Intro to Deep Learning with applications in Medical Research

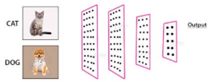
Jasper Zhang
Health Data Working Group

October 16, 2023

Health Data Working Group

- The Health Data Working Group is an initiative led by members of the Dalla Lana School of Public Health community to facilitate the uptake of programming-related tools for the benefit of public health research.

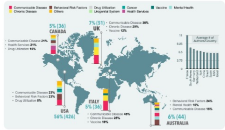
Workshops



Intro to Deep Learning using Keras

Monday, April 24, 2023

Cats and Dogs Classification from Scratch



Visualizing Research Data

Monday, February 27, 2023

A Hands-On Session for Creating Engaging and Compelling Charts

quarto

Building your first website using R and RStudio

Monday, November 28, 2022

This workshop introduces some tools for building and publishing your own website directly from RStudio.

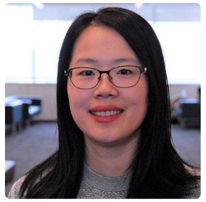
GitHub Collab

Monday, October 31, 2022

This workshop introduces Git concepts for purposes of collaborating with your colleagues.

Faculty Members

Aya Mitani



Kuan Liu

Osvaldo Espin-Garcia

Zahra Shakeri Hossein Abad

Student Members

Larry Dong

Alex Ji

Lorraine Lu

Jasper Zhang

Yushu Zou

Jasper Zhongyuan ZHANG
张 中原

PhD student at UofT
MSc. BMath.

Toronto, ON, Canada

Email

LinkedIn

Github

About Me

Hi, I'm **Jasper Zhang**, a **Biostatistics Ph.D. student** at the University of Toronto and University Health Network (UHN), under the supervision of **Dr. Wei Xu**. I also have the privilege to work as a **Research Analyst** at the Child Health Evaluative Sciences program, The Hospital for Sick Children (SickKids), with **Dr. Petros Pechlivanoglou**. Additionally, I'm a student member of the **Health Data Working Group** at the Dalla Lana School of Public Health.

My research spans areas like **Cancer Genomics, Survival Analysis, Deep Learning, and Health Econ**. For 2023-2024, I'm grateful to be supported by the **DLSPH Graduate Awards in Data Science for Public Health and Health Systems**.

I currently serve as the **Biostatistics section editor (since 2022)** for the University of Toronto Journal of Public Health (UTJPH) and was the **co-president (2022-2023)** of the Biostatistics Union of Graduate Students (BUGS).

I was fortunate to earn my **MSc in Biostatistics** from Toronto in 2021, where I was supervised by **Dr. Wei Xu** and **Dr. Pingzhao Hu**. Before that, I obtained my **B.Math** in **Computer Science and Statistics** from Waterloo in 2019. From 2018 to 2019, I worked with **Dr. Depeng Jiang** at Manitoba, a period during which my passion for health science research blossomed.

Beyond academics, music has always been a source of joy for me. I sing as a **Tenor** with the **University of Toronto Allegro Choir**. Our most recent performance was in April 2023 at Hart House, Toronto. If you're interested, it's available for viewing on bilibili and YouTube.

Outline of the Workshop

Part 1:

- Intro to Deep Learning and Image data
- Coding Task 1: Cats and Dog Classification CNN

Part 2:

- Deep Learning in Medical Research Applications
- Q and A

Components of Workshop

- Introduction to Basic Knowledge in Deep Learning
- Sample Interview Questions for Machine Learning Engineer Positions
- Hands-on Code Example to Demonstrate Practical Implementation
- Strategies for Adapting Deep Learning Techniques to Image Data in Your Research

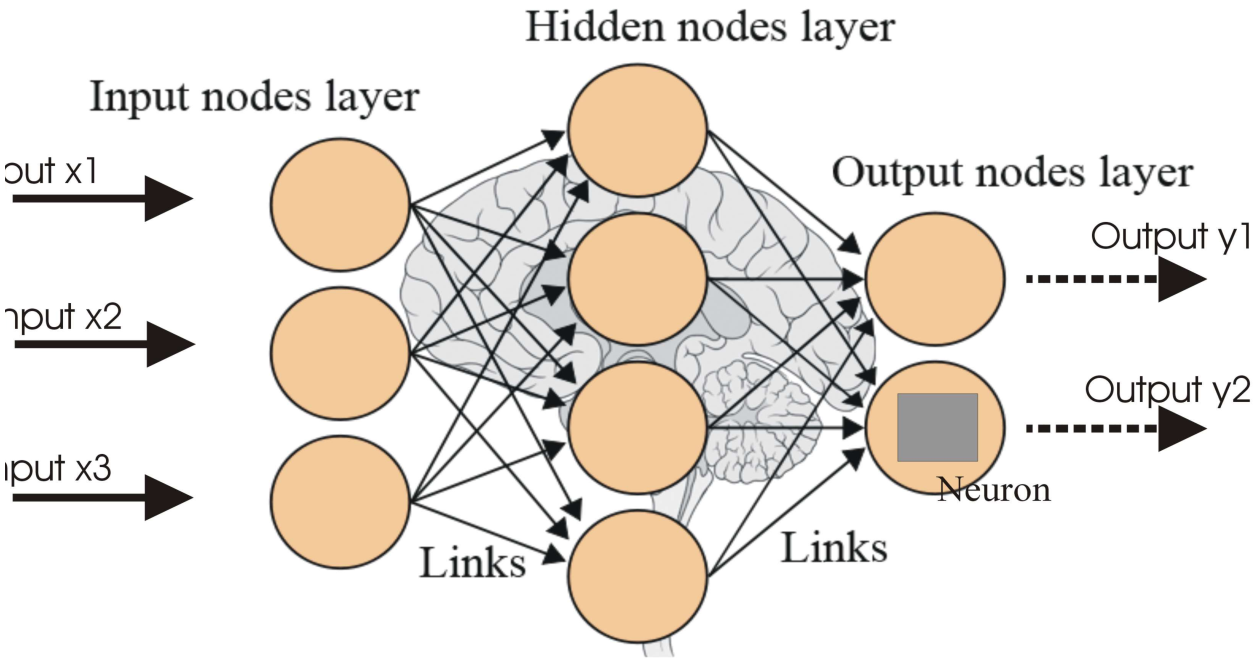
Interview Questions

Part 1: Intro to Deep Learning

Introduction to Deep Learning

- Deep learning, A subset of machine learning
- Inspired by the structure and function of the human brain
- Automatic feature learning
- Traditional machine learning: hand-engineered features by domain expertise, time consuming.

