International Journal of Research in Engineering and Applied Sciences(IJREAS)
Available online at http://euroasiapub.org/journals.php
Vol. 13 Issue 10, October -2023,
ISSN(O): 2249-3905, ISSN(P) : 2349-6525 | Impact Factor: 8.202| Thomson Reuters ID: L-5236-2015

Learning Disease-Relevant Representations from Visual Similarity Triplets in Chest CT Scans Using Convolutional Neural Networks

Patil Manjusha Manikrao
(Research Scholar)
Dr. R. Purushotham Naik (Professor)
(Research Supervisor)
Glocal School Of Technology and Computer Science

Abstract

The use of convolutional neural networks (CNNs) for supervised feature learning improves the speed and accuracy of disease-specific medical picture representation. On the other hand, labelled images are essential for convolutional neural network (CNN) training. When it comes to medical picture annotations, the process is laborious and full of controversy among experts. The degree of similarity between images could be more useful than illness identification or severity grading. This has the power to lower rater variability, making it easier for non-experts to participate, and reveal new patterns. It can be important to review chest computed tomography images for signs of emphysema. A convolutional neural network (CNN) is employed to construct a low-dimensional embedding after visually evaluating the amount of emphysema to provide visual similarity triplets. Our findings show that CNNs trained with 973 photos may successfully learn disease-related feature representations using generated similarity triplets. Similarity triplets provide a feature format for embedding unseen test pictures, making them the first of their type in medical picture applications.

Keywords: Attribute Extraction; Similarity or Proximity Triplets; Emphysema Evaluation

1. Introduction

Convolutional neural networks (CNNs) have shown encouraging achievements in medical image processing in recent years. The necessity for annotated picture data poses a substantial challenge to CNN training. It takes a lot of expertise, time, and effort to annotate medical photos. Annotations might vary greatly not just between but even within raters (Chollet, 2021). Task specification, time restrictions, and rater competency are some of the reasons why rater annotations might vary. Annotations such as "segment the tumor," "count nodules," or "assess emphysema extent" sometimes involve assigning absolute assessments. Psychological studies have shown that when given the option to rank items relative to one another instead of in absolute terms, people perform better (Jones, & Wheadon, 2015). If annotation assignments were updated to include relative comparisons, rater agreement may be enhanced.

As a general rule, the total lung capacity is visually assessed to determine the severity of emphysema (Ash, et al 2021). Instead of visually comparing several volumes, which might result in worse rater agreement, it is more beneficial to evaluate each 3D lung volume separately. 2D slice visual similarity comparisons are simpler and may be performed by non-experts with sufficient training. It is possible to save time and money by having crowdworkers act as medical experts. According to Juan et al. (2020), crowdsourced photo similarity has been useful for classifying birds, categorizing foods, and evaluating emphysema.

Lately, there has been a flurry of study on learning from similarities gained by absolute labeling. In their study, Cahyono, Wirawan, and Rachmadi showed that learning from similarities rather than labels alone can lead to better results (2020). “*These works use triplet learning to learn from visual image similarity when ratings for a triplet of pictures* (x_i, x_j, x_k) *are given in the manner of* x_i *is more similar to* x_j *than to* x_k”

2. Methodology

In order to learn a mapping from source pictures to a low-dimensional representation that mimics the characteristics of visual similarity metrics, this research sets out the parameters of the triplet learning issue and proposes a CNN based approach.

The Language Difficulty of the Triplets

Let $\mathbf{X}$ be an image space and let $x_i \in \mathbf{X}$.. A triplet of images (x_i, x_j, x_k)that aligns with the triplet criterion, where

$$\delta(x_i, x_j) \leq \delta(x_i, x_k) \ldots\ldots (1.1)$$

Equation (1.1) uses the undetermined value of δ, which is the measure of dissimilarity. Let us pretend that the set of ordered triplets $\mathbf{T} \subseteq \mathbf{X}^3$. satisfies (1.1). In order to minimize the anticipated number of violated triplets, a mapping $h^*: \mathbf{X} \to \mathbb{R}^d$, may be found, which maps image space to a low dimensional embedding space. This mapping should be defined as

$$h^* = \arg\min_h \mathbb{E}_{(i,j,k)\in\mathbf{T}} \left[\mathbb{1}\left\{ \tilde{\delta}\left(h(x_i), h(x_j)\right) \leq \tilde{\delta}\left(h(x_i), h(x_k)\right) \right\} \right] \ldots\ldots (1.2).$$

the indicator function and $\tilde{\delta}: \mathbb{R}^d \to \mathbb{R}$ having a known dissimilarity

Learning a Mapping

Because the subgradient is not defined, we cannot directly optimize (1.2) using gradient descent, the optimization method for convolutional neural networks (CNNs). The degree to which a triplet is satisfied or violated is often used to define a loss function.

$$L\left((x_i, x_j, x_k)\right) = \max\left\{0, \tilde{\delta}\left(h(x_i), h(x_j)\right) - \tilde{\delta}(h(x_i), h(x_k) + C)\right\} \ldots\ldots (1.3)$$

for every t triplet restrictions, where C is a constant offset that discourages oversatisfying and prevents solutions that are too easy. The optimizer may get fixated on outliers if major violations control the loss (1.3). We examine a variation of (1.3) that limits the loss on both sides because we anticipate some inconsistencies in the similarity triplets.

