UNIVERSIDADE FEDERAL DE OURO PRETO
INSTITUTO DE CIÊNCIAS EXATAS E BIOLÓGICAS
DEPARTAMENTO DE COMPUTAÇÃO

HUGO EDUARDO ZIVIANI
Orientador: Prof. Dr. Guillermo Cámara Chavez
Coorientador: Prof. Msc. Mateus Coelho Silva

BONE SUPPRESSION USING DEEP LEARNING

Ouro Preto, MG
2021

UNIVERSIDADE FEDERAL DE OURO PRETO
INSTITUTO DE CIÊNCIAS EXATAS E BIOLÓGICAS
DEPARTAMENTO DE COMPUTAÇÃO

HUGO EDUARDO ZIVIANI

BONE SUPPRESSION USING DEEP LEARNING

Monografia apresentada ao Curso de Ciência da Computação da Universidade Federal de Ouro Preto como parte dos requisitos necessários para a obtenção do grau de Bacharel em Ciência da Computação.

Orientador: Prof. Dr. Guillermo Cámara Chavez
Coorientador: Prof. Msc. Mateus Coelho Silva

Ouro Preto, MG
2021

MINISTÉRIO DA EDUCAÇÃO
UNIVERSIDADE FEDERAL DE OURO PRETO
REITORIA
INSTITUTO DE CIENCIAS EXATAS E BIOLOGICAS
DEPARTAMENTO DE COMPUTACAO

FOLHA DE APROVAÇÃO

Hugo Eduardo Ziviani

BONE SUPRESSION USING DEEP LEARNING

Monografia apresentada ao Curso de Ciência da Computação da Universidade Federal
de Ouro Preto como requisito parcial para obtenção do título de Bacharel em Ciência da Computação

Aprovada em 22 de Junho de 2022.

Membros da banca

Guillermo Cámara Chávez (Orientador) - Doutor - Universidade Federal de Ouro Preto
Matheus Coelho Silva (Coorientador) - Mestre - Universidade Federal de Ouro Preto
Rodrigo César Pedrosa Silva (Examinador) - Doutor - Universidade Federal de Ouro Preto
Elton José da Silva (Examinador) - Mestre - Universidade Federal de Ouro Preto

Guillermo Cámara Chávez, Orientador do trabalho, aprovou a versão final e autorizou seu depósito na Biblioteca Digital de Trabalhos de Conclusão de Curso da UFOP em 22/06/2022.

Documento assinado eletronicamente por **Guillermo Camara Chavez**, **PROFESSOR DE MAGISTERIO SUPERIOR**, em 21/06/2022, às 19:01, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do Decreto nº 8.539, de 8 de outubro de 2015.

A autenticidade deste documento pode ser conferida no site http://sei.ufop.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0 , informando o código verificador **0346397** e o código CRC **C4807522**.

R. Diogo de Vasconcelos, 122, - Bairro Pilar Ouro Preto/MG, CEP 35400-000
Telefone: 3135591692 - www.ufop.br

Agradecimentos

Agradeço à Denise Conceição das Graças, minha mãe, que sempre me apoia em estudar, ser uma pessoa cada dia melhor e contribuir para uma sociedade igualitária e justa. Aproveito para agradecer meu irmão Victor Pedro que sempre foi meu exemplo de vitória e superação. Não menos importante agradeço à UFOP por me receber em dois cursos superiores e no mestrado. Consequentemente sou grato ao Departamento de Computação, onde tive ótimos professores como Elton Silva, que me apoiou, estimulou e publicamos nosso primeiro artigo acadêmico. Aproveito para agradecer aos outros professores do DECOM que de forma direta contribuiram para a minha formação acadêmica e profissional. Agradeço à Mariana e à Viviane, secretárias do departamento que sempre me proferiam palavras de apoio e motivação. Agradeço ao Mateus Coelho, co-orientador e amigo, pelo suporte e apoio pleno. Agradeço ao iMobilis, e todos seus integrantes, laboratório o qual me ajudou a crescer academicamente. Sem deixar de lembrar, sou grato ao meu companheiro de luta, das discuções políticas e noites sem dormir para estudar e fazer trabalhos: Frederico Martins. Finalizo este agradecimento com carinho especial à minha companheira e futura esposa Natália Oliveira.

Resumo

A proposta deste trabalho é revisar técnicas de supressão óssea em imagens de tórax. A forma mais comum, porém não acessível, é através da Subtração de Dupla Energia (DES). Essa técnica exige um hardware específico para a gerar e receber diferentes níveis de energia capaz de diferenciar os materiais pelo número atômico. Baseado em deep learning, nosso método utiliza redes adversárias, conhecidas como GAN's. Especificamente utilizamos um tipo de GAN condicional (cGAN). Para quantificar os resultados foi necessária a revisão das principais métricas utilizadas pelo estado da arte. A busca se deu por trabalhos que utilizassem o mesmo dataset da proposta ou técnicas similares. Com o dataset da JSRT, alcançamos um indice PSNR de 35.604, que se apresentou melhor que o revisto na literatura e um coeficiente de similaridade, conhecido como SSIM de 0.9703. Quanto ao loss, calculado pelo MS-SSIM, obtivemos o menor em comparação aos trabalhos revisados.

Palavras-chave: Supressão óssea. Redes Neurais Adversárias. Aprendizado Profundo. Redes Neurais Profundas.

Abstract

The purpose of this work is to review bone suppression techniques in chest imaging. The most common but not accessible way is through Double Energy Subtraction (DES). This technique requires specific hardware to generate and receive different levels of energy capable of differentiating materials by atomic number. Based on deep learning, our method uses adversarial networks, known as GAN's. Specifically we use a conditional GAN type (cGAN). To quantify the results, it was necessary to review the main metrics used by the state of the art. The library search was made looking for works that used the same dataset as the proposal or similar techniques. With the JSRT dataset, we reached a PSNR index of 35.604, which was better than that reviewed in the literature, and a similarity coefficient, known as SSIM of 0.9703. As for the loss, calculated by MS-SSIM, we obtained the lowest compared to the reviewed works.

Keywords: Bone Suppression. GAN. Deep Learning. Deep Neural Networks.

Lista de Algoritmos

Pix2Pix - Image-to-Image Translation with Conditional Adversarial Networks

Lista de Abreviaturas e Siglas

AP - Anteroposterior radiographs

AUC - Area Under the receiver operating characteristic Curve

AUC - Area under the ROC curve

ASM - active shape model

BSR - Bone suppressed rate

CLAHE - Contrast Limited Adaptive Histogram Equalization

CXR - Abbreviation for Chest X-Ray

cGAN - Conditional Generative adversarial network

DES - Dual energy Subtraction

DE - Dual Energy images

DICOM - Digital Imaging and Communications in Medicine

GAN - Generative adversarial network

GHT - Generalized Hought Transform

JSRT - Japanese Society of Radiological Technology

NMAE - Normalized Mean Absolute Error

MS-SSIM - Multi-scale Structural Similarity Index Measure

MSE - Mean squarred error

MAE - mean absolute error

Multi-scale Conditional Adversarial Network (MCA-Net)

sMAE - Scaled Mean Absolute Error

PA - Posteroanterior radiographs

PSNR - Peak Signal-to-Noise Ratio

SSIM - Structural Similarity Index Measure

IEEE - Institute of Electrical and Electronics Engineers

CNNs - Convolutional Neural Networks

Contents

1 Introduction . . . 1
1.1 Justification . . . 2
1.2 Objectives . . . 2
1.3 Work Organization . . . 3
1.3.1 Monography Structure . . . 3
2 Literature Review . . . 4
3 Materials and Methods . . . 10
3.1 Methodology . . . 10
3.1.1 Generative Adversarial Network (GAN) . . . 10
3.1.2 Condictional Generative Adversarial Networks (cGAN) . . . 11
3.2 Model and develop environment . . . 11
3.2.1 Building the Model . . . 12
3.2.1.1 Downsample - Encoder layer . . . 13
3.2.1.2 Upsample - Decoder layer . . . 14
3.2.2 Generator Loss . . . 14
3.2.3 Discriminator . . . 14
3.2.4 Discriminator loss . . . 14
3.2.5 Training the model . . . 14
3.2.5.1 Generate images and Test . . . 15
4 Results and Discussion . . . 16
4.1 Dataset . . . 16
4.2 Experimental protocol . . . 17
4.3 Metrics . . . 18
4.3.0.1 Mean Square Error (MSE) . . . 18
4.3.0.2 Peak Signal-to-Noise Ratio (PSNR) . . . 18
4.3.0.3 Structural Similarity Index Measure (SSIM) . . . 18
4.3.0.4 Multi Scale Structural Similarity Index Measure (MS-SSIM) . 20
4.4 Experimental results . . . 20
5 Conclusions and Future Works . . . 24
5.1 Future Works . . . 24
5.2 Published Works . . . 25

Bibliography . . . 26

1 Introduction

The computer's power and advanced methods are growing, and it is helping daily of various professionals with their tasks. The techniques such as optimized algorithms and distributed processing are advancing and can be used to auxiliary on difficult and tiring human functions. Sometimes a professional spends hours making analyses and diagnoses of medical exams. The technology, and computer-aided systems, can help these professionals work spending some computer processing.

