139P: Identifying Structural Damage within Reinforced Concrete by CNN

Jonathan E. Dominguez

joedomin@ucsc.edu

Prof. Kawamoto

University of California, Santa Cruz

Abstract

In recent years, several apartment complexes and bridges have collapsed, some of which stood supported by reinforced concrete, resulting in the loss of lives. Damages such as delamination or cracks within the cement can pose a significant threat to structural safety. The proposed Convolutional Neural Network (CNN) can help prevent these devastating events. The CNN will look at radiographs, also known as x-rays, of reinforced concrete and identify structural damages within. The proposed CNN will be compared to how humans approach a similar task, such as radiologists in the medical field who use radiographs to identify minute cracks in bones. The learning algorithm chosen in this proposal is a backpropagation algorithm. The CNN will use the activation function known as Softmax, which aids in classifying the image. The CNN architecture begins with the input layer, two Convolutional Layer and the Pooling Layer groups, and finally, an output layer. Other CNNs take upon a similar task at looking at reinforced concrete, such as using thermography to identify damage to rebar within a floor and using a CNN to decide how damaged a reinforced concrete pillar is in post-disaster events. There will also be a discussion on the benefits of taking a CNN and Deep Learning approach. Such as the speed at which the CNN gains training and experience compared to radiologists. There is a discussion of the limitations of deep neural networks, such as training time due to the number of hidden layers and mapping the correct output.

139P: Structural Failure within Reinforced Rebar Identified by CNN

Introduction

Reinforced concrete is a common technique used to construct structures such as buildings and bridges. In the wake of the recent 12-story building collapse of the Champlain Towers South in Surfside, Florida, that killed 98 people, there has begun a fear in the people leading to implementation to prevent such circumstances from occurring again, calling for more monitoring for structural health (Ruiz-Goiriena & Jervis, 2021). In this paper, we propose a CNN based on deep learning architecture to tackle the problem of identifying the type of damage within reinforced concrete. The proposed CNN takes a radiograph, also known as x-rays, of reinforced concrete which then classifies the image as one of four classes: cracks in the rebar, delamination of the rebar, cracks in the cement, and undamaged reinforced concrete as the CNN is classifying the entire image. The convention in which humans and their natural intelligence would identify stress fractures in reinforced concrete could not be found in the real world as no real-world specific jobs exist that would do this similar task. It would then come down to different areas of expertise to aid in this endeavor, such as radiologists and structural engineers. Specifically, radiologists can identify minute fractures in bones and identify larger objects such as tumors simply by radiographs. In addition, some structural engineers work in this field and have seen firsthand what stress fractures look like after damage or what standard concrete looks like before it cracks. However, the underlying relationship between these two areas involves looking at small cracks. Deciding whether the presence or absence of the interested target is too subtle to notice is known as signal detection theory. In this situation, the radiologist is always using signal detection theory to decide whether what they are looking for, such as a crack in a bone, is present in the image or not. Identifying the wanted target can become challenging to distinguish from background noise, creating a bias when looking for a target (Chen & Howe, 2016, p. 1). Much can go wrong when looking for minute cracks, as expressed by the concepts of signal detection theory. With signal detection theory, there are four possible outcomes when examining something in an image: hit, miss, correct rejection, and false alarm. Identifying that something is in the image results in a 'hit' while missing the target is a 'miss' (Sumner & Sumner, 2020, p. 2). Identifying that there is nothing to see is known as 'correct rejection' while remarking something is there when nothing is there is a 'false alarm' (Sumner & Sumner, 2020, p. 2).

Doctors in the medical field are trained and spend ample time identifying tumors, specifically radiographs, to gain experience as there is a correlation between more time experience and better accuracy. Radiologists are taught and trained that the fundamental processes of detecting a target in an image are visual inspection and interpretation. The cognitive processes that radiologists go through are as follows: visual inspection followed by detection if something is spotted; if not, no further investigation is required. If something is detected, radiologists attempt to recognize what it could be by characterizing it (Waite et al., 2019, p. 2). However, if the radiologist cannot detect something even if something is present in the image, then the rest of the steps cannot be completed. Thus, the problem of natural human intelligence in this process is the bias of radiologists' interpretation, which is why error rates among radiologists are at 4% per year (Waite et al., 2019, p. 2).

