves as a reference to compare the accuracy of the novel method based on deep learning. The outcomes obtained with this deconvolution were also used as inputs to the Convolutional Neural Network to even further enhance its performance.

Looking back at equation (10), as previously said, λ , representing the step size, is an hyperparameter that needs to be determined. The LSTV deconvolution was performed with a large range of different λ in order to estimate the most appropriate one. The average SNR score for the diverse λ and all the possible level of blur and noise are presented on Figure 7. Changing the λ does not seem to have a big impact on the model performance. However, taking a value too big (e.g. 0.1 or 0.01) degrades quite a lot the outcomes. On the other hand, if it is too small, the time to reach the stopping criterion is bigger and not even for a great improvement. For those reasons, a λ of 4e-4 was chosen, which is located in the middle of the proposed range.

Once the hyperparameter λ defined, the LSTV deconvolution can be performed over the entire testing set (50 images with all the generated degradation levels). The metrics RMSE and PSNR are used to evaluate the performance of this approach and are represented on Figure 8. The reconstruction was efficient for small level of noise and blur but seems worse as the cutoff frequency or the SNR rate decreases.

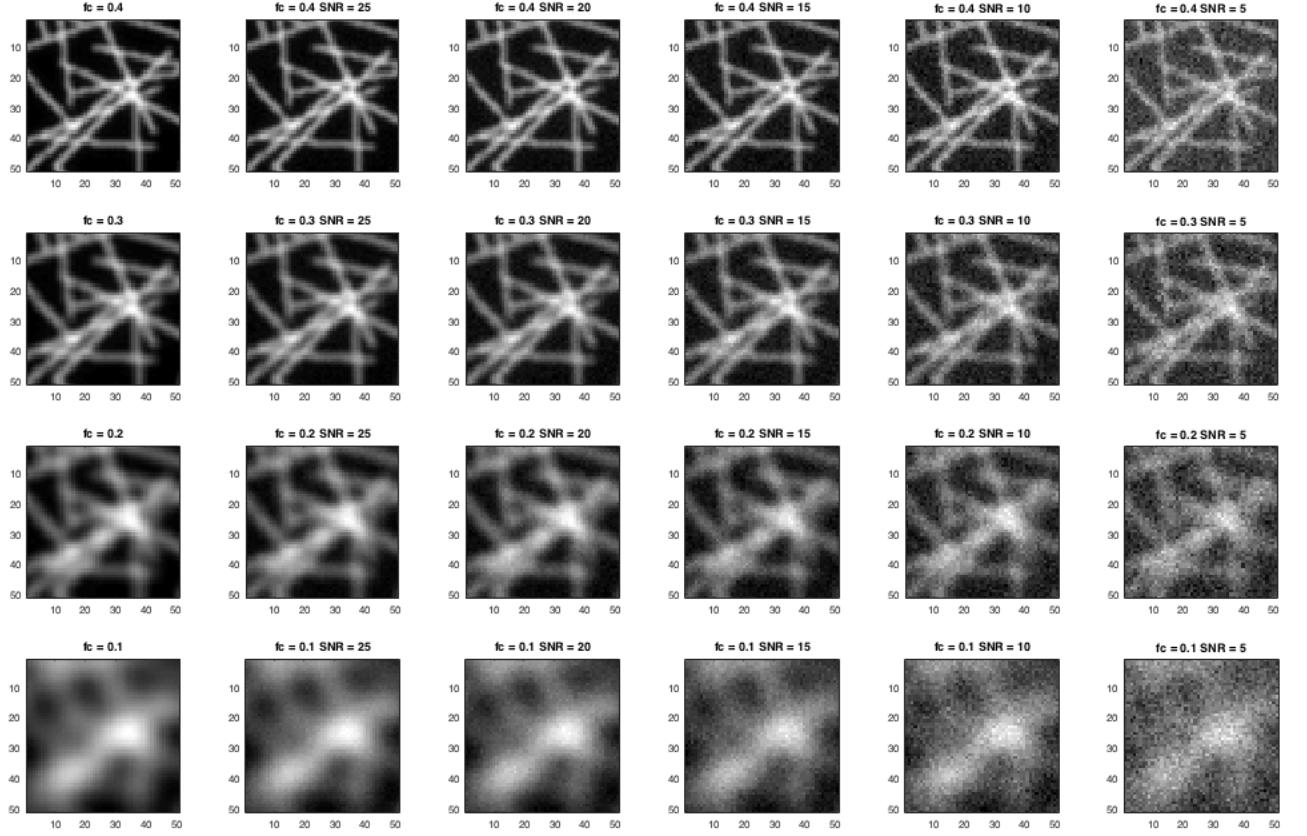


Figure 6: Zoom on various level of blur and noise

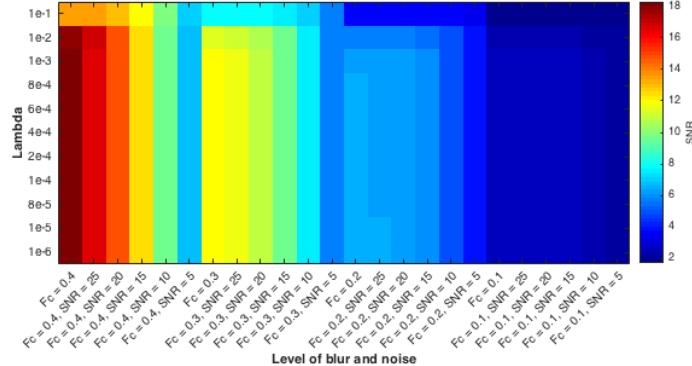


Figure 7: Various lambda for different level of blur and noise

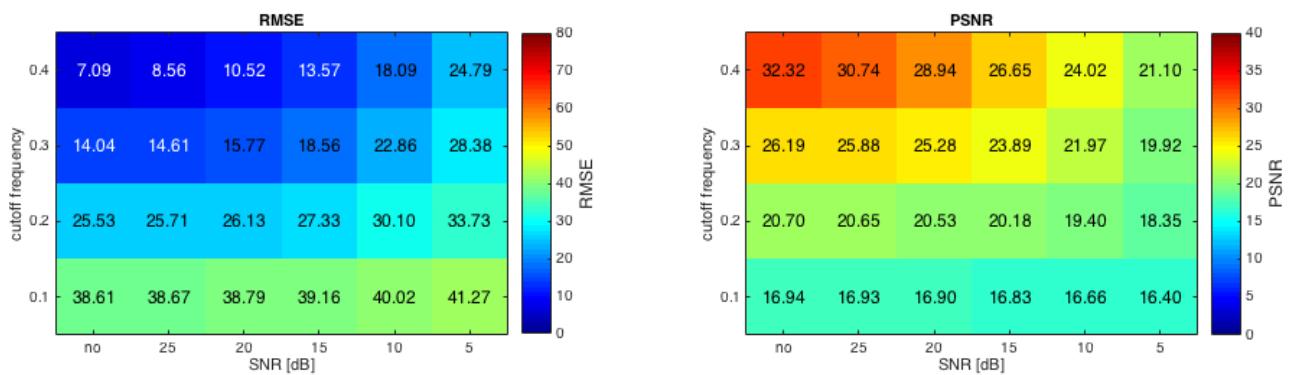


Figure 8: RMSE and PSNR metrics for LSTV reconstruction

The LSTV reconstructions are illustrated on Figure 9, here with a zoom on the region containing high frequency components. The worst outcomes are the ones obtained from inputs with the smallest cutoff frequency ($f_c = 0.1$).

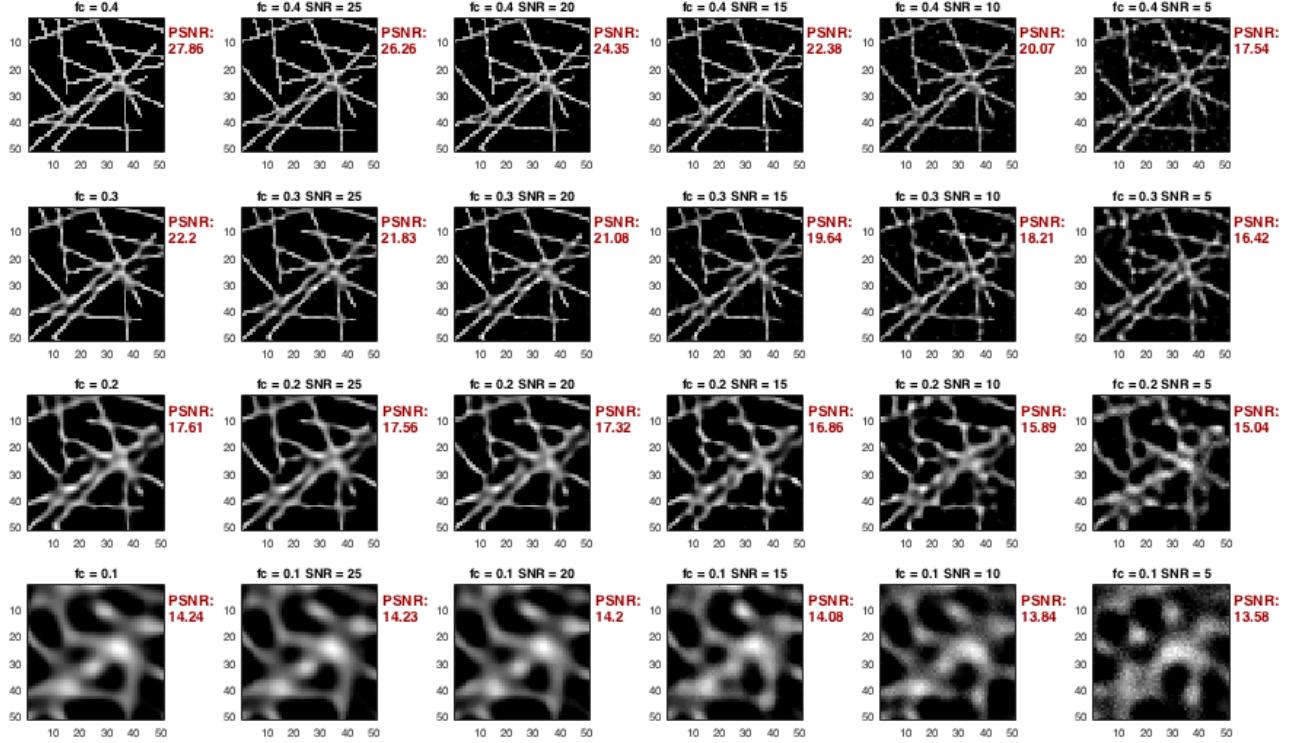


Figure 9: Zoom on LSTV reconstruction for levels of blur and noise

3.3 Convolutional Neural Network (CNN)

Now that a reconstruction comparison is made, the question of deep Convolutional Neural Network to solve inverse problem can be treated. As said previously, the CNN architecture had already been implemented (K.H. Jin et al.[15]). The next step of this project consisted then to understand the proposed network and estimate the most accurate training configuration, as the model performance is highly influenced by the type of data given to train it.

A model should remain the most simple as possible, while being well generalizable, which means getting good testing accuracy on new data (other level of degradation or different types). If it is not the case, the system will tend to overfitting, predicting pretty well the training images but poorly new ones. Then, which level of blur or noise should be given to train the network? The training is first performed on a restrained set containing 450 images. Once it is done, the model is tested on 50 new images for all generated levels of blur and noise. The RMSE and the PSNR gives then an estimation of the system performance and help to find out if it is well homogenous.

3.3.1 Specific cutoff frequency and SNR

To begin with, images degraded with a specific cutoff frequency (f_c) and SNR were used as training set. It will give us a first impression of the network competences. A condition with middle ranged degradation level was tried. All the training images were deteriorated with a cutoff frequency of 0.2 and a SNR of 15 dB (Figure 10).

The two scores (Figure 11) reveals that a network trained with a unique level of blur and noise is most effective on testing data containing the same or a lower degree of degradation. A good

homogeneity across the SNR rate can be observed, which is unfortunately not the case regarding the cutoff frequencies. The reconstructions (Figure 12) are really of poor quality for the most blurry input images ($fc = 0.1$). Similar constatations for cutoff frequencies of 0.4 and 0.3, also not considered in the training, even if the results are less degraded.