According to (WANG et al., 2019) a challenge for medicine and computer-aided diagnosis is to make the lung analysis. Detecting any disease, for example, pneumonia, tumors, or any illness's evolution, can be better without bones shadow. In especial, for the analysis of the soft tissue, it would be perfect if, in an image, all bones could be suppressed. A notorious company in the medical image area is Carestream (MATTERS, 2014), in 2014, was published in their White Paper Journal the benefits of the analysis of body images without bones. According to the interviewed doctors, this image processing facilitates detecting and observing every possible lesion or disease.

One type of image to make the bone suppression on lung images is called *Dual-Energy*. We can get an image with high intensity and another with low intensity of x-ray signal from the image generator system. These two signals allow us to simultaneously get the different intensities of pixels from the same object. The attenuation coefficient of a bone is greater than the other chest parts. Because of this, we can calculate the difference between atomic number bone, denser, and lung soft-tissue, less dense. Regardless of costs, the only requirement of a *dual-energy* approach is a machine able to emit and receive this kind of signal and generate the image. The technique that uses dual-energy images to eliminate bone shadow is called Dual Energy Subtraction (DES). Since 1981 we can find studies that broach this technique, such as (KRUGER et al., 1981), and with the X-ray digital images quality improving, new ways to do bone subtraction emerged.

With the growth of *Graphic Process Unity* power (GPU), techniques using Machine Learning and Deep Learning are being very frequently studied and researched to solve medical problems such as (RAJARAMAN et al., 2021a; SUJATH; CHATTERJEE; HASSANIEN, 2020; OLIVEIRA et al., 2021). The bone shadow elimination by computer processing is a way not to expend much money buying dual-energy hardware and is an alternative to updating single-energy equipment to do the task. In this study, we will quickly review the main techniques for bone elimination on chest images. These strands will be analyzed and related to the primary purpose of this work. A study as (JUHÁSZ et al., 2010) relates an experiment using image processing in a GPU and an object detection approach to eliminate the shadows. We can use artificial intelligence techniques, such as deep learning, as it is a specific task. In this way, a more recent work solved

the task in (GUSAREV et al., 2017), which uses deep learning models to clean the soft-tissue image.

1.1 Justification

According to Andrew and Isaac S., leading researchers in the medical area, they relate humans and machines solve health problems with algorithms. On the comment in (BEAM; KOHANE, 2018), it is possible to comprehend how GANs are less human-dependent if they are correctly applied. We can go further in the theme and find the importance of bone suppression in the assistance to analyzing chest images such (BERG et al., 2016), how to suppress the bones on clinical images improve the analysis and helps on the daily tasks of health professionals. Diseases such as nodules, COVID-19, and any lung infection or lesion have been a significant problem to humanity in the last few years. Creating public research on this area will be an essential contribution to society. This work can be used in clinical research and improve any part of the lung diagnosis system that requires a clean soft-tissue chest image.

1.2 Objectives

This work aims study computational techniques for bone shadow elimination in chest x-ray through deep learning and image processing. The target is generate by software chest images without bones. Comparing the Figures 1.1 and 1.2, it is better to detect any anomaly on soft tissue if the bones are removed. To realize this task, we used a GAN to generate less noise in soft lung tissue. We used a DES dataset, a bone-suppressed lung image, as a ground truth for the training process.

This approach can be applied to soft-tissue lung analysis in any computer-aided diagnosis (CAD) system. The final model will be able to make bone suppression to other applications that use the lung image. With this system, we will be able to endorse the necessity of systems to clean the chest radiography decreasing errors in the final diagnosis.

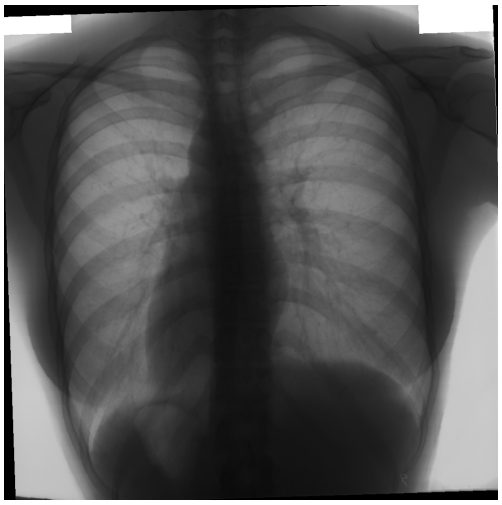

Figure 1.1 – Complete rib cage

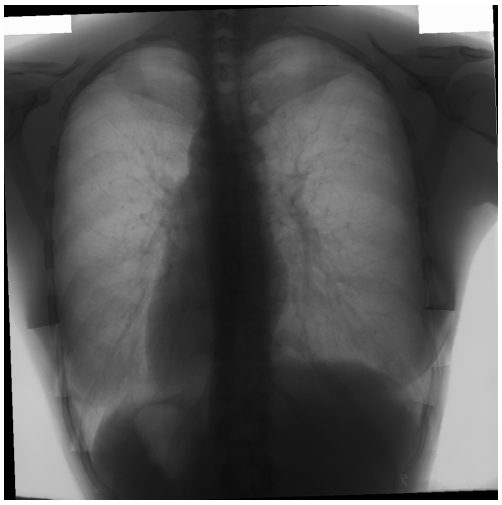

Figure 1.2 – Bone suppressed

Figure 1.3 – Comparing the possibility of views

1.3 Work Organization

The composition of this work is divided into chapters and subsections. In Chapter 2, we can find state-of-the-art techniques, theoretical references, and related works with similar approaches for bone suppression. Chapter 3 contains the used methodology, the description of a Generative Adversarial Network used, and its Conditional version. After that, we describe the model and the development environment and together we explain the GAN architecture. In Chapter 4, we start talking about the dataset. After, we describe the experimental protocol used to reach our results. After, we also present the used metrics to measure the model quality and how to calculate other examples of metrics mentioned in the library review. After all of that, we present the obtained results from experiments made with the test dataset. Finally, on the Conclusions in Chapter 5, we finalize talking about our strong and weak points of this work, related to the literature and we close with new possible works.

1.3.1 Monography Structure

Chapter 1: Introduction

Chapter 2: Literature Review

Chapter 3: Materials and Methods

Chapter 4: Results and Discussion

Chapter 5: Conclusion and Future Works

2 Literature Review

The literature review was done by research studies, articles, and books that assess bone shadow elimination or bone suppression. Techniques were found using dual-energy subtraction, deep learning, auto-encoder, convolutional networks, and adversarial networks. The research was carried out looking for works with outstanding results and a similar scenario to ours. In this way, it allows us to verify the results obtained according to the data set used. We considered the results that preserved the maximum of the features presented in the soft-tissue lung part.

There are different approaches developed for bone suppression. We established a chronological order to show the techniques.

The first of them is related to lung disease detection and monitoring. (HORVÁTH et al., 2009) proposed a process to suppress the bone on the lung image. The proposal is composed of three processes to detect the edge rib on the cage. The gradient is calculated to find the edges. After that, a statistical bone model is built and recalculated the image from the modified gradient, reducing the shadow's intensity inside the approximated border. Their innovation does not use any previous information or datasets, and the algorithm eliminates the shadow caused by the ribs of each image.

Another way to eliminate bone shadows on lung radiography is presented in (JUHÁSZ et al., 2010). The authors find the contours of the bones, including clavicles and rib cage, the heart contours and try to eliminate the shadows of these parts. The active shape model (ASM) segmented the chest areas and eliminated the shadows. After being determined by points, the object's boundary is described by a vector. This vector contains pairs of coordinates of the curve points. Each contour point is perpendicular to the contour. These positions were estimated and decided using a gradient method. They used the Mahalanobis distance to find the new point. The process to match the best contour representation is repeated until the image's resolution is increased.

Another study made by (LEE et al., 2012) has eliminated rib shadows as a process composed of lung field identification, rib segmentation, rib intensity estimation, and finally, the suppression task. The authors used a classical technique called Active Shape Model (ASM) to identify the ribs boundary. To reduce the noise during the estimation was applied a bilateral filter. So, they used a Generalized Hough Transform (GHT) to accurately the segmentation finding the upper and lower rib border. Then, the Real-coded Genetic Algorithm (RCGA) was used to estimate the rib gray-scale intensity. This model estimates the ribs edge based on the image gradient information. After identifying the rib borders, the corresponding intensity was subtracted from the original chest radiography. The performance was evaluated quantitatively using a Normalized Mean Absolute Error (NMAE) between rib-removed images and the corresponding ground-truth

images. The NMAE index was 0.03, better than 0.06 presented in a previous study.

