

Universidade Federal de Minas Gerais
Instituto de Ciências Exatas
Departamento de Ciência da Computação

DAVID ALEXANDER ANTUNES DE OLIVEIRA

MONOGRAFIA DE PROJETO ORIENTADO EM COMPUTAÇÃO II

**DATA AUGMENTATION PARA
TREINAMENTO DA CODAGAN 3D**

Belo Horizonte
2019

Universidade Federal de Minas Gerais
Instituto de Ciências Exatas
Departamento de Ciência da Computação
Curso de Bacharelado em Ciência da Computação

**DATA AUGMENTATION PARA
TREINAMENTO DA CODAGAN 3D**

por

David Alexander Antunes de Oliveira

Monografia de Projeto Orientado em Computação II
Apresentado como requisito da disciplina de Projeto
Orientado em Computação II do Curso de Bacharelado em
Ciência da Computação da UFMG

Prof. Dr. Jefersson Alex dos Santos
Orientador(a)
Hugo Neves de Oliveira
Co-orientador(a)

Belo Horizonte
2019

AGRADECIMENTOS

Agradeço à minha mãe, pelo amor incondicional, por ser a mulher mais batalhadora que já vi nesta terra, e por ser a melhor mãe do mundo.

RESUMO

O objetivo deste trabalho é utilizar uma arquitetura de Deep Neural Network (DNN) que consiga através de aprendizado supervisionado fazer a predição de diversas segmentações diferentes em imagens médicas.

Para isso, fazemos um estudo de uma das mais recentes inovações no campo de Deep Learning, as chamadas GANs. Uma nova proposta deste modelo, CoDAGANs, utiliza uma combinação de técnicas para alcançar resultados promissores em segmentação de imagens médicas.

Visando contribuir para a CoDAGANs, este trabalho propõe o uso de técnicas de processamento de imagem tradicionais para aperfeiçoar as entradas do modelo. Com o objetivo de melhorar ainda mais os resultados obtidos, aplicamos técnicas de data augmentation nos datasets de treino.

Palavras-chave:

Deep Learning, Imagens médicas, Segmentação Semântica, Aprendizado Semi-supervisionado, Processamento de Imagens Tradicional.

ABSTRACT

The objective of this work is to use a Deep Neural Network (DNN) architecture that can through supervised learning to predict several different segmentations in medical images.

To do this, we study one of the most recent innovations in the field of Deep Learning, called GANs. A new proposal for this model, CoDAGANs, uses a combination of techniques to achieve promising results in segmentation of medical imaging.

Aiming to contribute to CoDAGANs, this work proposes the use of traditional image processing techniques to improve model inputs. In order to further improve results, we apply data augmentation techniques to training datasets.

Keywords:

Deep Learning, Biomedical Images, Semantic Segmentation, Semi-Supervised Learning, Traditional Image Processing.

LISTA DE FIGURAS

FIGURA 1 - ARQUITETURA DE CODAGANS	11
FIGURA 2 - DRRs E SEUS RESPECTIVOS RÓTULOS	12
FIGURA 3 - FROM 3D TO 2D: A TRANSFER LEARNING CASE STUDY FOR BONE SEGMENTATION IN X-RAYS	13
FIGURA 4 - EXEMPLOS DE FLIP	14
FIGURA 5 - EXEMPLOS DE ROTATION	15
FIGURA 6 - EXEMPLOS DE PERSPECTIVE TRANSFORMATION	15

LISTA DE SIGLAS

CoDAGANs	Conditional Domain Adaptation GANs
CT	Computer Tomography
CXR	Chest X-Ray
DL	Deep Learning
DRR	Digitally Reconstructed Radiograph
GANs	Generative Adversarial Networks

SUMÁRIO

RESUMO	4
ABSTRACT	5
LISTA DE FIGURAS	6
LISTA DE SIGLAS	7
1 INTRODUÇÃO	9
2 REFERENCIAL TEÓRICO	10
3 TRABALHOS ANTERIORES (POC I)	12
4 DATA AUGMENTATION PARA TREINAMENTO DA CODAGAN	14
5 CONCLUSÕES	16
6 REFERÊNCIAS	17

1 INTRODUÇÃO

A indústria médica é uma das indústrias mais proeminentes onde o Deep Learning (DL) pode desempenhar um papel enorme, especialmente quando se trata de imagens médicas. Existem inúmeros trabalhos que comprovaram a eficiência e eficácia do uso de DL no campo. Alguns destes métodos são formas promissoras de auxiliar um médico a fazer um diagnóstico mais rápido e melhor, ou mesmo, encontrar respostas que um médico talvez não encontraria.

