Rapid AI

Challenging Misconceptions About Neural Network Models

Jon Neff
AI, Analytics and Innovation
Department
The Aerospace Corporation
El Segundo, California, USA
jon.m.neff@aero.org

ABSTRACT

Neural networks have had great success in computer vision and natural language processing in recent years. However, there is a misconception that neural networks require enormous amounts of data, specialized expertise and a lot of time to develop. The purpose of this paper is to dispel these misconceptions and demonstrate that excellent results can be obtained quickly for image and text classification problems given a suitable labeled data set.

The application chosen for this study is humanitarian assistance and disaster relief, one of the high priority national mission initiatives of the recently established Department of Defense Joint Artificial Intelligence Center. Satellite and airborne imagery are important data sources for situational awareness after disasters. Several studies have shown that social media such as Twitter and Facebook can also be important sources of information during humanitarian assistance operations.

The Aerospace Innovation Laboratory funded this study to investigate the use of AI to automatically classify satellite imagery and social media data to provide situational awareness during disaster relief operations. Data sets included satellite imagery from Hurricane Harvey in 2017, Twitter text data from the 2014 California earthquake, and images posted on social media from the Nepal earthquake in 2015.

This study shows that state of the art results can be obtained using off the shelf technologies with little effort in a very short time.

CCS CONCEPTS

• General conference proceedings • Natural language processing • Computer vision • Applied computing

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).
EVOKE CASCON 2020, November 2020, Toronto, Canada.
© 2020 Copyright held by the owner/author(s). 978-1-4503-0000-0/18/06...$15.00
https://doi.org/10.1145/1234567890

KEYWORDS

Neural networks, artificial intelligence, classification, transfer learning, data augmentation, satellite imagery, social media, humanitarian assistance

ACM Reference format:

Jon Neff. 2020. Rapid AI: Challenging Misconceptions About Neural Network Modes. In *Proceedings of the 30th Annual International Conference on Computer Science and Software Engineering (EVOKE CASCON 2020). ACM, New York, NY, USA, 6 pages.* https://doi.org/10.1145/1234567890

1 Introduction

Significant and rapid advances in software and hardware technology over the last ten years have made it possible to implement neural networks that can automatically classify images and text with a high degree of accuracy. [4] However, there is a persistent perception that neural networks require enormous amounts of data, specialized expertise and a lot of time to develop. [14] The purpose of this paper is to dispel these misconceptions and demonstrate that excellent results can be obtained quickly for image and text classification problems given a suitable labeled data set.

The application chosen for this paper is humanitarian assistance and disaster relief, one of the high priority national mission initiatives of the recently established Department of Defense Joint Artificial Intelligence Center (JAIC). The JAIC was established in 2018 to advance artificial intelligence (AI) applications for national security. The objective of the humanitarian initiative is to, “Reduce the time associated with search and discovery, resource allocation decisions, and executing rescue and relief operations to save lives and livelihood during disaster operations.” [1] Humanitarian assistance is a DOD core capability and has been exercised in disaster areas many times in recent years, including the 2011 tsunami in Japan, Typhoon Haiyan in the Philippines, and the 2015 earthquake in Nepal. [2]

Collecting information about structural damage and human needs can be challenging for disaster relief operations. News media may not have enough reporters to cover the affected area and normal communication channels may be unavailable. Satellite imagery has been used extensively for many years to track and evaluate the effects of natural disasters. More recently, imagery from unmanned aerial vehicles (UAVs) has also been employed to help with relief operations. [3] Commercial satellite imagery with 1-meter resolution and 1-day revisit time is available from Planet and other providers. Satellite imagery is usually georeferenced to latitude and longitude using image formats such as GeoTIFF, making this data valuable for damage assessment as well as providing a "big picture" overview of the affected area.

Several studies have shown that social media such as Twitter and Facebook can also be important sources of information about infrastructure damage as well as injured people needing medical attention and missing persons. [3] Social media users can tag images and text with precise location data using GPS receivers that are built into modern cell phones. Thus, these data sources have the potential of providing very detailed information to rescue and relief operations. The usefulness of these sources is limited by the sheer volume of data and the lack of available human analysts to review and classify them.