Another proposal similar to (HORVÁTH et al., 2009), but with some modifications, is (HORVÁTH et al., 2013). Both works are from the same author, but the second one is updated and has some improvements. The work uses a Support Vector Machine to eliminate the rib's shadow. The algorithm essence is thought that the CNN filter establishes an adaptable bone edge model to be subtracted from the original image. After the image gradient is formed along the bone edge, the bone cage model is finally built based on the curve equation found by the model; the rib upper and lower neighborhoods are defined. The article evaluates and makes an overview of rib-cage segmentation. The study applied two different procedures for each part of the rib cage. This study used the Jaccard index to measure the intersection between the area marked by an observer and algorithm marks. The authors used a threshold of 0.55 of the Jaccard index to detect the rib. A metric used was the Distance of Found, which is calculated by pairing the points of the reference borders with the segmented borders. The study shows that the results of disease detection get better with bone suppression, which could eliminate around 52% of the false positives detection of rib borders. As a result of bone shadow elimination, they used a metric based on the intersection of the marked area predicted by the model. They reached a method that can eliminate approximately 80% of bone shadows in the chest frontal and 84% for the posterior part.

(SUZUKI et al., 2006) proposed a model trained by generated images "bone-like-images" thought dense linear layers, and the bone-suppressed images can be computed by subtracting the generated bone image from the corresponding original. The Massive Artificial Neural Network (MATANN) predicts the bone image from a standard chest radiograph, which can be subtracted from a similar image resulting in a soft-tissue image. The proposed MATANN is a high nonlinear filter that can be trained using chest radiographs and the corresponding target images. The pixel value is extracted from a chest X-ray and is input into the Artificial Neural Network (ANN), also called a neural filter. It can be trained with input and corresponding images, which will learn.

In 2016, (BERG et al., 2016) proposed a method for bone shadow suppression in radiography and preserving the original signal. The technical resource was developed using a neural network trained by subtracting dual-energy images. The model trains with 146 images to extract the rib cage. Their study used Spatial Transformation and did not present any clear metric to measure the accuracy of bone suppression. The main challenge was maintaining the intercostal unaffected because the front signal could overlay the backward representation. Their differential showed it is possible to separate the lung and remove the shadow ribs even with a pacemaker.

The study developed by (GUSAREV et al., 2017) was made with a non-identified dataset. The dataset was composed of 35 different sources. From this amount of data, they generated an augmented dataset with 4000 images. Just 10 images, not derived from the stage of data augmentation, were used to test. A Contrast Limited Adaptive Histogram Equalization (CLAHE) was applied to improve the quality of training images and feature extraction. This study proposed

a Convolutional Neural Network (CNN) with 6-layers filters. The input layer of the model has a 440×440 pixels dimension. The Loss of Multi-Scale Structural Similarity Index (MS-SSIM) obtained was 0.093. The authors did not calculate the PSNR. Other work related to our research was (OH; YUN, 2018). They used a no-identified dataset composed of 348 paired images of bones and suppressed. The division of train and test was not specified and the model architecture used is composed of a CNN, GAN, and *Haar Wavelets*. The model input size was differential because of the 1024×1024 pixels resolution. The MS-SSIM reached was 0.930 and the PSNR 24.08.

(YANG et al., 2017) test their proposal on the same dataset. They reached 0.976 of SSIM and 38.7 of PSNR. They propose a Deep Learning method for bone suppression in a single X-ray using cascade architecture of deep Convolutional Neural Networks (ConvNets) to map the bone's gradient domain. The main idea was to fuse with multi-scale bone gradients to improve prediction quality. More specifically, they used a Cascade of Multi-scale ConvNets (CamsNet). Their method does not require teaching from DES, but it requires segmentation and the border locations of bony structures. A positive point in their research is that their method works and performs considerably with different types of X-ray sources.

In 2018, (ZHOU et al., 2018) divided the dataset into 170 images to train and 40 to test. They used a Multi-scale Conditional Adversarial Network (MCA-Net). Their process produces a virtual high-resolution chest soft-tissue image from a synthetic rib chain generated. The process consists of two parts; the first generates bone images using a multi-scale fully convolutional network, and the second generates soft-tissue chest images using bone suppression of the standard CR with the virtual bone image generated. The model was tested with the JSRT dataset, and the images were divided into 170 for train and 40 to test. The model reached a PSNR of 39.7 and an SSIM of 0.884.

(OH; YUN, 2018) present two approaches; the first uses a conditional generative adversarial network, and the second a Haar 2D wavelet decomposition. They used the Euclidean distance between pairwise outputs to calculate the final result precision. They add to the experiments adversarial training to maintain the sharpness of specific lesions and avoid suppressing them. The main objective was to minimize the pixel-wise differences in bone suppression. The objective was to propose an image-to-image translation better than disposed on literature. Furthermore, they used a 2D wavelet decomposition as a perceptual guideline to minimize generic and ground truth differences. Finally, it is proposed a rigorously evaluated model to suppress bones from Dual Energy X-rays (DXRs). The Haar 2D wavelet decomposition is designed to have specific features for a signal process. A domain and frequency are specified similarly to Fourier Transform. The proposed architecture is based on (ISOLA et al., 2017), they changed the input system to grayscale and the input size of the generator to 1024×1024. They did the experiments with a non-identified dual-energy dataset. All the images were in DICOM format and were divided 80% for the train and 20% for the test. According to the quantitative results, they used PSNR, NPS, and

SSIM. For comparing the generated image and the target, histogram matching was used. The best result was using CNN with GAN and Haar Wavelets. The PSNR index was 24.080, and on the lungs region, they got 28.582. For the SSIM on the lung region, they got 0.93. For bone detection, several images were observed, and so was perceptible parallel ribs arrangement. The proposed model estimates an angle position for each image based on bone position. After this, they could use the ribs contour estimation to determine the structure edges. Dynamic programming was used to optimize the approximate curves, estimating the ribs' lower and upper outlines. They got a result of subtracting the clavicles as well. The shadow elimination is based on creating intensity profiles on subtraction of vertically differentiated images, which can return to the original domain by integration.

(ZARSHENAS et al., 2019) propose a study that has significant results. They propose generating virtual dual-energy images, separating ribs and clavicles from soft-tissue chest radiographs. The proposal is an Orientation-frequency-specific Deep Neural Network Convolution. They evaluated the proposal in their dataset, composed of 118 chest images, and reached a PSNR index of 29.82 and as SSIM of 0.912.

(CHEN et al., 2019) proposed a Cascaded Convolutional Network Model in Wavelet Domain Decomposition to do the bone shadow elimination using 504 images from a proprietary dataset, divided into 404 for training and 100 for testing. The trained network is used to predict the wavelet coefficients of the bone images. Thus, the predicted bone image is subtracted from the source, generating a bone-free image to train the model. Their trained model reached an SSIM of 0.977 and a PSNR of 39.7.

In 2019, (MATSUBARA et al., 2020), proposed a Convolutional Neural Filter (CNF) for a spatial filtering via CNN regression. This filter outputs a value for the bone component according to the neighborhood of the target pixel. In this process, a bone image is generated and subtracted from the original chest X-ray image. The images used in the study were obtained from Computer Tomography (CT) data. These CT images were converted in isotropic voxels, projecting them in the ventral-dorsal direction and applying a nonlinear transformation for bone enhancement. After that, the filter is applied, isolating the bone-specific signal. Finally, the bone-extracted image is obtained by subtracting the bone isolated from the original chest X-ray. Using the Japanese Society of Radiological Technology (JSRT) to evaluate the trained model, they reached a PSNR of 36.23 and an SSIM of 0.96.

Another study is (LIANG et al., 2020), the proposal is based on a Generative Adversarial Network (GAN) that learns bone suppression from dual-energy chest radiographs. A GAN is composed of two networks: a generator and a discriminator. The former creates images similar to the training set, while the latter discriminates them, classifying them as real or artificial. The authors evaluate two variations of GANs, namely Pix2Pix (ISOLA et al., 2017) with paired radiographs and Cycle-GAN with unpaired radiographs. With a private dataset composed by 1,867 anonymized dual-energy the data was divided in 70% to train, 20% to test, and 10% to

validation. The authors got an SSIM of 0.867 and a 36.078 on PSNR index for the suppression task.

Another approach that makes bone segmentation for diagnoses uses a neural network to segment the chest region. (ESLAMI et al., 2020) propose a multitask model that does organ segmentation, and one of those processes on their pipeline is bone shadow elimination. A CNN-based PatchGAN is used as a model architecture to do the bone suppression task. This architecture produces a matrix of size $k * k * 1$ from an input tensor where k is the size of the image. They used the JSRT dataset with 247 CXRs, including lung nodule images. All the images have the 2048×2048 pixel dimension, which was resized to 512×512 to adapt to the model entrance. The architecture used to translate the images was Pix2Pix (ISOLA et al., 2017). The results were evaluated by the SSIM, looking for similarity estimation and the MSE to measure the difference between the predicted and ground truth values. Calculating the MS-SSIM the authors got state-of-the-art results of 0.97.