Along with the bias the radiologist has, there are two main types of errors, errors of human's natural intelligence, that occur: perceptual errors and cognitive errors. Cognitive errors are those in which the radiologist lacks knowledge or faulty reasoning, while perceptual is the missing of a target. Although these are just listed errors, these are the natural intelligence approaches radiologists are taught and practice.

Doctors are commonly taught different search patterns for examining; an example is looking at a radiograph of lungs and reading the right and left lungs individually in a downward 'z' pattern. The radiologist then processed a second pass to compare the right and left lung symmetry (Waite et al., 2019, p. 5). Many processes are going on that involve natural human intelligence. Radiologists search through a radiograph using a combination of bottom-up and top-down attention, using what they have learned from the past (Waite et al., 2019, p. 4). The reason for this is that the brightened or darkened areas of an x-ray draw the engagement of a radiologist resulting in a bottom-up approach. However, if the radiologist knows what it is looking for, they have an intended target which is a top-down approach. The experience of seeing the target in the past and the expectation of what it looks like results in a top-down process (Waite et al., 2019, p. 3). When radiologists know what they are looking for, they learn to ignore irrelevant areas and focus their attention on relevant areas they know where something should be.

Simulation, Deep Learning Approach

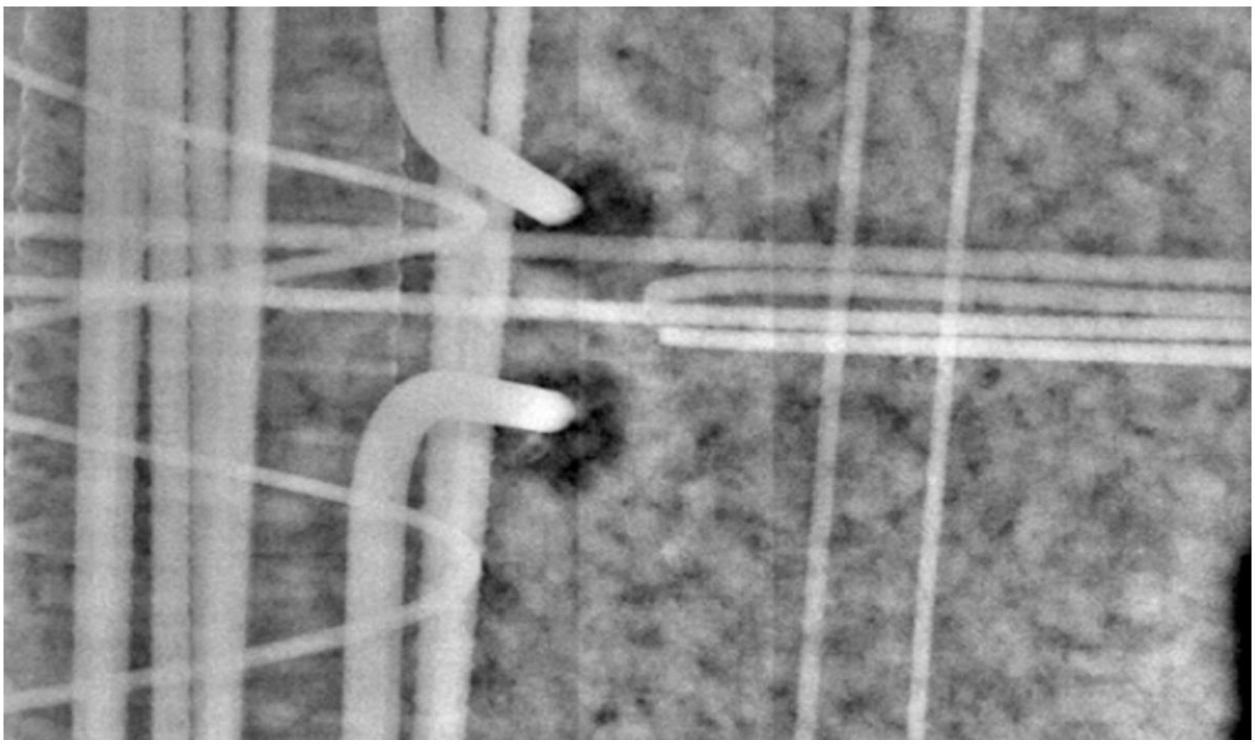
In its simplest terms, an artificial neural network (ANN) was created to model the human brain, such as the concepts of neuron connections like synapses. An ANN can be broken down into these components at its most basic architecture: a weight input layer, a hidden layer, and an output layer. In which their connections between layers represent the synapses in a brain. The input layer is where the data is fed, while the hidden layer is where features extraction occurs. The relationship is known as weights, in which their function determines the strength of a connection. The higher the value of the weight, the stronger the connection. While the lower the value of a weight, the weaker the connection. ANN can be applied in computer vision. A limitation to ANN is the backpropagation algorithm, as it must learn two things: the valuable features of the input set and making sure it is mapped to the correct output. The benefits of ANN are immense using an ANN; the network can be taught to simulate the relationship between the data by training it. Resulting in ANNs learning to develop a correct relationships response, in other words, a correct map from the input to output. Correctly mapping out from the input to the output means creating the correct relationship between weights, resulting in learning. Layers are connected to the next layer, and weights are associated with the connections to demonstrate the connection strength of each layer; the weights are learned by using training data as well as a backpropagation algorithm. To correctly map out means the weights are at their optimal from the beginning to the end of the networks, the weights changes at it uses backpropagation to learn from its errors, creating a generalized curve.

