Deep Learning Architecture

ALI SALEH

Agenda

- What is Deep Learning
- Brief History of deep learning
- Convolution Deep belief networks for detection and classification
- Other successful Deep learning Methods
- Criticism to Deep learning methods
- Conclusion

What is Deep Learning

- Trainable Feature detector units arranged in hierarchy of layers
- Lower layers detect simple features (Local Features)
- Higher layers detect complex features (Global Features)
- Each layer provide it's output as input for the next layer

Brief History of Deep Learning

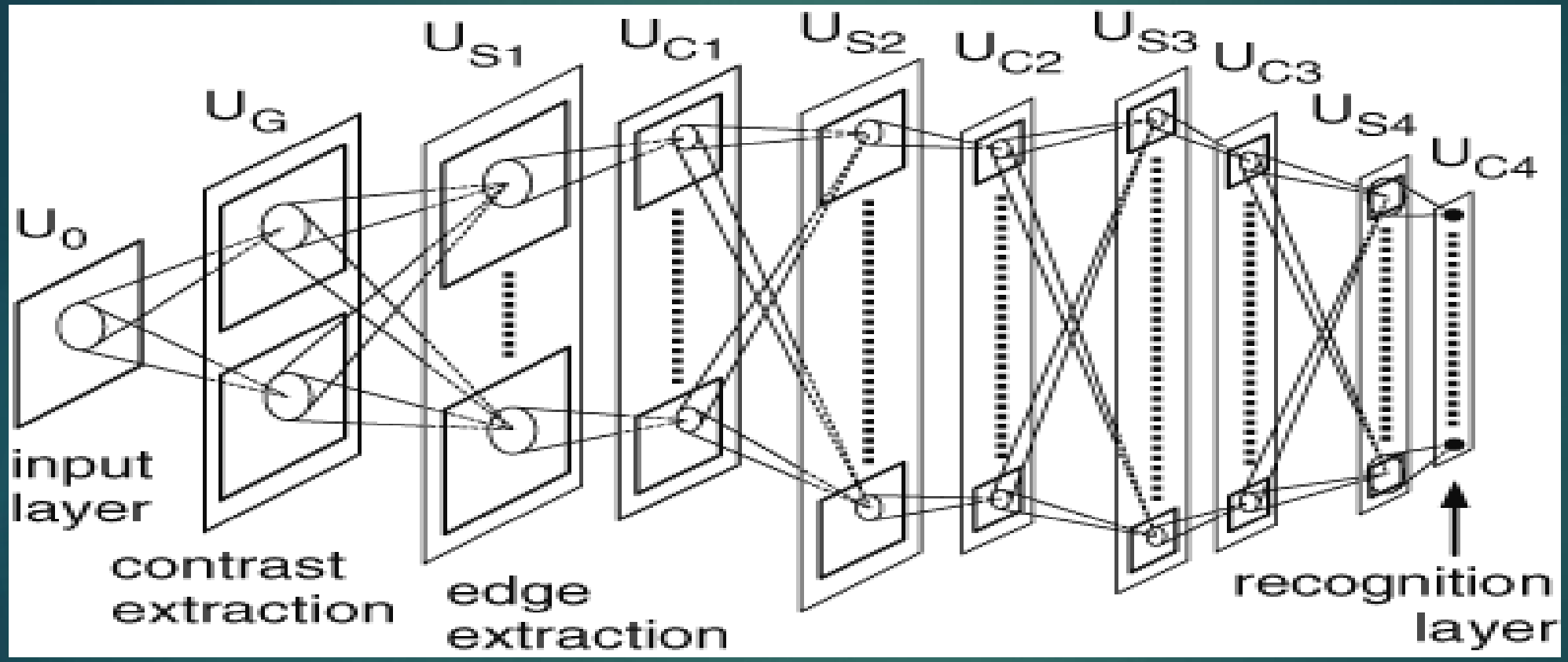
- Pandemonium architecture (late 1950s)
  - Inspired by Hubel and Wiesel work on vision system
  - Four different groups of detectors each assigned to a recognition stage (Data, Feature, Cognitive, Decision)
  - Was able to recognize characters
  - Needed large number of training exambles

Brief History of Deep Learning

- Neocognitron (Fukushima 1980s)
  - Alternate C-Cells and S-Cells Layers
  - S-Cells are for feature extraction
  - C-Cells are for position tolerance
  - Learning is only for the S-Cells planes.