On (SIRAZITDINOV et al., 2020) was used the ChestX-ray-14, a public dataset provided by (GUSAREV et al., 2017). They used 24 images for the train, 7 for the test, and 4 for the validation, with different models and architectures such as *autoencoder, U-Net, cGAN*. With all of them, the best precision was reached by the U-Net approach with 0.95 SSIM and 33.45 PSNR. Another study we can present is (ESLAMI et al., 2020), which used the augmented dataset of JSRT. Composed by 1,235, the study did not mention the division size of the train and test for the bone suppression part, just for the lung segmentation, but this is not our focus. For the bone suppression part, the architecture was pix2pix, the model input size was 512x512, and they calculated just the MS-SSIM, which was around 0.96 and 0.97.

Other similar work we can find on (ZHOU; ZHOU; SHEN, 2020). They propose a neural network model for bone suppression based on image-to-image translation. The model consists of dilated convolutions to avoid contextual information loss. Furthermore, the proposed method enforces pixel intensity similarity to improve the suppression quality using a deep convolutional network between the generated chest X-ray and the ground truth. The *dilated convolution* obtains a more effective receptive field, improving the performance of the conditional GAN. The extraction of visual features to generate the synthetic image was improved by expanding the receptive field to preserve the resolution. This module was plugged into the encoder and decoder and is composed based on U-Net architecture. Thus, the model buffer zone is filled with zeros to complete this process. The model of (ZHOU; ZHOU; SHEN, 2020) consists of a generator and a discriminator. For the generator, they use a U-net-like architecture with dilated convolutions/deconvolutions. The discriminator is based on PatchGAN to enforce the similarity of high-level feature representations. Using the JSRT, divided into 192 images to train and 42 to test, they reached 0.97 for SSIM index and 33.5 for PSNR.

(GOZES; GREENSPAN, 2020) presented a different approach, building their dataset from a Digital Reconstructed Radiographs (DRR) from a 664 Computer Tomography from a

cancer dataset, the LIDC-IDRI. The division had 386 images to train, 129 to test, and 129 to validate. The model and technique used were based on segmenting the bone structures in the CT domain to generate a bone-suppressed image to train a Fourier Convolutional Neural Network (FCNN) model available on (PRATT et al., 2017). The input of the trained network is 512×512, and they got 0.7 on the SSIM index and 22.6 for PSNR. The differential was to apply a dilated convolution and build an own dataset from a different source.

(RAJARAMAN et al., 2021b) used the JSRT dataset available at (HYUNH, 2021). They enhanced the contrast of the pixel values by 1%. The dataset composed of 4500 images was divided into 90% to train, 10% to test, and 10% for validation. The proposed architecture is a Residual Network model (ResNet-BS), where BS means Bone Suppression. The image size is 256×256, they experimented with 4 different architectures with the same dataset. For this last study, they got 0.9492 of SSIM and 34.0678 of PSNR.

To compare different views and approaches for bone suppression, we established some categories and main points to compare. One of them is the metrics, we focus on the PSNR index, SSIM, and derivatives. The other was observing the used dataset and pre-processing of images.

3 Materials and Methods

This Development section starts by describing the used methodology. After that, we briefly explain a Generative Adversarial Network (GAN). After this stage, we present how works a *conditional* GAN, known by cGAN. After that, we describe how is the training model and how to generate the images during the training. In the end, we present the complete model composition process and a brief description of each process on the full pipeline to the image processing.

3.1 Methodology

In this section, we present how Adversarial Networks are built and how a conditional variation of a GAN works. After this, we present the step-by-step used to build the model and illustrate the general architecture.

3.1.1 Generative Adversarial Network (GAN)

This work uses Generative Adversarial Network (GAN) as architecture. GAN is a Machine Learning (ML) framework that uses two neural networks. These networks are called Generative Network (Generator) and Discriminative Network (Discriminator). The Generator is a Convolutional Neural Network (CNN), and the Discriminator is a Deconvolutional Neural Network (DNN). The Generator aims to produce data as close as possible to the train data, and the Discriminator classifies the generated data. To illustrate the context in this work, we use images without bones so that the Generator would generate artificial images like that. The Discriminator will work on classifying these generated images as real or fake. To illustrate, we show Figure 3.1.

To better illustrate the architecture of a GAN, we can cite (OH; YUN, 2018). In an adversarial relationship, we identified **G** as the generator and **D** as discriminators. The *Generator* is a differentiable function, which takes the z variable for initial input for the model, so the sample $G(z)$ intent draw with the same distribution as observed in variable x. The *Discriminator* is a differentiable function called D as a binary classifier taking x and $G(z)$ and it outputs the probability. The discriminator is trained for a real and fake sample differently from traditional supervised learning. Finally, it generates a cross-entropy using a Sigmoid function. Below we can see the illustration on 3.1

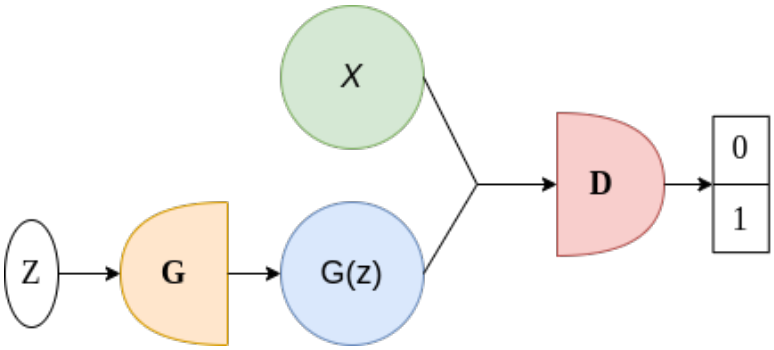

Figure 3.1 – GAN schema based on (OH; YUN, 2018)

3.1.2 Condictional Generative Adversarial Networks (cGAN)

A Conditional GAN can be comprehended according to (MIRZA; OSINDERO, 2014). To illustrate, for example, in a traditional GAN, we do not have control over the generated data, this is called an unconditioned generative model. Nevertheless, we can direct the model predictions to the objective, establishing conditions and class labels or part of the target data. In the figure below based on (OH; YUN, 2018), we can illustrate, through the 3.2, the discriminator **x** and **y** presented as inputs to a discriminative function.

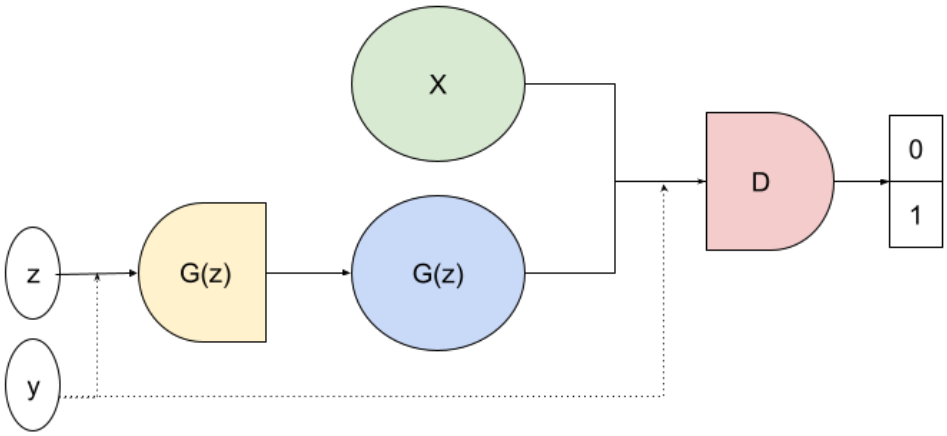

Figure 3.2 – Condictional GANs schema

As described, our proposal is based on a Conditional GAN, because the generator and discriminator are conditioned to an auxiliary image target.

3.2 Model and develop environment

We present a diagram that expresses the model composition and explains each development step in this section. First, we create the functions and elements for the Generator. It is

composed by *Downsample* and *Upsample* layers. To complete the process, we create the *loss* function that will be useful on Generator adjustment weights. After that, we create the Discriminator with its internal layers. Finally, we define the *checkpoints* rules and the Train and Test functions. This can be illustrated by a figure 3.3.

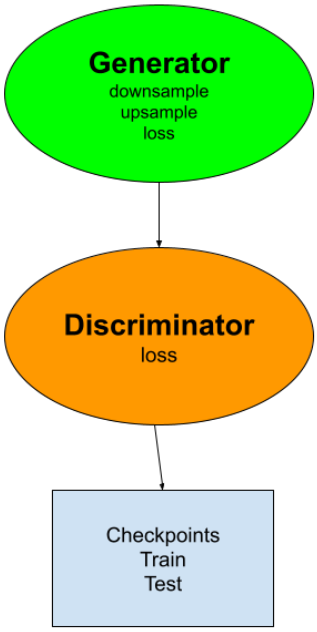

Figure 3.3 – Model description

3.2.1 Building the Model

The training process can be divided into two steps. The first moment, we created the folders to dispose of the data to be consumed by the train part of the model, and after creating the folder to receive the model checkpoints. During the training process, a set of few images are getting to readjust the model weights called validation set.