Deep learning is an artificial neural network in which deep refers to the depth or multiple hidden layers. The more hidden layers in a deep learning network, the more higher-level features can be extracted. Deep learning architecture is quite similar to ANN. It contains an input layer, hidden, and output layer; however, the architecture of a deep learning model contains multiple hidden layers. As the input data goes through each layer, a feature is extracted from the image. Learning in this context means the weight connection between layers is adjusted as data passes through. Further learning is done by backpropagation as weights are adjusted for errors as the network attempts to learn the relationship of the data to create a stronger relationship between the weights and layers. Deep learning has many strengths, from image classification to localization and object detection. One of the limitations of deep learning is that the more hidden layers there are, the longer a training time is required as it goes through all the layers, which takes time.

There is no preexisting dataset of x-rayed reinforced concrete in buildings or bridges that would be required to train and test the proposed CNN; thus, it would need to be gathered. One method in which this data could be collected is by going around existing buildings and bridges that incorporate reinforced concrete pillars and having them x-rayed individually. The images from the dataset would look (see Figure 1).

Figure 1

Example of Dataset Image



Note. *X-Ray of Rebar*. (n.d.). Conview Company Inc. Retrieved February 25, 2022, from https://concrete-view.com/service/x-ray-services/.

It would contain a grayscale picture of rebar within the cement and some cement pocket air bubbles, and anything not within the reinforced concrete would be black. One of the benefits of using radiographs is that they come in high-resolution with a pixel size of 175 μm to 100 μm, thus creating an excellent resolution quality for the CNN to examine (Huda et al., 2015, p. 396). Radiologists would label the images of the dataset. The radiologists would first need to be taught by foundation repair professionals to know what to look for in the image, such as delamination and cracks. Then having the radiologists identify those markers in the image. By default, the dataset images would be grayscale as that is how x-rayed images output after taking pictures. After gathering all the photos, the dataset would then be labeled and divided into the following four categories: rebar delamination, crack in the rebar, cracks in the cement, and unaffected reinforced concrete. It would be ideal to obtain 1,500 images for each of the four categories, from which 500 of each of those categories would then be used to test the CNN. In terms of image resolution, in which radiographs can capture some great quality images, for this dataset the images will have a resolution of 2000 × 2500 translating to five-million-pixel (Huda & Abrahams, 2015, p. 396).

