# Application of Deep Learning Algorithms to Image Classification

## Progress Presentation

---

*J.D. Gallego-Posada*[†] *D.A. Montoya-Zapata*[*]
*D.E. Sierra-Sosa*[§] *O.L. Quintero-Montoya*[‡]

{[†]jgalle29, [*]dmonto39, [§]dsierras, [‡]oquinte1} (at)eafit(dot)edu(dot)co

Research Group on Mathematical Modeling
School of Sciences
Universidad EAFIT

15/04/2016

# Introduction

# Introduction

*How can we teach computers to locate faces in an image?*

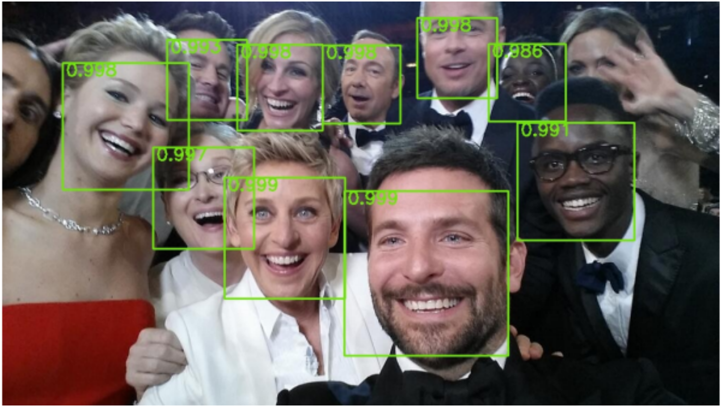



Image retrieved on 17/02/2016 from `http://www.ukprogressive.co.uk/wp-content/uploads/2015/02/face-algorithm.png`

*How can we teach computers to understand our voices?*


Image retrieved on 17/02/2016 from http://www.psfk.com/2014/12/voice-recognition-software-translates-words-from-those-with-speech-disorders.html

# Introduction

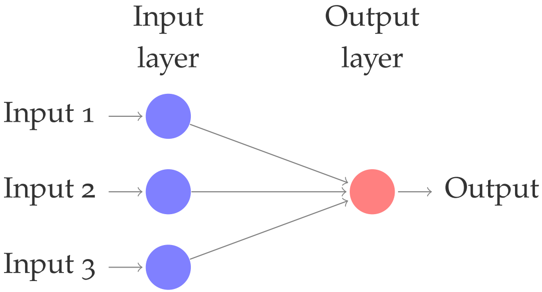
*How can we teach computers to recognize characters?*


Image retrieved on 18/02/2016 from `http://teaching.paganstudio.com/digitalfoundations/wp-content/uploads/2013/09/lpr_software_1.jpg`

# Inspiration

Image retrieved on 17/02/2016 from `http://cosmonio.com/Research/Deep-Learning/files/small_1420.png`

