Research on Musculoskeletal Radiographs Abnormality Detection

Using Deep Learning

Shiva Prasad Kudikala (120cs0016)
Surya Teja Regalla (120cs0021)

Under the guidance of
Dr. K. Nagaraju
CSE Department

November 30, 2023

Indian Institute of Information Technology Design and Manufacturing
Kurnool, Andhra Pradesh, India.

Outline

Introduction

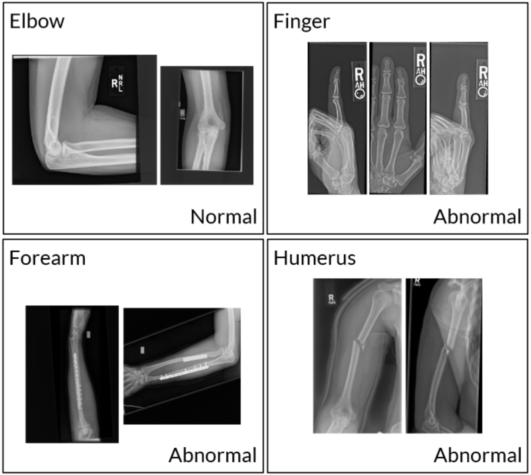
What are Musculoskeletal Disorders?

- Musculoskeletal disorders are conditions that affect the muscles, bones, tendons, ligaments, and other parts of the musculoskeletal system.
- These disorders can lead to pain, reduced mobility, and limitations in daily activities, affecting a person's overall quality of life.

What is Musculoskeletal Abnormality Detection?

- Musculoskeletal abnormality detection refers to the process of identifying deviations or anomalies in the structure of the musculoskeletal system through medical imaging like X-rays.
- It plays a crucial role in early detection and diagnosing musculoskeletal conditions like fractures, tumors, joint diseases.

Introduction



Source: MURA: Large Dataset for Abnormality Detection in Musculoskeletal Radiographs Research article [2]

Figure: Radiograph images from MURA dataset

Problem Statement

Statement

Given a dataset of X-ray images, let X represent the set of images and Y represent the corresponding labels, where $Y_i \in \{0, 1\}$ indicates whether the body part in image X_i is normal (Y_i=0) or abnormal (Y_i=1).

Literature Study

Paper-1 Name: MURA: Large Dataset for Abnormality Detection in Musculoskeletal Radiographs.(2018) [7]
Description:

- The paper introduces MURA dataset. The dataset is used to train a 169-layer DenseNet baseline model to detect and localize abnormalities.
- They have randomly chose 3 of their radiologists of gold standard and assessed the performance of both radiologists and their model on the test set.
- The model's performance is compared to that of radiologists using the Cohen's kappa [1] statistic.
- The radiologists achieved their highest performance on either wrist studies or humerus studies, and their lowest performance on finger studies. The model also achieved its highest performance on wrist studies and its lowest performance on finger studies.

Framework: 169-layer DenseNet

Results:

	Radiologist 1	Radiologist 2	Radiologist 3	Model
Elbow	0.850 (0.830, 0.871)	0.710 (0.674, 0.745)	0.719 (0.685, 0.752)	0.710 (0.674, 0.745)
Finger	0.304 (0.249, 0.358)	0.403 (0.339, 0.467)	0.410 (0.358, 0.463)	0.389 (0.332, 0.446)
Forearm	0.796 (0.772, 0.821)	0.802 (0.779, 0.825)	0.798 (0.774, 0.822)	0.737 (0.707, 0.766)
Hand	0.661 (0.623, 0.698)	0.927 (0.917, 0.937)	0.789 (0.762, 0.815)	0.851 (0.830, 0.871)
Humerus	0.867 (0.850, 0.883)	0.733 (0.703, 0.764)	0.933 (0.925, 0.942)	0.600 (0.558, 0.642)
Shoulder	0.864 (0.847, 0.881)	0.791 (0.765, 0.816)	0.864 (0.847, 0.881)	0.729 (0.697, 0.760)
Wrist	0.791 (0.766, 0.817)	0.931 (0.922, 0.940)	0.931 (0.922, 0.940)	0.931 (0.922, 0.940)
Overall	0.731 (0.726, 0.735)	0.763 (0.759, 0.767)	0.778 (0.774, 0.782)	0.705 (0.700, 0.710)

Figure: Comparing radiologists and their model on the Cohen's kappa statistic. They highlight the best (green) and worst (red) performances on each of the study types and in aggregate.