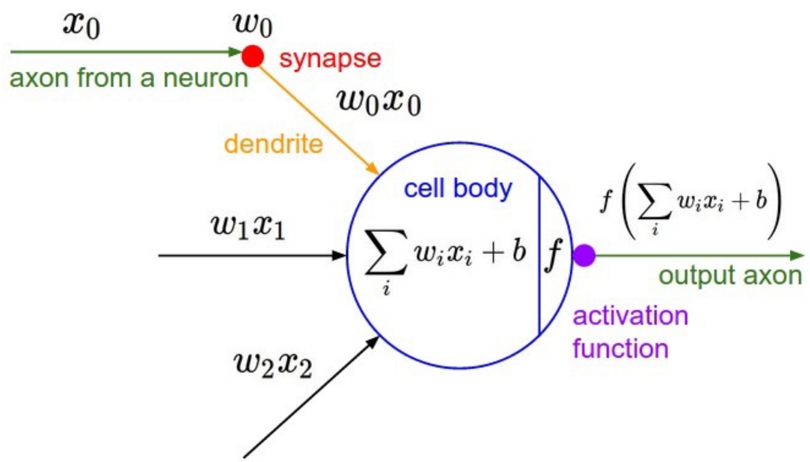
What is Deep Learning?
How it differs from Machine Learning?



Activation Function

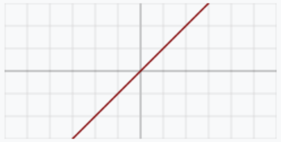
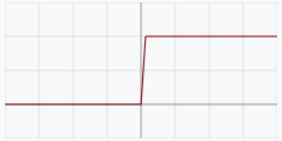
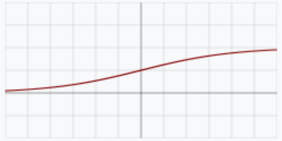
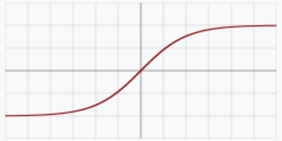
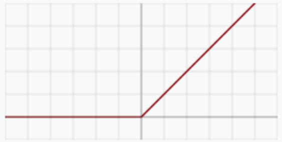
What is activation function?

- Activation functions decide the output of a neural network neuron based on its inputs.
- They add non-linearity to the network, empowering it to learn from complex and intricate data patterns.
- The choice of the activation function, such as Sigmoid, ReLU, or Tanh, greatly impacts the learning capability and performance of the neural network.



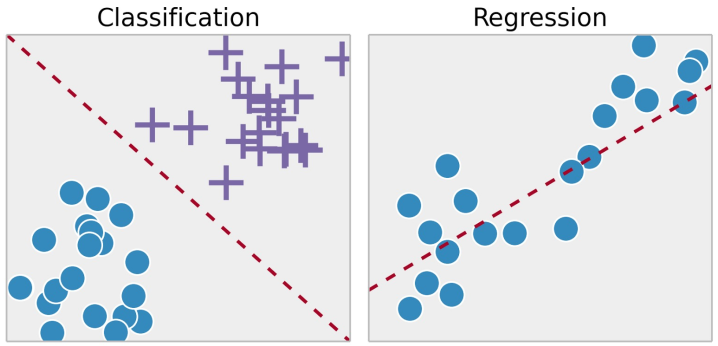
Common Activation Function

Difference Between sigmoid and tanh?

Name	Plot	Function, $f(x)$	Derivative of f, $f'(x)$	Range
Identity		x	1	$(-\infty, \infty)$
Binary step		$\begin{cases} 0 & \text{if } x < 0 \\ 1 & \text{if } x \geq 0 \end{cases}$	$\begin{cases} 0 & \text{if } x \neq 0 \\ \text{undefined} & \text{if } x = 0 \end{cases}$	$\{0, 1\}$
Logistic, sigmoid, or soft step		$\sigma(x) = \frac{1}{1+e^{-x}}$ [1]	$f(x)(1 - f(x))$	$(0, 1)$
tanh		$\tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}$	$1 - f(x)^2$	$(-1, 1)$
Rectified linear unit (ReLU)[11]		$\begin{cases} 0 & \text{if } x \leq 0 \\ x & \text{if } x > 0 \end{cases} = \max\{0, x\} = x\mathbf{1}_{x>0}$	$\begin{cases} 0 & \text{if } x < 0 \\ 1 & \text{if } x > 0 \\ \text{undefined} & \text{if } x = 0 \end{cases}$	$[0, \infty)$

Supervised Learning

What is Supervised Learning?



- Supervised Learning Task
  - Known labels/outcome
- Classification vs Regression
  - Categorical Target Variable: filter incoming emails, Spam or not
  - Continuous Target Variable: housing price

Deep Learning Tools

Do you have any experience using the DL framework?

