LAWRENCE
LIVERMORE
NATIONAL
LABORATORY

LLNL-TR-711858

The Livermore Brain: Massive Deep Learning Networks Enabled by High Performance Computing

B. Y. Chen

November 29, 2016

Disclaimer

This work performed under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under Contract DE-AC52-07NA27344.

The Livermore Brain: Massive Deep Learning Networks Enabled by High Performance Computing

Barry Chen (14-ERD-100)

Abstract

The proliferation of inexpensive sensor technologies like the ubiquitous digital image sensors has resulted in the collection and sharing of vast amounts of unsorted and unexploited raw data. Companies and governments who are able to collect and make sense of large datasets to help them make better decisions more rapidly will have a competitive advantage in the information era. Machine Learning technologies play a critical role for automating the data understanding process; however, to be maximally effective, useful intermediate representations of the data are required. These representations or "features" are transformations of the raw data into a form where patterns are more easily recognized. Recent breakthroughs in Deep Learning have made it possible to learn these features from large amounts of labeled data. The focus of this project is to develop and extend Deep Learning algorithms for learning features from vast amounts of *unlabeled* data and to develop the HPC neural network training platform to support the training of massive network models. This LDRD project succeeded in developing new unsupervised feature learning algorithms for images and video and created a scalable neural network training toolkit for HPC. Additionally, this LDRD helped create the world's largest freely-available image and video dataset supporting open multimedia research and used this dataset for training our deep neural networks. This research helped LLNL capture several work-for-others (WFO) projects, attract new talent, and establish collaborations with leading academic and commercial partners. Finally, this project demonstrated the successful training of the largest unsupervised image neural network using HPC resources and helped establish LLNL leadership at the intersection of Machine Learning and HPC research.

Background and Research Objectives

As vast amounts of data in the form of images, video, audio, text, etc. are collected; the need for computer-assisted analysis of such data grows. This data deluge presents both daunting challenges as well as remarkable opportunities. For Machine Learning researchers, the excitement in big data comes from the tantalizing potential for improved system performance resulting from training larger more expressive models on vast amounts of data. Instead of hand-engineering feature representations, learning discriminative features directly from massive datasets is one of the primary drivers for the "super-human" image recognition results using deep convolutional neural networks on the ImageNet Large Scale Visual Recognition Challenge (*Russakovsky2014, Krizhevshy2012, Szegedy2015, He2015*). These results depend on the availability of not only big data but also big labels. Unfortunately, the volume and variety of data collection far outpaces human annotation abilities, and while crowd-sourcing the labeling problem may be a viable approach for generating more labels, many data are

unsuitable for crowd-sourcing due to their sensitive nature or the required technical expertise to generate high-quality labels. As scientists and engineers continue to develop new sensor and measurement devices that are able to collect new forms of data, the importance of *unsupervised feature learning* becomes tantamount.

The goal of this LDRD is to research and develop new high performance computing (HPC) enabled Deep Learning algorithms for the unsupervised learning of transferrable feature representations of images and video from massive unlabeled datasets. This goal can be broken up into several sub-objectives:

1) The development of scalable neural network training algorithms and software that take advantage of HPC architectures
2) The creation of new unsupervised feature learning algorithms that can learn image and video features with high quality transfer performance on new tasks
3) The curation of datasets for training and quantifying algorithm performance and the evaluation of our algorithms on these datasets

Each of these objectives was achieved over the course of this three-year project. In the next section, we describe the fulfillment of each of these objectives in more detail.

Scientific Approach and Accomplishments

In order to support research in large-scale Deep Learning algorithms and scalable training frameworks, we worked with our collaborators at the International Computer Science Institute (ICSI) and Yahoo! to create a massive dataset for model training and evaluation. The resulting dataset, called YFCC100M or the Yahoo! Flickr Creative Commons 100M (*Thomee2016*), became the world's largest open/no-strings-attached[1] image and video dataset. YFCC100M consists of about 99.2 million images and 0.8 million videos uploaded to Flickr between 2004 and 2014 under the Creative Commons license. The majority of the images and video also come with associated metadata consisting of content keywords, time/date stamps, locations, and camera types. To expand the metadata of YFCC100M, our ICSI subcontractors also spearheaded an annotation effort that produced event labels for videos in YFCC100M. This multimedia event-detection dataset is called YLI-MED (*Bernd2015*) and is also freely available to the research community for analysis and publication. YFCC100M and YLI-MED provided the large-scale data required for testing the scalability and performance of our deep feature learning algorithms in this LDRD.