As we can see in the reference study, our proposal is composed of a generator based on a U-Net architecture and a discriminator represented by a PatchGAN, as proposed by (ISOLA et al., 2017). In a few words, the PatchGAN is a type of discriminator that only penalizes the scale of local image patches. Each patch of images is classified as whether a sample is real or fake.

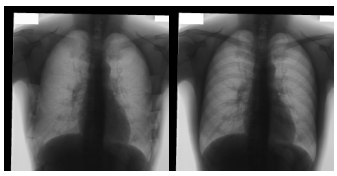

Figure 3.4 – Model description

The input image is composed of 2 parts. According the 3.4 The left is the chest x-ray without the bones, which is our ground truth. The right one is the source image with the bones. The process starts dividing the image input to encoding each half to the format **uint8** tensor. This is the common format that is usually used on Tensorflow framework. After loading the pair of images, they are separated into two parts to train the model. After that, we establish the buffer size, the batch size, and the image dimensions in 256×256.

The generator is based on a U-net architecture composed of an encoder (down-sampler) and a decoder that makes the up-sampling. The encoder block has the convolution process after a batch normalization and has as activation function Leaky ReLU. We know that activation functions can be changed based on our purpose. ReLU function does not activate all neurons simultaneously, and it makes the network sparse.

Inside the Generator, the architecture is composed by down stack and up stack. The first one is formed by 8 downsample layers and the second one is formed by 7 upsample layers with dropout. On the Discriminator layers we have 3 downsample sequential operations with Leaky Relu as activation function. To comprehend details we ilustrate the architecture on 3.5.

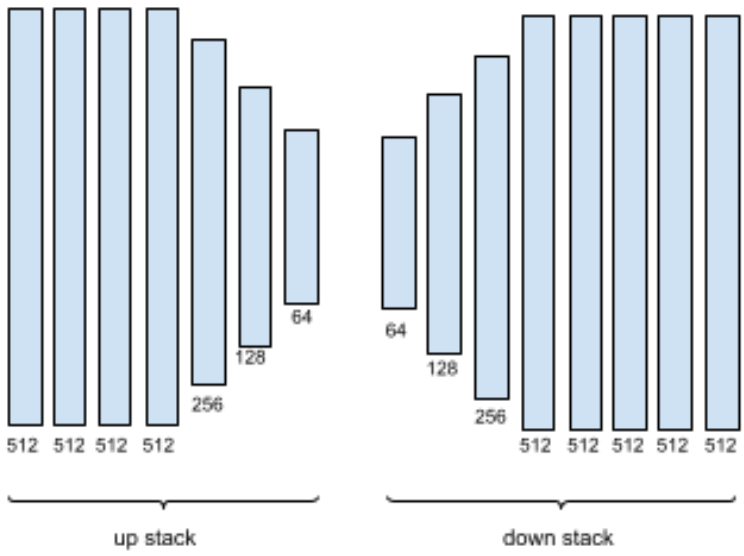

Figure 3.5 – Upsample and Downsample

3.2.1.1 Downsample - Encoder layer

The Downsample model, starts with a random normal initialize and is composed by Sequential Keras layer, which uses a standard 2D convolutional operations. We deactivated the Batch normalization in the first convolutional block, but the other layers are activated by default. The last step is applied and the Leaky ReLU. The down_stack is composed of 8 layers, and it

starts with the shape (256, 256, 3). In the future, we can adapt the model to a single-channel image. However, initially, just for tests and study, we decided to leave it.

3.2.1.2 Upsample - Decoder layer

This layer starts with a random normal initialization and is composed of a Sequential Keras model that also uses a standard 2D convolutional network. As on the first Sequential process mentioned, both were built to provide training and inference features on the model. Some of them were applied the *dropout* on intention to reduce the processing time.

3.2.2 Generator Loss

As described in (ISOLA et al., 2017), GAN's learn a loss and adapt it to the data. So the output which is distant from the target is penalized. A sigmoid cross-entropy represents the generator loss in this study. Another metric used was an L1 loss, calculated between the generated image and the target image based on mean absolute error (MAE). The formula to calculate it proposed by the authors was:

$$Generator\_loss = gan\_loss + LAMBDA * l1\_loss$$

3.2.3 Discriminator

The Discriminator proposed in pix2pix is a convolutional PatchGAN classifier. It tries to classify each image as artificial or real. The steps are composed of Convolution layers, Batch normalization, and a Leaky ReLU activation layer. The classifier receives the target and generated images, and the Discriminator classifies both images.

3.2.4 Discriminator loss

To evaluate the discriminator, we calculate its loss function. It means how the model is performing classifying real and artificial images. Basically, we pass to the discriminator loss function the real images and the generated images and we do the data verification for each image that was classified correctly by the discriminator. The final loss is the sum of the real and the generated losses. The real and the generated loss, are calculated using the Sigmoid Cross-entropy.

3.2.5 Training the model

We feed the input and target images into the network in the training step. After that, the generator calculates the discriminator loss. The gradient loss is optimized over each interaction. To train our GAN we use a loop interaction, this loop, involves the process of generate, discriminate, and validation during the train. Every 1k step is displayed the generated images to show the progress, and every 5k step is saved the checkpoints.

3.2.5.1 Generate images and Test

Finally, with the generator compiled and trained, we can use it to generate artificial images. We load the model, input the test data, and compare visually the Input Image, the Ground Truth, and the Predicted Image.

After all this process, we have a model ready to be used to predict lung images with the bone suppressed. This saved checkpoint can be found in a specific folder indicated on the code. It is possible to measure the checkpoint's quality and take the best one. We need to separate a known set of images source and targets for this process.

4 Results and Discussion

In this chapter, we will discuss about the dataset and results obtained from this study in comparing other works with a similar approach.

4.1 Dataset

The dataset used in our research comprises 240 pairs of dual energy chest radiography.. Available on (HYUNH, 2021), is a set of DES images. This dataset is available on the Institute of Electrical and Electronics Engineers (IEEE) site. The Japanese Society of Radiological Technology (JSRT) dataset is available with augmented data such as rotation and horizontal and vertical position changes between -30 and 30 degrees. These images were prepared, with each correspondent paired with bones and the other with removed shadows. These 240 pairs were rotated less than 90 degrees to augment the data.

For example we present some images below from the JSRT dataset.