No entanto, DL têm uma forte demanda por dados, especialmente no caso de dados médicos, basicamente por dois fatores: casos médicos anormais são geralmente muito mais raros do que os casos normais, é muito demorado coletar uma quantidade razoável para todos os casos. Além disso, para abordagem supervisionada, a quantidade de dados coletados também deve ser rotulada por médicos qualificados.

Conditional Domain Adaptation GANs (CoDAGANs), proposta por Oliveira et al., é um tipo de rede neural que pode ser usada para inferir rótulos de imagens médicas, já que as vezes os dados até existem, porém não são rotulados.

O objetivo específico desta Monografia de Projeto Orientado em Computação II é prover uma implementação voltada a *Data Augmentation* dos datasets usados no treinamento da CoDAGAN, de forma a buscar resultados ainda melhores.

Outras seções neste documento são organizadas da seguinte forma. A Seção 2 apresenta os trabalhos anteriores que abriram caminho para a proposta das CoDAGANs. A Seção 3 fala sobre o trabalho realizado até aqui no POC I. A Seção 4 faz um detalhamento técnico nas técnicas de processamento de imagens utilizados. A Seção 5 discute algumas conclusões.

2 REFERENCIAL TEÓRICO

Generative Adversarial Networks

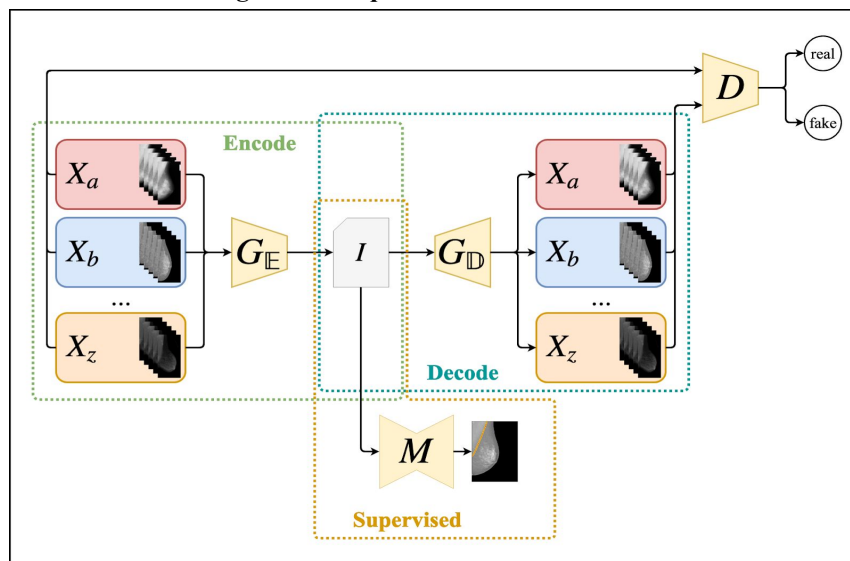
Generative Adversarial Networks (GANs) são um tipo de Deep Neural Networks (DNNs) proposto por Goodfellow *et al.* que consiste basicamente em duas redes neurais. A ideia é treinar simultaneamente essas duas redes neurais. A primeira, chamada Discriminador - vamos denotar $D(Y)$ - recebe uma entrada (por exemplo, uma imagem) e gera um escalar que indica se a imagem Y parece “natural” ou não. Em um exemplo de treinamento adversarial, $D(Y)$ pode parecer um tipo de função de energia que leva um valor baixo (por exemplo, próximo a 0) quando Y é uma amostra real (por exemplo, uma imagem de um dataset) e um valor positivo quando não é (por exemplo, se é uma imagem estranha ou com ruídos). A segunda rede é chamada de gerador, denotada por $G(Z)$, onde Z é geralmente um vetor aleatoriamente amostrado em uma distribuição simples (por exemplo, Gaussiana). O papel do gerador é produzir imagens de modo a treinar a função $D(Y)$ para obter a forma correta (valores baixos para imagens reais, valores maiores para todo o resto). Durante o treino, é mostrada uma imagem real à D , que ajusta seu parâmetro para tornar sua saída mais baixa. Então é mostrada uma imagem produzida a partir de G à D que ajusta seus parâmetros para tornar sua saída $D(G(Z))$ maior (seguindo o gradiente de alguma função predefinida objetiva). Mas $G(Z)$ irá treinar-se para produzir imagens de modo a enganar D em pensar que elas são reais. Ela faz isso obtendo o gradiente de D em relação a Y para cada amostra que produz. Em outras palavras, está tentando minimizar a saída de D enquanto D está tentando maximizá-la. Daí o nome treinamento adversarial.