In addition to being a great source of training data, YFCC100M also embodies one of the primary big data challenges – how does one find all the data associated with a desired event or concept of interest? Optimistically, it would take over 9,000 hours for a person to look at all of the images and video data in YFCC100M to find all relevant data associated with a single query. This LDRD supported some of ICSI's initial research in

[1] "no-strings-attached" refers to the unrestrictive nature of the dataset that makes it possible for researchers to freely publish their research findings using the data.

automatic event detection on web-scale datasets. ICSI developed the Evento360 system to automatically detect events from images and videos by clustering data based on their timestamps and keyword tags (*Choi2015*). The Evento360 system successfully demonstrated the utility of simple Machine Learning algorithms for helping users organize and query YFCC100M and was the winning system for the ACM Multimedia 2015 Grand Challenge on Event Detection and Summarization.

The Evento360 system is an important first step toward a generalized solution to finding relevant content from large image and video datasets, but its reliance on keyword tags limits performance to well-tagged data. Ultimately, to enable machine-automated search of massive datasets like YFCC100M, we need to build a feature representation for images and video and associate them with semantic information about their contents. The first step toward realizing this vision is building universal (i.e., transferable from one data genre to another) feature representations for images and video which is the core technical contributions of this LDRD. In particular, this LDRD developed the Deep Learning algorithms and architectures for learning transferable feature representations for images and video along with the HPC algorithms that enable training large models using massive amounts of data.

Motivated by the tremendous success of deep convolutional neural networks or CNNs (*LeCun1998*) on image recognition tasks such as ImageNet (*Russakovsky2014, Krizhevsky2012, Szegedy2015, and He2015*), we investigated the transferability of CNN image features learned via supervised training on ImageNet. Our general approach involved specifying various neural network architectures, training the network weights optimizing for object classification performance on ImageNet, and using the outputs of various CNN layers (except for the last layer which was tuned specifically for ImageNet classification) as features for other image recognition tasks. One of the main findings from this LDRD is that this supervised feature learning approach trained using consumer generated photos effectively generalizes to many different image types and applications.

In particular, we demonstrated that CNNs trained on ImageNet consumer generated photos could be adapted/transferred to accurately detect and count cars in overhead imagery (*Mundhenk2016*), detect and localize specific aircraft, classify various vehicle types in synthetic aperture radar (SAR) images, and detect remnant damage sites in images of National Ignition Facility (NIF) optics with high performance (Figure 1). Furthermore, we also investigated the transferability of these features for tagging images by extending the popular TagProp algorithm (*Makadia2008*) using CNN features. We showed that CNN image features achieved better tagging metrics than hand-engineered features by between 8.1% and 16.1% (*Mayhew2016*) on benchmark image tagging datasets IAPR-TC12 (*Grubinger2006*) and ESP Games (*VonAhn2004*).

We also explored various CNN architectures varying the size and depth of network layers and developed a hybrid network architecture that blended the best features from two industry leading CNNs: the multi-resolution convolutional kernels of Google's

Inception network (*Szegedy2015*) and the residual shortcut connections of Microsoft's ResNet (*He2015*). By blending these two architectures, our ResCeption architecture (*Mundhenk2016*) is able to learn image features of varying sizes and allows for rich hierarchies of deeply stacked layers. We demonstrated that ResCeption networks achieve large performance improvement versus the AlexNet architecture (*Krizhevsky2012*) and a modest one over Inception (19.3% and 0.89% relative gain respectively) on LLNL's overhead car counting dataset (*Mundhenk2016*).

The next significant accomplishment on this LDRD was the investigation and development of new unsupervised feature learning algorithms that are able to learn useful feature representations without the need for labeled data. This project investigated three different approaches for unsupervised feature learning: 1) extensions of autoencoder networks, 2) self-supervised context prediction networks, and 3) generative adversarial networks.

The standard Deep Learning approach to unsupervised feature learning is the autoencoder (*Rumelhart1986*). Autoencoders project input data into intermediate feature representations from which the original input can be well reconstructed by the output layer. Working with our Stanford subcontractors, we extended their work on the Google Brain (*Le2012*), which famously learned image concepts such as human faces, silhouettes of people, and cat faces, using 1 million unlabeled YouTube images. We adapted Stanford's model-parallel Deep Learning training software for our Graphics Processing Unit (GPU) supercomputing cluster and trained an autoencoder network similar to the one presented in (*Le2012*) except with 15 times more network parameters, 100 times more training data using YFCC100M (*Ni2015*), and using only 100 GPU compute nodes. The resulting autoencoder was the largest unsupervised neural network for images ever trained, and it too learned interesting concepts like airplanes, fireworks, latticed towers, and pictures with text all without labeled image data (Figure 2). While we were able to train this massive network in eight days, we realized our existing codebase would not scale beyond 100 nodes due to communication bottlenecks arising from the serial staging of input data and diminishing amount of work per node resulting from a limited form of parallelization tied to input image sizes. This led us to develop the Livermore Big Artificial Neural Network (LBANN) training toolkit.