The layers of this CNN are as follows: input layer, convolution layer with the Rectified Linear Unit function (ReLu), max-pooling layer, followed by another convolution layer with a ReLu activation function, max-pooling layer, and finally a fully connected layer using the Softmax function. In the input layer, the input image is represented as the pixel matrix of the image. In the first convolutional layer, the features extract of the image occurs at a low level, in this case, edges, lines, and corners. The size of the kernel for the convolution kernel size of 3x3x3. While the later convolutional layer extract high-level with the aid of the low-level features, which also uses a 3x3x3 kernel. The output layer uses the softmax activation to aid in predicting what the image is, which adds up to 1. The output layer is also fully connected with four outputs. In the pooling layer, the specification area is a 2x2x2 area with a stride of 1. For some more context of the CNN, the learning algorithm used in this proposed CNN is backpropagation. The activation function that the CNN uses is the Softmax activation function (1), also known as the normalized exponential function, as it is used to identify many different categories. The Softmax activation function (1) is used at the fully connected layer, while the convolution layers use the ReLU functions as the activation functions (2). The ReLu is a piecewise linear function that will output the input directly if it is positive; if not, it will output zero. It has become the default activation function for many types of neural networks because a model that uses it is easier to train and often achieves better performance.

$$\sigma\left(\underset{z}{\rightarrow}\right)_i = \frac{e^{z_i}}{\sum_{j=1}^{K} e^{z_j}} \tag{1}$$

$$y = \max(0, x) \tag{2}$$

The following is a CNN simulation starting with the input layer taking in a radiograph image of cracked rebar with the matrix resolution size of 2000 × 2500. The radiograph can be considered a three-dimensional input image, in terms of the RGB, as it is grayscale and not completely back and white. Since the image is grayscale, the CNN can still take in three inputs units: red, green, and blue. This will aid in the spatial location. The kernel is in the first input data to the first convolution layer, which size is a 3x3x3 matrix. The kernel will pass through the entire image to find low-level details: the reinforced concrete's edges, lines of the rebar, and corners. This occurs as the kernel passes the values of the image pixels multiplied by the kernel filter weight, resulting in a single value. The convolutional will be using the ReLu activation function. As the ReLu will output the given inputs either as positive or zero. From those single values from the convolution layer, the image is now at the first pooling stage. The pooling area will use a 2x2x2 MAX pool filter, in which the pooling looks for the highest number in the 2x2x2 area. The pooling area uses a stride of 1, which means it slides one matrix area moving the filter. After the pooling area, the data is now on to the next convolution layer, in which it would the kernel will pass through the data to gather high-level features: the areas of which are cement, where the rebars are, cracks in both rebar and cement. The same ReLu activation function is used once again in this stage. The data then passes the 2x2x2 MAX pool filter with a stride of 1. Finally, in the output layer size is four outputs, we have a fully connected layer using the Softmax activation function. In the output layer, there are four outputs, as discussed before: rebar delamination, crack in the rebar, cracks in the cement, and unaffected reinforced concrete. In the fully connected, the combination of the features further aids in helping to predict the category. The input of the cracked rebar now comes to the fully connected layer, which outputs the following: rebar delamination 0.02, crack in the rebar 0.91, cracks in the cement 0.06, and unaffected reinforced concrete 0.01. The prediction of the softmax function results must add to 1.

Overfitting occurs when there is high variance in the complex model. The high variance in the data causes the learning algorithm to miss finding the underlying relationship between the input and the output. Understanding the underlying relationship means creating a curve that does not fit all the data exactly but can create a curve that can generalize the training data. When the training does not understand the underlying relationships, the CNN fails to learn to generalize the data in the training stage. If the curve was trying to hit and fit all the data points in the training, then there is no point in learning; it might as well create a table and look up the exact data pairs. The CNN is after generalization and not fitting the data perfectly as there will always be noise in the data; thus, CNN must learn to deal with the noise optimally. Overfitting can become a significant problem and detrimental to the CNN learning process, leading to inadequate interpretation of the data. For example, suppose the network learned the wrong underlying relationship between the data during the training and testing phases. When it is tried in the real world with new input, the data might fall either within or out of the curve giving incorrect outputs. Many curves are being created to satisfy the noise in the data instead of the intended output, which would underestimate the relationship between connections. Increasing the number of parameters reduces the Sum Squared Error (SSE), which means that noise is being predicted, making the predicted curve have too much variance as it fits the data and not the underlying relationship. The CNN is after the relationship of the data and not trying to predict it exactly, which would lead to errors.