The function

$$L\left((x_i, x_j, x_k)\right) = \text{clip}_{l,u}\left(\tilde{\delta}\left(h(x_i), h(x_j)\right) - \tilde{\delta}\left(h(x_i), h(x_k)\right)\right) \ldots\ldots (1.4).$$

where the function clip

$$\mathrm{clip}_{l,u}(x) = \begin{cases} 0 & \text{if } x < l \\ 1 & \text{if } x > u \\ \dfrac{x-l}{u-l} & \text{otherwise} \end{cases} \quad \ldots..(1.5)$$

Using an increasing number of filters in each layer, one CNN architectural arrangement is mostly based on VGGnet. In the alternative CNN configuration, the number of filters used in each layer is constant. A layer is always composed of maxpooling, 3×3 convolution, and zero padding, regardless of the situation. After the last layer, we add d fully linked units and a global average pooling layer to obtain an input embedding with d dimensions. We use squared Euclidean distance, which is represented as as $\tilde{\delta} = \|\cdot\|_2^2$.to assess the degree of dissimilarity.

Data

Brodersen et al. (2020) and Cahyono, Wirawan, and Rachmadi (2020) cite a countrywide study that screened for lung cancer in 1947 participants using computed tomography (CT) scans, segmented lung masks, and ocular assessment of emphysema severity. "*Emphysema is assessed using a six-point extent scale that covers the six regions of the lungs: the upper, middle, and lower portions of both the right and left lung. Here, the portion of the right lung just above the carina—the upper right area—is all that matters. The six-point extent scale is formed by the intervals{0, 1-5%,6-25%,26-50%,51-75%,76-100%}*." While a tiny percentage(13%) had emphysema scores ranging from 1% to 5%, the vast majority(73%) have no score at all. Various stages of emphysema are illustrated in Figure 1.1 by representative slices.

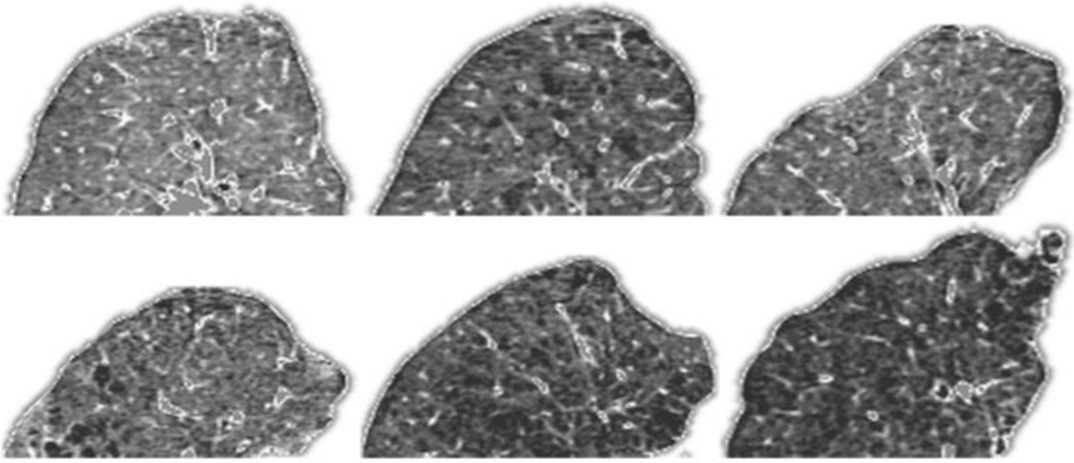

Fig.1.1 The amount of emphysema can be visually judged as follows: 0%, 1-5%, 6-25%, 26-50%, 51-75%, and 76-100%, from top left to bottom. There is a 780 HU level and a 560 HU width window.

3. Results

We used a random number generator to divide the participants into two groups: one for training and another for testing. Afterwards, for each trial, we randomly split the training group in two, with each half being used for validation and half for training. We present the median value derived from these ten runs, as each experiment was repeated ten times. All experiments utilize the same clip function, which is [−0.01,0.1]. This table 1.4 provides the validation and test performance, summarizing median epochs and violations across both untrained and trained models with different emphysema-based sampling.