Brief History of Deep Learning

- Neocognitron (Fukushima 1980s)



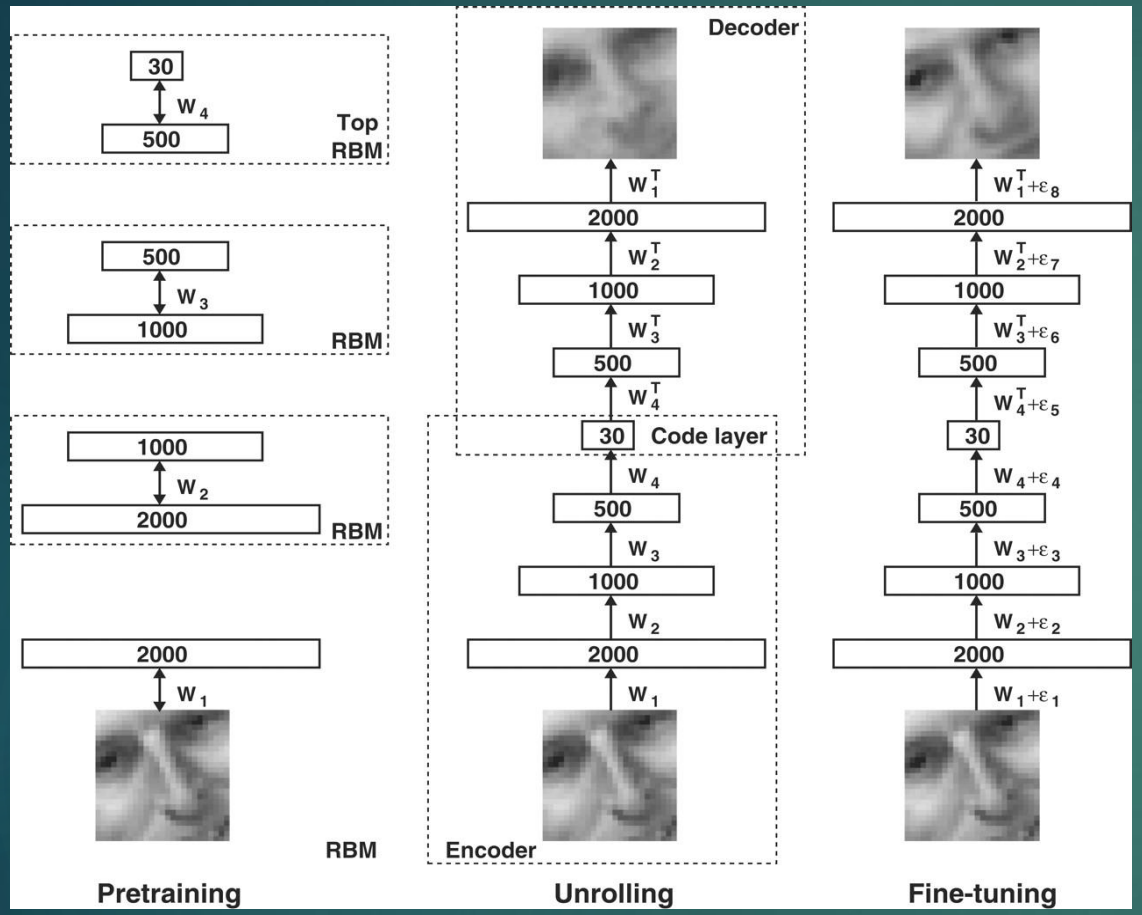
Brief History of Deep Learning

- Deep belief network (Hinton 2006)
  - Un-supervised training for a Neural Network
  - The network is a set of RBMs stacked one over the other
  - Training one layer at a time
  - Use the output of each layer as input to the next layer
  - Gave 1.25% error rate for the MNIST dataset
  - Training time was about 1 week

