Q: Anyone using Caffe on Windows Server?
A: I have not tried it. There is an official Windows version you could try this on.

Q: I was under the impression that Caffe can be installed on OSX without a Nvidia GPU but can then be run with -cpu only flag? That is - it does not need to be compiled with cpu only flag.
A:Yes set the CPU flag in the Makefile.

Q: Would CUDA in anyway help in deep learning?
A: CUDA and cuDNN are used in a large number of Deep Learning applications and framework. Answer is yes.

Q: Does it need labeled data? Or can it also train unsupervised?
A: Yes it does. You need to have data labeled, (show slide). I have seen some interesting work where folks have used feature responses to help them classify objects the network has not been trained on. You may want to look at that -http://yosinski.com/media/papers/Yosinski__2015__ICML_DL__Understanding_Neural_Networks_Through_Deep_Visualization__.pdf

Q: Is there any way to extract weights of the trained net from a .caffemodel file without changing caffe sources?
A: Yes, you can use get.data(). See the interfaces here:
http://caffe.berkeleyvision.org/tutorial/interfaces.html

Q: Can trained models be shared between Caffe, Torch, and Theano?
A: Yes on Torch, you will find this listed on the Torch page
https://github.com/torch/torch7/wiki/Cheatsheet and the github site for the caffe module is
https://github.com/szagoruyko/loadcaffe

Q: Are caffemodel files going to be backward compatible as new versions of caffe come out?
A: That was the case in past and it is expected to be the case in the future in caffe. But this will depends and is supported by caffe. The current implementation has an on-the-fly upgrade model mechanism for this.

Q: Can we use caffe by a distributed way on a computer network?
A: Depends on what you mean. Caffe is open source so you can fork and modify it as you want. There is a multi-GPU caffe branch- this allows you to perform a single training on more than one GPU. People have discussed this - https://github.com/BVLC/caffe/issues/653. This has also been a discussion on DIGITS - https://github.com/NVIDIA/DIGITS/issues/108.

Q: why caffe or other major deep learning framework don't provide any interface with R? Are there any plans to provide it?
A: This is a great question. I know that they have a matlab and python interface. You can also program in C++. We work with the Caffe folks on things such a cuDNN. One thing you can do is

reach out to them directly through their users group or github. One thing that is nice about these open forums is you can view what other researchers are working on and post your own questions.

Q: So far, images seems to have been mentioned very often.. Is caffe very image centric? Is it just as easy to handle say, text input? (for NLP type things?)
A: It is image centric, but if you can input your data in the form of a array, like an image it may work. I have not done this myself, but I believe if you can input the data as a 1x100 or what ever array you can use it in Caffe. We have been asked about signals using DIGITS and they is one way we handle this. Rather than your input images being 256x256 they will be 1x###. Remember, to include these changes in your deploy file.

Q: is there a way to find how many layers should we use??
A: this is an experimental process. Some are using very deep networks. Here is a presentation by Google on their ImageNet competition -
http://image-net.org/challenges/LSVRC/2014/slides/GoogLeNet.pptx

Q: Does caffe support using 3D volumes (ie medical images) as an input?
A: It supports this via the color channels already. I noticed on github - there was a discussion about adding a `vol2col` analogue of `im2col` for an explicit 3D convolution.
-https://github.com/BVLC/caffe/pull/1486

Q: Is it true that all the weights and activations are stored in single precision floating point format?
A: Yes, everything is stored in single precision format

Q: Can categories be inferred from directory naming conventions?
A: Yes, with Caffe the categories can be specified by the naming convention of the directories, or by supplying a text file listing the categories and images in those categories

Q: Is there a tool to visualize the protobuf files? Please provide a link.
A: Yes, the NVIDIA Digits toolkit includes functionality to visualize networks described in a Caffe prototext file. https://developer.nvidia.com/digits

Q: In the covolution moves over the image, does the value of its kernel change as it moves or stays fixed for the entire image?
A: The data in kernels or filters stays constant during the convolution, while the weight of that filter convolved with the data is the output for the activation image in the next layer

Q: Does Caffe in DIGITS (SW) support more than one DIGITS Box i.e. does it also work over Network?
A: Currently, DIGITS 2.0 using Caffe supports multi-gpu data parallel training on one node, but there is no support for training over multiple nodes over the network.

Q: How does stride affect the result?
A: Setting a stride other than 1 in the convolutional layer causes the output images to be smaller by a factor of the stride. This is useful in the first layers when using large input images, as it reduces the computational workload in later layers

Q: Is the Titan Z supported, or is the Titan X just as fast for training?
A: Both the Titan Z and Titan X are supported. Titan Z has two Kepler GPUs. Titan X has a single Maxwell GPU (newer), and is about 40% faster for single GPU training.