# Single-Layer Perceptron



# The XOR Problem

*What about non linear-separable groups?*

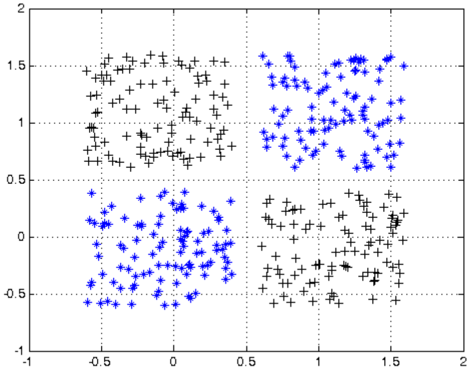
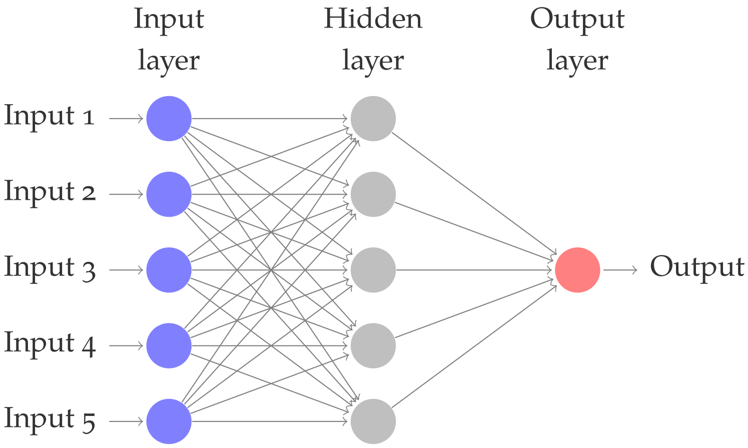
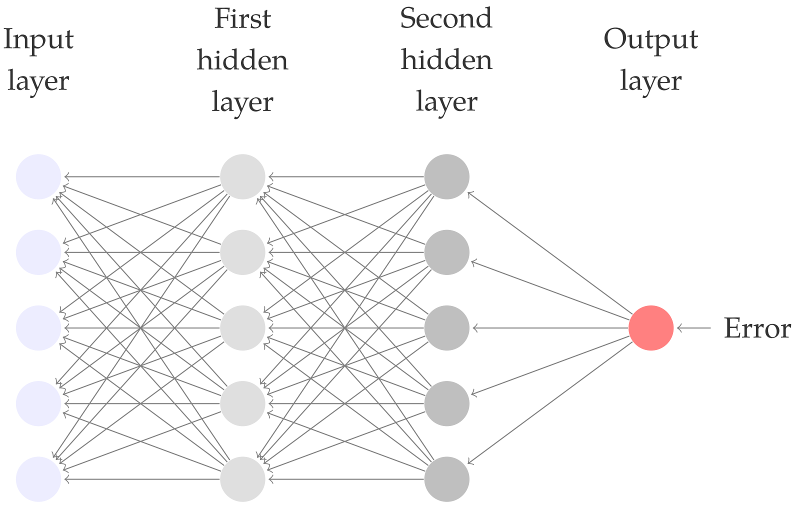



Image retrieved on 18/02/2016 from http://lab.fs.uni-lj.si/lasin/wp/IMIT_files/neural/nn06_rbfn_xor/html/nn06_rbfn_xor_3_newpnn_01.png

# Neural Network - Multilayer Perceptron



# Back-propagation



# State of the Art

# State of the Art

2004 — Methods based on BoW for image classification problems [5]

# State of the Art - Bag of Words

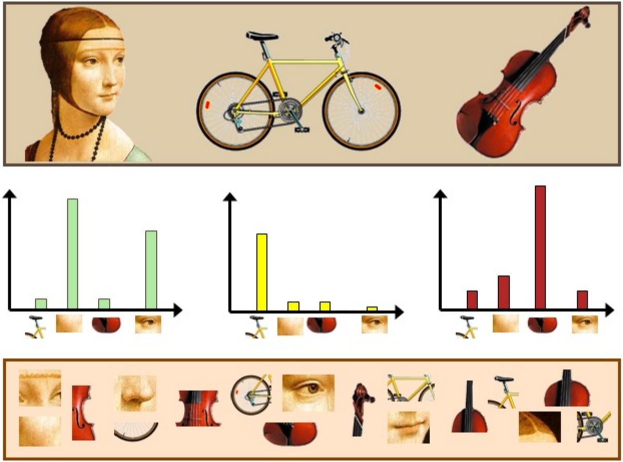


Image retrieved on 04/04/2016 from
https://gilscvblog.files.wordpress.com/2013/08/figure31.jpg

# State of the Art

2004 Methods based on BoW for image classification problems [5]

2006 Incorporating spatial geometry to BoW models [7]

# State of the Art

2004 — Methods based on BoW for image classification problems [5]

2006 — Incorporating spatial geometry to BoW models [7]

2006 — Deep Learning success: Hinton [11], LeCun [12] and Bengio [13]