Paper-2 Name: Person Re-Identification With Triplet Focal Loss (2018) [10]

Description:

- In this paper, motivated by focal loss designed for the classification model, they propose the triplet focal loss for person Re-ID.
- So they have initially used triplet loss where it takes an image triplet as input, which is called anchor sample, positive sample and negative sample respectively.
- They have later used batch hard triplet loss, which chooses the hardest positive and negative sample in the mini-batch.
- They later proposed Triplet Focal Loss which can up-weight the hard triplets training samples and relatively down-weight the easy triplets adaptively.

Literature Study

Paper-3 Name: Hard negative examples are hard, but useful (2020) [9]
Description:

- So in their previous research they found out that optimizing with the hardest negative examples leads to bad training behavior. These hardest negatives are literally the cases where the distance metric fails to capture semantic similarity.
- In this paper, they have derived why hard negatives make triplet loss training fail. And they have offered a simple fix to the loss function and with this fix optimizing with hard negatives examples becomes feasible which lead to more generalizable features.
- To avoid this bad effect on training, authors have developed alternative approaches, such as semi-hard triplet mining [8].

Literature Study

Loss function

The loss function they have used is, as shown in equation below

$$\mathcal{L}(S_{\text{ap}}, S_{\text{an}}) = -\log\left(\frac{e^{S_{\text{ap}}}}{e^{S_{\text{ap}}} + e^{S_{\text{an}}}}\right)$$

where S_{ap} is a similarity metric of anchor-positive pair, and S_{an} of anchor-negative pair.

The solution for the challenge with hard negative triplets is to decouple them into a-p pairs and a-n pairs, and ignore the a-p pairs, and introduce a contrastive loss that penalizes the a-n similarity. They named this Selectively Contrastive Triplet loss L_{SC}, and defined this as follows:

$$L_{SC}(S_{\text{ap}}, S_{\text{an}}) = \begin{cases} \lambda S_{\text{an}}, & \text{if } S_{\text{an}} > S_{\text{ap}} \\ L(S_{\text{ap}}, S_{\text{an}}), & \text{otherwise} \end{cases}$$

Limitations

- Finger part has the least model performance scores compared to other parts.
- The model performance was lower than the worst radiologist's performance in some body parts.

Motivation

Origin of our problem statement:

- Radiologists worldwide are reading an increasing number of cases and problems are even worse for those in underserved communities.
- Determining whether a study is normal / abnormal is a critical task for radiologists.

Motivation:

- A tool that can highlight the abnormality in an image can draw the attention of the radiologist, which could potentially reduce errors, speed up image interpretation.
- It can be used for worklist prioritization.
- We make the scenario where cases that seemed abnormal can be prioritized so that sickest patients get diagnosis first.

Work Plan

Division of Work :

Name	Roll No.	Role
K. Shiva Prasad	120cs0016	1. Literature Study 2. ResNet50 Implementation 3. Triplet Loss 4. Triplet Focal Loss
R. Surya Teja	120cs0021	1. Literature Study 2. DenseNet-169 Implementation 3. CLAHE Preprocessing 4. Batch Hard Triplet Loss 5. Hard Examples Training

Approach

Objective

The primary objective of this project is to enhance the performance of abnormality detection in Finger MuRA dataset using a Siamese network architecture by studying and implementing various loss functions and some preprocessing techniques.

Dataset : MURA

- MURA dataset is collected from the Stanford Hospital. The dataset consists of 40,561 multi-view radiographs between 2001 to 2012. Each study was manually labeled as normal or abnormal by board-certified radiologists from the Stanford Hospital at the time of clinical radiographic interpretation.

Study	Train		Validation		Total
	Normal	Abnormal	Normal	Abnormal	
Elbow	1,094	660	92	66	1,912
Finger	1,280	655	92	83	2,110
Hand	1,497	521	101	66	2,185
Humerus	321	271	68	67	727
Forearm	590	287	69	64	1,010
Shoulder	1,364	1,457	99	95	3,015
Wrist	2,134	1,326	140	97	3,697
Total number of studies	8,280	5,177	661	538	14,656

Table: MURA Dataset[7]

CLAHE Pre-processing

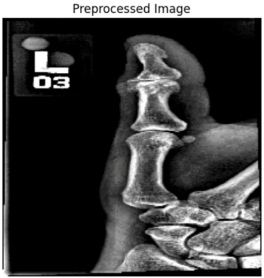
Contrast Limited Adaptive Histogram Equalization (CLAHE) [6] is used to enhance the contrast of an image by changing values in the intensity of it, and improve the visibility of details [5]. It is particularly effective in **medical imaging**, where subtle details in X-rays or MRIs may be crucial.