- What are Deep Learning Tools? These are software libraries and frameworks that streamline the process of creating, training, and implementing deep learning models by providing pre-coded functions, algorithms, and high-level interfaces.
- Why Do We Need Them? Deep learning tools make the model-building process more accessible by handling lower-level details, optimizing computations, and offering GPU acceleration and distributed computing capabilities for large-scale models.

What is Keras?

Keras

- High-level neural networks API
- Easy-to-use and beginner-friendly
- Tensorflow backend

Strength of Deep Learning Python Library

Coding Task 1: Cats and Dog Classification using CNN

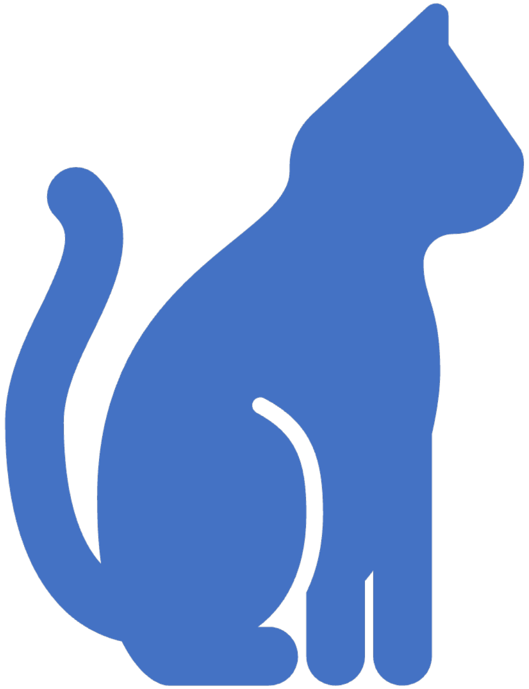


Image Classification Tasks

What is image classification, and can you provide some practical examples of its applications?

- Assigning labels to images
- Applications: object recognition, facial recognition, medical imaging

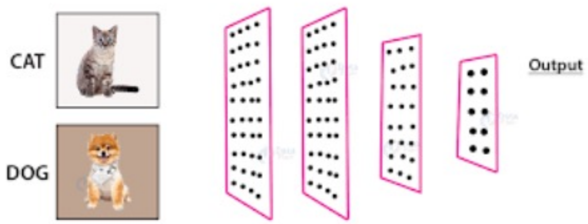


Image Data

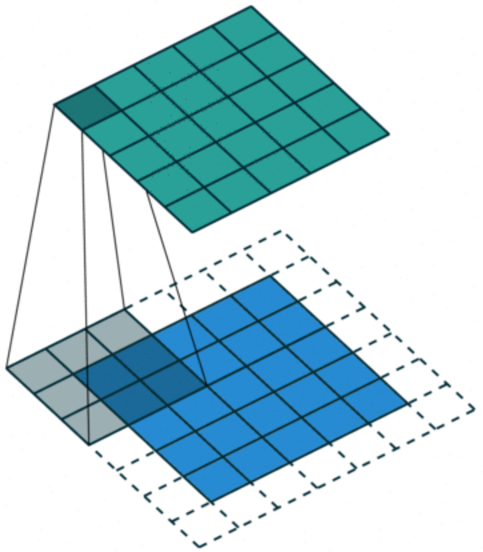
What's the difference between image data and tabular data?

- RGB Scale: Red Green Blue
  - 3 Channels
  - [0-255]
  - Example.
    - An image with 200 pix length and 100 pix height
    - $200 \times 100 \times 3 = 60{,}000$ entries
- Other Scale: Hue Saturation Lightness

Convolutional Neural Networks

- Convolutional Neural Networks (CNNs) are a type of deep learning architecture designed to process grid-like data, such as images, by using specialized layers like convolutional and pooling layers.
- CNNs excel at **learning spatial features** and are highly effective for image recognition, classification, and computer vision tasks.
- Example
  - Input 5 x 5
  - Kernel 3 x 3
  - Padding: 1
  - Stride:1
  - **Output: 5 x 5**
- O = (W - K + 2 * P) / S + 1   5 = (5 – 3 + 2 * 1) + 1

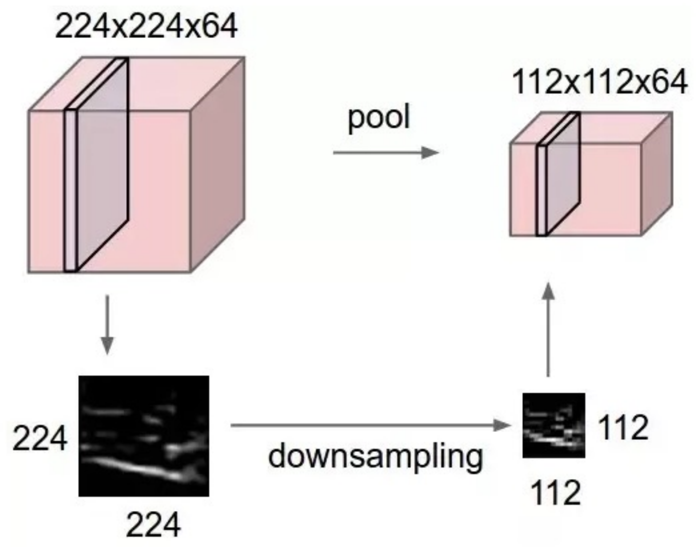
What is CNNs?