Brief History of Deep Learning

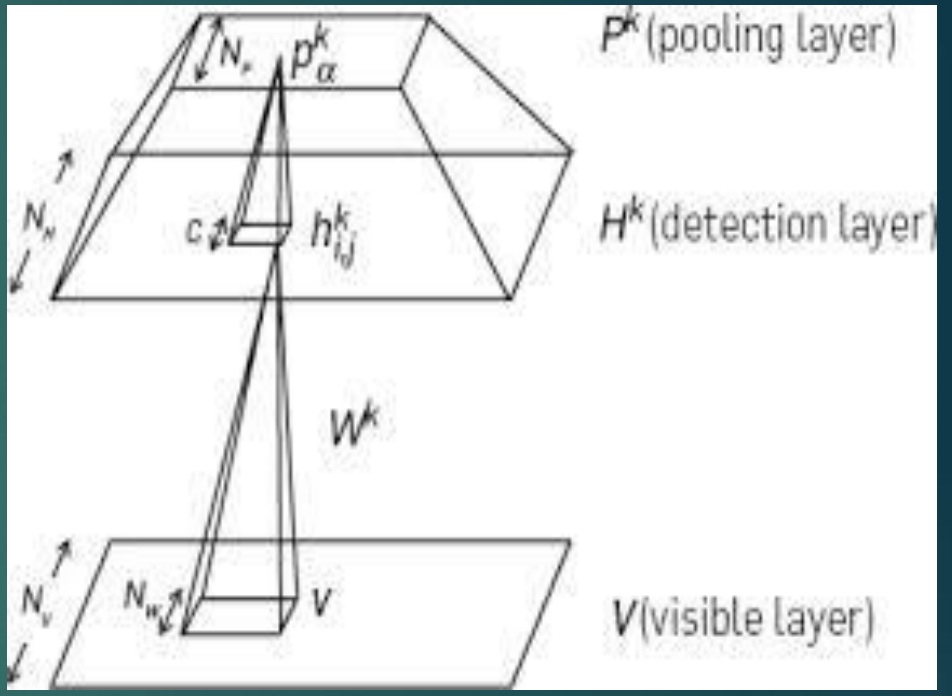
- Autoencoders for dimensionality reduction (Hinton 2006)
  - Similar technique like the deep belief networks.
  - Backpropagation is used through the whole system to fine tune the weights
  - Pre-training gives better results.
  - Give better results Than PCA variants on MNIST dataset.

Brief History of Deep Learning



Convolution Deep belief networks for detection and classification

- Stable to the change of the object place in the image
- Uses CRBMs instead of RBMs
- Weights are shared among the whole layer
- Allows detectors to learn the same object in different places
- Max pooling is used to help tolerating changes in position of features
- Experiment showed that max pooling makes the algorithm 10 times faster
- Uses layer by layer training as in the original model.



A single CRBM

Convolution Deep belief networks for detection and classification

- Experimental results
  - Caltech-101 object classification task

	CDBN	shape-context	SIFT
15 image per class	57.7 ±1.5%	59.0 ± 0.56%	54.0%
30 image per class	65.4 ± 0.5%	66.2 ± 0.5%	64.6%

  - MNIST Dataset
    - 2 layers of CRBMS to prepare a feature vector
    - SVM classifier over the features gives 0.82% error rate.

Other successful Deep learning Methods

- Google Brain (detecting cats and human faces in random videos)
- Facebook Ads improvement
- Microsoft speech recognition
- Microsoft Bing speech translation

Criticisms to Deep learning methods

- lack of theory for most of the methods.
- Most results are experimental not theoretical.
- Neural Networks takes a long time for training.

References

- A Fast Learning Algorithm for Deep Belief Nets : Geoffrey E. Hinton et al
- Learning multiple layers of representation : Geoffrey E. Hinton
- Learning Deep Architectures for AI :Yoshua Bengio
- Neocognitron: A Self-organizing Neural Network Model for a Mechanism of Pattern Recognition Unaffected by Shift in Position: Kunihiko Fukushima
- Building High-level FeaturesUsing Large Scale Unsupervised Learning : Quoc V. Le et al
- Reducing the Dimensionality of Data with Neural Networks G. E. Hinton and R. R. Salakhutdinov
- Representation Learning: A Review and New Perspectives : Yoshua Bengio, Aaron Courville, and Pascal Vincent
- Convolutional Deep Belief Networks for Scalable Unsupervised Learning of Hierarchical Representations :Honglak Lee et al