Procedure

- The procedure begins by converting the input image from the RGB color space to the LAB.
- CLAHE is applied specifically to the L channel, representing the lightness of the image. The processed L channel is merged back with the original A and B channels.
- The resulting LAB image with enhanced contrast is then converted back to the RGB color allowing for adaptive contrast enhancement while preserving color information.

CLAHE Pre-Processing

[Normal]



[CLAHE]

[Normal]

[CLAHE]

Approach

DenseNet-169

- DenseNet-169 [3] is named after its dense connectivity pattern.DenseNet connects each layer to every other layer in a feedforward fashion.
- DenseNet-169's dense blocks facilitate the extraction of rich and hierarchically organized features. In DenseNet, each layer's output is concatenated with the outputs of all previous layers, creating a dense feature map.
- DenseNet-169 achieves remarkable performance while being relatively parameter-efficient.
- DenseNet-169 can handle variations in the input data.

Approach

DenseNet-169

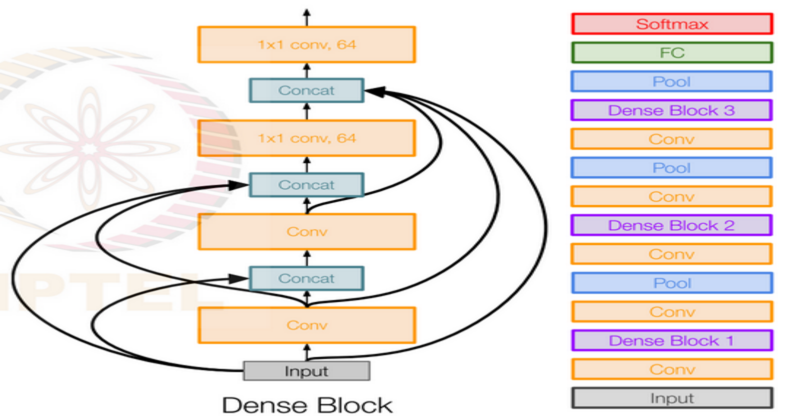


Figure: Dense Block and a 3-layer Dense Architecture

Source: NPTEL course: Deep Learning for Computer Vision.

Approach

A Siamese network is a type of neural network architecture consists of two identical subnetworks (twins) sharing the same parameters. These networks process input data independently and produce embeddings, which are then compared to measure the similarity between the inputs.

Siamese Networks

- We will measure the distance between these two vectors and if the distance between these is small then the vectors are similar or of the same classes and if the distance between is larger then the vectors are different from one another, based on the score.
- By connecting the output embeddings to a classifier layer, the Siamese network becomes versatile, enabling not only similarity evaluations but also effective classification.

Approach

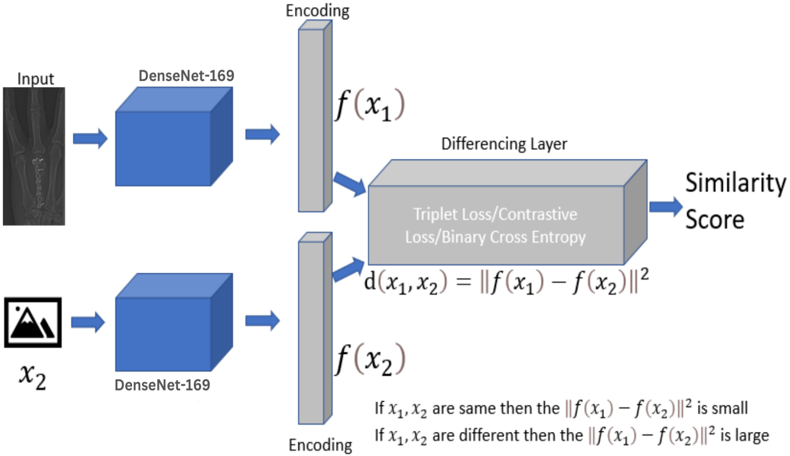


Figure: Siamese Network Architecture (medium.com)

Triplet loss function

- In the context of a Siamese network, which has twin subnetworks sharing parameters, triplet loss involves training the network to minimize the distance between an anchor sample and a positive sample while maximizing the distance between the anchor and a negative sample.
- This encourages the network to map similar samples close together in the embedding space and dissimilar samples farther apart.

