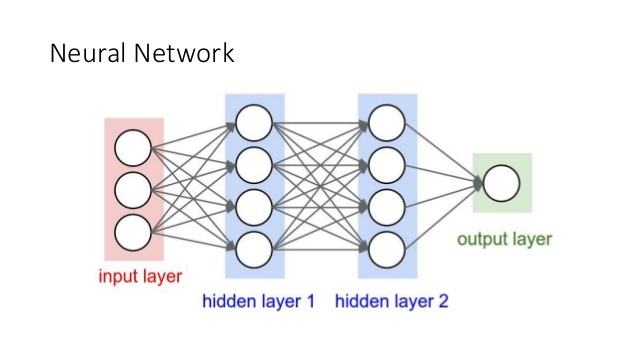
* report’s purpose
* results
* conclusions
* recommendations
* provide logical connections or transitions between the information included
* understandable to a wide audience

Should contain the following information:

* communicate specific information from the paper
* include the purpose, methods, and scope of the paper
* provide the paper’s results, conclusions, and recommendations

Outline

1. Introduction
2. Body 1/ History
3. Body 2
4. Body 3
5. Body 4
6. Body 5
7. Conclusion (results/ recommendations?)

* most of the machine learning methods are based on supervised learning
* deep learning is primarily about neural networks where a network is an interconnected web of nodes and edges.
* neural nets were designed to perform complex tasks, such as the task of placing objects into categories based on a few attributes
* neural nets are highly structured networks, and have three kinds of layers – an input, an output, and so called hidden layers, which refer to any layers between the input and output layers.
* each node (also called a neuron) in the hidden and output layers have a classifier.
* 
* input -> feature representation -> learning algorithm
* feature extractors require expert knowledge and time-consuming hand-tuning
* neural networks / neural network architecture
* for approximately 20 years, attempts were made to train deeper neural networks (with more than one hidden layer), however rarely with benefits (vanishing gradient).
* in 2006, a major breakthrough was made in deep architectures following three key principles:
  + unsupervised learning of representations is used to pre-train each layer
  + unsupervised training of one layer at a time, on top of the previously trained ones. The representation learning at each level is the input for the next layer.
  + use supervised training to fine-tune all the layers (in addition to one or more additional layers that are dedicated to producing predictions)
* after the 2006 breakthrough, a lot of ideas were also developed. Nowadays, pre-training is almost obsolete.
  + new activation functions
  + regularization methods
  + initialization methods
  + data augmentation
  + optimization techniques
* another reason on why deep learning is possible, is the availability of lots of data (i.e. ImageNet)
* companies are working on solutions for deep learning acceleration:
  + NVIDIA – created an entire platform stack dedicated to work with deep learning, called DIGITS. Their GPUs are widely used in deep learning.
  + Amazon – Amazon AWS also created EC2 instances with NVIDIA GPUs (with 4GB of memory and 1536 CUDA cores). Lots of AMIs with Deep Learning software ecosystem already installed.
  + Microsoft - announced that it will offer NVIDIA GPUs on its Azure cloud platform
* convolutional neural networks
* based on following principles: local receptive fields, shared weights, pooling (or down-sampling)
* this special neural network architecture takes advantage of the spatial structure of data
* shared weights and bias are called kernel or filter
* convolutional layers provide translation invariance since these filters work on every part of the image, they are searching for the same feature everywhere in the image
* pooling layers are usually present after a convolutional layer. They provide a down-sampling of the convolution output
* different types of pooling, the most used are max-pooling and average pooling
* pooling layers downsamples the volume spatially, reducing small translations of the features. They also provide a parameter reduction
* max pooling is how the network ask whether a feature is found anywhere in some region of the image. After that, it will lose the exact position. (-> classification after this)
* we have defined all of the components required to create a convolutional neural network but you will rarely see a shallow convolutional neural network like that
* actually, experiments demonstrated that the replication of convolutional and pooling layers produce better results the deeper you go
* the dropout technique helps with the overfitting, especially on dense layers. Drop occur only at training time, not on test time.
* data augmentation can help with overfitting and ill certainly improve results
* features learned by convolutional neural networks on larger dataset problem can be helpful on different problems. It’s very common to pre-train a convolutional neural network on imagenet and then use it as a fixed feature extractor or as an initialization.(transfer learning)