The Aerospace Innovation Laboratory funded this study to investigate the use of AI for humanitarian assistance and disaster relief operations. Data sets included satellite imagery from Hurricane Harvey in 2017, Twitter text data from the 2014 California earthquake, and images posted on social media from the Nepal earthquake in 2015. This study shows that state of the art results can be obtained using publicly available unclassified data and off the shelf technologies with little effort in a very short time.

2 Computer Resources

All research was conducted on an Nvidia P100 GPU server with 56 cores and 16 GB of memory. This class of hardware costs about $100,000 and can also be accessed on a "pay as you go" basis from cloud providers such as Amazon Web Services. The fastai and PyTorch Python libraries were used to construct, train and validate neural networks while Pandas was used to read and manipulate the image data. Computations were performed in a Jupyter notebook. All software libraries used are free and open source.

3 Social Media Image Classification

3.1 Data

The social media image data used in this study were obtained from the CrisisNLP website [7]. According to [6], images were collected from social media platforms and labeled as either severe damage, mild damage, or little-to-no damage by volunteers. We used 15,283 labeled images from the 2015 Nepal earthquake. 7,142 images were labeled "severe," 1,805 were labeled "mild" and 6,336 were labeled "none." Images were roughly square and ranged from 250 to 700 pixels per side. Images were scaled to 224x224 pixels for training and validation. The data were randomly split into 80% training set and 20% validation set. Sample images are listed below.

Figure 1: Sample images from the 2015 Nepal earthquake data set

3.2 AI Model and Results

A residual convolutional neural network (CNN) architecture known as ResNet-34 was used for this data set (Figure 3). Convolutional neural networks are used with data that consist of multiple matrices or tensors such as images. Unlike fully connected layers, convolutional layers connect a node in one layer to a small number of local nodes in the previous layer. Convolutional networks are easier to train and more accurate than fully connected networks for image classification problems and have revolutionized computer vision over the last ten years. [4] Residual networks add identity mappings that skip layers, enabling the network to train on residuals. Identity mappings improve accuracy in deep networks without additional parameters or increased computational complexity. [9]

CNNs can require large amounts of data and compute time to train from scratch and our data set is relatively small. Therefore, we used a transfer learning approach with a pretrained network based on the ImageNet data set. In transfer learning, the output layers are altered to fit the number of classes in the current data set and the weights are fine-tuned during training. ImageNet consists of over a million images from everyday life labeled with about a thousand classes.

We also used data augmentation to regularize our model and prevent overfitting. Data augmentation expands the training set by transforming the images. The transforms used for this data set are generally good for normal, everyday photographs. Transformations included horizontal flips, rotations by up to 10 degrees, zoom up to 110%, lighting and contrast changes and warping to alter apparent perspective.

As is common with neural network models, we used back-propagation with mini-batch stochastic gradient descent to train the model with a batch size of 64. We used the "1cycle" learning rate policy for training. [8] This approach uses a variable learning rate that increases linearly for the first half of a cycle and decreases linearly in the last half. Experiments have shown that this learning rate schedule can reduce the iterations required to converge on an accurate solution by almost 90%.

We used event-specific data only, i.e., data from the Nepal earthquake itself. In a real-world scenario it is more likely that a classifier would be trained on a combination of event-specific data and cross-event data, which consists of data from previous disasters.

Results from the validation set are shown in Figure 2. We achieved 84% accuracy, which equals the accuracy of the best model in [6]. Our model does well in distinguishing between severe damage and no damage but struggles with mild damage. A similar effect was noted for the models in [6]. Note that these results were obtained using less than 30 lines of code and less than 4 minutes of GPU time.