Loss Function

$$L_{\text{triplet}}(A, P, N) = \max\left(0, \|f(A) - f(P)\|^2 - \|f(A) - f(N)\|^2 + \alpha\right)$$

where A, P, and N represent the anchor, positive, and negative samples, respectively.
$f()$ represents the mapping of a sample to its embedding in the Siamese network, and α is a margin.
$\|.\|$ denotes the Euclidean distance.

Approach

Other Loss Functions

- **Batch Hard Triplet Loss :** Batch Hard Triplet Loss combines the triplet generation step with the training process and to mine the hard triplet samples within each mini-batch.
- Each image in the mini-batch is in turn treated as an anchor sample, Batch Hard Triplet Loss tries to choose the hardest positive and negative sample in the mini-batch.

$$\mathcal{L}_{\mathsf{BHTL}}(\theta) = \sum_{i=1}^{P} \sum_{a=1}^{K} \max(0, D^*_{a,p} - D^*_{a,n} + m)$$

where $D^*_{a,p}$ and $D^*_{a,n}$ denotes the hardest positive and negative sample corresponding to anchor sample x_a .

Approach

Other Loss Functions

- **Triplet Focal Loss:** Triplet Focal Loss which can up-weight the hard triplets training samples and relatively down-weight the easy triplets adaptively.
- The formulation of Triplet Focal Loss based on the Batch Hard Triplet Loss is as follows,

$$\mathcal{L}_{\mathsf{TFL}}(\theta) = \sum_{i=1}^{P} \sum_{a=1}^{K} \max\left(0, \exp\left(\frac{D^{*}_{a,p}}{\sigma}\right) - \exp\left(\frac{D^{*}_{a,n}}{\sigma}\right) + m\right)$$

where σ is a constant used to control the spread of the exponential function.

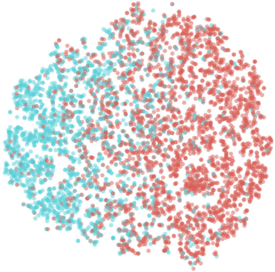
Observed Results

Output-Layer	ResNet50		DenseNet169		DenseNet169-CLAHE	
	Accuracy	Val-acc	Accuracy	Val-acc	Accuracy	Val-acc
256-Layer	0.635	0.496	0.929	0.722	0.9375	0.780
128-Layer	0.678	0.668	0.946	0.744	0.9508	0.752

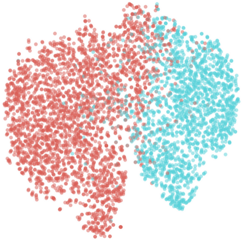
Table: Accuracy of classifier models

Output-Layer	ResNet50		DenseNet169		DenseNet169-CLAHE	
	Train-score	Valid-score	Train-score	Valid-score	Train-score	Valid-score
256-Layer	0.053	0.082	0.851	0.449	0.868	0.504
128-Layer	0.093	0.333	0.887	0.491	0.896	0.476

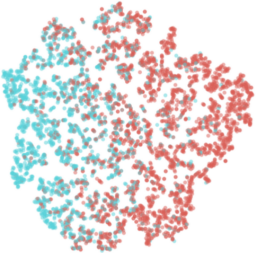
Table: Cohen Kappa Scores



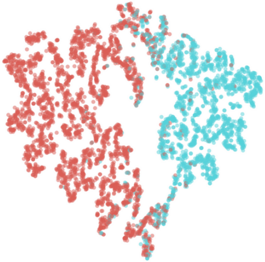
[Densenet-256]



[CLAHE]



[Densenet-128]



[CLAHE]

Observed Results

Triplet Focal Loss - DenseNet169				
Output-Layer	Acc	Val-acc	Cohen-kappa Train Score	Cohen-kappa Val Score
256-Layer	0.746	0.679	0.403	0.374
128-Layer	0.719	0.692	0.408	0.374

Table: Results of Triplet focal loss

Our Model Design [4]