LBANN addresses the scaling challenges by providing staged and distributed data ingest and by leveraging the Elemental library (*Poulson2013*) to distribute both the model parameters and data matrices. We focused on optimizing LBANN for massive deep fully-connected neural networks and achieved efficient weak and strong scaling performance (*VanEssen2015*). LBANN achieves this via model parallelism where individual compute nodes are responsible for different subsets of the model and data parallelism where different nodes compute gradients for different mini-batches of data. While LBANN has efficiently distributed the computation of gradient descent optimization at the heart of neural network training, its data parallelism functionality is tantamount to computing a gradient on a very large batch of data very quickly. Unfortunately, large batch sizes do

not necessarily lead to faster overall convergence times for training. Developing data parallel training algorithms that lead to faster training convergence remains an open research topic.

The second major approach to unsupervised feature learning is the self-supervised context prediction networks (*Doersch2015*). The idea is to train neural networks to learn to predict the structure of the co-occurrence of input data. In the realm of images, an example context prediction task is to classify the location of a randomly selected image patch relative to another randomly selected patch. (Figure 3) illustrates an example of this – the job of the neural network is to learn that the ear patch is northwest of the nose patch. In order to perform the context prediction task well, the network must learn feature representations that capture the inherent structure of the data, and the hypothesis is that these features are generalizable to other image recognition tasks. The key benefit of context prediction is that it is self-supervised, i.e., no human generated labels are required because the computer controls what contexts to present for training and knows the relationship of these contexts.

In this LDRD, we extended (*Doersch2015*) by exploring new network architectures like our ResCeption network and new context prediction tasks such as classifying the orientation of three image patches instead of two. One of the drawbacks of two image patch contexts is that often two image patches randomly come from uninformative regions of the image, e.g., two patches of the sky. Our triple patches strike a good balance between computational complexity and informative patch selection. We tested the transferability of the features learned via context prediction with those of the gold-standard supervised CNN features and found that the context prediction features achieved impressive transfer performance. For example, when training the features on ImageNet data and fine-tuning networks on the CompCars classification dataset, we found that our triple-patch features outperformed Doersch's double-patch features (87.2% vs 84.8%). Compared to the supervised CNN features, this result is only 3.2% absolute worse, which is quite remarkable given the fact that our triple-patch features are learned without costly hand labels. We also performed some initial experiments with training on the much larger YFCC100M unlabeled images and found that increasing the amount of training data leads to some improvements in classification performance, but unfortunately we were not able to fully test out the performance increases from also increasing model sizes due to time constraints. Follow-on work will explore this avenue, build an understanding of what makes a context prediction task more transferrable to other image recognition tasks, and generalize this method to video data.

The final approach for unsupervised feature learning that we extended was the generative adversarial network (GAN) training. GANs (*Goodfellow2014*) turns network training into a game played by two separate neural networks: one learns feature representations useful for generating realistic fake data, while the other learns features for catching forgeries. Like the other unsupervised feature learning techniques, GANs take advantage of large amounts of unlabeled data. Furthermore, the GANs game can

be played almost endlessly until both networks achieve an equilibrium point where both are winning as much as losing. Achieving this equilibrium can be challenging when one of the networks overpowers the other resulting in suboptimal features learned by both. We developed new adversarial training algorithms that stage the game in successively harder games for both networks to prevent one network from dominating. We demonstrated that this "curriculum learning" approach leads to dramatically more stable learning that converges more quickly to the equilibrium point. We applied our curriculum learning approach to learning features for images as well as video data, and found that our GANs features outperformed the standard autoencoder based features (Figure 4). Furthermore, GANs features excelled in cases when labeled data for the subsequent task adaptation was limited. This is a particularly important characteristic for unsupervised feature learning approaches given the costly nature of obtaining labeled data.