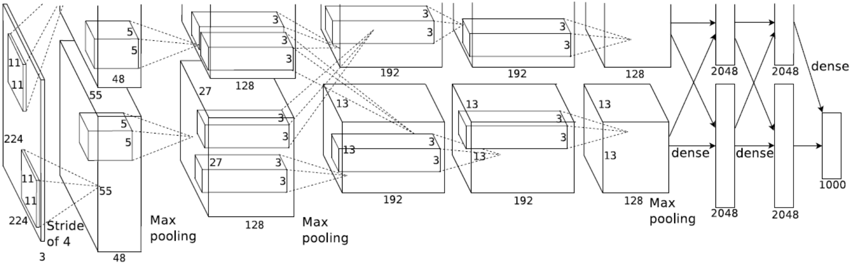
# State of the Art

- 2004 — Methods based on BoW for image classification problems [5]
- 2006 — Incorporating spatial geometry to BoW models [7]
- 2006 — Deep Learning success: Hinton [11], LeCun [12] and Bengio [13]
- 2010 — Sparse coding for the image classification problem [6]
- 2011 — Extracting high-order statistics - Fisher kernel + SVMs [4]

# State of the Art

- 2004 — Methods based on BoW for image classification problems [5]
- 2006 — Incorporating spatial geometry to BoW models [7]
- 2006 — Deep Learning success: Hinton [11], LeCun [12] and Bengio [13]
- 2010 — Sparse coding for the image classification problem [6]
- 2011 — Extracting high-order statistics - Fisher kernel + SVMs [4]
- 2012 — CNN for image classification problem (AlexNet) [9]

# State of the Art - AlexNet Architecture

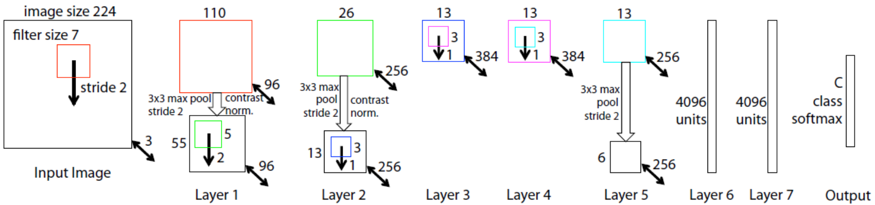



Source: Krizhevsky et al., 2012

# State of the Art

- 2004 — Methods based on BoW for image classification problems [5]
- 2006 — Incorporating spatial geometry to BoW models [7]
- 2006 — Deep Learning success: Hinton [11], LeCun [12] and Bengio [13]
- 2010 — Sparse coding for the image classification problem [6]
- 2011 — Extracting high-order statistics - Fisher kernel + SVMs [4]
- 2012 — CNN for image classification problem (AlexNet) [9]
- 2013 — Development of a new visualization strategy (Clarifai) [2]

# State of the Art - Clarifai Architecture




Source: Zeiler et al., 2014

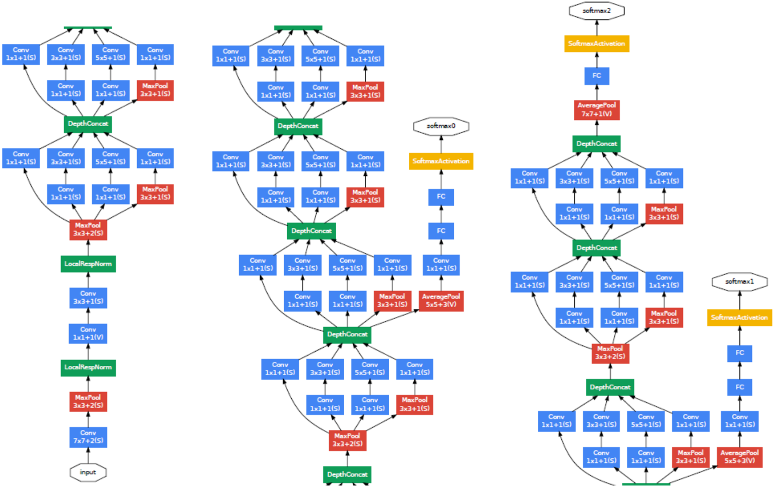
# State of the Art

- 2004 – Methods based on BoW for image classification problems [5]
- 2006 – Incorporating spatial geometry to BoW models [7]
- 2006 – Deep Learning success: Hinton [11], LeCun [12] and Bengio [13]
- 2010 – Sparse coding for the image classification problem [6]
- 2011 – Extracting high-order statistics - Fisher kernel + SVMs [4]
- 2012 – CNN for image classification problem (AlexNet) [9]
- 2013 – Development of a new visualization strategy (Clarifai) [2]
- 2014 – New pooling technique: not fixed input size required (SPP) [14]