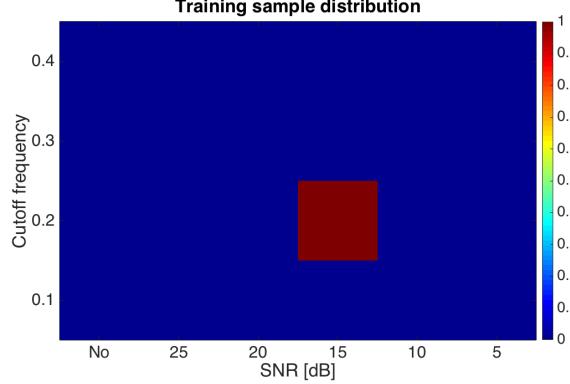


Figure 10: training sample distribution ($fc = 0.2$ and $SNR = 15$)

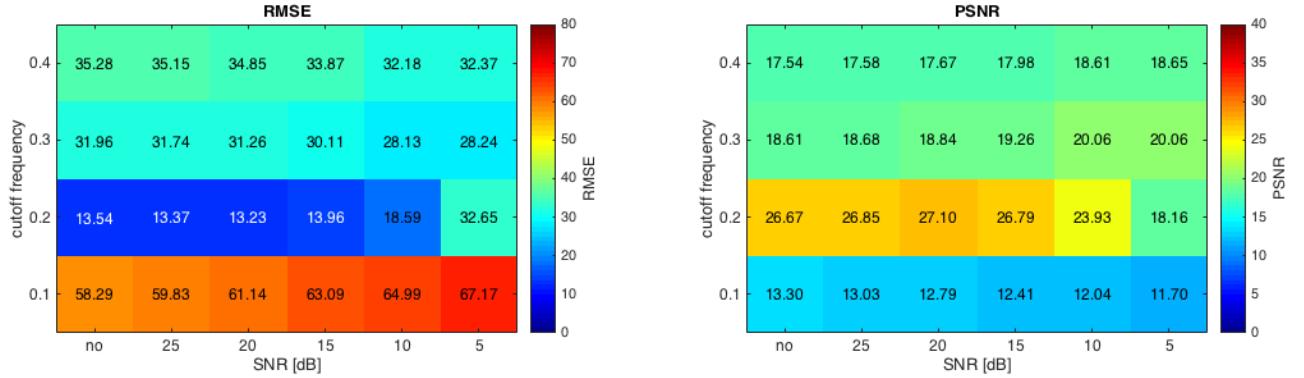


Figure 11: RMSE and PSNR metrics for CNN reconstruction ($fc = 0.2$ and $SNR = 15$)

This first observation leads to the belief that best training configuration would contain of the most degraded images. The next training set consisted then to the extreme case, with all images corrupted with a cutoff frequency of 0.1 and a SNR of 5 dB (Figure 13).

Those results are presented on Figure 14 and illustrate a more generalized but less efficient network. Despite the fact that the values seems more similar across all fc and SNR levels, the PSNR are lower and the RMSE worse than with the previous training. The reconstructions (Figure 15) are far from the reference images, with a slight improvement for the ones with a cutoff frequency of 0.1.

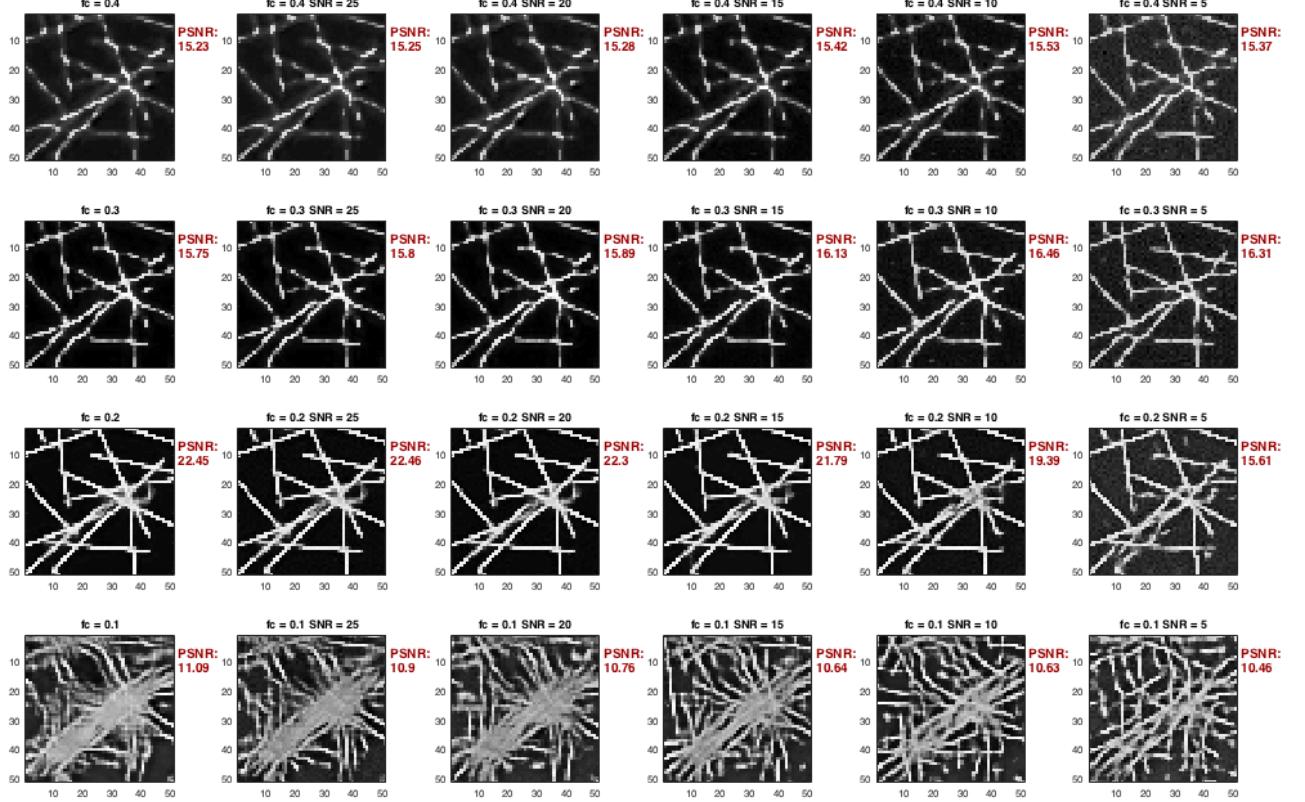


Figure 12: Zoom on CNN reconstruction ($fc = 0.2$ and $SNR = 15$)

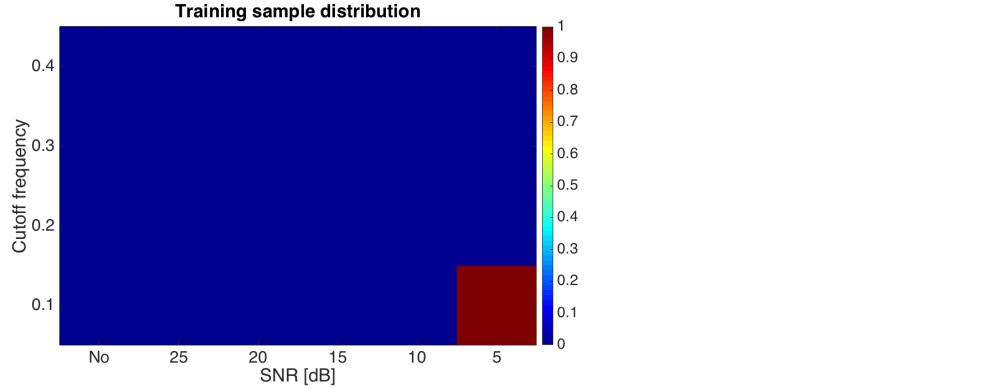


Figure 13: training sample distribution ($fc = 0.1$ and $SNR = 5$)

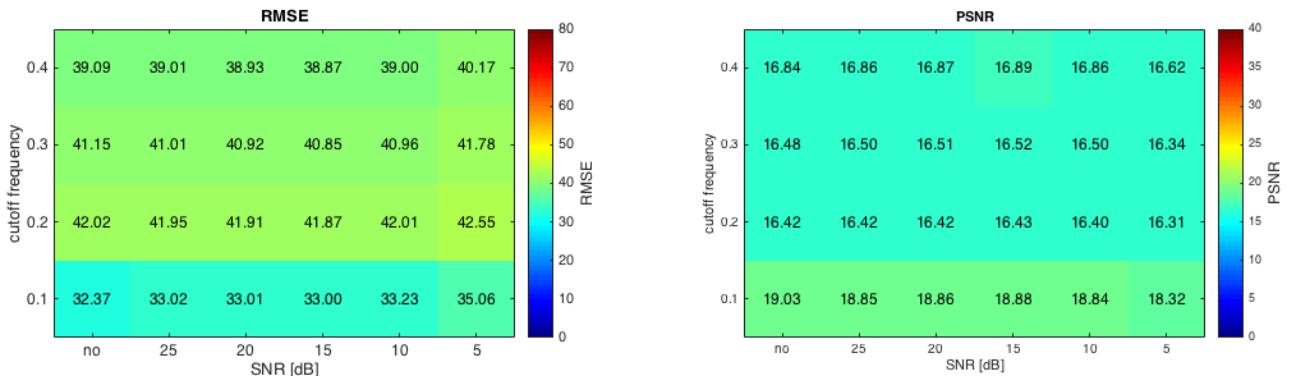


Figure 14: RMSE and PSNR metrics for CNN reconstruction ($fc = 0.1$ and $SNR = 5$)

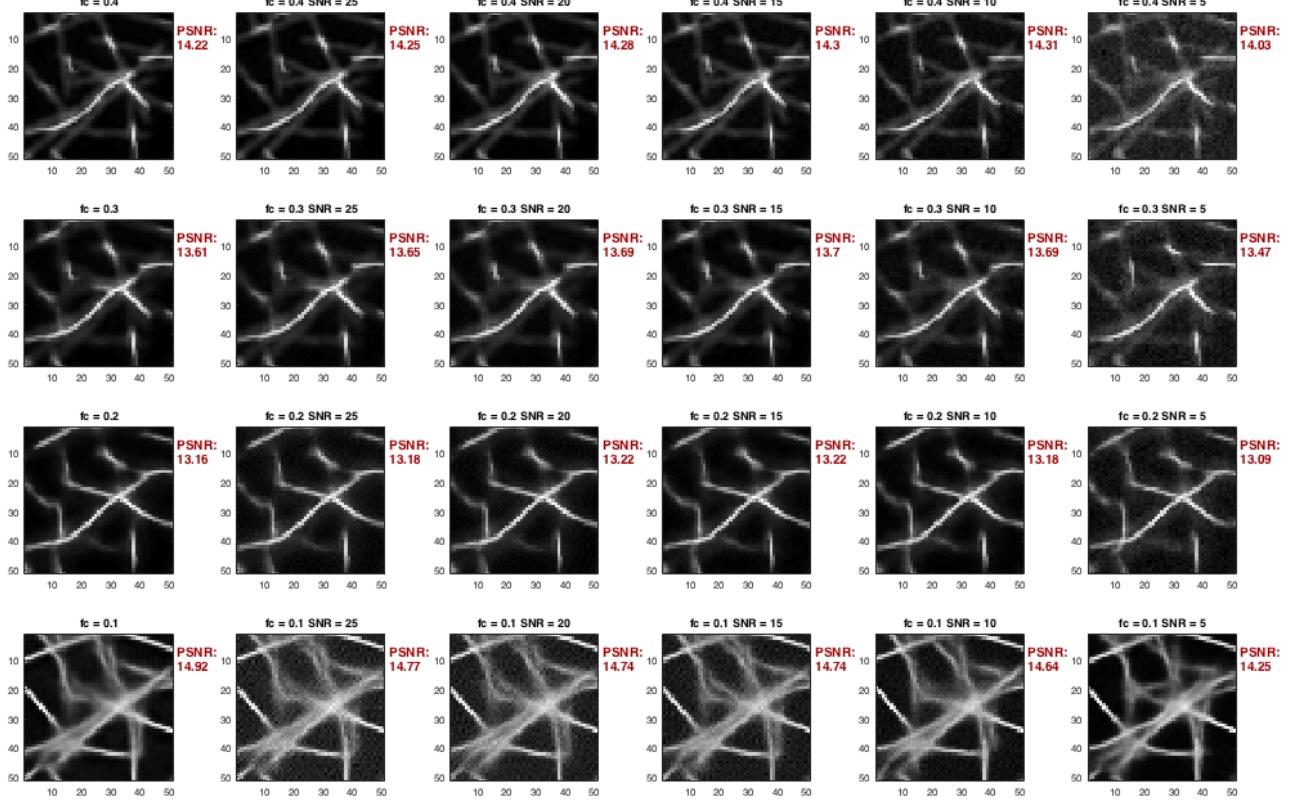


Figure 15: Zoom on CNN reconstruction ($fc = 0.1$ and $SNR = 5$)

3.3.2 WienerConvNet

K.H. Jin et al.[15] showed that by training the network with the outcomes of a deconvolution instead of directly the degraded images goof the reconstruction performance can be achieved. As a first deconvolution approach, here, a simple Wiener filter was applied to the inputs before the training (450 images degraded with cutoff frequency of 0.2 and SNR of 15 dB, Figure 16).

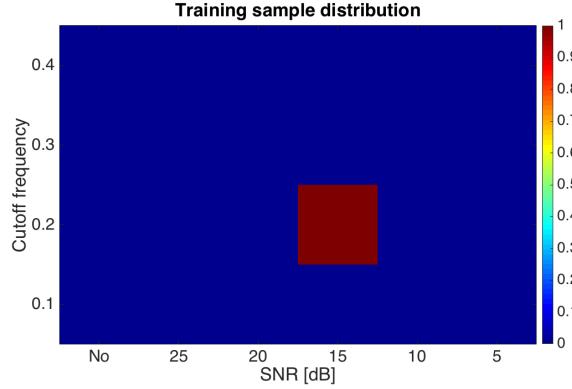


Figure 16: training sample distribution ($fc = 0.2$ and $SNR = 15$)

Not a net improvement of the model accuracy was achieved (Figure 18). RMSE and PSNR are slightly better or lowered depending on the blur and noise level (Figure 17). Those results can suggest that performing a deconvolution before the CNN might be relevant. But in order to obtain better scores it might be necessary to use another method than the Wiener filter.