Our final significant scientific accomplishment on this LDRD is the development of a new learning algorithm for mapping image and text features into a shared feature space where images and text of related concepts are proximal. Such a shared feature space enables image tagging and image search from either keywords or query images. Our bimodal learning algorithms start from separate text and image neural networks pre-trained in the various ways described earlier. For the text network, we pre-trained Word2Vec (*Mikolov2013*) or GloVe (*Pennington2014*) on 20 years of New York Times, while the image network is a standard CNN like VGG16 (*Simonyan2015*) trained on ImageNet. Finally, we learn several mapping layers from the penultimate layer of VGG16 to the Word2Vec feature space using a smaller image tagging dataset like ESP Games or IAPR-TC12. This bimodal learning system, called Image2Vec (*Boakye2016*), maps images and text into a shared semantic feature space (Figure 5). We benchmarked the performance of Image2Vec on the image tagging problem and showed that the performance was only slightly worse than the baseline TagProp systems explored in (*Mayhew2016*), but unlike TagProp, Image2Vec was able to perform zero-shot learning. Zero-shot learning in our context is achieved when the system is able to associate correct tags to images without ever seeing examples of them during training. For example, Image2Vec was able to tag images of "missiles" despite never having seen images tagged with "missile" in the training data. These developments lay the ground work for future research on generalizing these approaches to multimodal feature space learning (for unifying the features for images, text, video, etc.) that will enable generalized multimodal search.

Impact on Mission

The new capabilities developed in this LDRD have broad applicability to almost every laboratory program that collects and use imagery and video. In particular, the LDRD technologies have helped spawn several new projects: six in Global Security (GS) totaling 4 FTEs, one in NIF, one in Advanced Manufacturing, and another in the Biological Applications of Advanced Strategic Computing (BAASiC) program. Specifically, our research on scaling Deep Learning frameworks, transferring image and video features

for overhead recognition tasks, and bimodal feature learning frameworks led to the successful capture of the new projects in GS. NIF invested in developing a new system for detecting remnant damage in their optics based on our Deep Learning technologies, and our pilot experiments for classifying laser weld quality in selective laser sintering devices will be the basis of a new Machine Learning LDRD for Advanced Manufacturing. Finally, the unsupervised feature learning techniques at scale serves as one of the key technologies for understanding large scale image and video datasets for several projects in BAASiC.

This LDRD sits at the very exciting intersection of HPC and Machine Learning and was instrumental in recruiting seven new research staff including a world leader in multimedia retrieval research. Our project has also attracted over half a dozen summer researchers. Additionally, this project helped LLNL establish new strategic partnerships with academia and industry including Stanford, UC Berkeley, ICSI, In-Q-Tel, IBM, and NVIDIA. Our researchers have served on prestigious national supercomping panels and conference organizing committees in Machine Learning and HPC. This LDRD has helped establish LLNL as a leader in large-scale Deep Learning research.

Conclusion

This LDRD successfully developed and demonstrated several major new capabilities: first, we developed a scalable neural network training toolkit for fitting multi-billion-parameter neural networks to massive datasets; second, we created several unsupervised Deep Learning algorithms for images and video and demonstrated high-quality transferability of these feature to new tasks; third, we invented a bimodal feature learning algorithm that successfully mapped image and text features into a single feature space enabling semantic image search and tagging; finally, we helped create the largest open image and video dataset for multimedia research. These capabilities lay the foundation for further research on learning universal multimodal feature representations at massive scales and support the development of a new generation of situational awareness, nuclear nonproliferation, and counter-WMD programs.

References

O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. S. Bernstein, A. C. Berg and F.-F. Li (2014), "ImageNet Large Scale Visual Recognition Challenge," *Computing Research Repository (CoRR)*, arXiv:1409.0575.

A. Krizhevsky, I. Sutskever and G. Hinton (2012), "ImageNet Classification with Deep Convolutional Neural Networks," *Advances in Neural Information Processing Systems (NIPS)*.

C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke and A. Rabinovich (2015), "Going Deeper with Convolutions," *Proceedings of the Conference on Computer Vision and Pattern Recognition (CVPR)*.

K. He, X. Zhang, S. Ren and J. Sun (2015), "Deep Residual Learning for Image Recognition," *Computing Research Repository (CoRR)*, arXiv:1512.03385.

B. Thomee, D. Shamma, G. Friedland, B. Elizalde, K. Ni, D. Poland, D. Borth, L. Li (2016), "YFCC100M: The New Data and New Challenges in Multimedia Research," *Communications of the ACM (Commun ACM)*.

J. Bernd, D. Borth, B. Elizalde, G. Friedland, H. Gallagher, L. Gottlieb, A. Janin, S. Karabashlieva, J. Takahashi, J. Won (2015), "The YLI-MED Corpus: Characteristics, Procedures, and Plans", *Computing Research Repository (CoRR)*, arXiv:1503.04250.

J. Choi, B. Thomee, G. Friedland, L. Cao, K. Ni, D. Borth, B. Elizalde, L. Gottlieb, C. Carrano, R. Pearce, D. Poland (2014), "The Placing Task: A Large-Scale Geo-Estimation Challenge for Social-Media Videos and Images", *Proceedings of the Workshop on Geotagging and Its Applications in Multimedia (GeoMM)*.