Q: Does installing Digits automatically install caffe? I am a little bit overwhelmed trying to understand which parts of CUDA 7.5, CUDNN 3, DIGITS 2 and Caffe are included in each other.
A: Yes, installing Digits also installs the correct version of Caffe and cuDNN automatically.

Q: Does Caffe work with CUDA 7.5?
A: CUDA 7.5 production version hasn't been released yet. But generally speaking, when new versions of the CUDA toolkit are released, they are supported by the major frameworks such as Caffe.

Q: can i train Caffe on a non-recent nvidia card like the geforce gtx-480?
A: According to the Caffe documentation you can, but you might sacrifice the ability to use newer features of Caffe/CUDNN that are only supported in newer GPUs.

Q: Is the Titan Z supported, or is the Titan X just as fast for training?
A: Titan Z is supported.

Q: I want to extend a published model, let's say AlexNet, and when training the new model, I want to keep old parameters in AlexNet unchanged, and only want to update newly added parameters, is there any way I can do it in caffe?
A: You can adjust/modify the parameters in any way you see fit. You could start with AlexNet, and then add more layers of your choosing, adjusting those parameters however you like.

Q: Can categories be inferred from directory naming conventions?
A: One can use separate directories or a text file which allows for easy categories setup (with a few shell commands for example).

Q: If I use python interface to describe a costum layer, will that code run on GPU if I use GPU mode?
A: If the custom layer can benefit from GPU acceleration, yes.

Q: What is the easiest way to use multi GPU ? Caffe or Theano or Torch ? thks

A: All three framework have multi-gpu implementations that are different so it will depend on the problem one is looking at. Using NVIDIA Digits makes multi-GPU easy for image classification on top of caffe.

Q: what does the parameter "lr_mult" specify?
A: lr_mult is the multiplier on the global learning rate for the parameter.

Q: Is there a minimum size of dataset needed in order to get a good speedup on Titan X GPU? In the past I have seen that the GPU pipelines need to be filled in order to get needed speedups
A: Good question. Generally the nature of DL requires an extensive training data set. That often implies you have ample work to keep one or more GPUs fully busy. The model parameters also impact the performance but for most cases you would likely see plenty of work to keep a Titan X busy. The good news is that caffe, cuDNN, and DIGITS all do a great job of making sure you are getting the maximum value out of whatever GPU resources you have available. In summary, use a framework that uses cuDNN and you should be seeing very good speedups with a Titan X.

Q: Does Caffe support multiple GPUs? Can caffe run on multiple GPU?
A: There is a github fork of caffe that does support multiple GPUs. Here's a link with more info:
https://github.com/BVLC/caffe/pull/1148
Work is ongoing and future releases of cuDNN will support multi-GPU training. Expect increased support for multi-GPU across many DL frameworks soon..

Q: If there are multiple GPU on single system, do we need to explicitly specify which GPU to use? Is there any plans for unified models?
A: Great question. In general, applications (like caffe) will use first available GPU following the order of enumeration of the devices. assuming linux, you have system tools (nvidia-smi, emv variable CUDA_VISIBLE_DEVICES) that can control what GPUs are available for your application to use and in what priority order they will be access. Here's a link to the caffe documentation that might help you as well:
http://caffe.berkeleyvision.org/tutorial/interfaces.html

Q: could you elaborate on " fully connected layer " ?
A: here's a link to help: http://cs231n.github.io/convolutional-networks/#fc Simplistically, a fully connected layer connects all the activations of the previous layer to the "neurons" of the fully connected layer

Q: Can Caffe be coupled with DSP?
A: I'm not sure I fully understand the question but here's some comments. In general you can support a DSP PCIE card in the same system that you have GPUs for accelerating DL on caffe so you could use the DSP for some streaming processes and caffe and GPUs could then use the resulting output as input for DL based use. This would be an exercise left to the reader to

construct :^). If you are asking can a DSP card accelerate the training or classification tasks within caffe, I'm unaware of anyone who has forked a version of caffe that will use a DSP card.

Q: What are some server-grade GPUs?
A: All the GPUs in the Tesla line are server-grade
http://www.nvidia.com/object/tesla-supercomputing-solutions.html

Q: I would also like to know if Nervana Neon is recommended as well for university labs?
A: If Nervana Neon has the functionality you desire you could certainly try it. NVIDIA doesn't make any recommendations on Nervana Neon currently.

Q: What is the best framework to use for speech recognition ?
A: One popular framework for speech is Kaldi. http://kaldi.sourceforge.net/

Q: In practise, people use more Theano or Caffe, or some other framework?
A: It really depends what you want to do. Each framework has different strengths. We have seen lots of usage of Theano, Caffe and Torch, to name a few of the major frameworks.