# State of the Art

- 2004 Methods based on BoW for image classification problems [5]
- 2006 Incorporating spatial geometry to BoW models [7]
- 2006 Deep Learning success: Hinton [11], LeCun [12] and Bengio [13]
- 2010 Sparse coding for the image classification problem [6]
- 2011 Extracting high-order statistics - Fisher kernel + SVMs [4]
- 2012 CNN for image classification problem (AlexNet) [9]
- 2013 Development of a new visualization strategy (Clarifai) [2]
- 2014 New pooling technique: not fixed input size required (SPP) [14]
- 2014 Successful use of deeper architectures (VGG, GoogLeNet) [1] [8]

# State of the Art - GoogLeNet Architecture




Source: Simonyan et al., 2015

# State of the Art

- 2004 Methods based on BoW for image classification problems [5]
- 2006 Incorporating spatial geometry to BoW models [7]
- 2006 Deep Learning success: Hinton [11], LeCun [12] and Bengio [13]
- 2010 Sparse coding for the image classification problem [6]
- 2011 Extracting high-order statistics - Fisher kernel + SVMs [4]
- 2012 CNN for image classification problem (AlexNet) [9]
- 2013 Development of a new visualization strategy (Clarifai) [2]
- 2014 New pooling technique: not fixed input size required (SPP) [14]
- 2014 Successful use of deeper architectures (VGG, GoogLeNet) [1] [8]
- 2015 Strategies for avoiding overfitting and underfitting [3]

# State of the Art

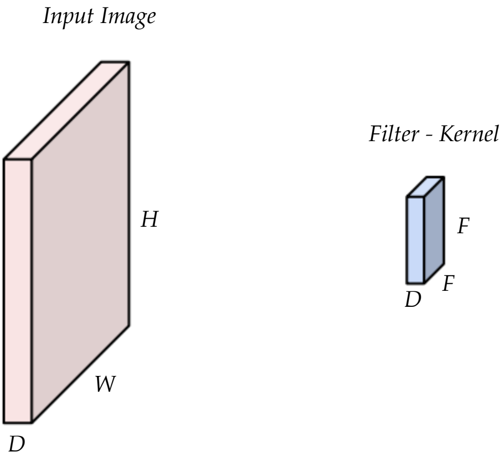
- 2004 – Methods based on BoW for image classification problems [5]
- 2006 – Incorporating spatial geometry to BoW models [7]
- 2006 – Deep Learning success: Hinton [11], LeCun [12] and Bengio [13]
- 2010 – Sparse coding for the image classification problem [6]
- 2011 – Extracting high-order statistics - Fisher kernel + SVMs [4]
- 2012 – CNN for image classification problem (AlexNet) [9]
- 2013 – Development of a new visualization strategy (Clarifai) [2]
- 2014 – Successful use of deeper architectures (VGG, GoogLeNet) [1] [8]
- 2015 – Strategies for avoiding overfitting and underfitting [3]
- 2016 – Representation learning for Deep Neural Networks [10]

*Not only improving performance, but also gaining a better understanding of DL and DNN.*

# Convolutional Neural Nets

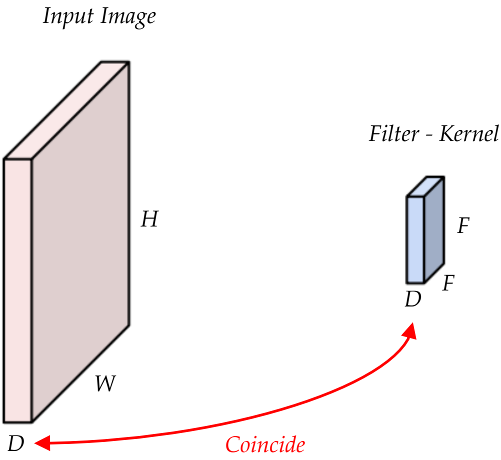
# Convolution Layer

*[Based on recent Li et al slides]*




*Convolve: Slide the filter over the image spatially computing dot products*

# Convolution Layer



# Convolution

*Consider for example the case W=H=5, D = 1, F=3,*

*Input Image*

| | | | | |
|---|---|---|---|---|
| 1 | 2 | 1 | 1 | 0 |
| 2 | 1 | 0 | 0 | 2 |
| 1 | 0 | 1 | 2 | 1 |
| 2 | 1 | 0 | 1 | 0 |
| 1 | 2 | 2 | 1 | 1 |