Y. LeCun, L. Bottou, Y. Bengio and P. Haffner (1998), "Gradient-Based Learning Applied to Document Recognition," *Proceedings of the IEEE*, 86(11).

T. Mundhenk, G. Konjevod, W. Sakla, and K. Boakye (2016), "A Large Contextual Dataset for Classification, Detection and Counting of Cars with Deep Learning," *Proceedings of European Conference on Computer Vision (ECCV)*.

A. Makadia, V. Pavlovic, and S. Kumar (2008), "A New Baseline for Image Annotation," *Proceedings of the European Conference on Computer Vision (ECCV)*.

M. Mayhew, B. Chen, and K. Ni (2016), "Assessing Semantic Information in Convolutional Neural Network Representations of Images via Image Annotation," *Proceedings of the International Conference on Image Processing (ICIP)*.

M. Grubinger, P. Clough, H. Muller, and T. Deselaers (2006), "The IAPR Benchmark: A New Evaluation Resource for Visual Information Systems," *Proceedings of the International Conference on Language Resources and Evaluation (LREC)*.

L. Von Ahn and L. Dabbish (2004), "Labeling Images with a Computer Game," *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI)*.

D. Rumelhart, G. Hinton, and R. Williams (1986), "Learning internal representations by error propagation," *Parallel Distributed Processing,* 1.

Q. Le, M. Ranzato, R. Monga, M. Devin, K. Chen, G. Corrado, J. Dean and A. Ng (2012), "Building high-level features using large scale unsupervised learning," *Proceedings of the International Conference on Machine Learning (ICML)*.

K. Ni, R. Pearce, K. Boakye, B. Van Essen, D. Borth, B. Chen, E. Wang (2015), "Large-scale deep learning on the YFCC100M dataset," *Computing Research Repository (CoRR)* arXiv:1502.03409.

J. Poulson, B. Marker, R. van de Geijn, J. Hammond, and N. Romero (2013), "Elemental: A new framework for distributed memory dense matrix computations," *ACM Transactions on Mathematical Software (TOMS)*, 39(2).

B. Van Essen, H. Kim, R. Pearce, K. Boakye, and B. Chen (2015), "LBANN: Livermore Big Artificial Neural Network HPC Toolkit," *Proceedings of the Workshop on Machine Learning in High-Performance Computing Environments (MLHPC)*.

C. Doersch, A. Gupta, and A. Efros (2015), "Unsupervised Visual Representation Learning by Context Prediction," *Proceedings of the International Conference on Computer Vision (ICCV)*.

I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville and Y. Bengio (2014), "Generative Adversarial Nets," *Advances in Neural Information Processing Systems (NIPS)*.

T. Mikolov, I. Sutskever, K. Chen, G. S. Corrado and J. Dean (2013), "Distributed representations of words and phrases and their compositionality," *Advances in Neural Information Processing Systems (NIPS)*.

J. Pennington, R. Socher and C. D. Manning (2014), "GloVe: Global Vectors for Word Representations," *Empirical Methods in Natural Language Processing (EMNLP)*.

K. Simonyan and A. Zisserman (2015), "Very Deep Convolutional Networks for Large-Scale Image Recognition," *Proceedings of the International Conference on Learning Representations (ICLR)*.

K. Boakye, C. Carrano, M. Gokhale, H. Kim, M. Mayhew, N. Mundhenk, B. Ng, K. Ni, R. Pearce, D. Poland, B. V. Essen, E. Wang, D. Widemann, A. Wilson, and B. Chen (2016), “A Deep Learning Framework for Multimodal Pattern Discovery,” *Lawrence Livermore National Laboratory Technical Report (LLNL TR)*.

Feature image/figures (with caption)