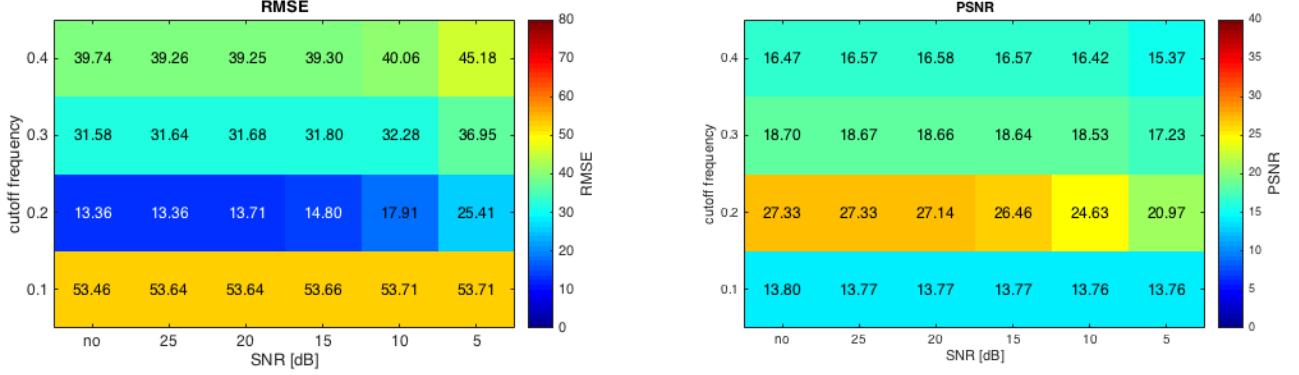


Figure 17: RMSE and PSNR metrics for WienerCovNet reconstruction ($fc = 0.2$ and $SNR = 15$)

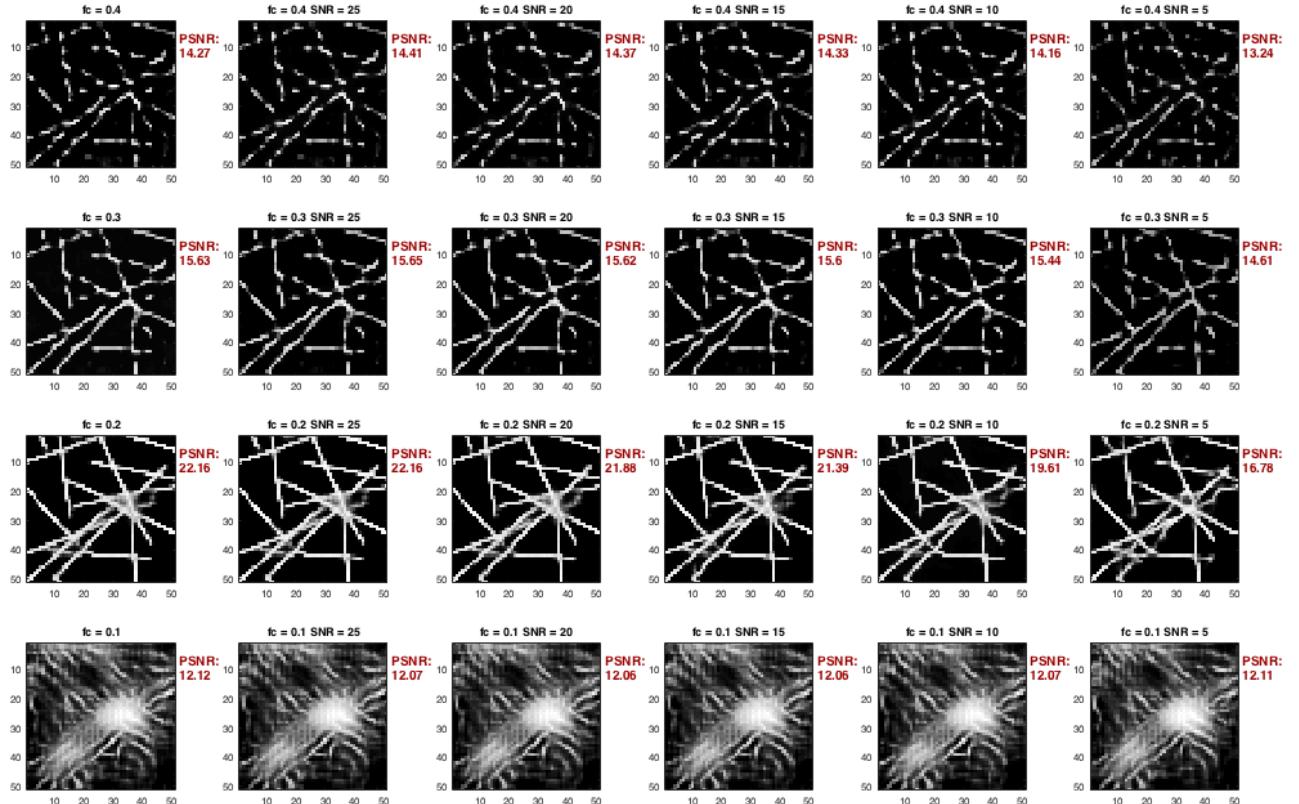


Figure 18: Zoom on WienerCovNet reconstruction ($fc = 0.2$ and $SNR = 15$)

3.3.3 Improve high blur level performance

As the model seems efficient over a broader ranger of noise level than uniquely the one on which it was trained, it does not look indispensable to mix SNR components for the training part. Nevertheless, using various cutoff frequency appears necessary to obtain a more generalizable model. So the next training set contained still 450 images, all with a SNR of 15 dB but split the most equally possible to cover the entire range of blur levels (Figure 19).

The outcomes are represented on Figure 20 and 21. This approach outperforms the precedent ones, especially for cutoff frequencies above 0.1 and SNR larger than 5 dB. Mixing the cutoff frequencies does not reduce the performance and results in more homogenous scores. However, achieving a Wiener deconvolution before the CNN with the same training sample configuration gives degraded results, which confirms the assumption that the Wiener methods is not the most pertinent one for this purpose.

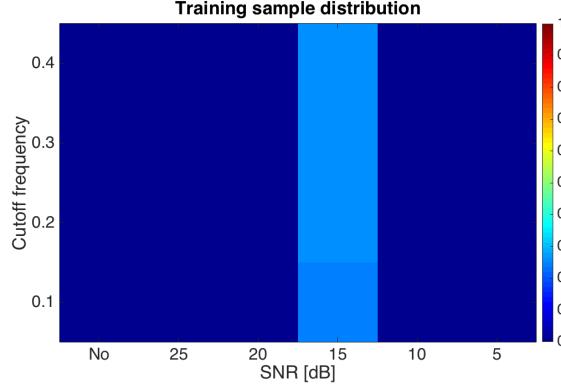


Figure 19: training sample distribution (all fc and SNR = 15)

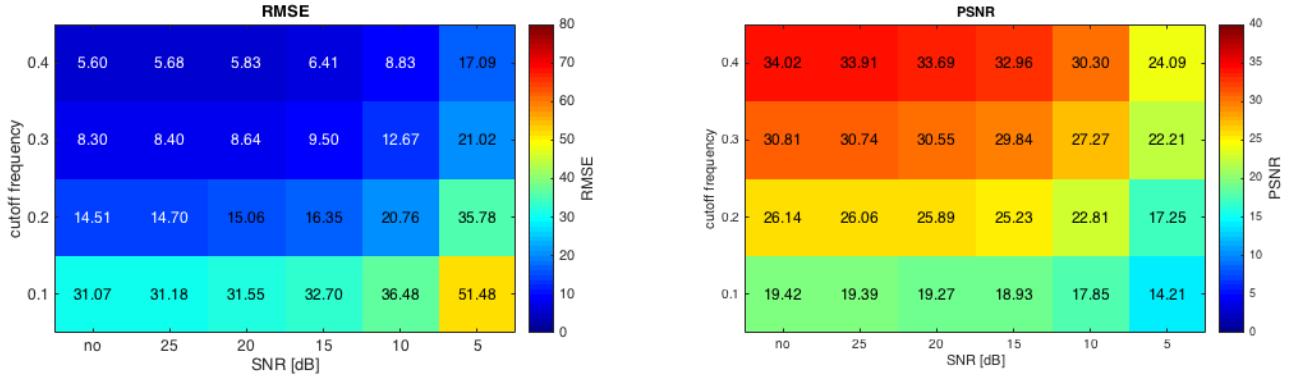


Figure 20: RMSE and PSNR metrics for CNN reconstruction (all fc and SNR = 15)

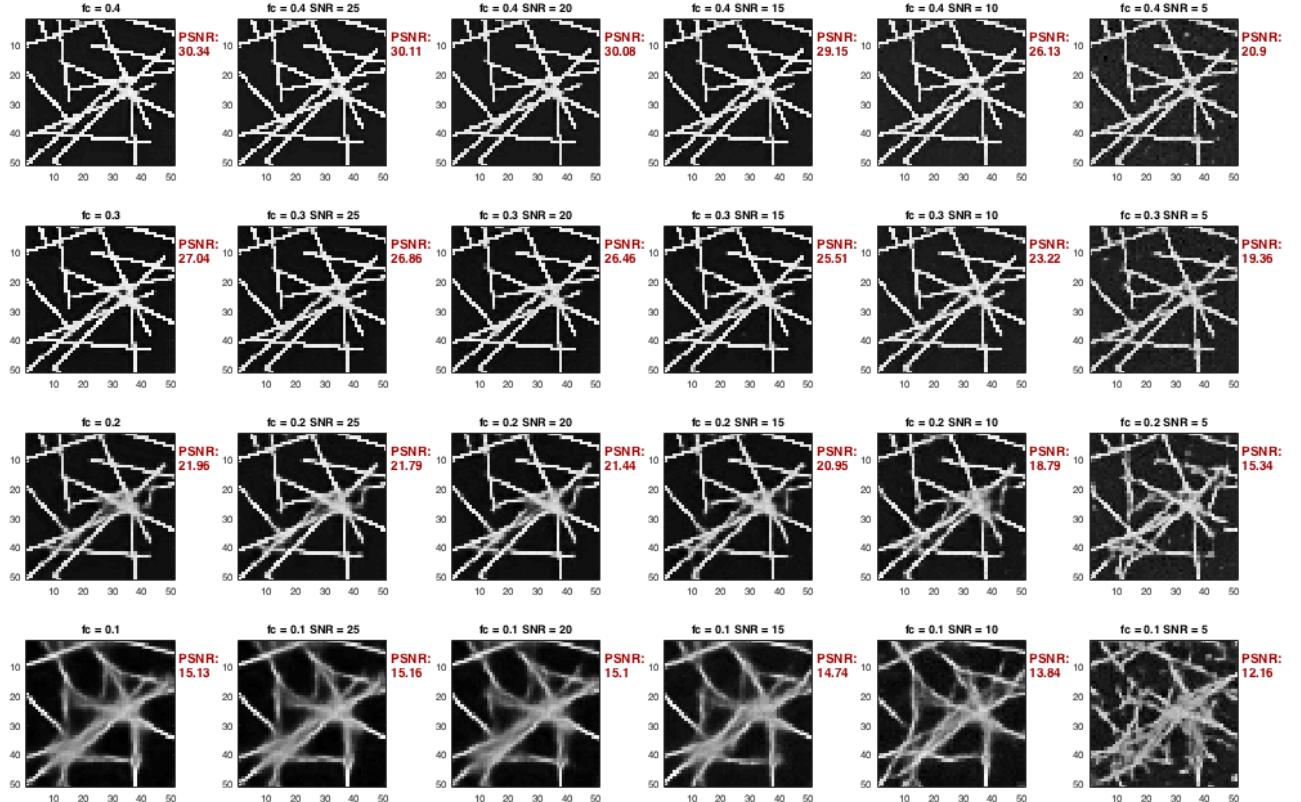


Figure 21: Zoom on CNN reconstruction (all fc and SNR = 15)

3.3.4 Bigger training set

The next question concerns the total amount of data given to train the CNN. Would increasing the number of training sample enhance the model performance? Instead of an overall of 450 images, here 450 images per cutoff frequency were used for a total of 1800 training samples (Figure 22).

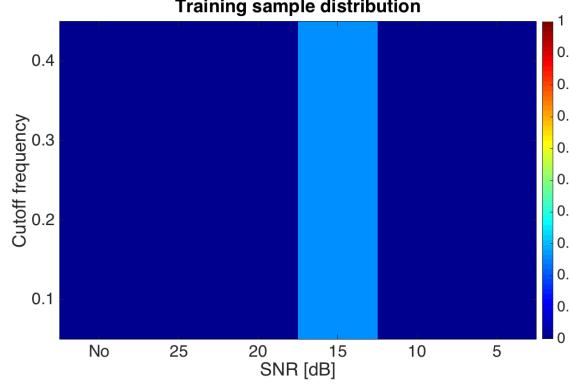


Figure 22: training sample distribution (all fc and SNR = 15)

Even if it significantly increases the time necessary for the training part of the model (14 hours), the augmentation of the training set leads to a nice improvement of the model performance. This progress is well illustrated on Figure 23 and 24, with a reduction of most RMSE values and an augmentation of the PSNR, reaching up 36.12 dB.