Pooling Layer

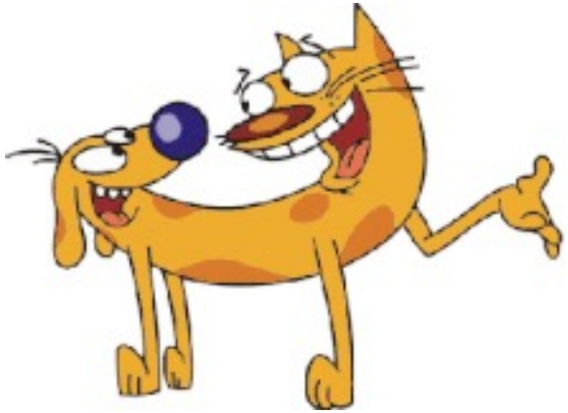
- **Dimensionality Reduction:** Pooling reduces the spatial size of the input, lowering computation, memory usage, and parameter count to help prevent overfitting.
- **Translation Invariance:** Pooling ensures feature detection is invariant to their position in the image, increasing model robustness to spatial transformations.
- **Feature Extraction:** Pooling helps extract dominant features, thereby contributing to the feature learning capacity of CNNs.

Pooing layer vs Convolution layer



Cats and Dogs Classification Keras Example

- Binary classification task
- Challenges: variations in appearance
- Precautions: data quality, class imbalance



Dataset Introduction

- Cats and dogs images
- Dataset source: Kaggle
- Format: labeled folders, JPEG images

1

0

Data Preprocessing

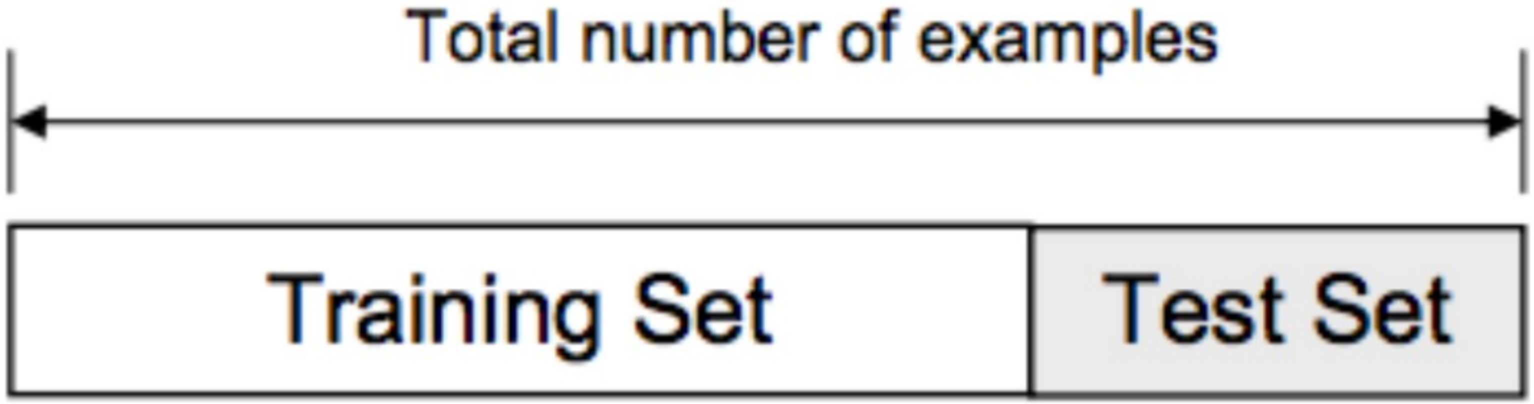
What is the purpose of image preprocessing in deep learning?

- Data cleaning: removing duplicates, corrupted images
- Data augmentation: rotation, flipping, zooming
- Resizing and normalizing images

Training, Testing, and Evaluation

- Splitting data: training and testing sets
- Evaluation metrics: accuracy

Why do you need to split train, test sets?



Batch size and Epoch

Epoch vs iterations?

- **The batch size** is the number of training examples used in one iteration of model training. The selection of batch size can impact both the speed and the stability of the learning process.
- **An epoch** is one complete pass through the entire training dataset. The number of epochs is a hyperparameter that determines how many times the learning algorithm will work through the entire training dataset.

Code Example

- Building the neural network with Keras
- Compiling the model
- Training and evaluating the model

Summary of the Keras Cats and Dogs Code Example

- Recap of the code walkthrough
- Model performance
- Potential improvements

This image is 99.80% cat and 0.20% dog.

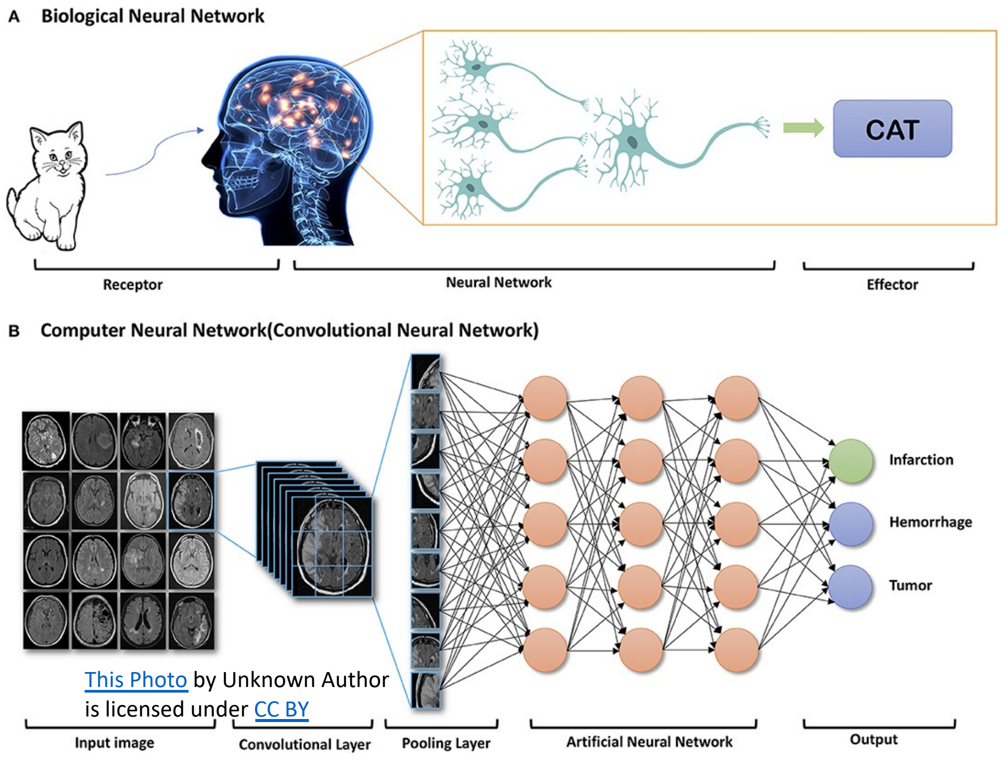
This image is 69.31% cat and 30.69% dog.