Q: Are caffemodel files going to be backward compatible as new versions of caffe come out?
A: That was the case in past and it is expected to be the case in the future in caffe. But this will depends and is supported by caffe. The current implementation has an on-the-fly upgrade model mechanism for this.

Q: can i use caffe without a GPU?
A: Yes. Performance will be different. Caffe will have to be compiled and installed with a CPU only flag.

Q: Do we also give negative samples as a negative class for training?
A: It can be done. It often improves classification for example.

Q: Why does LSTM exist as a pull request only? How do you compare it's readiness as compared to the LSTM implementation on torch?
A: We can't talk for the caffe developers but this might happen soon since tests passed very recently, see: https://github.com/BVLC/caffe/pull/2033
Torch LSTM implementation is different so the choice will be application dependent.

Q: is there any reason why one would work with theano over caffe?
A: The approach of both frameworks are very different. Caffe is a DL framework. Theano can be seen as a compiler. It will be application dependent.

Q: Does cuDNN require cuda 7.5?
A: No. cuDNN works with cuda 6.5 and cuda 7.0.

Q: does caffe support rnn layer?
A: See here: https://github.com/BVLC/caffe/pull/1873. It is working in caffe and could be integrated in caffe very soon since it passed the tests.

Q: could you give me a refrence in using CNNs in sensor fusing? it was asked already and was promised in the lecture.
A: Check this paper from Stanford: http://arxiv.org/pdf/1504.01716.pdf

Q: Can i read images from OpenCV and send it to Caffe in memory without using files on disk?
A: Yes. You can pass them or their pixel values/data as an array for example.

Q: What is strong typed?
A: Check here: https://en.wikipedia.org/wiki/Strong_and_weak_typing

Q: Are trained caffe models be used to extract features and pass it to other deep learning frameworks (like Theano?)
A: Yes. It can be converted to Lasagne for example, see https://github.com/kitofans/caffe-theano-conversion and then used with Lasagne+Theano

Q: It is possible to define custom layers and activation functions and use in the network?
A: Yes. See the caffe manual to do that here: http://caffe.berkeleyvision.org/tutorial/net_layer_blob.html

Q: Does a model learnt on one platform perform good on another platform. Is it portable ?
A: Yes. You can for example train on a supercomputer/workstation and use for inference on Jetson TK1. See http://devblogs.nvidia.com/parallelforall/embedded-machine-learning-cudnn-deep-neural-network-library-jetson-tk1/

Q: How many single-node GPUs does Caffe scale to ? Can it use GPUs on multiple nodes (with MPI for example) ?
A: The multi-gpu branch of Caffe can scale to as many GPUs as you can fit in a node. It does not currently scale across nodes.

Q: Can Caffe utilize depth images alongside RGB?
A: Yes, you just add the depth in as a 4th channel and set your input tensor accordingly

Q: If there is a preexisting model that identifies giraffes and another that identifies horses, how do you know which ones to choose if you want to transfer knowledge to identify cats, for example?

A: You can train a model on one set of images, giraffes and horses in this example, then modify the final layers to include new categories and re-train them on the new categories. This is called fine-tuning

Q: Can I deploy a trained model on an embedded system that's not a Jetson?
A: Yes you can, as the model is just a large matrix of numbers that represent the weights connecting the neurons. Classification just won't be as fast without a GPU

Q: Does nvidia have any cloud computing platform/service to train CNNS?
A: You can make use of Amazon Web Services to train on GPUs

Q: For speech, does storing spectrograms as images and training the network work?
A: Yes, you can store speech in spectrograms as images and apply convolutional neural networks in Caffe.

Q: is there a visualization tool for the layers files? they go bottom to top and not left to right
A: Our DIGITS software includes a tool for visualizing Caffe network prototext files.

Q: Should I upgrade to 980 GTX or Titan X now or wait for Pascal? Will pascal be indeed 10 x faster?
A: You could definitely benefit from using a TitanX for the time being. Pascal should have significant performance increases due to the fast stacked DRAM, NVLINK interconnect, and half-precision math capability

Q: Can I download the source code of cuDnn?
A: No, cuDNN is only available as a library at present

Q: What is the advantage of using batch size >1?
A: Using a larger batch size allows the GPU to train on multiple images at a time, greatly boosting performance.

Q: does caffe support online learning?
A: At this time Caffe uses SGD, NAG and AdaGrad for learning. You can edit the solver.prototxt file to change the batch size and rate that weights are updates.

Q: what does the "shufflng" tool do?
A: The data is shuffled within the training, validation, and test dataset. All of your images are still in that data set. If the data is in a certain order it can sometimes bias the gradient and lead to poor convergence.

Q: is there extensive inline comments in Caffee
A: Caffe is open source and there are some comments. If you find that you are having a hard time interpreting the code, you can always reach out to their Caffe-users google group.