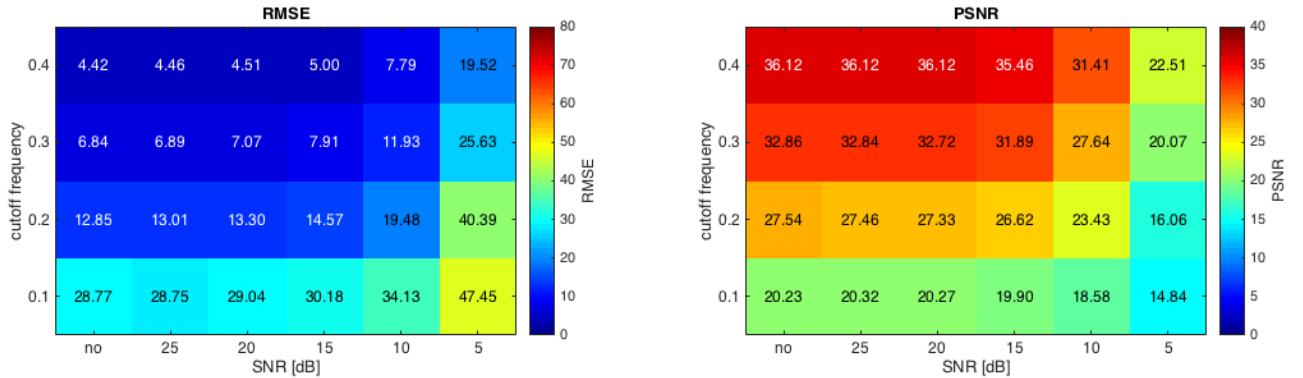


Figure 23: RMSE and PSNR metrics for CNN reconstruction (all fc, SNR = 15 and 450 training sample per fc)

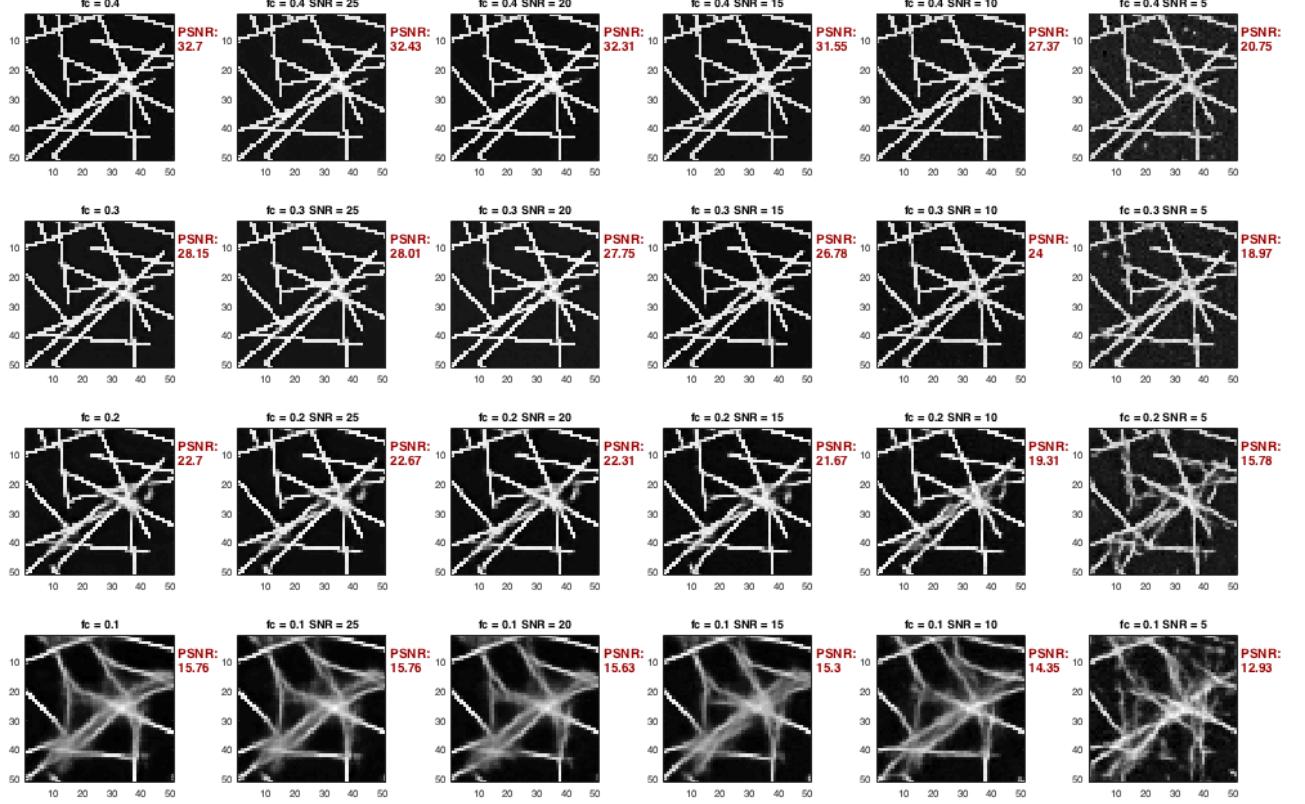


Figure 24: Zoom on CNN reconstruction (all fc, SNR = 15 and 450 training sample per fc)

3.3.5 Number of training sample proportional to cutoff frequency

Considering that the low frequency components get the worst reconstruction outcomes, rather than using an homogenous repartition of the training sample across the cutoff frequencies, it might be relevant to attribute more weight to the most blurry levels. Therefore, the distribution of the training images starts from 300 samples with an increment of 100 as the cutoff frequency goes down (Figure 25).

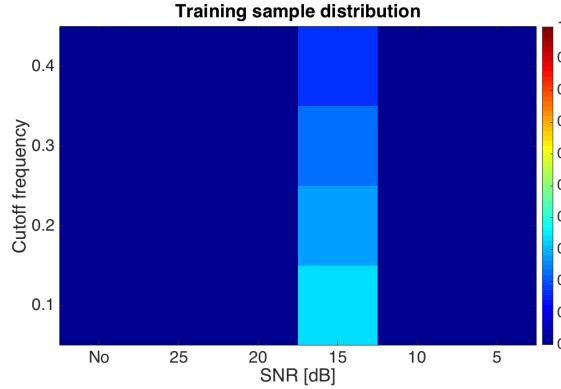


Figure 25: training sample distribution (all fc and SNR = 15)

As it can be seen on Figure 26 and 27, it allows a slight improvement of the low cutoff frequency elements without affecting much the high ones.

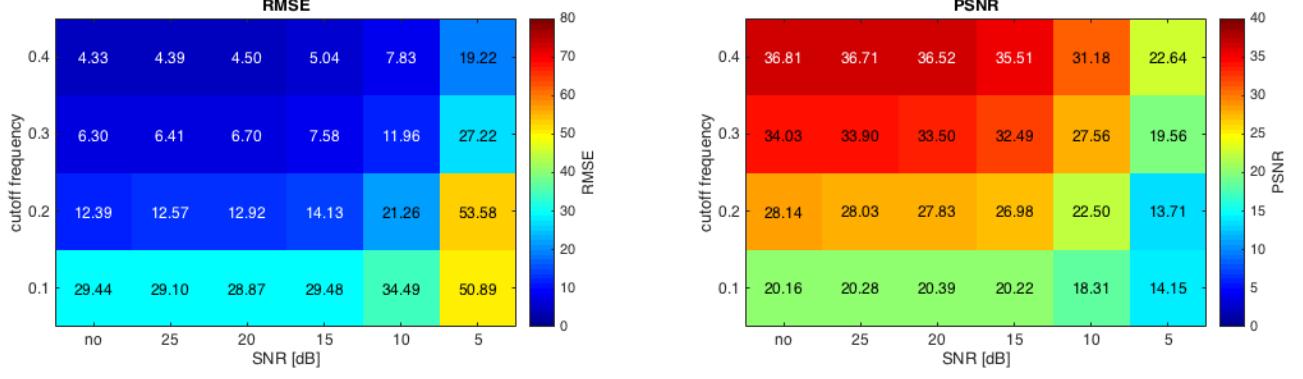


Figure 26: RMSE and PSNR metrics for CNN reconstruction (all fc, SNR = 15 and an increasing training sample number as the fc goes down)

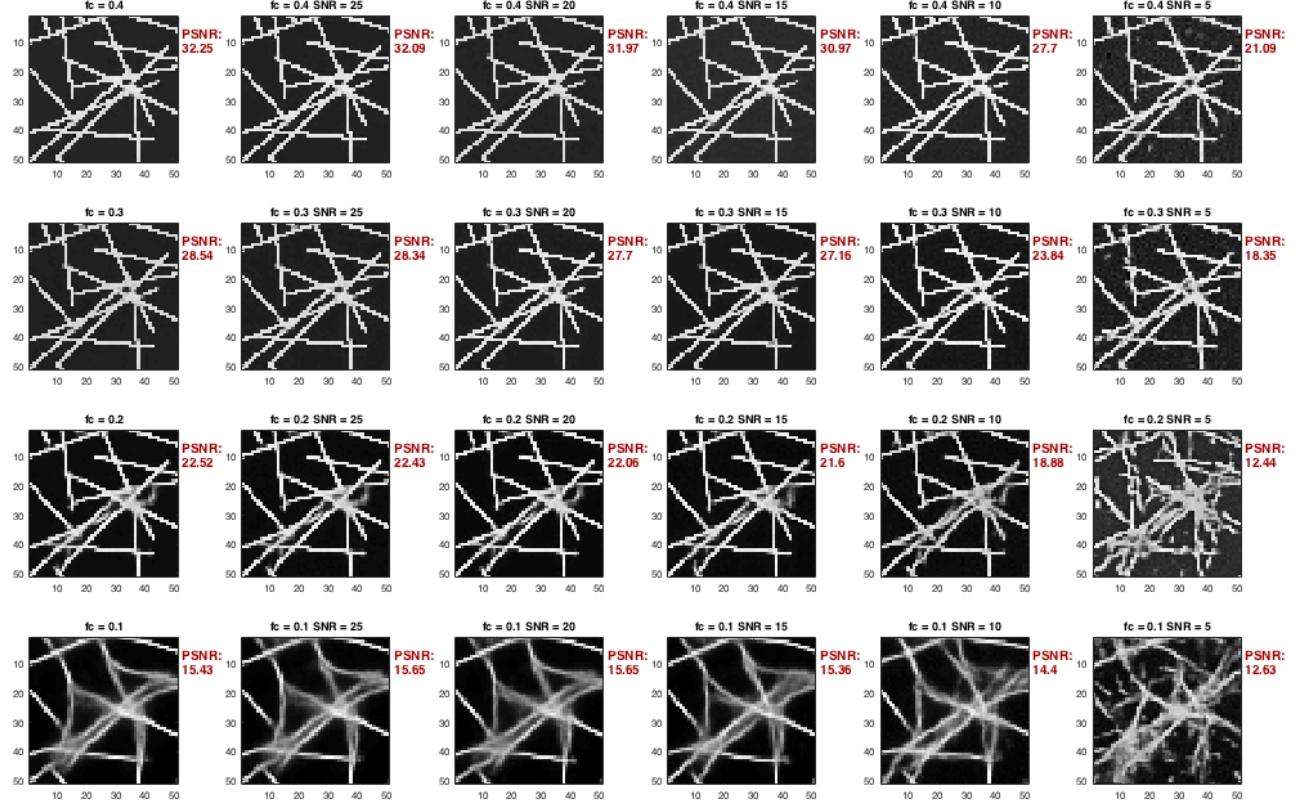


Figure 27: Zoom on CNN reconstruction (all fc, SNR = 15 and an increasing training sample number as the fc goes down)

3.3.6 Improve high noise level performance

Now, in order to improve the performance over the most noisy images, the idea might be to shift the training band, here from a SNR of 15 dB to 5 dB (Figure 28). The results are represented on Figure 29 and 30. Unfortunately, even if a net improvement in the last column of the degradation grid can be observed, the scores for the lower level of noise are worse than what previously obtained.

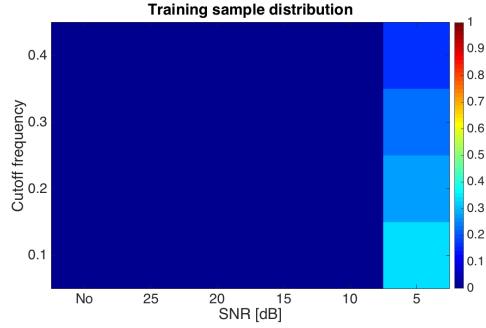


Figure 28: training sample distribution (all fc and SNR = 5)

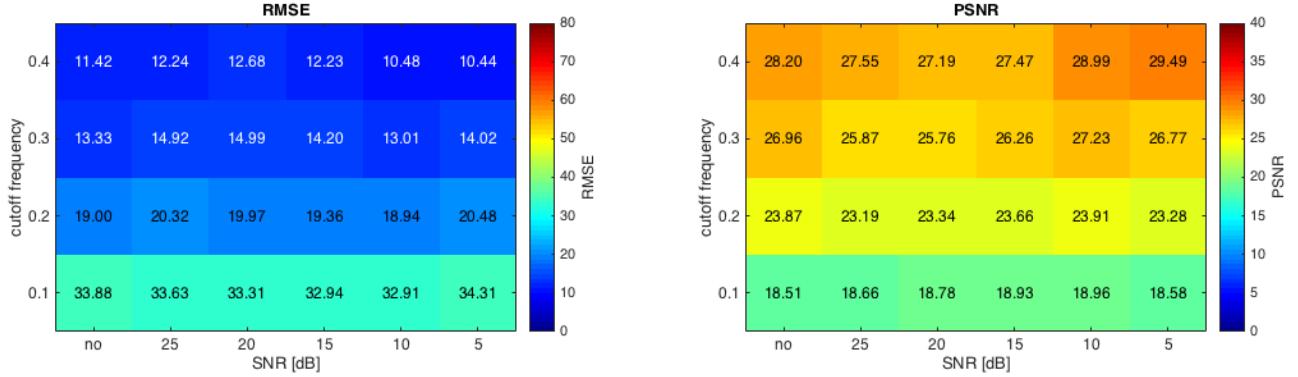


Figure 29: RMSE and PSNR metrics for CNN reconstruction (all fc, SNR = 5 and an increasing training sample number as the fc goes down)