Part 2: Deep Learning in Medical Research Applications

Image Data in Medical Research

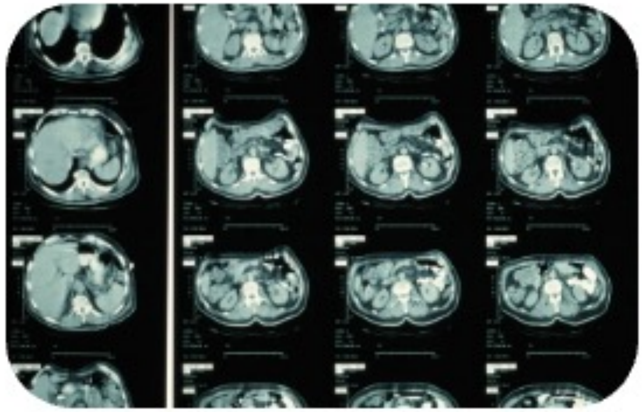
- Medical imaging techniques: MRI, CT scans, X-rays
- Disease detection and diagnosis
- Patient monitoring

CNNs applications in medical research



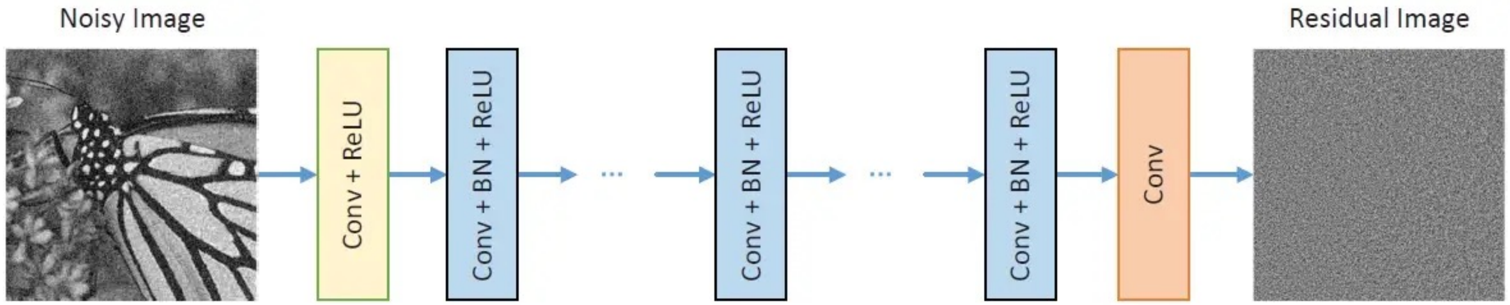
Low Dose Computed Tomography Denoising

- Course Project by Zou and Zhang in 2022 Fall.
- Computed Tomography (CT) is a medical screen tool using X-ray technology to produce high-quality cross-sectional images of the body.
- **Radiation toxicity** of CT scans can cause radiation syndrome, radiation poisoning and increases the risk of developing fatal cancer.
- Compared with Full-dose CT (FDCT) at a regular dose level**, Low-dose CT (LDCT)** lowers the radiation exposure significantly. However, LDCT scans are often too noisy for diagnostic use.
- Apply computational denoising methods in multiple approaches to increase the rate of LDCT adoption.