Figure 4.1 – Complete rib cage

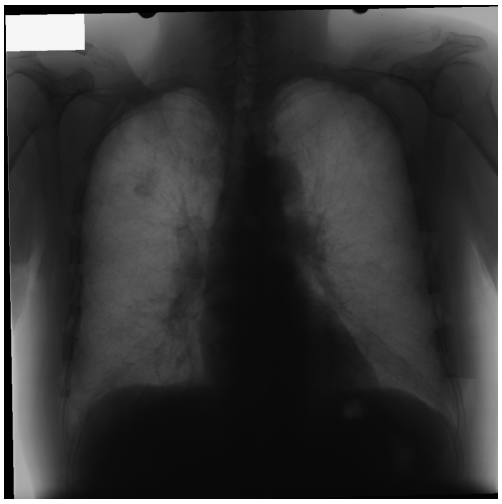

Figure 4.2 – Bone suppressed

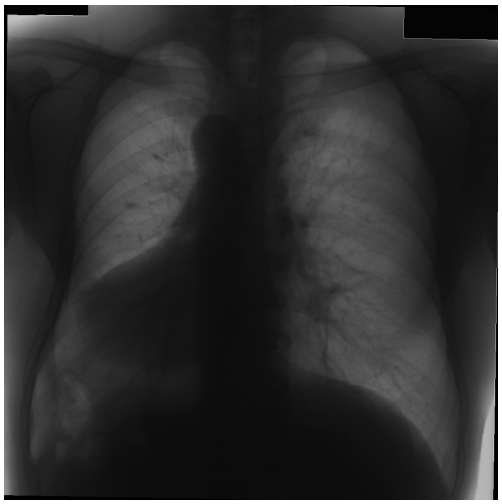

Figure 4.3 – Complete rib cage

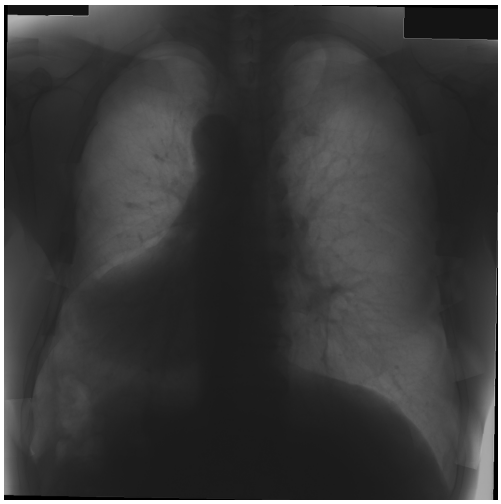

Figure 4.4 – Bone suppressed

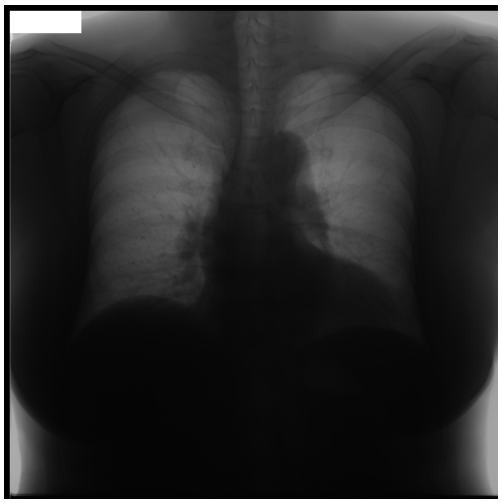

Figure 4.5 – Complete rib cage

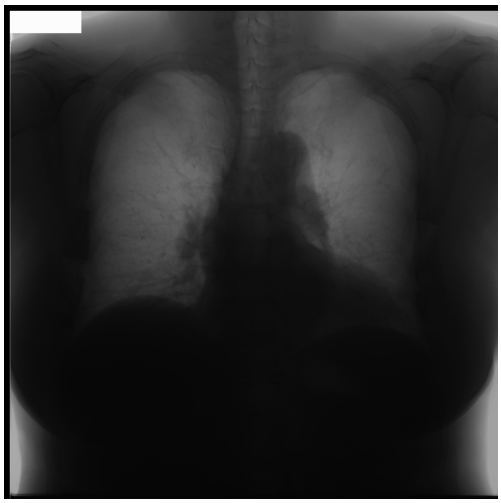

Figure 4.6 – Bone suppressed

4.2 Experimental protocol

To reproduce these experiments is necessary to use the (ISOLA et al., 2017) framework for cGAN architecture. If is used a GPU with 16G of RAM, on this code the buffer size has to be 400. The batch size needs to be 1. The image input size can be 256×256, 512×512 or 1024×1024. Inside the Generator or Discriminator, the activation function is a Leaky ReLU with Batch normalization on each convolution. The Generator optimizer is the Adam with a learning rate of 2^{-4} and a beta value of 0.5. The Discriminator optimizer is also the Adam optimizer with a learning rate of 2^{-4} and a beta value of 0.5. For the training phase, we used 45K interactions.

We used a total of 4,080 images. To do our experiments, we divided the image pairs into three categories: 3223 for the train (79%), 816 for the test (20%), and 40 for validation (1%). The training set was used to extract the features and balance the weights of the proposed cGAN. During the training, the validation set was used to check the train's model output. Finally, the test set was used to measure how accurate the model was.

4.3 Metrics

Below, we presented a brief description and explanation of the most used metrics to measure the model quality on the bone suppression task and make the literature review. In our proposal, we decided to use PSNR, SSIM, and MS-SSIM.

4.3.0.1 Mean Square Error (MSE)

Before understand Peak Signal-to-Noise Ratio (PSNR), we must explain the Mean Square Error (MSE). Given the images $\boldsymbol{f}$ and $\boldsymbol{g}$, both of $M \times N$ size. The relation is a squared difference between two pixels value and can be expressed by the Equation 4.1.

$$MSE(f,g) = \frac{1}{MN}\sum_{i=1}^{M}\sum_{i=1}^{N}(f_{ij} - g_{ij})^2 \tag{4.1}$$

4.3.0.2 Peak Signal-to-Noise Ratio (PSNR)

Given the images $\boldsymbol{f}$ and $\boldsymbol{g}$, both of $M \times N$ size (HORE; ZIOU, 2010) calculate the PSNR index with the Equation 4.2.

$$PSNR(f,g) = 10 * log_{10}(\frac{255^2}{MSE(f,g)}) \tag{4.2}$$

In other words, according to (RAJARAMAN et al., 2021b), the Peak Signal-to-noise Ratio is a metric that computes pixel values between the predicted and ground truth images. This ratio is used to provide a quantitative assessment of the predicted image. A higher value for PSNR indicates a higher quality of prediction. Concluding, when PSNR approches infinity, MSE approaches zero.

4.3.0.3 Structural Similarity Index Measure (SSIM)

Measuring image quality is a challenging problem without a reference. A metric that measures an image's content is needed independent of the difference of color intensity or luminosity. To solve this problem in 2004, (WANG et al., 2004) developed an index called Structural

Similarity Index Measure (SSIM). This index compares pixel intensities' local patterns with the normalized luminance and contrast. We must have the same image size to measure the signal quality. For example, we have two images $\boldsymbol{x}$ and $\boldsymbol{y}$, respectively. First, we measure the luminance, contrast, and structure. Finally, the three components are combined to yield an overall similarity measure. For more details of this metric, please review (WANG et al., 2004).

To understand the SSIM, we need to know some details and variables that compose the formula. The luminance is calculated by Eq. 4.3:

$$l(x,y) = \frac{2\mu_x\mu_y + c_1}{\mu_x^2 + \mu_y^2 + c_1} \tag{4.3}$$

The contrast is calculated by Eq. 4.4:

$$c(x,y) = \frac{2\sigma_x\sigma_y + c_2}{\sigma_x^2 + \sigma_y^2 + c_2} \tag{4.4}$$

The structure is calculated by Eq. 4.5:

$$s(x,y) = \frac{\sigma_{xy} + c_3}{\sigma_x\sigma_y + c_3} \tag{4.5}$$

The SSIM index can be calculated using the luminosity, the contrast, and the structure like below on Eq. 4.6.

$$SSIM(x,y) = [l(x,y)]^{\alpha}.[c(x,y)]^{\beta}.[s(x,y)]^{\gamma} \tag{4.6}$$

Using 4.6 to form the main equation, we just set the weights α, β, γ to 1, the formula can be reduced to Eq. 4.7:

$$SSIM(x,y) = \frac{(2\mu_x\mu_y + c_1)(2\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)} \tag{4.7}$$

To better understand the variables, c_1, c_2 and c_3 are constants to stabilize the division and avoid a null denominator. Those μ_x and μ_y are the average of x and y signal values. The σ_x and σ_y are the variance of x and y signal values. We use in our work, the SSIM according to (RAJARAMAN et al., 2021b), the index provides a measurement of the similarity between the ground truth and predicted images.

4.3.0.4 Multi Scale Structural Similarity Index Measure (MS-SSIM)

Another version of SSIM, called Multi Scale Structural Similarity Index Measure (MS-SSIM), was created to solve some specific class of measurement. According to (WANG et al., 2004), this index is conducted over multiple scales to evaluate overall image quality. This index is calculated based on SSIM, and the equation is:

$$PSNR(f,g) = 10 * log_{10}\left(\frac{255^2}{MSE(f,g)}\right) \tag{4.8}$$

On Equation 4.8, the variables x_j and y_j are the content of the signal extraction window. The M, on the equation, means the number of local signal extraction windows of the image.

4.4 Experimental results

We made a table with all results to be analyzed easily. Then, we will discuss and compare our results with all the cited works. More specifically, we established the SSIM and PSNR as basic metrics to compare. Another comparison key is the size of the digital image generated from the models, which are 256×256, 512×512, and 1024×1024. Furthermore, we directly compared our work and the approaches that used the same GAN architecture as ours.

Figure 4.7 display a sample generated from our GAN model after 45K training iterations. The output is of the same size as the input algorithm. It has a size of 256×256. The model is prepared to receive three channels, but our source is gray-scale. It generates the same color output. Next, we present the Table 4.4 summarizing the main results.

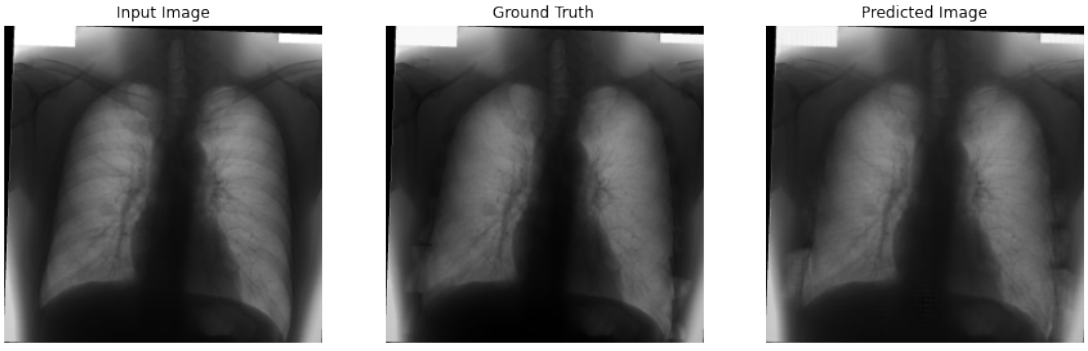

Figure 4.7 – Input image (Source), Target image (Ground truth) and Predicted images