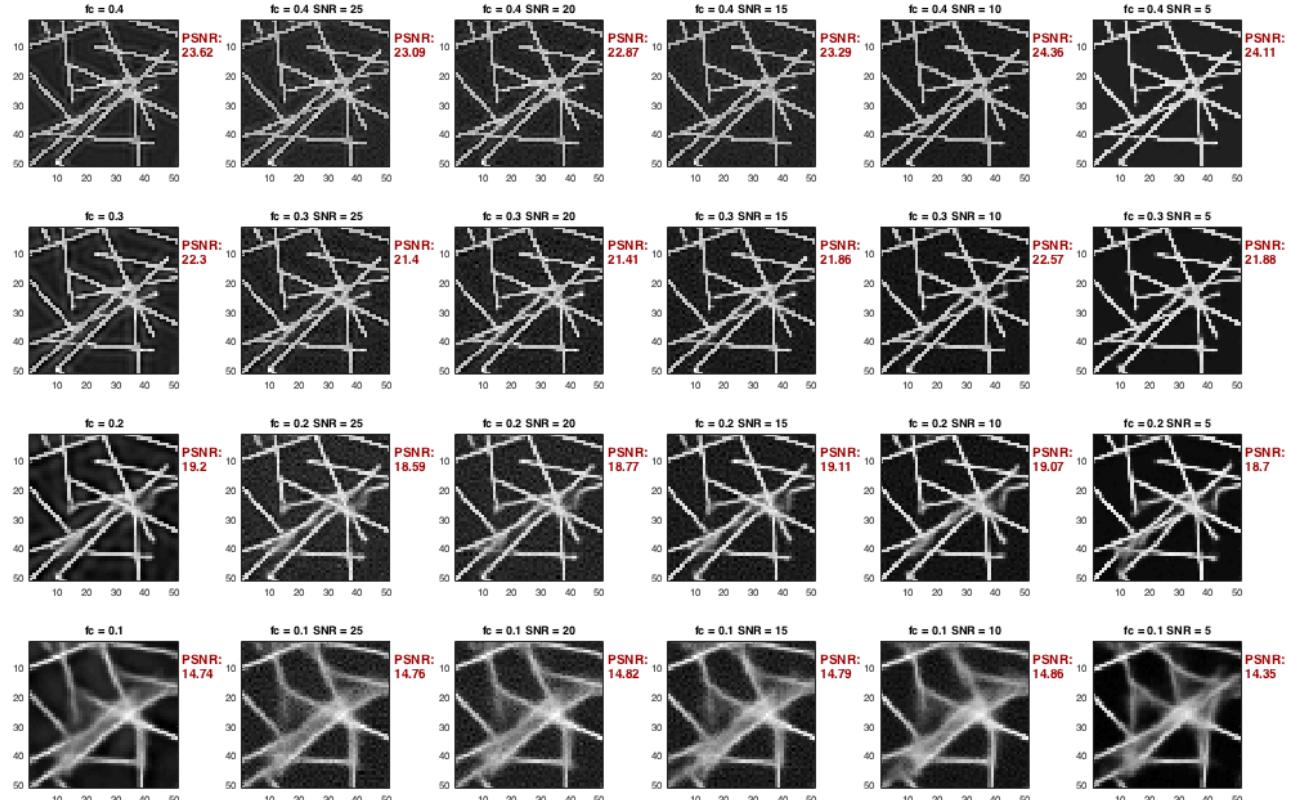


Figure 30: Zoom on CNN reconstruction (all fc, SNR = 5 and an increasing training sample number as the fc goes down)

Since shifting the training band to the biggest level of noise was not relevant, it is here moved to 10dB (Figure 31).

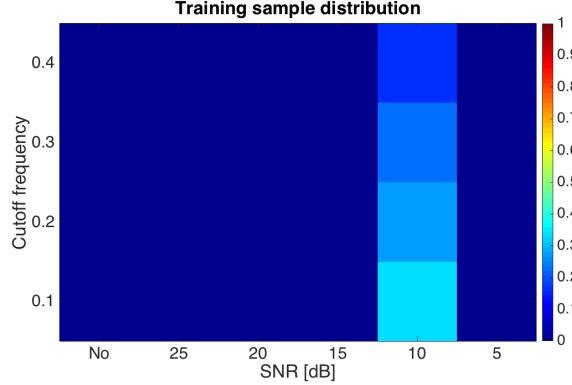


Figure 31: training sample distribution (all fc and SNR = 10)

Even if some RMSE and PSNR do not attain the previous scores (Figure 32), the lower noise level are not too degraded and the values in the two last SNR columns are improved. Overall, it is a quite performant and more generalizable model, two important aspect of such system. Looking at the reconstruction (Figure 33), the results seem quite good except for the lowest cutoff frequency components (last raw).

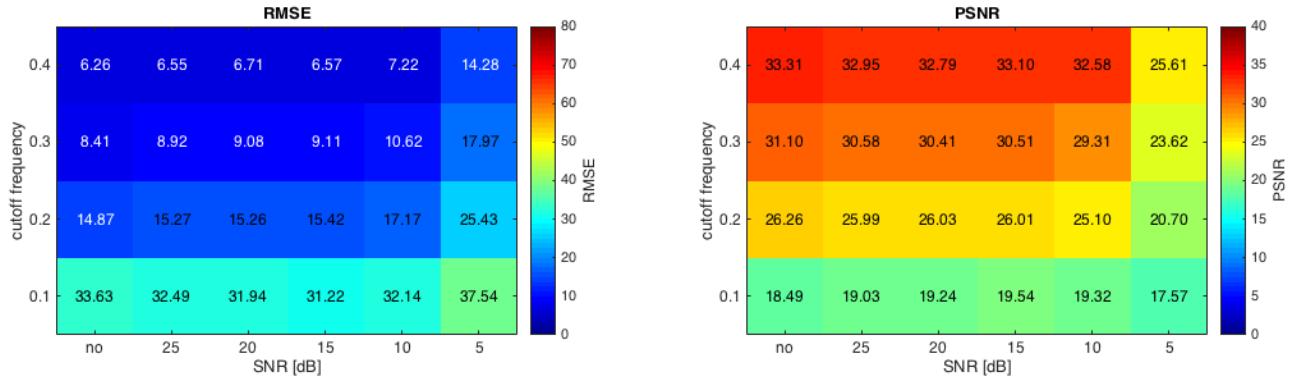


Figure 32: RMSE and PSNR metrics for CNN reconstruction (all fc, SNR = 10 and an increasing training sample number as the fc goes down)

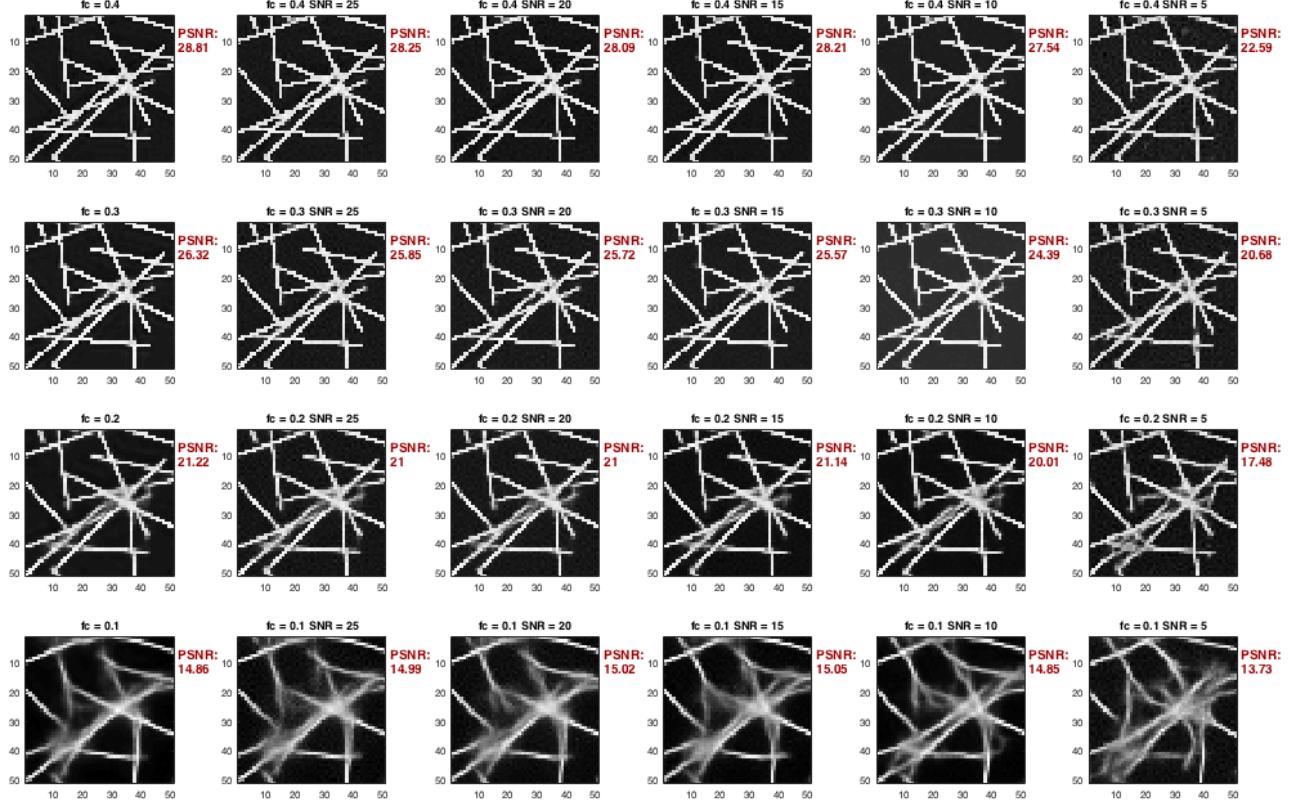


Figure 33: Zoom on CNN reconstruction (all fc, SNR = 10 and an increasing training sample number as the fc goes down)

3.3.7 LSTVConvNet

Based on K.H. Jin et al.[15] findings and the results obtained when performing a Wiener deconvolution as a first step before the CNN, for the final implementation, reconstructions obtained by a least-squares deconvolution with total variation regularization (LSTV) serves as inputs to the network.

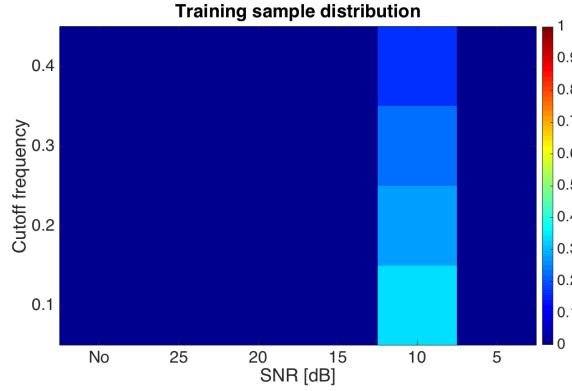


Figure 34: training sample distribution (all fc and SNR = 10)

Globally no net increase of the accuracy is observed, but once more, the scores seems smoothed over the blur and noise levels (Figure 35). The combination of the CNN with a deconvolution seems to induce a slight reconstruction improvement of images degraded with either the lowest cutoff frequency or SNR (last column and last row) (Figure 36).

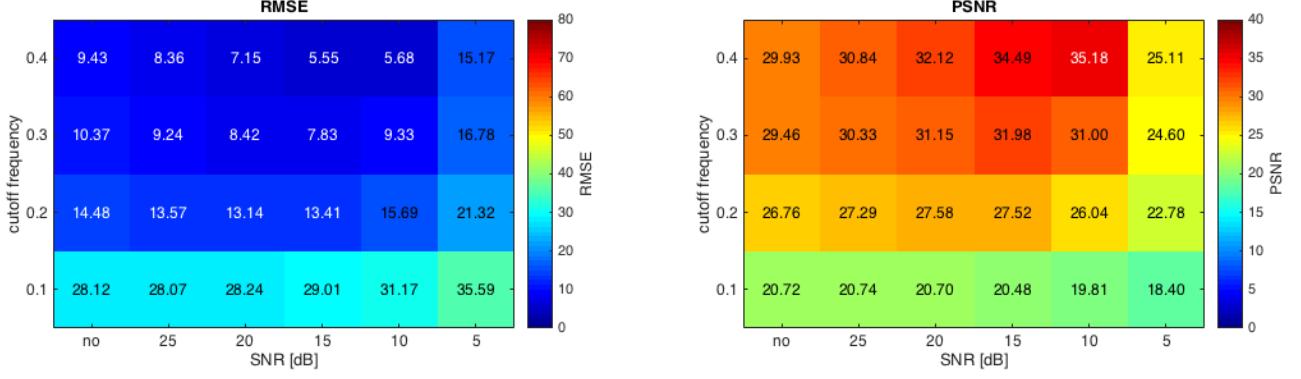


Figure 35: RMSE and PSNR metrics for LSTVCovNet reconstruction (all fc, SNR = 10 and an increasing training sample number as the fc goes down)