There are a couple of ways to tackle the overfitting problem, such as dataset augmentation. The given stimulus, dataset image, can be modified in different ways to increase the number of stimuli used in training. Modifications such as shifting, blurring, changing pixels, and rotating an image can increase randomness in the training set. An idea already proposed to use in this CNNs dataset. Other solutions involve the weights of the network, such as dropout, simply dropping a random weight in the hopes of seeing if it helps or not, which is reducing the number of parameters as it reduces the number of connection weights. Another weight solution is weight decay, which refers to the decay over time of a weight to zero, eliminating connections. Early stopping is when the CNN trains on the training set data while simultaneously testing that data. At the end of an epoch, a round of training, the output is tested to see how well it did. It stops the training when the error on the test set starts to increase. Essentially, the CNN maximizes the training by finding the sweet spot at the middle point before too much error occurs.

Discussion

Overall, CNNs would be better at carrying out the given task compared to the natural intelligence of humans. One such reason is the speed at which a CNN can become an expert at identifying a wanted target. Radiologists take years to gain experience; they can still create errors even then. Compared to a human natural intelligence approach, radiologists tend to ignore irrelevant areas and focus their attention on relevant areas they know where something should be. However, this leads to them not examine the whole image equally, which may result in missing unexpected findings. Unlike the radiologist, the CNN does not ignore any part; instead, it looks and processes the entire image. Thus, the natural intelligence approach of the top-down process fails here, unlike for the CNN, which uses experiences from the training stage, it still processes the whole image, not missing any vital areas (Waite et al., 2019, p. 4). The natural intelligence of humans also fails them compared to CNN's due to what is known as cognitive errors in which the radiologist lacks in knowledge or faulty reasoning leads to a mistake (Waite et al., 2019, p. 2). Compared to natural intelligence, a CNN can create the same errors from signal detection, such as a miss or false alarm (Sumner & Sumner, 2020, p. 2). This is because radiologists would have labeled the dataset; thus, whatever the radiologists missed or incorrectly saw, then CNN would produce the same outcome.

There is no other current CNN in published research papers identical to this proposed CNN; however, there are CNNs that revolve around looking at reinforced concrete. Cheng et al. (2019) looked to identify delamination of rebar in bridge decks using thermography, which is helpful but is not the central part of a bridge's structural integrity (p. 1). Cheng et al.'s approach also comes with problems such as thermography, which relies heavily on temperature; however, temperature differs depending on their location, environment, and weather. An example of this is taking a thermography reading of reinforced rebar of a bridge deck in New York's winter months and then comparing that to the summer months reinforced rebar of a bridge deck in sunny Los Angeles (Cheng et al., 2019). This paper's proposed CNN can take images regardless of location, environment, and weather as the radiographs come back in grayscale. Also, the inputs would be cleaner and more consistent to read compared to Cheng et al. (2019) CNN and dataset. Comparing the architectures of the CNN of Cheng et al. (2019) and this paper's proposed CNN, CNN of Cheng et al. (2019) uses three dense blocks followed by three upscale structures, which this paper's proposed CNN does not use. Another similar approach to reinforced rebar damage identification done by Pan & Yang (2020) looked at reinforced concrete post-disaster to detect damages and estimate repair (p. 1). Thus, it is trying to determine how damaged a reinforced concrete pillar is once the damage is visible. This could be a future added benefit to the proposed CNN in this paper as it would help to tell engineers if the crack or delamination in the rebar is something to be concerned about and if it would need immediate action. The CNN architecture between both are similar regarding the use of convolution; however, they differ in that this paper's proposed CNN does not use batch normalization. The most significant distinction between these two CNNs is that Pan & Yang's (2020) CNN is a dual CNN as they need one to identify and localize items of the reinforced concrete and another to categorize it into a damaged state (p. 3-6). The proposed CNN in this paper would not have to worry about coming into play in a post-disaster event because it would be able to find the damage before it went to that extreme.