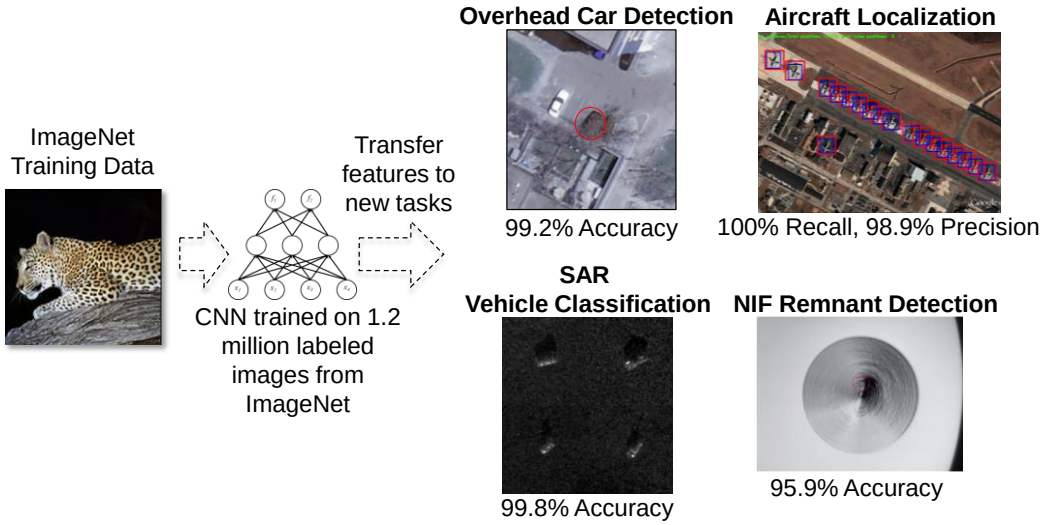

Figure 1 – Convolutional neural network (CNN) image features learned via supervised training on crowdsourced images (ImageNet) can be adapted for high performance classification on recognition tasks in very different image genres. We have demonstrated the transferability of such image features for overhead car detection, aircraft localization, vehicle classification in SAR imagery, and defect remnant detection in NIF optics.

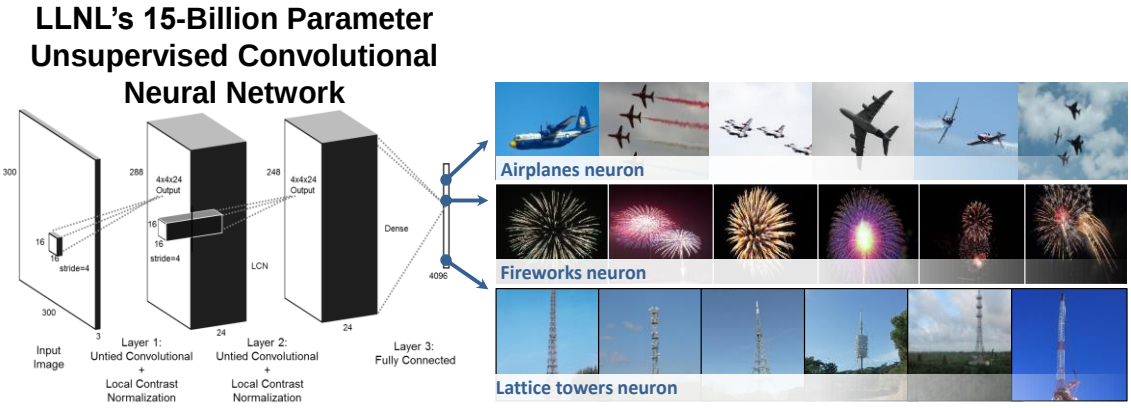

Figure 2 – In collaboration with Stanford, LLNL trained the largest convolutional neural network (15-billion weights and biases) for learning unsupervised image features from ~100 million images in YFCC100M. Training completed in 8 days using 100 nodes of LLNL's GPU HPC cluster. This network learned interesting high-level concepts including airplanes, fireworks, lattice towers, etc. without being provided explicit training images labeled with these concepts.