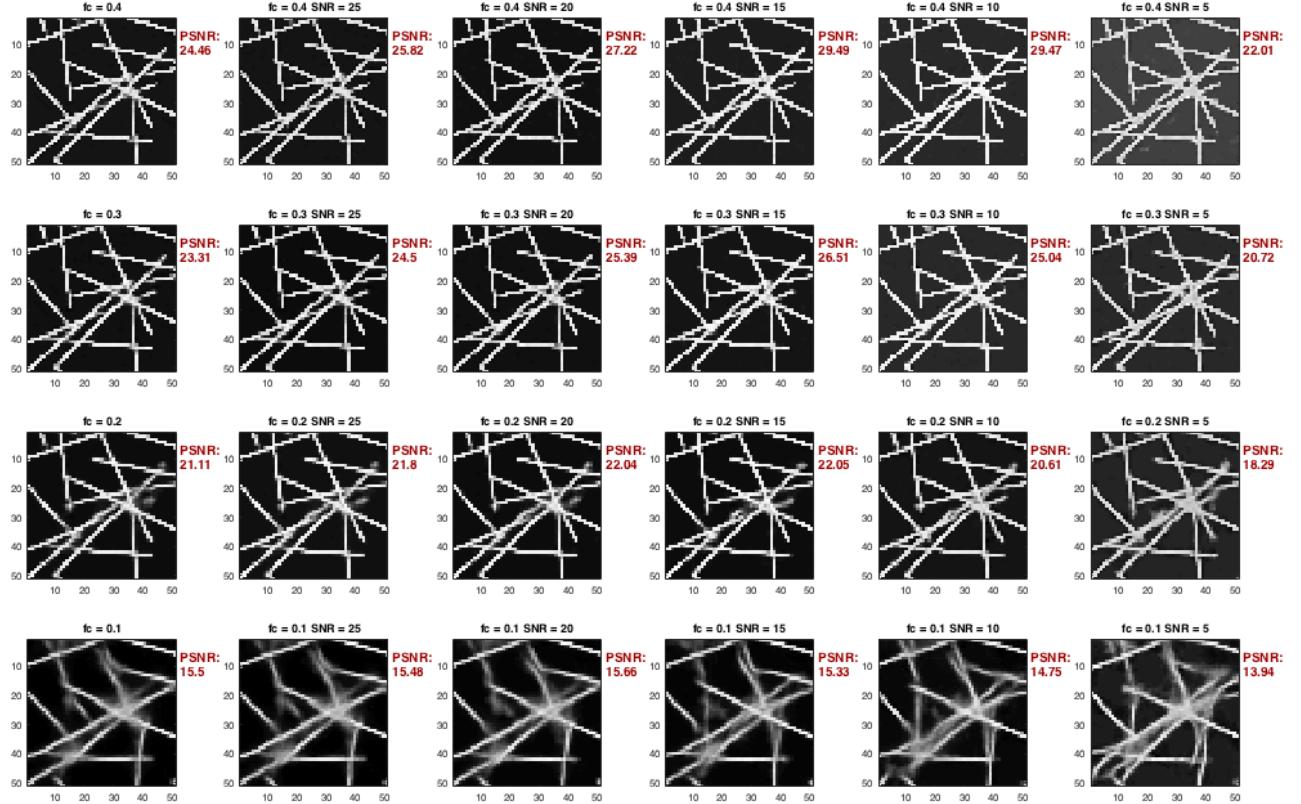


Figure 36: Zoom on LSTVCovNet reconstruction (all fc, SNR = 10 and an increasing training sample number as the fc goes down)

3.4 Model comparison

The PSNR scores of the LSTV deconvolution, the most performant CNN and the LSTVCovNet are regrouped on Figure 37 to better visualize the impact of the network. Achieving the reconstruction through the neural network leads to a net improvement comparing to the LSTV deconvolution. The maximum PSNR is a little bit higher. But most importantly, the performance is enhancement for a wide range of degradation level. Combining both techniques in a LSTVCovNet model seems to bring a small improvement to the lowest cutoff frequency and SNR level, without affecting much the other degree of deterioration.

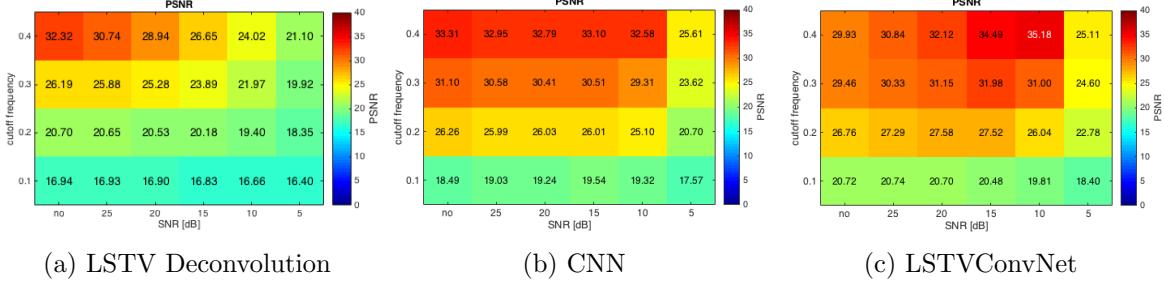


Figure 37: PSNR scores comparison of LSTV deconvolution, CNN and LSTVConvNet

However, it remains difficult to determine which of the CNN or LSTVConvNet is the most appropriate for this inverse problem task. It should be kept in mind that to perform a LSTV deconvolution some prior knowledges about the PSF are required, which is not the case when the reconstruction occurs simply through a neural network.

The constatations just made can be confirmed when looking at the reconstruction outcomes of a specific degradation level (cutoff frequency of 0.2 and SNR of 10 dB) (Figure 38).

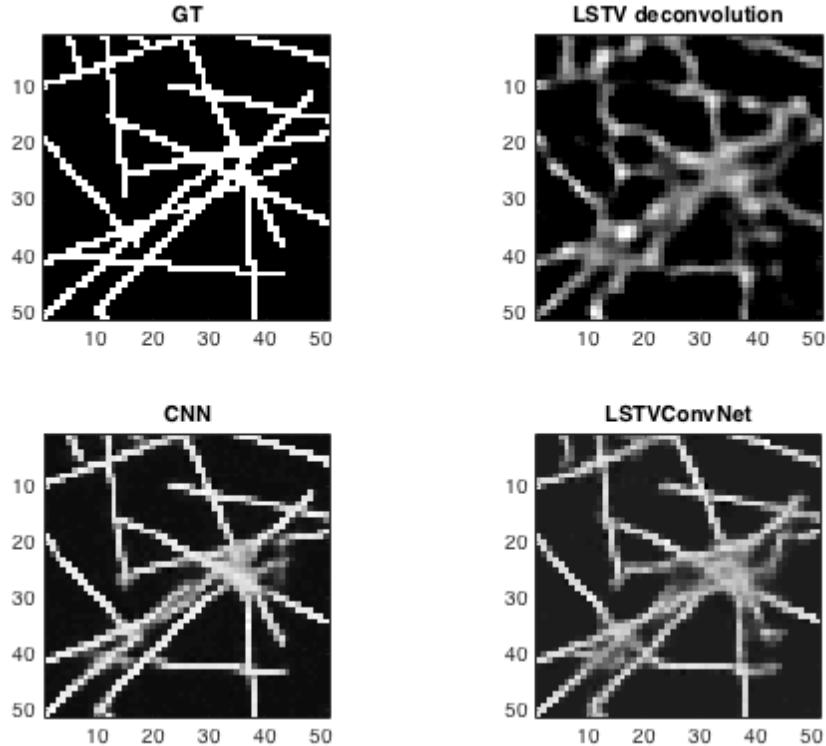


Figure 38: Reconstruction outcomes of diverse methds from input with $fc = 0.2$ and $SNR = 10$ dB

It is also interesting to look back in frequency domain. Figure 39.b illustrates well that the components about a frequency of 0.2 are lost. Using a LSTV deconvolution (39.c), high frequency compounds are only partially recovered. But if the reconstruction is performed through a CNN (39.d and .e), more components are regained.

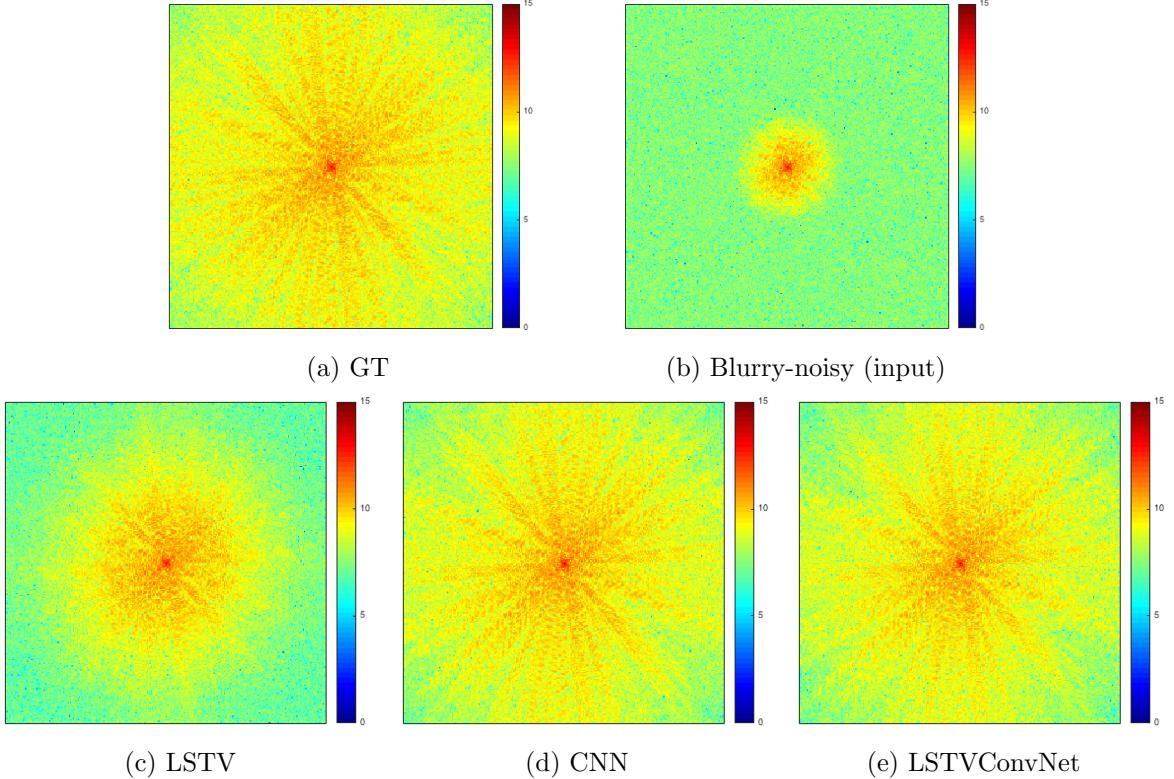


Figure 39: Reconstruction in frequency domain of diverse methods from input with $fc = 0.2$ and $\text{SNR} = 10 \text{ dB}$

3.5 Testing model on other data

After the investigations regarding the understanding of the proposed network and the training set configuration, It is interesting to verify how the most effective model responds to completely new types of data.

3.5.1 Different segment configuration

To begin with, the last CNN (training band at 10dB) is tested over 50 images of all the different segments configurations generated (Figure 2). The PSNR scores are represented on Figure 40. The model seems well efficient if the segments are present in low density in the images, independently of the distribution. It is even possible to obtain higher scores than the one accessed with the segments configuration used to train the model (Gaussian ring, low density and large length). The worst cases are when the segments of large length were generated in high density, which might be due to the presence of more high frequency components. But globally, the model performance is positive over the different segments configurations.

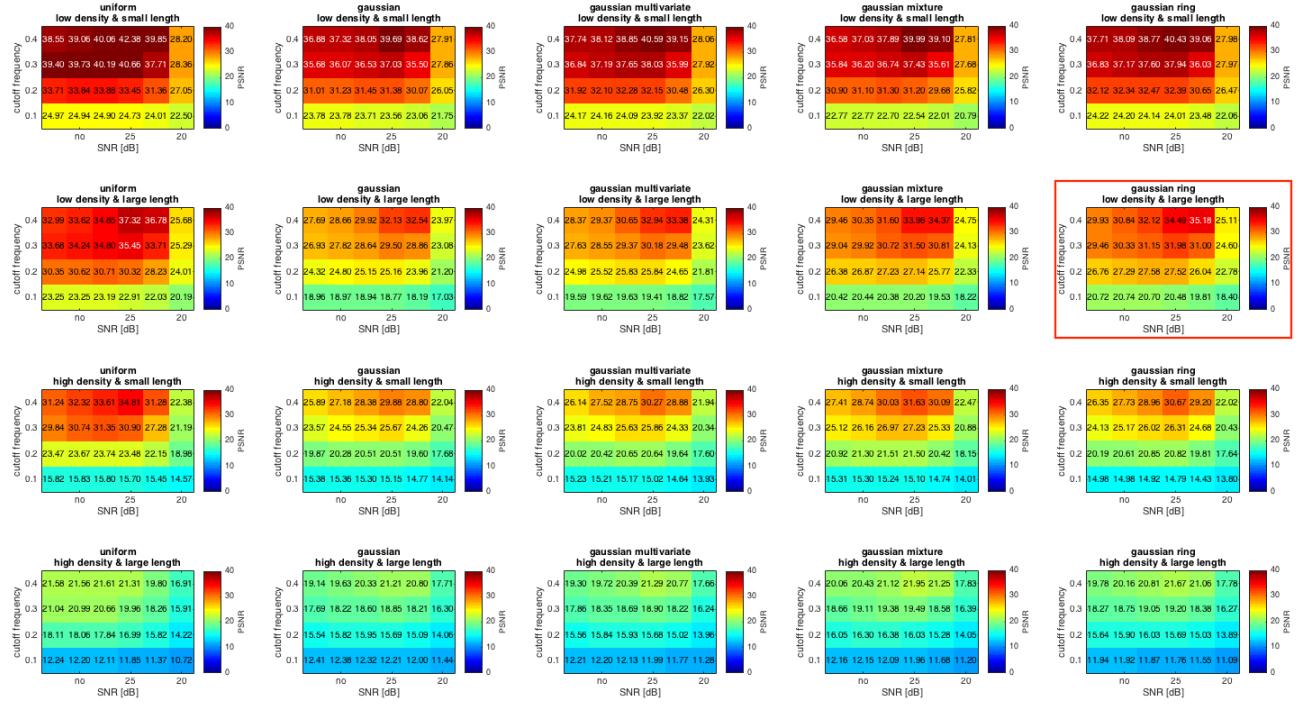


Figure 40: PSNR scores concerning the reconstruction of the 20 different segments configurations generated

3.5.2 Different segment intensity

Then the question regarding the segments intensity comes. Up to now, the data generated were all binary, either the pixel is a segment and take a value of 1, or it represents the background and is set to 0. But this concept is far from what is commonly observable. Would the model trained on binary data be capable to reconstruct images containing segments of varying intensities?