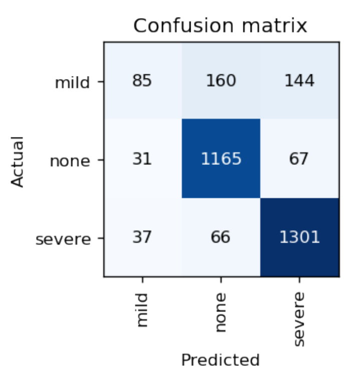

Figure 2. Confusion matrix for social media image model

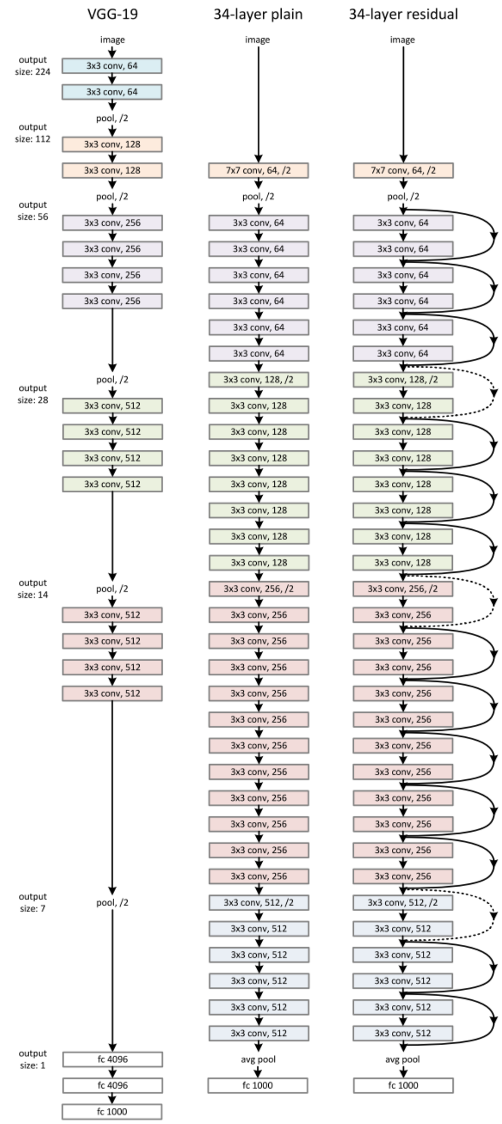

Figure 3. ResNet-34 architecture (right) compared to VGG-19 and 34 layer plain CNN [9]

4 Satellite Image Classification

4.1 Data

Satellite imagery data were obtained from [10]. As described in [5], Digital Globe provided satellite images of the Houston, Texas area following Hurricane Harvey in 2017. 128x128 pixel tiles were extracted from the satellite images around coordinates of known buildings and low-quality images were discarded. Crowdsourced volunteers labeled the data as either "damaged/flooded building" or "undamaged building." The training set consisted of 5,000 images of each class (total of 10,000 images) and the validation set consisted of 1,000 images of each class (total of 2,000 images). Sample images are show in Figure 4 below.

Figure 4: Sample images from the 2017 Hurricane Harvey data set

4.2 AI Model and Results

A ResNet-50 residual convolutional neural network was used for this problem. ResNet-50 is similar to the ResNet-34 architecture shown in Figure 3 but has 50 layers. For this model we used the same advanced techniques listed for the social media image model except that the transforms for data augmentation were selected for satellite imagery. The image transformations included horizontal and vertical flips, lighting and contrast changes, and zoom up to 105% but no warping. We also used an approach to setting the learning rate hyperparameter described in [11]. Specifically, we started with a small learning rate and increased it for each mini-batch. The results were plotted and we picked the highest learning rate for which the loss is still decreasing rapidly. We trained for 5 epochs.

Results from the validation set are shown in Figure 5. We achieved 98% accuracy, which matches the accuracy of the best model in [5], which was custom designed specifically for this data set. These results were obtained using less than 30 lines of code and about 90 seconds of GPU time.