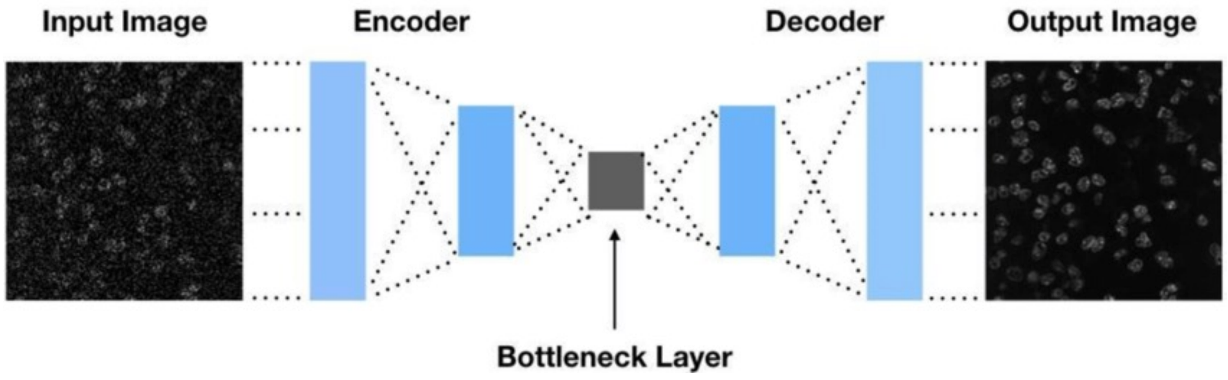


LDCT Method

DnCNN



AutoEncoder



LDCT Result

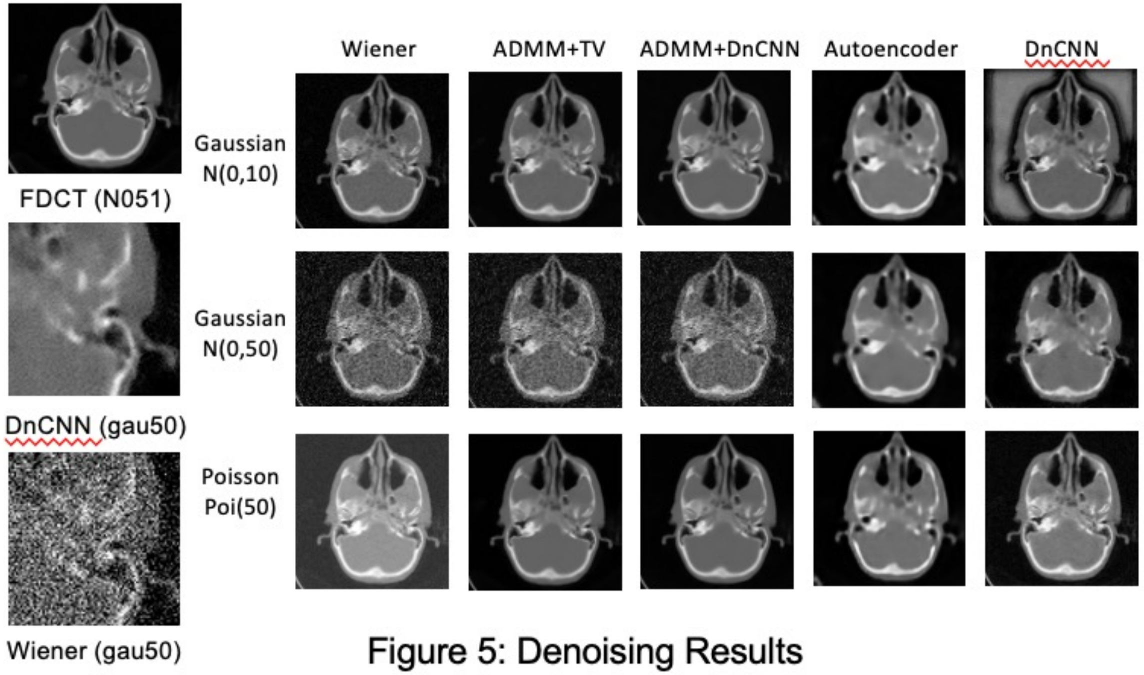


Figure 5: Denoising Results

- **PSNR**

Noise Type	Wiener	ADMM+TV	ADMM+DnCNN	Autoencoder	DnCNN
Gaussian: N(0,10)	**47.2**	30.89	29.23	29.79	15.57
Gaussian: N(0,50)	**47.29**	26.05	29.03	28.67	27.75
Poisson: Poi(50)	14.14	14.16	14.18	**29.94**	27.79

- **MSE**

Noise Type	Wiener	ADMM+TV	ADMM+DnCNN	Autoencoder	DnCNN
Gaussian: N(0,10)	**0.00002**	0.00082	0.00120	0.00114	0.03034
Gaussian: N(0,50)	**0.00002**	0.00248	0.00125	0.00147	0.00170
Poisson: Poi(50)	0.03858	0.03839	0.03815	**0.00109**	0.00167

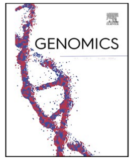
CNNs in Genomics and Microbiome Data

Genomics 113 (2021) 2023–2031

Contents lists available at ScienceDirect

Genomics

journal homepage: www.elsevier.com/locate/ygeno



ELSEVIER

Check for updates

Original Article

ChrNet: A re-trainable chromosome-based 1D convolutional neural network for predicting immune cell types

Shuo Jia [a], Pingzhao Hu [a,b,c,*]

[a] Department of Biochemistry and Medical Genetics, University of Manitoba, Winnipeg, MB, Canada
[b] Department of Computer Science, University of Manitoba, Winnipeg, MB, Canada
[c] Research Institute in Oncology and Hematology, CancerCare Manitoba, University of Manitoba, Winnipeg, MB, Canada

Bioinformatics, 36(17), 2020, 4544–4550
doi: 10.1093/bioinformatics/btaa542
Advance Access Publication Date: 25 May 2020
Original Paper

OXFORD

Genome analysis

TaxoNN: ensemble of neural networks on stratified microbiome data for disease prediction

Divya Sharma[1], Andrew D. Paterson[1,2] and Wei Xu[1,3,*]

[1]Division of Biostatistics, Dalla Lana School of Public Health, University of Toronto, Toronto, ON, Canada M5T 3M7, [2]Genetics and Genome Biology Program, The Hospital for Sick Children, Toronto, ON, Canada, M5G 1X8 and [3]Department of Biostatistics, Princess Margaret Cancer Center, University Health Network, Toronto, ON, Canada, M5G 2C1

*To whom correspondence should be addressed.

Associate Editor: Pier Luigi Martelli

Received on March 10, 2020; revised on May 8, 2020; editorial decision on May 18, 2020; accepted on May 19, 2020

Data and Outcomes

Genomics:

- Gene Expression (scRNA--seq)
- Cell type multi-class **classification**

- Microbiome:
  - microbial Operational Taxonomic Units (OTUs)
  - Type 2 diabetes (T2D) binary **classification**

What are the Spatial Correlations?

Genomics - ChrNet Approach:

- Input: Gene by cell matrix with TPM (Transcripts Per Million) expression values.
- Gene Alignment: Values are aligned based on their linear positions on the human genome reference (hg19). Early starting position genes are prioritized when considering overlaps.
- Key Insight: Genes adjacent on the same chromosome exhibit similar expression patterns - beneficial for feature extraction.