50 images of segments having different intensities were generated. They were then deteriorated with all the possible levels of blur and noise. The reconstruction was performed with the final CNN and the RMSE and PSNR scores are presented on Figure 41. The results are poor compared to the case where all segments had the same and highest intensity.

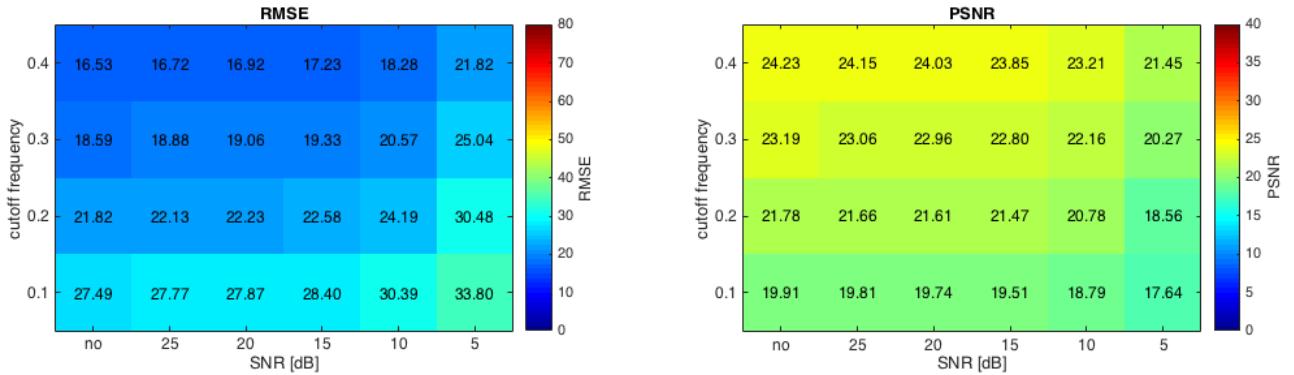


Figure 41: PSNR scores concerning reconstruction of segments with different intensities after degradation with all the level of blur and noise generated

The noise and the blur have mostly disappear. But the reduced performance probably comes from the fact that the network had learnt to output something binary, as it is only on this type

of data that it was trained. On the reconstructions on Figure 42, all the segments appear white and not of diverse intensities as expected. Furthermore, the background does not look as black as it should be.

This outcome suggests that the data generated to train the CNN were apparently too specific. The model will perform well on new data having similar characteristics than the training set.

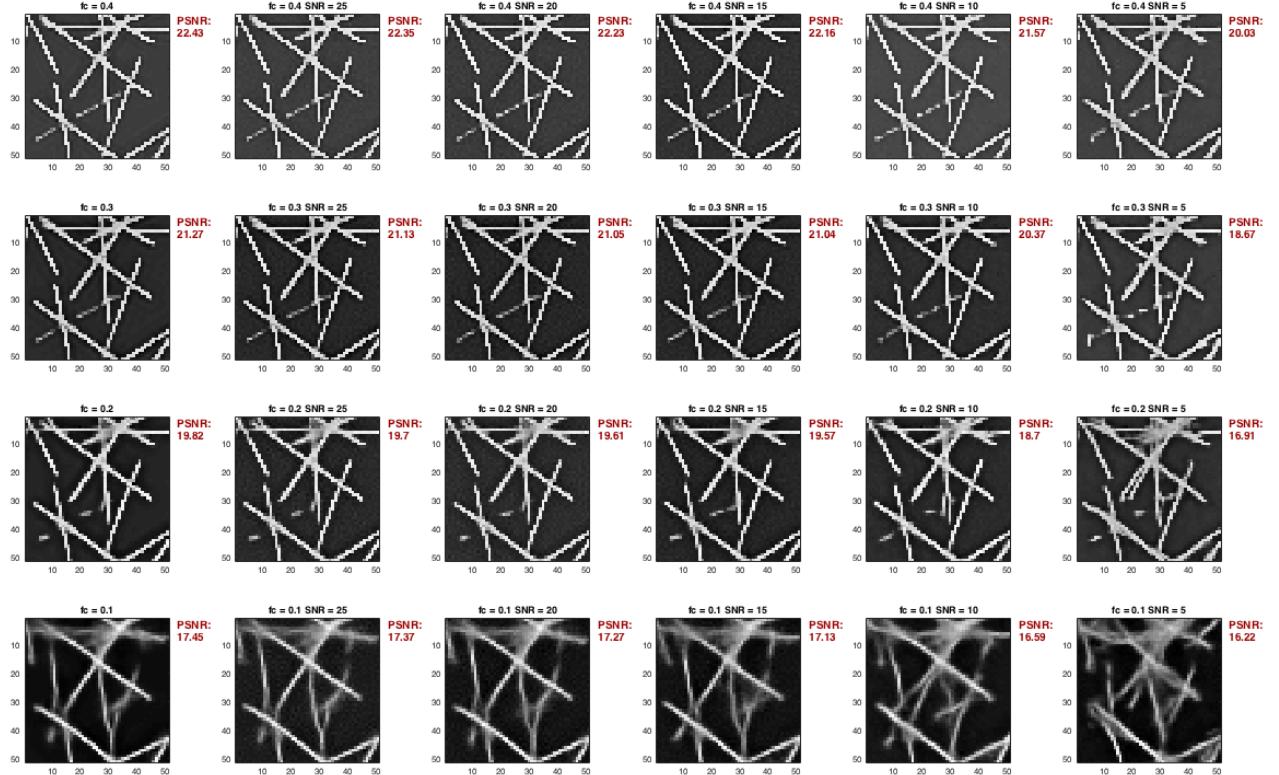


Figure 42: reconstruction of segments with different intensities after degradation with all the level of blur and noise generated

3.5.3 Lena

As this project is touching the question of image restoration and by curiosity, it would be interesting to see how the model would perform on standard test images commonly used in the field of image processing, such as Lena (image 43). An image of Lena in gray scales and with the same dimensions as the training images (256 x 256) was first degraded with once more with all the level of blur and noise generated.



Figure 43: Lena

Based on the previous results and as mostly expected, the reconstructions are pretty bad (Figure 44), only the shape of the image components can be discerned. It seems like the network had learnt to rebuild the edge of the objects. Here again, the model tends to output binary values. In the reconstructions, either the pixel is light or dark depending on its intensity in the original image.

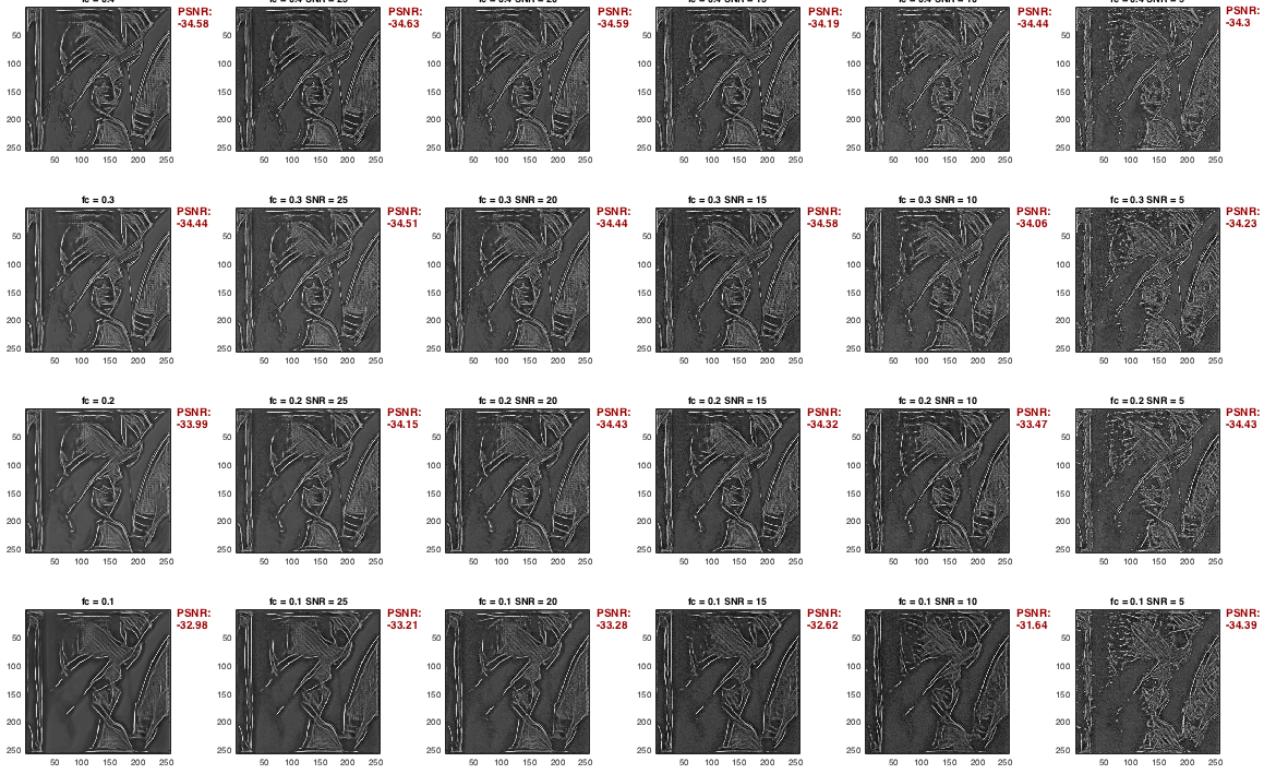


Figure 44: Lena reconstruction after degradation with all the levels of blur and noise generated

3.5.4 SIM data

The whole project was performed with the goal to challenge deep learning and test the feasibility regarding inverse problems solving. The final hope is to reduce the number of acquisition frames for the reconstruction of a super-resolution image by processing through a similar network. Knowing that the CNN was not trained on appropriate data, how would it react and perform if a microscopy wide-field is given to it as input? Figure 45 illustrates a wide-field (.a) and a super-resolution SIM image (.b) that are used here.

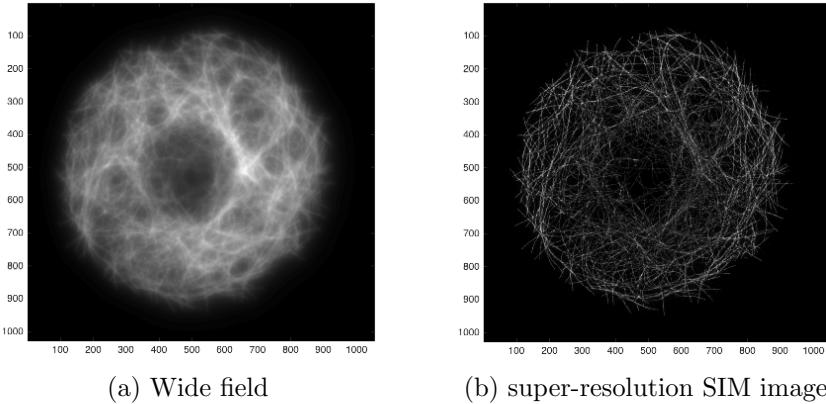


Figure 45: Wide field and super-resolution SIM [23]

The image was cropped in order to have a size equivalent to the training samples. First a wide-field was given as input (Figure 46.b) to the network and the resulting reconstruction is shown on Figure 46.c. Even if this image illustrates structures that looks a bit similar to the simulated images (segments crossing each others), this restoration is far from the super-resolution SIM image (Figure 46.a). This poor outcome might probably have multiple causes. To begin with, the segments here are not appears as straight as in the training set. Differences in their intensity can be observed, which has just been shown as an aspect not well managed by the model. Furthermore, no assumption is made regarding the PSF of the wide-field. The level of blur and noise might be completely different from what the network was trained to.