Isso é interessante pois essa rede permite treinar um discriminador como um "estimador de densidade" não supervisionado, ou seja, uma função de contraste que nos dá um valor baixo para os dados e uma saída mais alta para todo o resto. Esse discriminador deve desenvolver uma boa representação interna dos dados para resolver esse problema adequadamente. Em seguida, ele pode ser usado como um extrator de recursos para um classificador, por exemplo.

Conditional Domain Adaptation GANs

CoDAGANs combinam aprendizado não supervisionado e supervisionado para executar Domain Adaptation (DA) entre dois ou mais conjuntos de imagens. A Figura 1 apresenta a arquitetura de CoDAGANs, um único par $\{D, G\}$ de DNNs é treinado em um cenário adversarial para realizar traduções entre os conjuntos de dados. Um único modelo supervisionado M é treinado para executar uma tarefa supervisionada em todos os conjuntos de dados.

Figura 1 - Arquitetura de CoDAGANs.



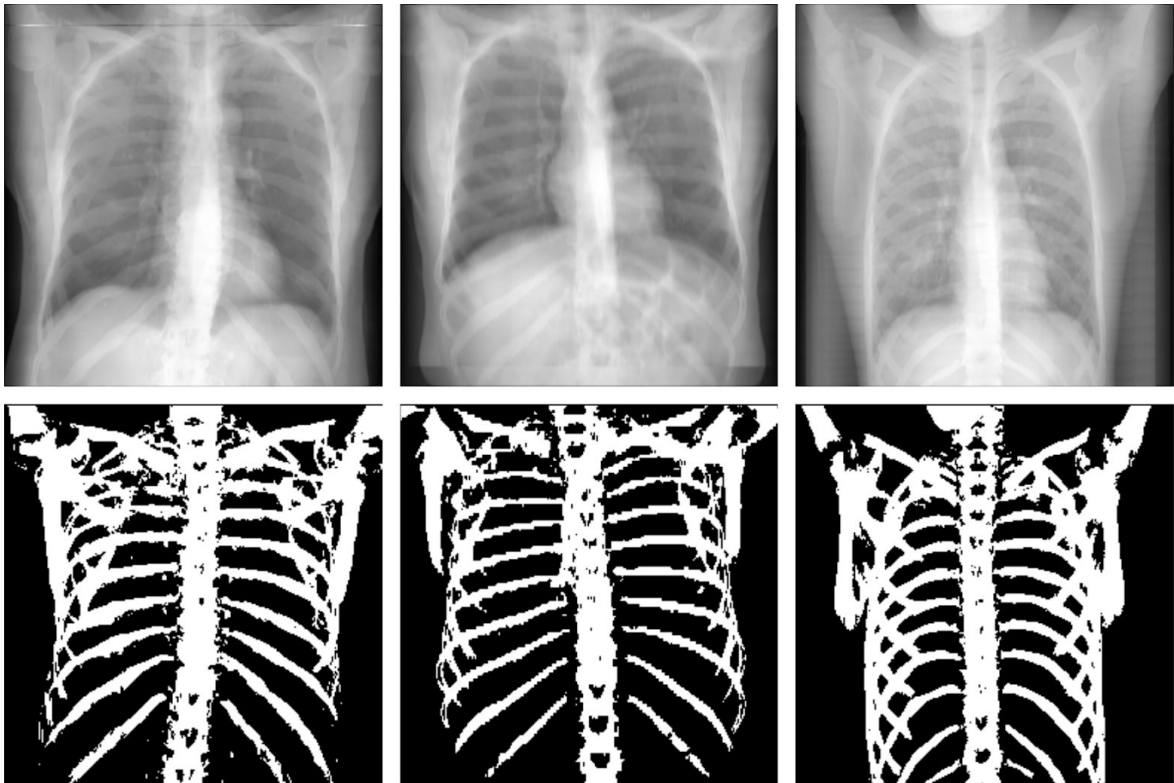
Fonte: Oliveira et al.

3 TRABALHOS ANTERIORES (POC I)

O objetivo do trabalho era o de estudar as CoDAGANs no caso de utilização para a transferência de conhecimento entre imagens radiológicas sintéticas e radiografias digitais. A rede é capaz de aprender um modelo de segmentação óssea em CXRs não rotuladas usando dados de DRRs. Uma DRR é uma radiografia sintética que pode ser gerada a partir de um CT scan.