*Filter - Kernel*

| | | |
|---|---|---|
| -1 | 0 | 1 |
| 0 | -1 | 1 |
| 1 | -1 | 1 |

$(1)(-1) + (2)(0) + (1)(1) + (2)(0) + (1)(-1) + (0)(1) + (1)(1) + (0)(-1) + (1)(1) + 0.5 = 1.5$

# Convolution

*Consider for example the case W=H=5, D = 1, F=3,*

*Input Image*

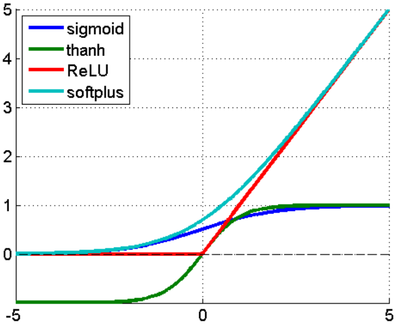
| | | | | |
|---|---|---|---|---|
| 1 | 2 | 1 | 1 | 0 |
| 2 | 1 | 0 | 0 | 2 |
| 1 | 0 | 1 | 2 | 1 |
| 2 | 1 | 0 | 1 | 0 |
| 1 | 2 | 2 | 1 | 1 |

*Filter - Kernel*

| | | |
|---|---|---|
| -1 | 0 | 1 |
| 0 | -1 | 1 |
| 1 | -1 | 1 |

$$\underbrace{(1)(-1) + (2)(0) + (1)(1) + (2)(0) + (1)(-1) + (0)(1) + (1)(1) + (0)(-1) + (1)(1)}_{\text{Dot product}} + \underbrace{0.5}_{\text{Bias}} = 1.5 \in \mathbb{R}$$

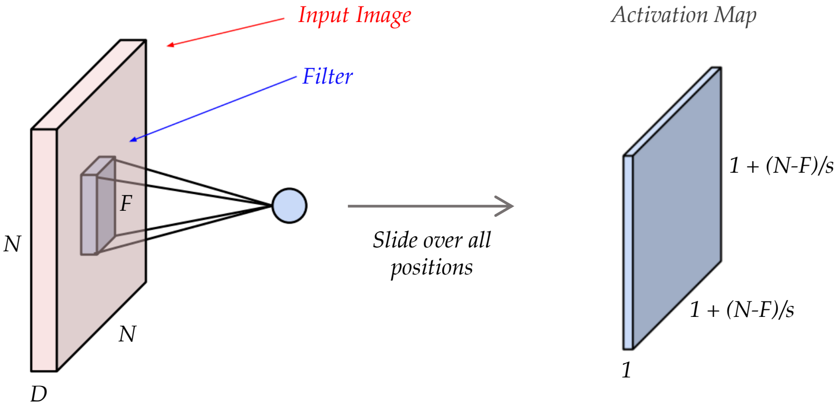
# ReLU



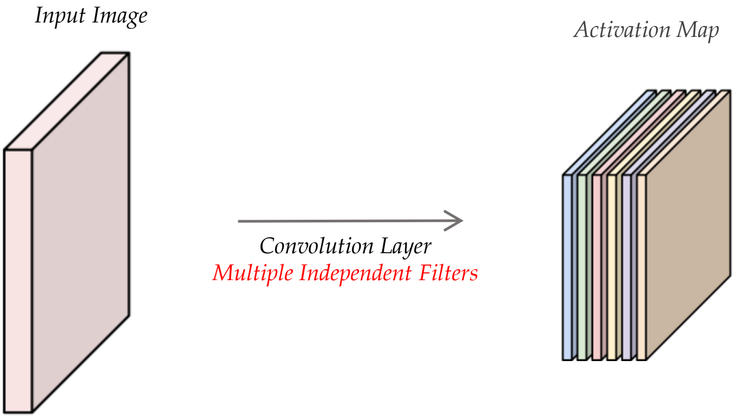

$$\underbrace{(1)(-1) + (2)(0) + (1)(1) + (2)(0) + (1)(-1) + (0)(1) + (1)(1) + (0)(-1) + (1)(1) + 0.5}$$