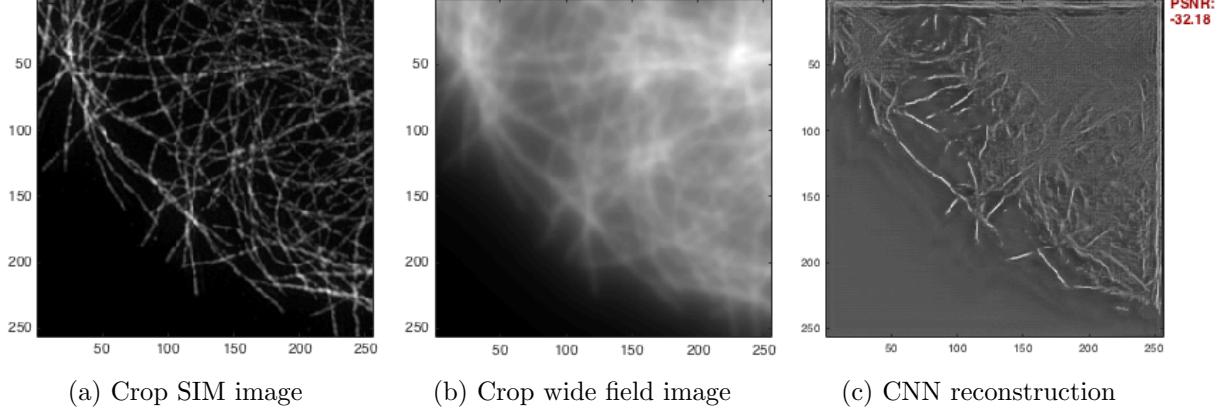


Figure 46: Network performance on real SIM data with wide-field as input

To overcome the problem of the unknown and new PSF, the super-resolution images was manually degraded with the different blur and noise levels (Figure 47). But previously, the background of the SIM super-resolution image was found not to be completely black, which one more time is not a configuration the model was trained to. For this reason, all the pixels having a value below a threshold of 0.08 were set to 0. This action did not seem to affect much the image quality, while bringing it closer to the training set. The degraded images were given as inputs to the CNN with the hope to obtained a better reconstruction.

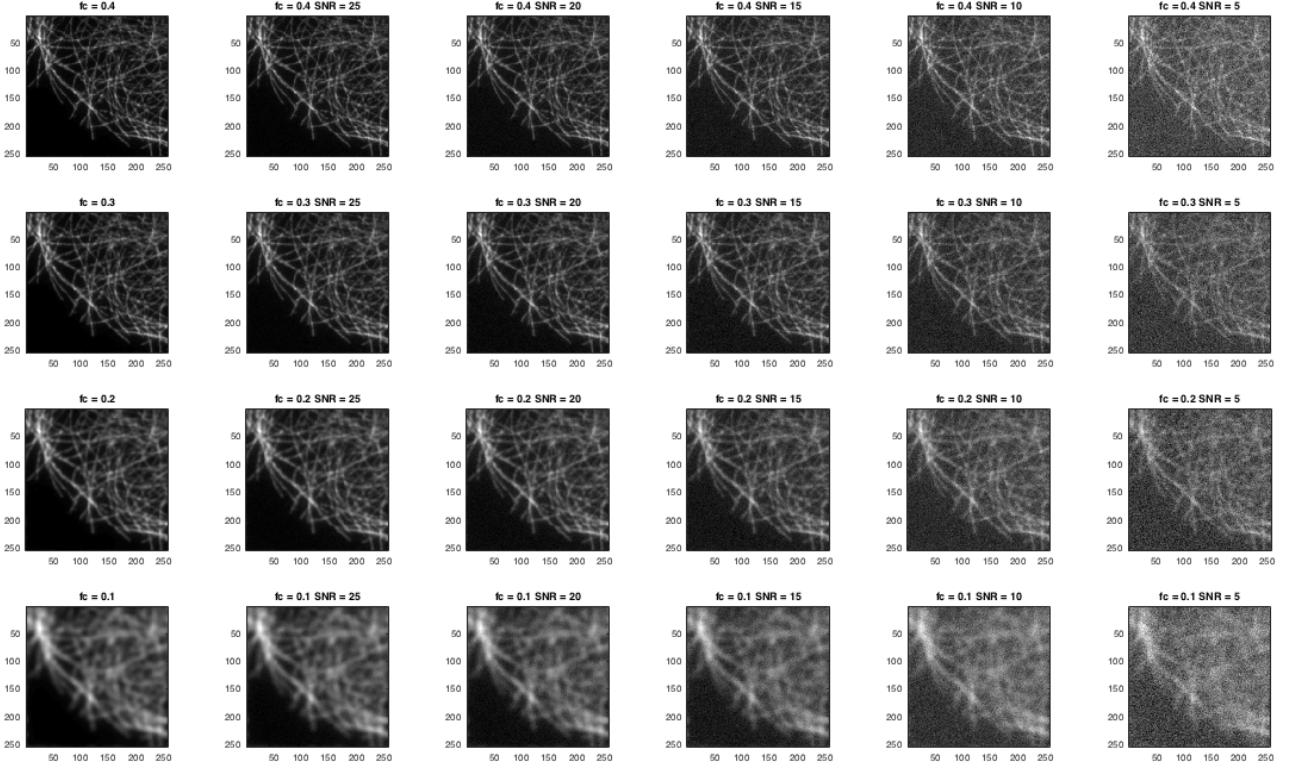


Figure 47: Crop of SIM image degraded with all the levels of blur and noise generated

The outcomes are illustrated on Figure 48. The problem of binarity has not disappear. The segments reconstructed have all the same intensity and the back ground is not black.

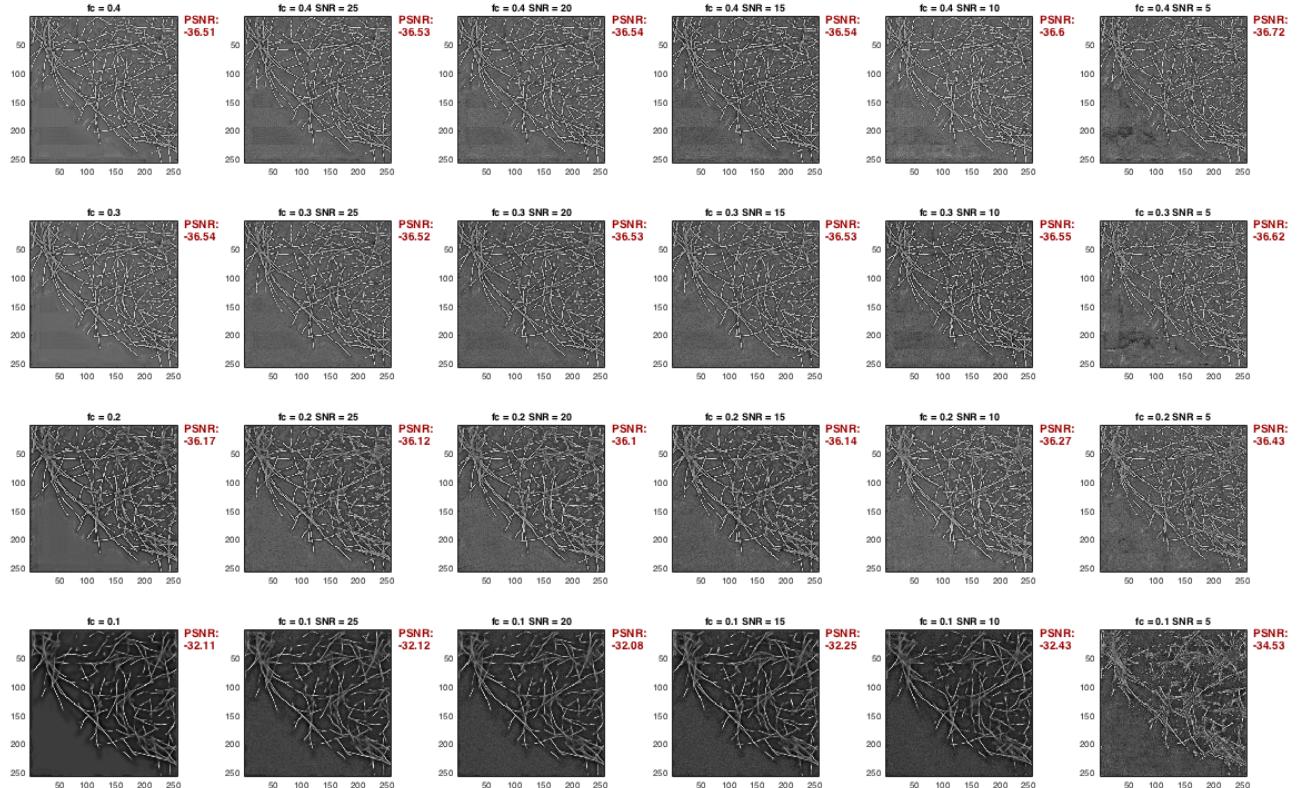


Figure 48: CNN reconstruction of degraded SIM after background set to 0

Beside the problems listed above, the network was, even so, capable to reconstruct the main components without too much blur or noise anymore. It is a positive and promising point for further development. As it was shown in this project, the performance of the model is really depending on the type of training data. The network will reconstruct well structures and conditions close to its training. However, knowing the type of structures mainly observed by the microscope and having an idea regarding the PSF of the device, a similar network could be properly trained. And this project leads to belief that a super-resolution image similar to SIM real data could be reconstructed from the wide-field or a reduced number of acquisitions, increasing then the temporal resolution.

4 Conclusion

Even though the dataset generated to train the Convolutional Neural Network was well specific with major differences comparing to real data, the results of this project illustrate the potential of deep learning regarding the question of deconvolution. The proposed CNN seems robust for different SNR levels. In order to further enhance the performance of the network over different blurring degrees without affecting the achievements, mixing the cutoff frequencies in the training set is a good suggestion. However, the power of the network seems strongly linked to the kind of images used for the training part. The model trained on basic simulated data was, for instance, not capable to reconstruct accurately real SIM images.

Such network might still have a great potential for the purpose of reconstruction SIM super-resolution images from wide-field or fewer acquisitions, assuming that the model is trained on adequate data. The temporal resolution could thus be enhanced, making the SIM a relevant technique for live-cell imaging and helping researchers to visualize and understand biological processes at microscopic level.

References

- [1] M. G. L. Gustafsson, *Surpassing the lateral resolution limit by a factor of two using structured illumination microscopy*, Journal of Microscopy, March 2000
- [2] T. Lukes, G. M. Hagen, P. Krizek, M. Klima, *Comparison of image reconstruction methods for structured illumination microscopy*, The International Society for Optical Engineering, April 2014
- [3] Y. Hirano, A. Matsuda, Y. Hiraoka, *Recent advancements in structured-illumination microscopy toward live-cell imaging*, Microscopy, June 2015
- [4] S. Labouesse, M. Allain, J. Idier, S. Bourguignon, A. Negash, P. Liu, A. Sentenac, *Joint reconstruction strategy for structured illumination microscopy with unknown illuminations*, IEEE Transactions on Image Processing, July 2016
- [5] F. Orieux, E. Sepulveda, V. Loriette, B. Dubertret, J-C. Olivo-Marin, *Bayesian Estimation for Optimized Structured Illumination Microscopy*, IEEE Transactions on Image Processing, July 2011
- [6] R. Heintzmann, *Structured Illumination Methods*, Handbook of Biological Confocal Microscopy, 2006
- [7] *Introduction to Superresolution Microscopy*, ZEISS Campus: Superresolution Microscopy
- [8] A. Diaspro, G. Chirico, C. Usai, P. Ramoino, J. Dobrucki, *Photobleaching*, Handbook of Biological Confocal Microscopy, 2006
- [9] R. Rottenfusser, E.E. Wilson, M. W. Davidson, *The Point Spread Function*, Zeiss webpage, Resource basic microscopy
- [10] S. F. Hamood, M. S. M. Rahim, O. Farook, D. Kasmuni, *A Survey on Various Image Deblurring Methods*, Journal of Engineering and Applied Sciences, 2016
- [11] M. Maule, A. L. Beling da Rosa, *Wiener Filter Applied to Deconvolution*, Fundamentos Matematicos para Processamento Grafico
- [12] M.B. Cannell, A. McMorland, C. Soeller, *Image Enhancement by Deconvolution*, Handbook Of Biological Confocal Microscopy, 2006
- [13] M. S. C. Almeida, M. Figueiredo, *Deconvolving Images with Unknown Boundaries Using the Alternating Direction Method of Multipliers*, IEEE Transactions on Image Processing, March 2013
- [14] F. Orieux, E. Sepulveda, V. Loriette, B. Dubertret, J.C. Olivo-Marin, *Bayesian Estimation for Optimized Structured Illumination Microscopy*, IEEE Transactions on Image Processing, Institute of Electrical and Electronics Engineers, 2012
- [15] K.H. Jin, M.T. McCann, E. Froustey, M. Unser, *Deep Convolutional Neural Network for Inverse Problems in Imaging*, ARXIV, November 2016
- [16] M. Nielsen, *Neural Networks and Deep Learning*, January 2016
- [17] J. Xie, L. Xu, E. Chen, *Image Denoising and Inpainting with Deep Neural Networks*, Advances in Neural Information Processing Systems, 2012
- [18] L. Xu, J. S. Ren, C. Liu, J. Jia, *Deep Convolutional Neural Network for Image Deconvolution*, Advances in Neural Information Processing Systems, 2014

- [19] O. Ronneberger, P. Fischer, T. Brox, *U-net: Convolutional networks for biomedical image segmentation*, International Conference on Medical Image Computing and Computer-Assisted Intervention, 2015
- [20] F. Li, A. Karpathy, J. Johnson, *Notes related to the Convolutional Neural Networks for Visual Recognition course*, 2016
- [21] K. He, X. Zhang, S. Ren, J. Sun, *Deep Residual Learning for Image Recognition*, arXiv preprint arXiv:1512.03385, December 2015
- [22] P. Ndajah, H. Kikuchi, M. Yukawa, S. Muramatsu, *An Investigation on The Quality of Denoised Images*, International journal of circuits, systems and signal processing, July 2011
- [23] V. Cazares, L. Chao, J. A. Galbraith ,C. G. Galbraith, *Microtubule cytoskeleton organization*, Cell imaging library, 2011