Uma imagem em 3D (CT volume) é processada com operações matemáticas simples. A DRR é gerada a partir de uma combinação linear de algumas características da projeção 2D do CT volume. Ao mesmo tempo é feita uma binarização por limiarização para obter a segmentação dos ossos. O resultado disso é mostrado na Figura 2.

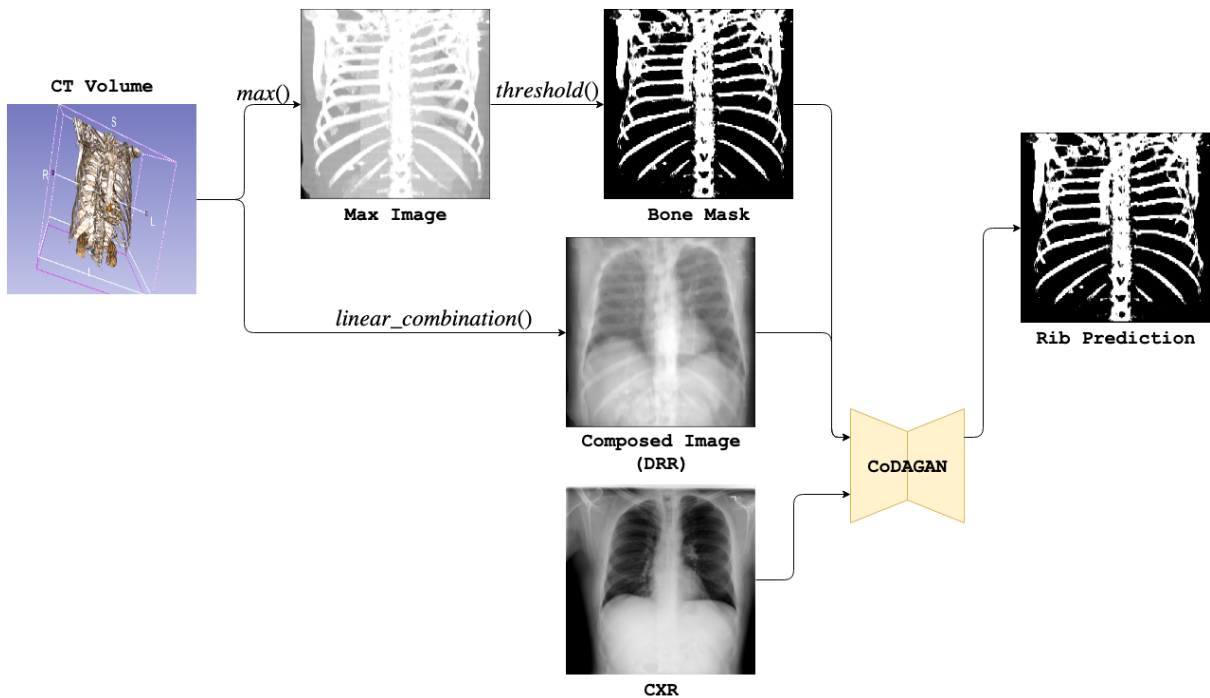
Figura 2 - DRRs e seus respectivos rótulos



Fonte: Oliveira et al.

A partir daí, a DRR e seu label são passadas para o treinamento da CoDAGAN que, depois de treinada, pode ser usada para fazer a segmentação dos ossos em um CXR real não visto antes. O esquema pode ser visto na Figura 3.

Figura 3 - From 3D to 2D: A Transfer Learning Case Study for Bone Segmentation in X-Rays.



Fonte: Oliveira et al.

Como é possível ver, ainda na Figura 2, a rotulação dos ossos das DRRs geradas por operações a partir do CT volume é ruidosa, por isso o uso de técnicas de processamento de imagens tradicionais valia de grande estudo neste caso. Conjuntos de operações podem ser construídos a partir das operações básicas, como a abertura e o fechamento. Estas técnicas combinadas podem ser utilizadas para selecionar ou suprimir detalhes de uma determinada forma, por exemplo, remover ruído de imagens ou selecionar objetos com uma direção específica.

4 DATA AUGMENTATION PARA TREINAMENTO DA CODAGAN

Nesta etapa do trabalho o foco foi passar os esforços para a versão 3D do treinamento da CoDAGAN. O treinamento supervisionado de uma rede neural requer muitos dados, por isso, é comum que os cientistas utilizem uma prática denominada *Data Augmentation*. Isso é uma estratégia que permite aumentar significativamente a diversidade de dados disponíveis para modelos de treinamento, sem realmente coletar novos dados.