Benefits and Limitations

The significant benefit of such a CNN would be to save lives and prevent homes from being destroyed by catching unseen damage to the reinforced cement before it becomes too late. Another benefit of the CNN is that there is no need to expose the rebar, as doing so would damage the integrity of the support pillar, thus making this CNN dataset noninvasive. If the dataset did not accumulate enough samples, the dataset size would become a fundamental limitation in this proposed CNN. If this CNN were to be tried in the real world by someone, then the cost issue would be raised as it would be expensive to move a heavy radiograph machine to a location simple to capture an image. It is also best to mention that with radiograph machines comes the potential to cause harm to those around due to the radiation emitted when capturing an image. Another limitation of CNN is being able to distinguish different angles, backgrounds noise, as well as lighting conditions. However, there is not much background noise for this proposed CNN as the images come in grayscale, nor are lighting conditions a factor at play due to the grayscale image. Finally, we could use the different angles problem to our advantage by giving images already seen by the CNN, rotating them, and feeding them back to the CNN as new inputs. One of the limitations of deep learning is having numerous hidden layers. As it requires very long training times thus, the deep learning network could be going through the given dataset for a prolonged time. Accompany the vast number of layers with numerous images in a given dataset, and the deep learning network could be there even longer. The reason if can have such a long training can also be due to the backpropagation algorithm as the network needs to learn the error. As the network's hidden layers are trying to learn the representation of the given inputs. To do this is it compute the error at the output units. Another limitation is whether or not the backpropagation algorithm would learn the valuable features of the input and if it could map the correct output. A limitation that could destroy the goal of the deep learning network before the training process even begins would be the labeling of the images in the dataset. The concept of garbage in garbage out condition would come into play, and the deep learning network would be learning only to give inaccurate outputs.

References

Chen, W., &; Howe, P. D. (2016). Comparing breast screening protocols: Inserting catch trials does not improve sensitivity over double screening. PLOS ONE, 11(10). https://doi.org/10.1371/journal.pone.0163928

Cheng, C., Shang, Z., & Shen, Z. (2019). CNN-based deep architecture for reinforced concrete delamination segmentation through thermography. *Computing in Civil Engineering 2019* (pp. 1–8).

Huda, W., & Abrahams, R. B. (2015). X-ray-based medical imaging and resolution. AJR. American journal of roentgenology, 204(4), W393–W397. https://doi.org/10.2214/AJR.14.13126

Pan, X., & Yang, T. Y. (2020). Postdisaster image-based damage detection and repair cost estimation of reinforced concrete buildings using dual convolutional neural networks. *Computer-Aided Civil and Infrastructure Engineering*, *35*(5), (pp. 495–510).

Ruiz-Goiriena, R., &; Jervis, R. (2021, December 16). 'save lives, not just money': Surfside collapse grand jury report calls for reforms, warns of further troubles. USA Today. Retrieved February 23, 2022, from https://www.usatoday.com/story/news/nation/2021/12/15/grand-jury-probe-into-surfside-condo-collapse-condo-safety-released/8906682002/

Sumner, C. J., &; Sumner, S. (2020). Signal detection: Applying analysis methods from psychology to animal behaviour. Philosophical Transactions of the Royal Society B: Biological Sciences, 375(1802), 20190480. https://doi.org/10.1098/rstb.2019.0480

Waite, S., Grigorian, A., Alexander, R. G., Macknik, S. L., Carrasco, M., Heeger, D. J., & Martinez-Conde, S. (2019). Analysis of perceptual expertise in radiology – current knowledge and a new perspective. *Frontiers in Human Neuroscience 2019*(pp. 1-21).