Approaches	**Model and techniques**	**SSIM**	**PSNR**
(GUSAREV et al., 2017)	Auto-encoder and convolutional layers, Auto-encoder and without down/up sample operations	0.907	-
(OH; YUN, 2018)	Haar 2d Wavelet decomposition and vanilla pix2pix	0.930	24.08
(ESLAMI et al., 2020)	Condictional GAN and dilated convolutions	0.97	
(ZHOU; ZHOU; SHEN, 2020)	Dilated convolution to expand the receptive field	0.97	33.5
(YANG et al., 2017)	Cascade of multiscale CNN	0.976	38.7
(ZARSHENAS et al., 2019)	Anatomy-specific orientation-frequency-specific deep neural network convolution	0.912	29.82
(CHEN et al., 2019)	Cascade of multiscale CNN & wavelet decomposition	0.977	39.40
(ZHOU et al., 2018)	Multi-scale and conditional adversarial network	0.884	39.7
(MATSUBARA et al., 2020)	Bone suppression for chest X-ray image susing a convolutional neural	0.930	24.08
(LIANG et al., 2020)	Cycle-GAN - Image-to-image translation	0.867	36.078
(SIRAZITDINOV et al., 2020)	autoencoder, U-net, FPN, cGAN	0.955	33.45
(OH; YUN, 2018)	CNN + GAN + Haar Wavelets	0.930	24.08
(GOZES; GREENSPAN, 2020)	Hounsfield unit (HU) based segmentation and FCNN	0.70	22.6
(RAJARAMAN et al., 2021b)	Residual Network Model (ResNet-BS), where BS means Bone Suppression.	0.9492	34.0678
Our approach	Conditional GAN 256×256	0.9402	34.604
Our approach	Conditional GAN 512×512	0.9629	35.575
Our approach	Conditional GAN 1024×1024	**0.9703**	35.604

Table 4.1 – Results from related works for bone suppression

In Table 4.4, we present the mean results of three executions . This table contains the metrics from the reviewed studies indicating their precision in lung soft-tissue image generation. At the end of the table, we can find our results according to the related metrics. We adapted our results to possibly compare them with studies with similar approaches and used datasets.

Metric/Input Size	256×256	512×512	1024×1024
PSNR	34.6068	35.5753	35.6047
SSIM	0.9402	0.9629	0.9703

All the reviewed studies have a trade-off between time to train the model, train data amount, and if it is a supervised or non-supervised method. Our model depends directly on the number of dual-energy images and the computational time training the cGAN.

In the reviewed literature, we find two studies with the input size of the image being 256×256. Using the dilated conditional GAN, (ZHOU; ZHOU; SHEN, 2020) with a smaller input image size, they reached an SSIM 0.970 and a PSNR 33.50. Compared to our study, we did not reach this accuracy using the same image size. In contrast, we proposed extending the image's dimensions and getting better accuracy. The (RAJARAMAN et al., 2021b) used a Residual Network and divided the dataset in 90% for train 10% test and 10% for validation. Enhancing 1% the pixel contrast values, they got an SSIM of 0.9492 and a PSNR of 34.0678 with a low-resolution image. Working with images of 256×256, we got results better than (RAJARAMAN et al., 2021b), and with our architecture, our work is more embracing using high quality images

than (ZHOU; ZHOU; SHEN, 2020). On the 256×256 size category we reached on this image size an SSIM of 0.940 and a PSNR of 34.606.

Comparing our proposal with the reviewed works that used images with 512×512 size dimensions (ESLAMI et al., 2020), used the same approach of image-to-image translation using GAN. But to measure their model got 0.970, of a multi-scale SSIM. With the same image size and neural architecture, (MATSUBARA et al., 2020) got an SSIM of 0.960 and a PSNR of 36.230. However, they set 19 images to train and just one for the test, which is small. Thus, we can have an idea of their Convolutional Filter of six layers, but we can not confirm that the model is accurate. Referring to that, we had more consistent results because we used more data in the training phase. The (LIANG et al., 2020) divided their image set in 70% train 20% test and 10% for validation, with a Cycle-GAN image-to-image translation, they got an SSIM of 0.867 and a PSNR of 36.078.

In the same way, (GOZES; GREENSPAN, 2020) used the same input size of images and got an SSIM of 0.70 and a PSNR of 22.6. The smallest size of 440×440 was used by (GUSAREV et al., 2017), and they reached a multiscale SSIM of 0.907. The accuracy can not be compared directly because of the different input sizes. However, comparing our study with all those with the same architecture and input size of 512×512, we got better results in this category.

Evaluating the performance of different approaches using a bigger image size 1024×1024 dimension, we can find (OH; YUN, 2018), the authors used a similar approach like us, using a CNN with GAN and wavelets. They reached an SSIM of 0.930 and a PSNR of 24.08 with 348 pairs of images. The dataset division was not detailed, and the source was not specified. On the literature with 1080×1080 image size, we got (YANG et al., 2017; CHEN et al., 2019; ZHOU et al., 2018; SIRAZITDINOV et al., 2020; ZARSHENAS et al., 2019) proposals. (YANG et al., 2017) reached an SSIM better than ours, but the source dataset is not public, so we cannot reproduce it. The (ZHOU et al., 2018) used 170 images to train and 40 to test. They used the same architecture as us, but due to the little data to train, they reached an SSIM of 0.884 and a PSNR of 39.70.

According to Table 4.4, we can observe that our results are close to the state-of-the-art results, even better than some methods. Papers like (ESLAMI et al., 2020; ZHOU; ZHOU; SHEN, 2020; ZHOU et al., 2018; RAJARAMAN et al., 2021b) and our, used the JSRT dataset with a similar approach. It is possible to see that our model achieved significant results. Studies like (ESLAMI et al., 2020; ZHOU; ZHOU; SHEN, 2020; YANG et al., 2017; CHEN et al., 2019) presented high results of similarity, but it is hard to reproduce their results because the dataset and hyperparameters are not specified on the paper.

Below we present a table of the works that used the same JSRT dataset and the same approach using GAN to solve the bone supression problem:

Approaches	Model and techniques	SSIM	PSNR
(OH; YUN, 2018)	Haar 2d Wavelet decomposition and vanilla pix2pix	0.930	24.08
(ESLAMI et al., 2020)	Condictional GAN and dilated convolutions	0.970	
(ZHOU et al., 2018)	Multi-scale and conditional adversarial network	0.884	39.7
(LIANG et al., 2020)	Cycle-GAN - Image-to-image translation	0.867	36.078
(SIRAZITDINOV et al., 2020)	autoencoder, U-net, FPN, cGAN	0.955	33.45
(OH; YUN, 2018)	CNN + GAN + Haar Wavelets	0.930	24.08
Our approach	Conditional GAN 256 × 256	0.9402	34.604
Our approach	Conditional GAN 512 × 512	0.9629	35.575
Our approach	Conditional GAN 1024 × 1024	0.9703	35.604

5 Conclusions and Future Works

In 2020, many researchers focused on lung research because of COVID-19; obtaining a soft-tissue chest image facilitates the analyses and diagnosis of the disease. As we can see in the literature review, the bones on the chest image can sometimes be noisy when the soft tissue is diagnosed. There are plenty of ways to attenuate the shadows to improve the medical analysis. Some of those require a specific type of equipment that the costs are not accessible and exposes the patient to a double x-ray emission. Other approaches use classical computer image processing, looking for contours and edges. Other approaches use neural networks as filters or feature extractors. The choice of our work was to use deep learning with Adversarial Networks.

Our proposal is a CNN-based solution that learns from a dataset source. Our solution uses a DES dataset to show the cGAN the source and the target, and the model will learn how to generate an artificial image like the target dataset as possible. In the literature review, we could observe some not covered points. Some studies did the experiments and did not mention how many iterations or code parameters were used. Our results were near similar to state-of-the-art methods using the same metrics. For example, our approach is the third-best using PSNR, and for SSIM, we got results closer to the average performances.

To conclude this study, we could analyze the literature review and compare our results. We could conclude that our approach is promising, getting better results than classical methods and other deep learning strategies. As we can see, the model is promising, and that it can be trained with real-size images such as 1024×1024 or bigger. Furthermore, the proposed model is a proof of concept that uses a public dataset and public code and shows better results than state-of-the-art methods.

5.1 Future Works

In the future, we aim to increase the input model for a high resolution, such as 2048×2048, as we can use it with real images in a Digital Imaging and Communications in Medicine (DICOM) format. We are planning, as well, to apply more image processing techniques to increase the image quality, reduce the noise, and attenuate the shadows. Another future work is to use a dual-energy dataset, make the bone subtraction with classic techniques, and train our cGAN-based model.

Also, we would like to leave this work open to be used in other research, such as detecting lesions, infections, or diseases in the lung.

5.2 Published Works

This work gave rise to the scientific article "Applying a Conditional GAN for Bone Suppression in Chest Radiography Images", published in 2022, at the Integrated Software and Hardware Seminar (SEMISH), one of the main events of the academic and technology community in the area of computing.

1. Hugo Ziviani. Applying a Conditional GAN for Bone Suppression in Chest Radiography Images. Niterói - Rio de Janeiro: Conferência, 2022.