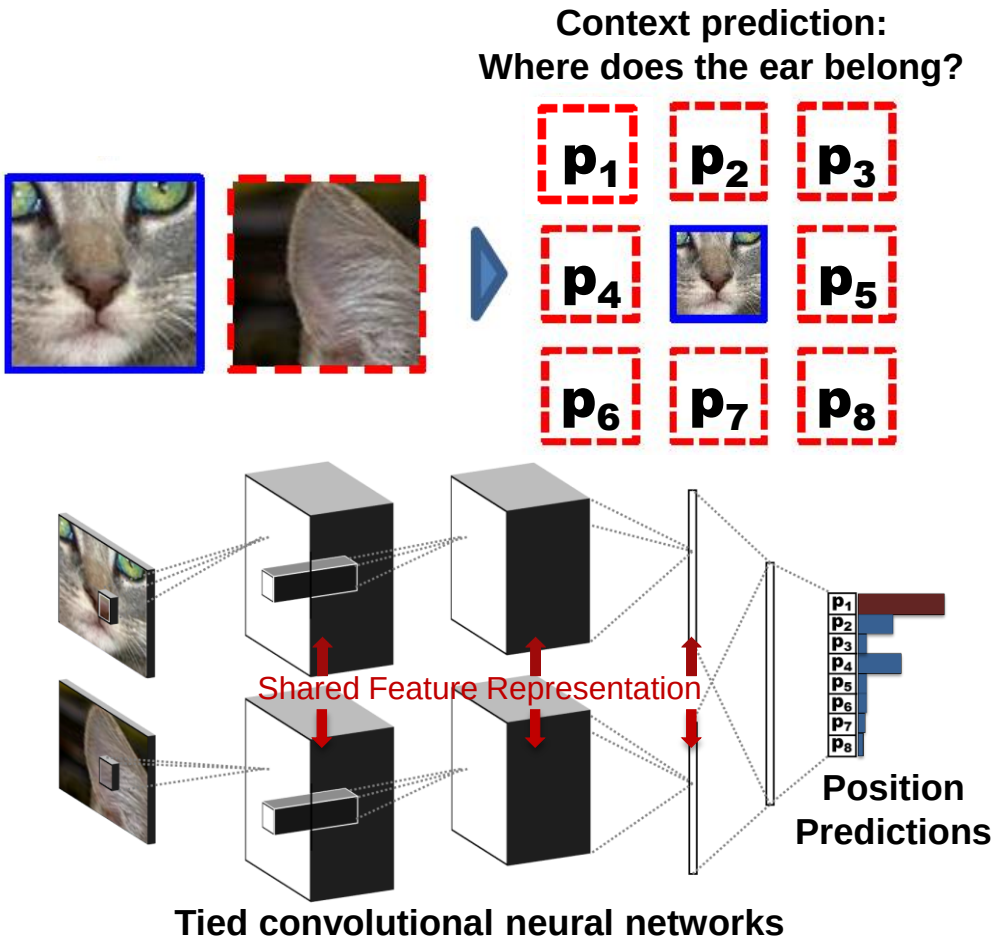

Figure 3 – (Adapted from *Doersch2015*) Unsupervised learning of image features via context prediction involves training convolutional neural networks (CNNs) to learn how to classify the position of a randomly selected image patch relative to another image patch. Upon successful training completion, the CNNs will have learned a shared feature representation of images that captures the structure of objects critical for high performance context prediction and is transferrable to other classification tasks.

Feature Learning on CFAR-10 transfer on MNIST (Image recognition error rates)	Labeled transfer training data	
	1k	10k
Autoencoder	7.23%	1.88%
GANs w/Curriculum Learning	**5.43%**	**1.73%**

Feature Learning on Sports-1M transfer to HMDB-51 (Video Action Classification)	**Accuracy**
Predictive Autoencoder	43.1%
GANs w/Curriculum Learning	**43.5%**

Figure 4 – Transfer classification results for standard autoencoders versus our generative adversarial networks (GANs) with curriculum learning for unsupervised feature learning. (Left) Image recognition results for features trained on CFAR-10 images and transferred to MNIST digit recognition demonstrate the effectiveness of GANs especially when the amount of labeled transfer training data is limited. (Right) Video classification results for GANs features trained on Sports-1Million data and transferred to video action classification on HMDB-51 outperform those of autoencoders.