Microbiome - TaxoNN Approach:

- The basis of correlation: Operational Taxonomic Units (OTUs) at the same taxonomy level inherently share correlations.
- Key Insight: Non-linear relationships exist between OTUs, especially evident during disease prediction.

ChrNet

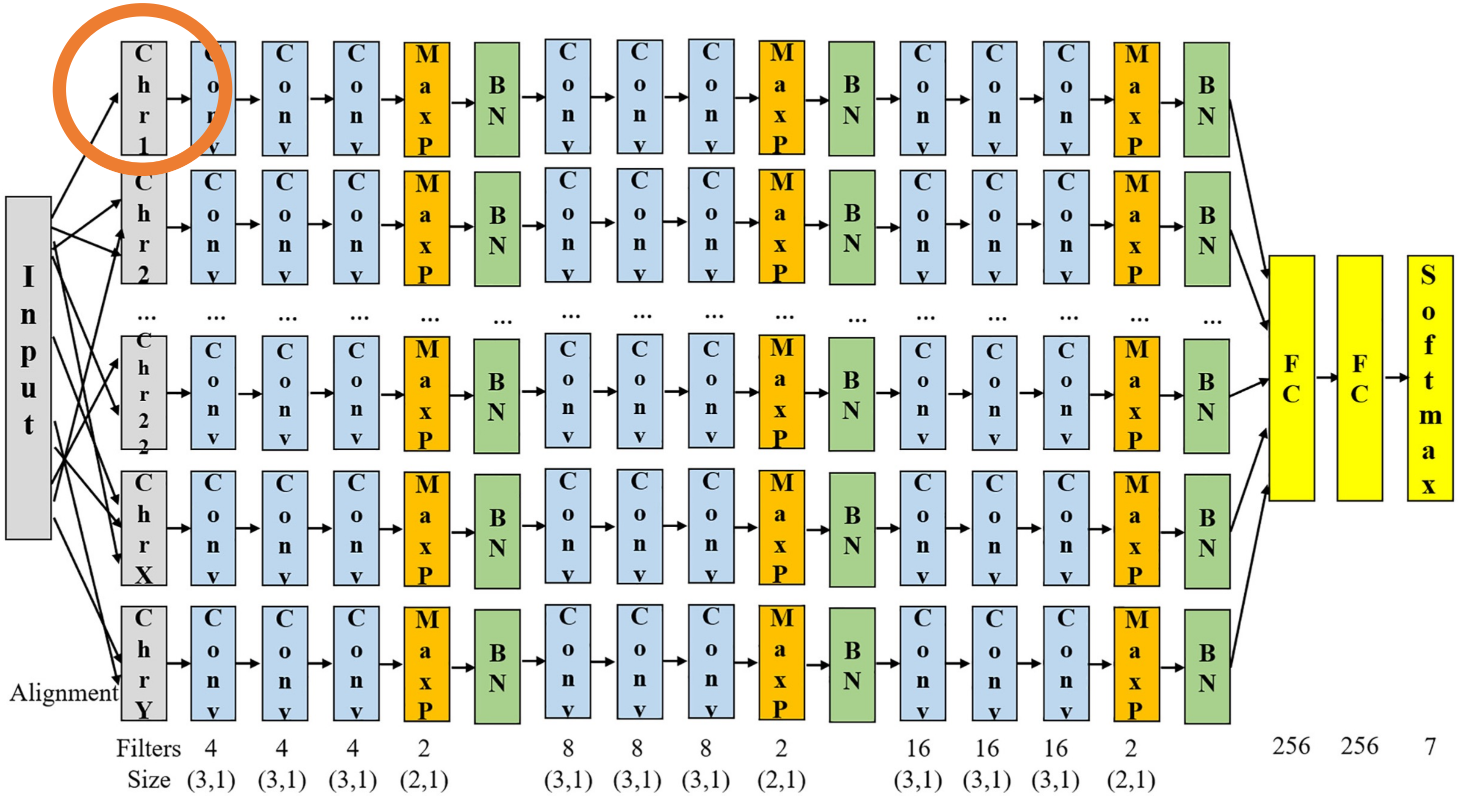


Fig. 1. The layout of ChrNet architecture. The input features are first selected and aligned to different chromosomes based on their genomic positions. 3 1D-CNN layers are stacked together, followed by a max pooling layer with a batch normalization layer. After the 3 similar structures, the net goes through two dense layers and then makes prediction for the cell types. Filters indicate the number of filters in that layer and size indicates the size of the filters in that layer.