O que é feito são pequenas alterações no dataset disponível, como girar, inverter, cortar partes da imagem, de forma que a rede que está treinando pense que estas são imagens totalmente diferentes, e aprenda a generalizar mais. Nesta seção faremos um resumo dos principais *augments* usados para expandir o dataset utilizado no treinamento da CoDAGAN, e alguns resultados.

Bibliotecas utilizadas

A principal biblioteca utilizada para o data augmentation realizado foi a “imgaug”¹ que contém várias transformações implementadas a partir de outras bibliotecas como scipy, Pillow e scikit-image.

Operações aplicadas

Flip - Girar a imagem pode ser de duas maneiras, horizontal e verticalmente. O flip horizontal (fliplr) é o mesmo que espelhar a foto. O flip vertical (flipud) é o mesmo que rotacionar a foto 180° e espelhá-la horizontalmente. No data augmentation feito, o flip tinha 10% de chance de acontecer horizontalmente, e 10% de chance de acontecer na vertical. Portanto, vez ou outra uma imagem do sample de treinamento era passada flipada.

Figura 4. Exemplos de flip

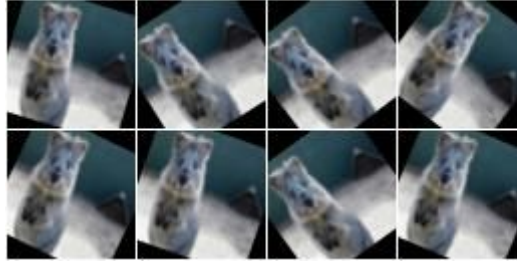


Fonte: imgaug.

Rotation - Um exemplo simples de transformações afins é a rotação, que é mais uma operação que pode ser utilizada para aumentar a quantidade de dados, provendo praticamente uma

imagem nova para a rede. No data augmentation feito, as imagens podiam ser rotacionadas aleatoriamente em qualquer valor variando de -90° a 90° .

Figura 5. Exemplos de rotation.



Fonte: imgaug.

Perspective transformation - essa operação aplica transformações aleatórias de perspectiva de quatro pontos às imagens. Cada um dos quatro pontos é colocado na imagem usando uma distância aleatória de seu respectivo canto. A distância é amostrada a partir de uma distribuição normal. Como resultado, a maioria das transformações não altera muito a imagem, enquanto algumas "focam" em polígonos bem dentro da imagem.

Figura 6. Exemplos de perspective transformation



Fonte: imgaug.

Permute e transpose - a permutação altera a ordem das dimensões também, ou seja, dos eixos. No data augmentation era feito de forma aleatória. Transposição é um caso especial de permutação que só se aplica em duas dimensões.

Random crops e resize - é feita uma amostragem aleatória de uma seção da imagem, em seguida a redefinição do tamanho das imagens para mandá-las ao treinamento.

5 CONCLUSÕES

Com algumas dessas operações aplicadas ao dataset foram feitos diferentes treinamentos da rede. Uma forma de verificar o quão bom está o treinamento da rede é através do método Jaccard, passando para o algoritmo os labels ground truth e os labels previstos pela rede, é possível verificar a similaridade dos dois conjuntos pela interseção e pela união.

O comparativo dos métodos usados neste trabalho (versão 1) contra outra implementação de data augmentation (versão 2) pode ser vista abaixo:

	VERSÃO 1	VERSÃO 2
Teste A	43.1597 +/- 6.5881	33.8036 +/- 6.1314
Teste B	5.7222 +/- 2.6871	12.1670 +/- 2.9734
Teste C	5.6867 +/- 3.2275	9.7440 +/- 5.8545
Teste D	6.1739 +/- 2.8451	4.9845 +/- 0.7793
Teste E	4.8599 +/- 0.2601	6.8690 +/- 1.1762
Teste F	0.0254 +/- 0.0584	0.0066 +/- 0.0104
Teste G	1.6824 +/- 2.1382	1.4534 +/- 1.5236

Como é possível observar, o método é melhor para alguns casos de teste, porém pior para outros. Com novas tentativas é possível melhorar os resultados.

6 REFERÊNCIAS

H. Oliveira, E. Ferreira, J. A. dos Santos, Conditional Domain Adaptation GANs for Biomedical Image Segmentation, arXiv preprint arXiv:1901.05553 (2019).

I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, Y. Bengio, Generative Adversarial Nets, in: NIPS, 2014, pp. 2672–2680.

GONZALEZ, R. C., & WOODS, R. E. *Digital image processing*. Upper Saddle River, N.J., Prentice Hall, 2002.