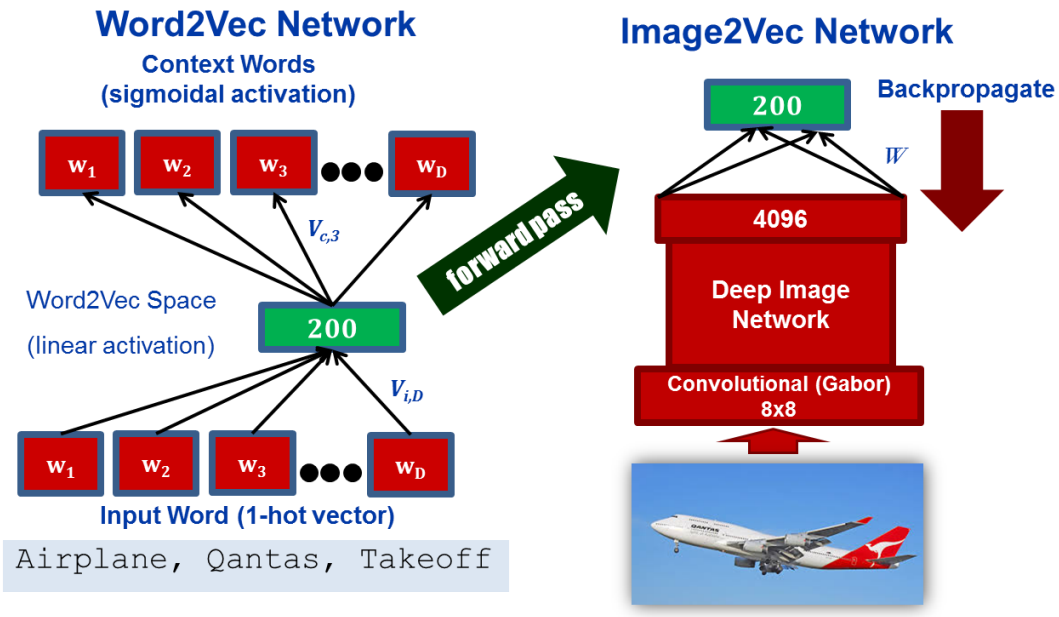

Figure 5 – Image2Vec learns to map image features to the semantic feature space learned by Word2Vec. The Word2Vec and image feature spaces are first separately pre-trained. Image2Vec then takes training examples of images with corresponding word tags, projects each tag into a 200-dimensional Word2Vec feature vector, and uses this feature vector as the training target for learning the image projection weights. The resulting image+text feature space enables zero-shot image tagging and image retrieval.

List of publications, patents, and awards attributable to LDRD

B. Huval, A. Coates, A. Ng (2013), "Deep learning for class-generic object detection," *Computing Research Repository (CoRR)*, arXiv:1312.6885.

J. Choi, B. Thomee, G. Friedland, L. Cao, K. Ni, D. Borth, B. Elizalde, L. Gottlieb, C. Carrano, R. Pearce, D. Poland (2014), "The Placing Task: A Large-Scale Geo-Estimation Challenge for Social-Media Videos and Images", *Proceedings of the Workshop on Geotagging and Its Applications in Multimedia (GeoMM)*.

K. Ni, R. Pearce, K. Boakye, B. Van Essen, D. Borth, B. Chen, E. Wang (2015), "Large-scale deep learning on the YFCC100M dataset," *Computing Research Repository (CoRR)* arXiv:1502.03409.

B. Thomee, D. Shamma, G. Friedland, B. Elizalde, K. Ni, D. Poland, D. Borth, L. Li (2015), "YFCC100M: The New Data and New Challenges in Multimedia Research", *Computing Research Repository (CoRR)*, arXiv:1503.01817.

J. Bernd, D. Borth, B. Elizalde, G. Friedland, H. Gallagher, L. Gottlieb, A. Janin, S. Karabashlieva, J. Takahashi, J. Won (2015), "The YLI-MED Corpus: Characteristics, Procedures, and Plans", *Computing Research Repository (CoRR)*, arXiv:1503.04250.

T. Narihira, D. Borth, S. Yu, K. Ni, T. Darrell (2015), "Mapping Images to Sentiment Adjective Noun Pairs with Factorized Neural Nets," *Computing Research Repository (CoRR)*, arXiv:1511.06838.

B. Van Essen, H. Kim, R. Pearce, K. Boakye, and B. Chen (2015), "LBANN: Livermore Big Artificial Neural Network HPC Toolkit," *Proceedings of the Workshop on Machine Learning in High-Performance Computing Environments (MLHPC)*.

J. Choi, E. Kim, M. Larson, G. Friedland, A. Hanjalic (2015), "Evento 360: Social Event Discovery from Web-scale Multimedia Collection," *Proceedings of the ACM Multimedia Conference (MM)*.

K. Boakye, C. Carrano, M. Gokhale, H. Kim, M. Mayhew, N. Mundhenk, B. Ng, K. Ni, R. Pearce, D. Poland, B. V. Essen, E. Wang, D. Widemann, A. Wilson, and B. Chen (2016), "A deep learning framework for multimodal pattern discovery," *Lawrence Livermore National Laboratory Technical Report (LLNL TR)*.

B. Thomee, D. Shamma, G. Friedland, B. Elizalde, K. Ni, D. Poland, D. Borth, L. Li (2016), "YFCC100M: The New Data and New Challenges in Multimedia Research," *Communications of the ACM (Commun ACM)*.

M. Mayhew, B. Chen, and K. Ni (2016), "Assessing Semantic Information in

Convolutional Neural Network Representations of Images via Image Annotation," *Proceedings of the International Conference on Image Processing (ICIP)*.

T. Mundhenk, G. Konjevod, W. Sakla, and K. Boakye (2016), "A Large Contextual Dataset for Classification, Detection and Counting of Cars with Deep Learning," *Proceedings of European Conference on Computer Vision (ECCV)*.

L. Jing, B. Liu, J. Choi, A. Janin, J. Bernd, M. Mahoney, G. Friedland (2016), "A Discriminative and Compact Audio Representation for Event Detection," *Proceedings of the ACM Multimedia Conference (MM)*.

J. Sawada, et. al. (2016), "TrueNorth Ecosystem for Brain-Inspired Computing: Scalable Systems, Software, and Applications," *Proceedings* of the *International Conference for High Performance Computing*, *Networking*, *Storage and Analysis (SC)*.