TaxoNN

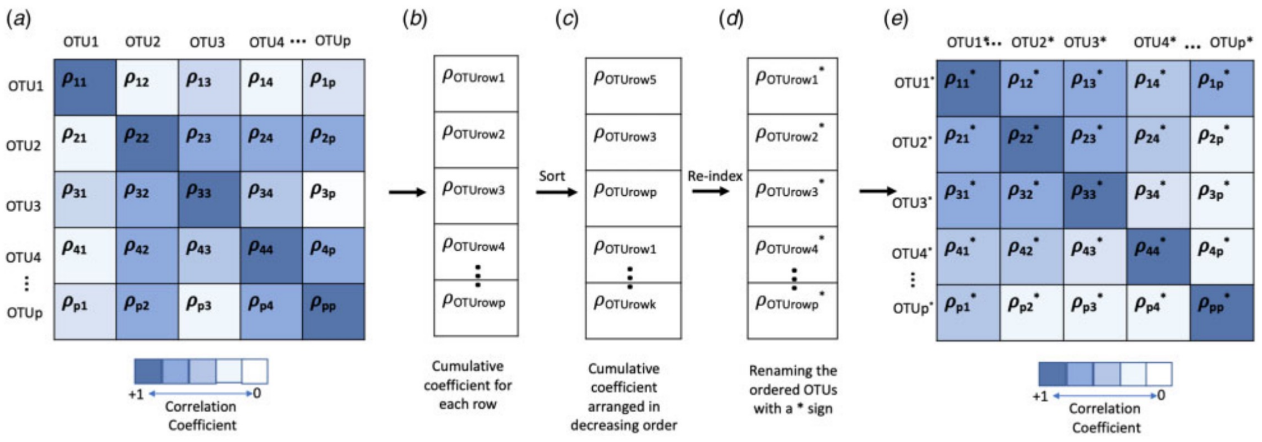
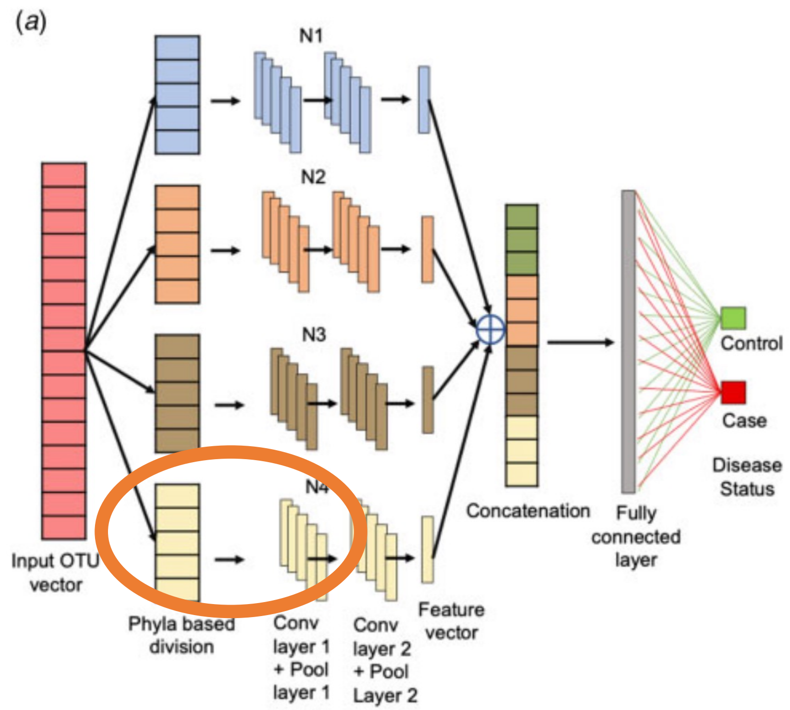


Fig. 1. An illustration of correlation-based ordering in the OTUs in a cluster. (**a**) Example heatmap obtained by plotting Spearman rank coefficients between positively correlated OTUs in a cluster. (**b**) Cumulative coefficient obtained with respect to each row of the heatmap matrix. (**c**) Vector of cumulative coefficients arranged in a decreasing order where, $\rho_{OTU_{row5}} > \rho_{OTU_{row3}} > \rho_{OTU_{rowp}} > \rho_{OTU_{row1}} > \rho_{OTU_{rowk}}$. (**d**) The cumulative coefficients are renamed as $\rho_{OTU^*_{row_i}}$ to represent that they are now arranged in a decreasing order. (**e**) Heatmap sorted based on the new order of cumulative coefficients, making the correlated terms concentrate in a space and arrange closer in the matrix



Build Your Deep Learning Skills

Math Skills: Calculus, Linear Algebra, Statistics

Programming Skills: Python, R, MATLAB, (SAS?)

Interview Questions list is a good guide to Learn key concepts

Research Projects

Dalla Lana School of Public Health

CHL5230H F

Applied Machine Learning for Health Data

Instructor: Dr. Zahra Shakeri

CHL7001H S7

Directed Reading: Introduction to Machine Learning for Big Data in Health Care

Instructor: Dr. Divya Sharma

Q and A

Acknowledgement

- Health Data Working Group, DLSPH, University of Toronto
- Xu Lab, Biostatistics Department, University Health Network
- Dr. Pingzhao Hu, The Visual and Automated Disease Analytics (VADA) graduate training program, University of Manitoba and University of Victoria.
- Keras team and code example authors: François Chollet, Amogh Joshi
- Cat: Melo

References

Chollet, F. (2022, November 10). Keras documentation: Image Classification From Scratch. Keras. https://keras.io/examples/vision/image_classification_from_scratch/

Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. Neural Computation, 9(8), 1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735

Jia, S., & Hu, P. (2021). ChrNet: A re-trainable chromosome-based 1d convolutional neural network for predicting immune cell types. Genomics, 113(4), 2023–2031. https://doi.org/10.1016/j.ygeno.2021.04.037

Joshi, A. (2021, June 5). Keras Documentation: Next-frame video prediction with convolutional lstms. Keras. https://keras.io/examples/vision/conv_lstm/

Parkhi, O. M., Vedaldi, A., Zisserman, A., & Jawahar, C. V. (2012). Cats and dogs. 2012 IEEE Conference on Computer Vision and Pattern Recognition. https://doi.org/10.1109/cvpr.2012.6248092

Ramachandran, P., Zoph, B., & Le, Q. V. (2017, October 27). Searching for activation functions. arXiv.org. https://arxiv.org/abs/1710.05941

Sharma, D., Paterson, A. D., & Xu, W. (2020). Taxonn: Ensemble of Neural Networks on stratified microbiome data for disease prediction. Bioinformatics, 36(17), 4544–4550. https://doi.org/10.1093/bioinformatics/btaa542

Sherstinsky, A. (2020). Fundamentals of Recurrent Neural Network (RNN) and long short-term memory (LSTM) network. Physica D: Nonlinear Phenomena, 404, 132306. https://doi.org/10.1016/j.physd.2019.132306