Q: Where can I learn how to format the database in order to correctly use Caffe?
A: Caffe has a lot of documentation on how to use their framework and format their data. Check out their training with ImageNet example.

Q: what does Level DB mean?
A: Leveldb is a database format that Caffe accepts. Your images are formatted into a database before being ingested by the network during training. Here is a github link to levedb -https://github.com/google/leveldb Caffe allows a few different formats in addition to leveldb, one of which is lmdb.

Q: is local batch size defined in the solver file?
A: Batch size is defined in your netwrok file, train_val.prototxt. You can define the batch size for training and testing in this file.

Q: How much is the maximum number of iterations to my network?
A: You can change the maximum number of iteration in your solver.prototxt file. The right number of iterations for you depends on your data and network configuration. I try to train until I stop seeing a significant improvement in performance.

Q: Does Caffe support unsupervised deep learning?
A: Not at this time. You need to label all of your input categories.

Q: I assume the nets can be accessed from a C++ based API?
A: There is a simple classification example provided with Caffe. Here is a link to an example for using it -http://caffe.berkeleyvision.org/gathered/examples/cpp_classification.html

Q: is there a comparison of the processing difference some commercial GPUs might give vs server-grade? I'd like to know how bog a difference there is between my GTX-660 and a Tesla. specifically in using Caffe, Torch or Theano. SOmething beyond simple flop comparison
A: I don't have a list of performance numbers that compare the GTX 660 to our Tesla line. However, one thing you can is build caffe on your machine and train with your GTC 660. Then you can compare your iterations times to others posted online. Caffe has included some training times with the K40 here - http://caffe.berkeleyvision.org/performance_hardware.html

Q: What is "Caffe Model Zoo"? Is it just a collection of pre-trained models included with Caffe?
A: Model Zoo is a collection of pretrainined caffe models. I have used them a bit for simple classification tests. If i have some data and am wondering how well this data might be classified by a network trained with the ImageNet data. I have also used these pretrained networks for fine-tuning.

Q: Any references to improve skills in finetuning a network?

A: I used this the Caffe example to help me get started when I was learning to use Caffe- http://caffe.berkeleyvision.org/gathered/examples/finetune_flickr_style.html.

Q: Can I use any kinds of neural network model with caffe framework
A: One nice thing about Caffe is it gives you flexibility create your own network, convolutional layers and their number of outputs. You also get to decide when and where you apply pooling, and if you want to maximum or average for your window. You also define your activation function and normalization. Caffe has a lot of flexibility when it comes to creating your network.

Q: How do we make sure that our batch size is appropriate relative to the capabilities of our GPU?
A: You can run nvidia-smi to check your GPU utilization. One thing you can do is increase or decrease batch size to maximize GPU utilization given the amount of memory on the board.

Q: What does "autoboost on" mean?
A: This means that NVIDIA GPU Boost is enabled by default. This ensures that out of the box, the Tesla K80 will always try to achieve the best possible GPU clock and maximize performance for a given workload, power and thermal condition.

Q: use-case 2 was about object recognition and localization, right? where do i start? most tutorials are about classification. ;)
A: One easy way to start with Caffe is to use this example -http://nbviewer.ipython.org/github/BVLC/caffe/blob/master/examples/detection.ipynb
If you are using theano, you may want to check out sklearn-theano - http://sklearn-theano.github.io/

Q: For BSc/MSc university labs / lectures, which framework would you recommend? (Caffe / Theano / Torch / other?)
A: Each framework offers differing capability and subject matter focus. Often Caffe is chosen as a starting point because it has a rich set of existing models, examples, and data. DIGITS + Caffe makes an excellent tool for teaching and learning.

Q: What does end-to-end deep learning mean?
A: End to end refers to both the training of models and then the use of the model for classification.

Q: good question about server-grade above. Will Pascal and Volta be server-grade?
A: Yes, there will be server-grade “Tesla” based products using Pascal and Volta generation GPUs.

Q: can it be practical to deploy a trained NN into a raspberry pi2, and get real time performance?

A: Deploying a trained model for classification on mobile platforms is pretty common. We have achieved very high levels of performance on the NVIDIA Tegra SOC system. Depending on what “real time” means to you, and how large the images/data you are classifying, you could achieve good results on a rasberry pi2 system…

Q: What is the max number of GPUs that Caffe can train and classify?
A: The multi-GPU support in Caffe is still evolving and you can find ongoing status at: https://github.com/BVLC/caffe/pull/1148

Q: is cudnn necessary to get GPU support in caffe?
A: cuDNN is not required to get GPU acceleration in caffe. cuDNN provides higher performance however so we definitely recomend you use cuDNN with caffe.