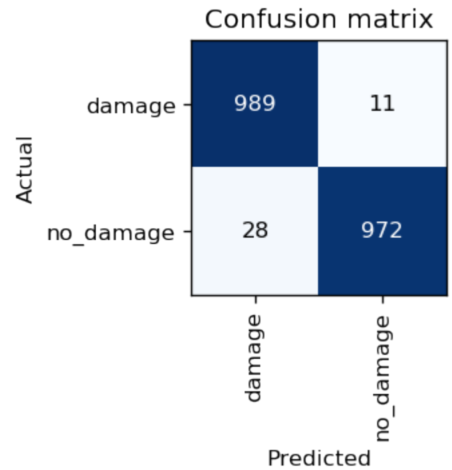

Figure 5: Confusion matrix for satellite imagery model

5 Social Media Text Classification

5.1 Data

The social media text data used in this study were obtained from the CrisisNLP website [7] and documented in [12]. For this problem we used text data collected from Twitter during the 2014 California earthquake and labeled by paid workers. The data set consists of 2,013 text blocks or "tweets" of 140 characters or less. The classes consist of "caution and advice"; "displaced people and evacuations"; "donations needs or offers or volunteering services"; "infrastructure and utilities damage"; "injured or dead people"; "missing trapped or found people"; "not related or irrelevant"; "other useful information"; and "sympathy and emotional support." The text blocks were tokenized and converted to integers prior to training. Vocabulary was limited to 60,000 words to avoid computation problems with sparse high dimensional data.

5.2 AI Model and Results

For this problem we started with a pre-trained language model, used inductive transfer learning to fine-tune the language model for our data set, then used the language model encoder in a text classifier that we trained. This approach is a recent innovation in natural language processing (NLP) and is documented in [13].

Language models represent the probability of a sequence of words and are widely used in NLP applications. Basically, a language model predicts the next word in a sentence with a certain accuracy based on the previous words. We started with an existing average stochastic weight dropped (AWD) long short-term memory (LSTM) recurrent neural network (RNN) language model trained on the Wikitext-103 data set. The Wikitext-103 data set consists of over 28,000 "large" Wikipedia articles with 103 million words. We then fine-tuned the language model to the 2014 California earthquake Twitter data set.

Unlike feedforward networks, RNN outputs from a previous time step as inputs to the current time step. LSTM adds memory nodes that allow the RNN to learn over many time steps. RNNs have been successful in AI applications for NLP and time series analysis. Our RNN consisted of three layers and the same AWD-LSTM RNN architecture was used for both the language model and the text classifier. Dropouts were used to regularize the model and prevent overfitting. We used a batch size of 48 for the language model and 96 for the classifier. We used the learning rate techniques discussed in the sections on image models and incrementally unfroze layers of the classifier to fine-tune performance. Results from the validation set are shown in Figure 6.

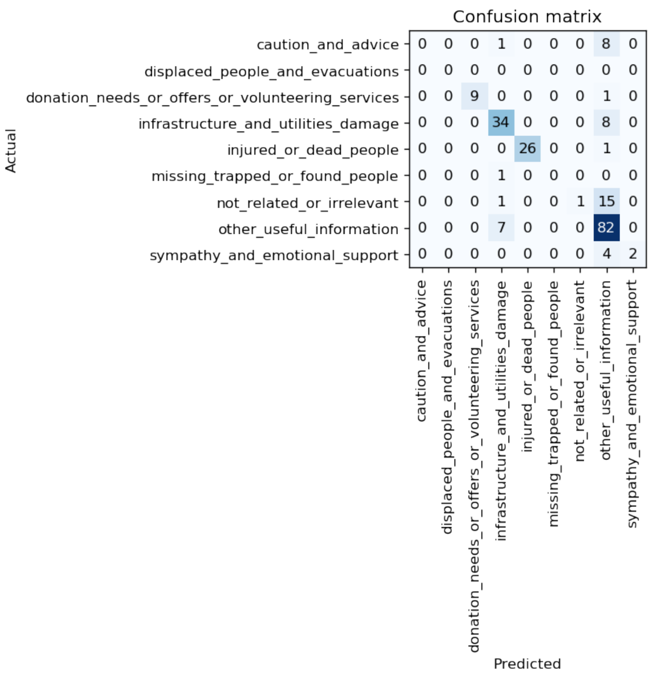

Figure 6: Confusion matrix for text classifier

Our model does well in classifying other useful information, injured or dead people, infrastructure and utilities damage, and donations or volunteering. The model does not do well in classifying irrelevant information or caution and advice. Note also that the data are skewed, with little or no data for some classes.