Bibliography

BEAM, A. L.; KOHANE, I. S. Big data and machine learning in health care. *Jama*, American Medical Association, v. 319, n. 13, p. 1317–1318, 2018.

BERG, J. von; YOUNG, S.; CAROLUS, H.; WOLZ, R.; SAALBACH, A.; HIDALGO, A.; GIMÉNEZ, A.; FRANQUET, T. A novel bone suppression method that improves lung nodule detection. *International journal of computer assisted radiology and surgery*, Springer, v. 11, n. 4, p. 641–655, 2016.

CHEN, Y.; GOU, X.; FENG, X.; LIU, Y.; QIN, G.; FENG, Q.; YANG, W.; CHEN, W. Bone suppression of chest radiographs with cascaded convolutional networks in wavelet domain. *IEEE Access*, IEEE, v. 7, p. 8346–8357, 2019.

ESLAMI, M.; TABARESTANI, S.; ALBARQOUNI, S.; ADELI, E.; NAVAB, N.; ADJOUADI, M. Image-to-images translation for multi-task organ segmentation and bone suppression in chest x-ray radiography. *IEEE transactions on medical imaging*, IEEE, v. 39, n. 7, p. 2553–2565, 2020.

GOZES, O.; GREENSPAN, H. Bone structures extraction and enhancement in chest radiographs via cnn trained on synthetic data. In: IEEE. *2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI)*. [S.l.], 2020. p. 858–861.

GUSAREV, M.; KULEEV, R.; KHAN, A.; RIVERA, A. R.; KHATTAK, A. M. Deep learning models for bone suppression in chest radiographs. In: IEEE. *2017 IEEE Conference on Computational Intelligence in Bioinformatics and Computational Biology (CIBCB)*. [S.l.], 2017. p. 1–7.

HORE, A.; ZIOU, D. Image quality metrics: Psnr vs. ssim. In: IEEE. *2010 20th international conference on pattern recognition*. [S.l.], 2010. p. 2366–2369.

HORVÁTH, Á.; ORBÁN, G. G.; HORVÁTH, Á.; HORVÁTH, G. An x-ray cad system with ribcage suppression for improved detection of lung lesions. *Periodica Polytechnica Electrical Engineering*, v. 57, n. 1, p. 19, 2013.

HORVÁTH, G.; ORBÁN, G.; HORVÁTH, Á.; SIMKÓ, G.; PATAKI, B.; MÁDAY, P.; JUHÁSZ, S. A cad system for screening x-ray chest radiography. In: SPRINGER. *World Congress on Medical Physics and Biomedical Engineering, September 7-12, 2009, Munich, Germany*. [S.l.], 2009. p. 210–213.

HYUNH, M.-C. *X-ray Bone Shadow Suppression*. IEEE Dataport, 2021. Disponível em: <https://dx.doi.org/10.21227/xnb5-hg35>.

ISOLA, P.; ZHU, J.-Y.; ZHOU, T.; EFROS, A. A. Image-to-image translation with conditional adversarial networks. *CVPR*, 2017.

JUHÁSZ, S.; HORVÁTH, Á.; NIKHÁZY, L.; HORVÁTH, G. Segmentation of anatomical structures on chest radiographs. In: SPRINGER. *XII Mediterranean Conference on Medical and Biological Engineering and Computing 2010*. [S.l.], 2010. p. 359–362.

KRUGER, R. A.; ARMSTRONG, J.; SORENSON, J.; NIKLASON, L. Dual energy film subtraction technique for detecting calcification in solitary pulmonary nodules. *Radiology*, v. 140, n. 1, p. 213–219, 1981.

LEE, J.-S.; WANG, J.-W.; WU, H.-H.; YUAN, M.-Z. A nonparametric-based rib suppression method for chest radiographs. *Computers & Mathematics with Applications*, Elsevier, v. 64, n. 5, p. 1390–1399, 2012.

LIANG, J.; TANG, Y.-X.; TANG, Y.-B.; XIAO, J.; SUMMERS, R. M. Bone suppression on chest radiographs with adversarial learning. In: INTERNATIONAL SOCIETY FOR OPTICS AND PHOTONICS. *Medical Imaging 2020: Computer-Aided Diagnosis*. [S.l.], 2020. v. 11314, p. 1131409.

MATSUBARA, N.; TERAMOTO, A.; SAITO, K.; FUJITA, H. Bone suppression for chest x-ray image using a convolutional neural filter. *Physical and Engineering Sciences in Medicine*, Springer, v. 43, n. 1, p. 97–108, 2020.

MATTERS, I. Carestream's new bone suppression software receives fda clearance, now available worldwide. 2014.

MIRZA, M.; OSINDERO, S. Conditional generative adversarial nets. *arXiv preprint arXiv:1411.1784*, 2014.

OH, D. Y.; YUN, I. D. Learning bone suppression from dual energy chest x-rays using adversarial networks. *arXiv preprint arXiv:1811.02628*, 2018.

OLIVEIRA, B.; ZIVIANI, H.; OLIVEIRA, J.; VIEGAS, A.; CALVO, D. Suporte para diagnóstico de covid-19 por meio de classificação automática de imagens de raio-x e modelos explicáveis. In: FILHO, C. J. A. B.; SIQUEIRA, H. V.; FERREIRA, D. D.; BERTOL, D. W.; OLIVEIRA, R. C. L. ao de (Ed.). *Anais do 15 Congresso Brasileiro de Inteligência Computacional*. Joinville, SC: SBIC, 2021. p. 1–8.

PRATT, H.; WILLIAMS, B.; COENEN, F.; ZHENG, Y. Fcnn: Fourier convolutional neural networks. In: SPRINGER. *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*. [S.l.], 2017. p. 786–798.

RAJARAMAN, S.; COHEN, G.; ANTANI, S. et al. A bone suppression model ensemble to improve covid-19 detection in chest x-rays. *arXiv preprint arXiv:2111.03404*, 2021.

RAJARAMAN, S.; ZAMZMI, G.; FOLIO, L.; ALDERSON, P.; ANTANI, S. Chest x-ray bone suppression for improving classification of tuberculosis-consistent findings. *Diagnostics*, Multidisciplinary Digital Publishing Institute, v. 11, n. 5, p. 840, 2021.

SIRAZITDINOV, I.; KUBRAK, K.; KISELEV, S.; TOLKACHEV, A.; KHOLIAVCHENKO, M.; IBRAGIMOV, B. Evaluation of deep learning methods for bone suppression from dual energy chest radiography. In: SPRINGER. *International Conference on Artificial Neural Networks*. [S.l.], 2020. p. 247–257.

SUJATH, R.; CHATTERJEE, J. M.; HASSANIEN, A. E. A machine learning forecasting model for covid-19 pandemic in india. *Stochastic Environmental Research and Risk Assessment*, Springer, v. 34, p. 959–972, 2020.

SUZUKI, K.; ABE, H.; MACMAHON, H.; DOI, K. Image-processing technique for suppressing ribs in chest radiographs by means of massive training artificial neural network (mtann). *IEEE Transactions on medical imaging*, IEEE, v. 25, n. 4, p. 406–416, 2006.

WANG, S.; YANG, D. M.; RONG, R.; ZHAN, X.; FUJIMOTO, J.; LIU, H.; MINNA, J.; WISTUBA, I. I.; XIE, Y.; XIAO, G. Artificial intelligence in lung cancer pathology image analysis. *Cancers*, Multidisciplinary Digital Publishing Institute, v. 11, n. 11, p. 1673, 2019.

WANG, Z.; BOVIK, A. C.; SHEIKH, H. R.; SIMONCELLI, E. P. Image quality assessment: from error visibility to structural similarity. *IEEE transactions on image processing*, IEEE, v. 13, n. 4, p. 600–612, 2004.

YANG, W.; CHEN, Y.; LIU, Y.; ZHONG, L.; QIN, G.; LU, Z.; FENG, Q.; CHEN, W. Cascade of multi-scale convolutional neural networks for bone suppression of chest radiographs in gradient domain. *Medical image analysis*, Elsevier, v. 35, p. 421–433, 2017.

ZARSHENAS, A.; LIU, J.; FORTI, P.; SUZUKI, K. Separation of bones from soft tissue in chest radiographs: Anatomy-specific orientation-frequency-specific deep neural network convolution. *Medical physics*, Wiley Online Library, v. 46, n. 5, p. 2232–2242, 2019.

ZHOU, B.; LIN, X.; ECK, B.; HOU, J.; WILSON, D. Generation of virtual dual energy images from standard single-shot radiographs using multi-scale and conditional adversarial network. In: SPRINGER. *Asian Conference on Computer Vision*. [S.l.], 2018. p. 298–313.

ZHOU, Z.; ZHOU, L.; SHEN, K. Dilated conditional gan for bone suppression in chest radiographs with enforced semantic features. *Medical Physics*, Wiley Online Library, v. 47, n. 12, p. 6207–6215, 2020.