Table 1.1 Training and Testing Performance Metrics for Triplet Selection Methods

Sampling scheme	Model type	Median training epochs	Median Validation Violation	Median Test Violations (at Emphysema Extent %)
Untrained	F3	-	46.80 ± 0.94	All: 48.5, 0%: 48.9, 0-5%: 44.3, 0-25%: 37.2, 0-50%: 29.2
Uniform	I4	23.0 ± 7.0	40.84 ± 0.71	All: 41.0, 0%: 40.2, 0-5%: 30.0, 0-25%: 19.0, 0-50%: 11.6
Extent	F4	18.0 ± 5.0	39.30 ± 0.58	All: 39.3, 0%: 39.0, 0-5%: 26.4, 0-25%: 14.6, 0-50%: 9.4

International Journal of Research in Engineering and Applied Sciences(IJREAS)
Available online at http://euroasiapub.org/journals.php
Vol. 13 Issue 10, October -2023,
ISSN(O): 2249-3905, ISSN(P) : 2349-6525 | Impact Factor: 8.202| Thomson Reuters ID: L-5236-2015

Figure 1.2 shows the test data embedding, with trained models producing clusters based on emphysema extent. A clear depiction of clustering would show the concentration of emphysema samples in distinct clusters, even when there's overlap with non-emphysema samples. The figure can represent a one-dimensional embedding showing data points distributed along the emphysema extent continuum, with different colors for varying emphysema severity (mild, moderate, severe)..

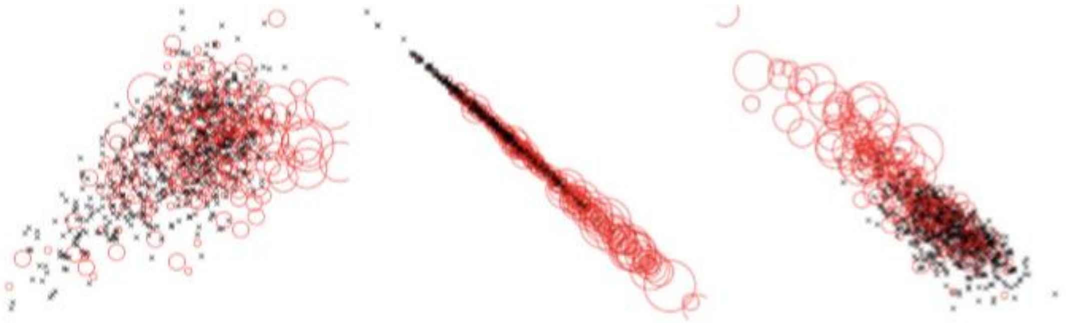

Fig.1.2An example of test data embedding.

Conclusion

Using convolutional neural networks (CNNs) trained on similarity triplets, we were able to reframe emphysema evaluation as a visual similarity job. Through the use of triplets derived from visual evaluations, we shown that feature separation may be accomplished with as little as one cropped 2D slice, even though there is still some overlap between emphysema patients and controls. Initial results indicate that enhancing the quality and quantity of these triplets is important for effective learning, however we did study the possibilities of crowdsourcing triplets for CNN training. Additionally, for moderate cases of emphysema, we discovered that selecting triplets based on the severity of the disease can marginally improve performance. The use of pulmonary function measurements in triplet selection was studied, but ultimately rejected. A potential avenue for future research is the study's suggestion that CNN can learn meaningful emphysema representations from relative scores, which emphasizes the necessity to investigate other concepts of visual similarity.

Reference

Chollet, F. (2021). *Deep learning with Python*. Simon and Schuster.

Jones, I., & Wheadon, C. (2015). Peer assessment using comparative and absolute judgement. *Studies in Educational Evaluation*, *47*, 93-101.

Ash, S. Y., San José Estépar, R., Fain, S. B., Tal-Singer, R., Stockley, R. A., Nordenmark, L. H., ... & COPDGene Investigators and the COPD Biomarker Qualification Consortium. (2021). Relationship between emphysema progression at CT and mortality in ever-smokers: results from the COPDGene and ECLIPSE cohorts. *Radiology*, *299*(1), 222-231.

Juan, D. C., Lu, C. T., Li, Z., Peng, F., Timofeev, A., Chen, Y. T., ... & Ravi, S. (2020, January). Ultra fine-grained image semantic embedding. In *Proceedings of the 13th international conference on web search and data mining* (pp. 277-285).

Cahyono, F., Wirawan, W., & Rachmadi, R. F. (2020, September). Face recognition system using facenet algorithm for employee presence. In *2020 4th international conference on vocational education and training (ICOVET)* (pp. 57-62). IEEE.

DECLARATION

I as an author of this paper / article, hereby declare that paper submitted by me for publication in the journal is completely my own genuine paper. If any issue regarding copyright/ patent/ other real author arises. The publisher will not be legally responsible. If any of such matters occur publisher may remove my content from the journal website/ updates. I have resubmitted this paper for the publication, for any publication matters or any information intentionally hidden by me or otherwise, I shall be legally responsible.

Name: Patil Manjusha Manikrao