Our text classifier achieved 76.6% accuracy for event-only data. The accuracy of the custom CNN classifier in [12] was 76.85% for the 2014 California earthquake using event-only data. While these results may not sound impressive given the accuracy of the image models, they are actually quite good for Twitter data. Tweets during disasters are difficult to classify even for human analysts due to the large amount of noise and ambiguity in the text data. Inter-annotator agreement (IAA) is the percentage that two human analysts labeled texts with the same class. The authors in [12] calculated the IAA for the California earthquake data set as 85%, which represents an upper bound for accuracy in this problem. The text model consisted of 40 lines of code and took 32 seconds of GPU time to train.

6 Summary

Highly accurate neural network models for image and text classification can be developed with modest amounts of data and compute time and without a large investment in specialized expertise. This entire project, including writing the paper, took forty hours of labor. The key to making progress quickly is to use existing labeled data sets, modern hardware and software, and state of the art techniques. This does not minimize the effort involved in collecting and labeling the data in the first place, which was outside the scope of this project. Also, image and text classification are well-researched problems with many advanced techniques already embedded in convenient software libraries. The data sets used in this study had few classes; accuracy would probably be lower for problems with many classes. Other types of problems such as time series anomaly detection with high dimensional data can be much more difficult to solve. Explainable and resilient neural network models require other advanced techniques beyond the scope of this study.

Artificial intelligence using neural networks is a fast-moving field. Maintaining an awareness of current best practices can produce excellent returns in the ability to rapidly train accurate models.

ACKNOWLEDGMENTS

The author would like to thank the Aerospace Innovation Laboratory, which provided funding for this project.

REFERENCES

[1] “Department of Defense’s Artificial Intelligence Structure, Investments, and Applications” Statement By Dana Deasy, Department Of Defense Chief Information Officer, Before The House Armed Services Committee Subcommittee On Emerging Threats And Capabilities. 11 December 2018.

[2] "https://www.health.mil/Military-Health-Topics/Health-Readiness/Global-Health-Engagement/Humanitarian-Assistance-and-Disaster-Relief" [Online].

[3] Said, Ahmad et al, "Natural Disasters Detection in Social Media and Satellite imagery: a survey". arXiv:1901.04277. 14 January 2019.

[4] LeCun, Bengio and Hinton, "Deep learning". doi:10.1038/nature14539. 28 May 2015.

[5] Cao and Choe, "Detecting Damaged Buildings on Post-Hurricane Satellite Imagery Based on Customized Convolutional Neural Networks". arXiv:1807.01688v2

[cs.CV]. 28 November 2018.

[6] Nguyen, Ofli, Imran, and Mitra, "Damage Assessment from Social Media Imagery Data During Disasters". 2017 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining. ACM ISBN 978-1-4503-4993-2/17/07. 31 July 2017.

[7] "Resources for Research on Crisis Informatics." https://crisisnlp.qcri.org [Online].

[8] Smith, "A Disciplined Approach to Neural Network Hyper-parameters: Part 1 – Learning Rate, Batch Size, Momentum, And Weight Decay." US Naval Research Laboratory Technical Report 5510-026. arXiv: 1803.09820. 24 April 2018.

[9] He et al, "Deep Residual Learning for Image Recognition." arXiv: 1512.03385. 10 December 2015.

[10] https://github.com/qcao10/DamageDetection [Online]

[11] Smith, " Cyclical Learning Rates for Training Neural Networks." arXiv:1506.01186. 4 April 2017.

[12] Nguyen et al, "Rapid Classification of Crisis-Related Data on Social Networks using Convolutional Neural Networks" In Proceedings of the 11th International AAAI Conference on Web and Social Media (ICWSM), 2017, Montreal, Canada. arXiv:1608.03902. 12 August 2016.

[13] Howard and Ruder, "Universal Language Model Fine-tuning for Text Classification." arXiv:1801.06146. 23 May 2018.

[14] Ramanathan, "5 Challenges for Deep Learning." EE Times. https://www.eetimes.com/author.asp?section_id=36&doc_id=1333535# [Online]. 31 July 2018.