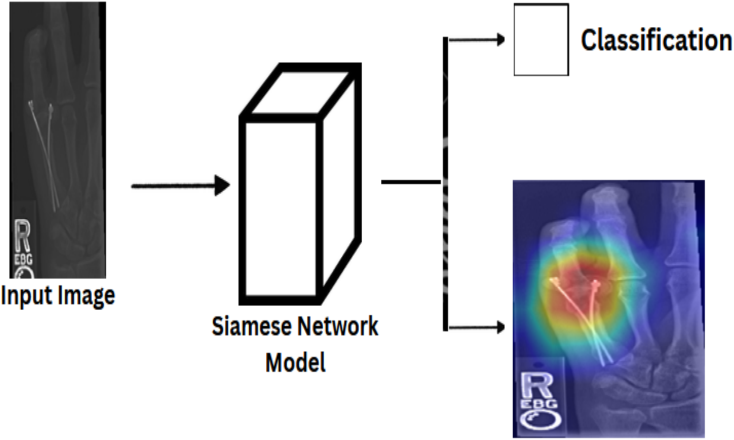
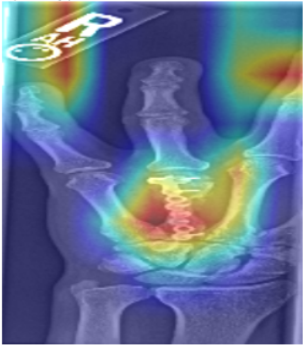


Figure: Representation of our model in a flow chart

```
(1, 224, 224, 3)
Smooth GRAD CAM
tf.Tensor([[0.33369687 0.6663031 ]],
Predicted Label: positive
50
(1664,)
```




```
(1, 224, 224, 3)
Smooth GRAD CAM
tf.Tensor([[0.32461086 0.6753891 ]],
Predicted Label: positive
50
(1664,)
```

Figure: Test case images of class positive

Test Cases

Preprocessed Image

+ Code   + Markdown

```
pred = classifier_model.predict(dense_model.predict(processed_image))
```

```
1/1 [==============================] - 0s 33ms/step
1/1 [==============================] - 0s 17ms/step
```

```
pred
```

```
array([[0.98739654, 0.01260341]], dtype=float32)
```

Figure: Test case images of class negative

Future Work

- Focused on abnormality detection in finger images, can be extended to images of different body parts present in the MuRA dataset.
- Can use HEF, UM preprocessing techniques to improve the model.

Scope:

- Domain Transfer of data.
- Dataset can be made diverse.

Bibliography I

[1] Namit Chawla and Nitika Kapoor. "Musculoskeletal Abnormality Detection in Humerus Radiographs Using Deep Learning.". In: *Rev. d'Intelligence Artif.* 34.2 (2020), pp. 209–214.

[2] Minliang He, Xuming Wang, and Yijun Zhao. "A calibrated deep learning ensemble for abnormality detection in musculoskeletal radiographs". In: *Scientific Reports* 11.1 (2021), p. 9097.

[3] Gao Huang et al. "Densely connected convolutional networks". In: (2017), pp. 4700–4708.

[4] Goodarz Mehr. *Automating Abnormality Detection in Musculoskeletal Radiographs through Deep Learning*. 2020. arXiv: 2010.12030 [eess.IV].

[5] Thi Phuoc Hanh Nguyen et al. "Pre-processing Image using Brightening, CLAHE and RETINEX". In: *arXiv preprint arXiv:2003.10822* (2020).

[6] Raul David Palma Olvera et al. "A feature extraction using SIFT with a preprocessing by adding CLAHE algorithm to enhance image histograms". In: *2014 International Conference on Mechatronics, Electronics and Automotive Engineering*. IEEE. 2014, pp. 20–25.

Bibliography II

[7] Pranav Rajpurkar et al. "MURA: Large Dataset for Abnormality Detection in Musculoskeletal Radiographs". In: (2018). arXiv: 1712.06957 [physics.med-ph].

[8] Florian Schroff, Dmitry Kalenichenko, and James Philbin. "Facenet: A unified embedding for face recognition and clustering". In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2015, pp. 815–823.

[9] Hong Xuan et al. "Hard negative examples are hard, but useful". In: *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XIV 16*. Springer. 2020, pp. 126–142.

[10] Shizhou Zhang et al. "Person Re-Identification With Triplet Focal Loss". In: *IEEE Access* 6 (2018), pp. 78092–78099. DOI: 10.1109/ACCESS.2018.2884743.

Thank You!