*Output = Activation(Induced Local Field) = Non-linearity(Dot product + bias)*

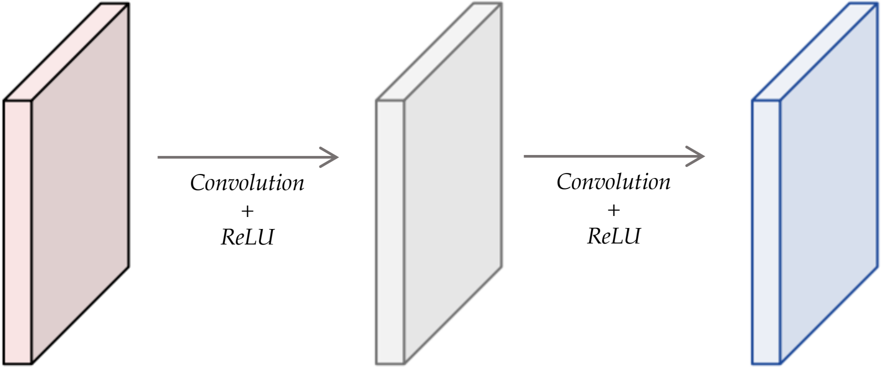
# Convolution Layer



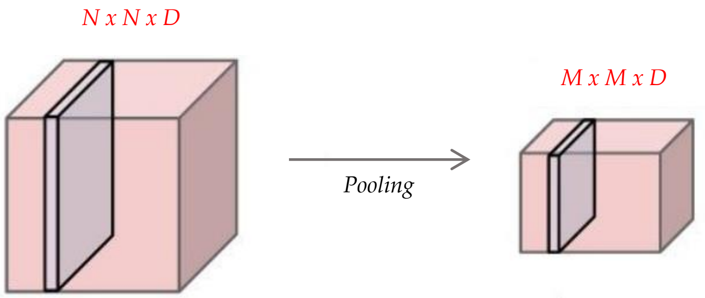
# Convolution Layer



# Convolution Layer



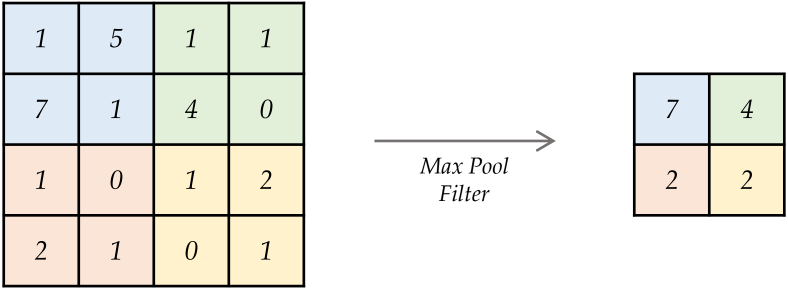
# Pooling Layer



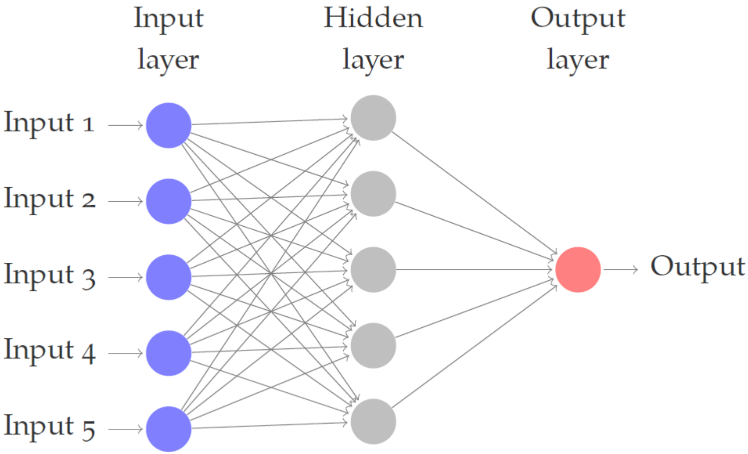

*Pool: Reduce the spatial dimension of the image in a controlled way*

$M < N$

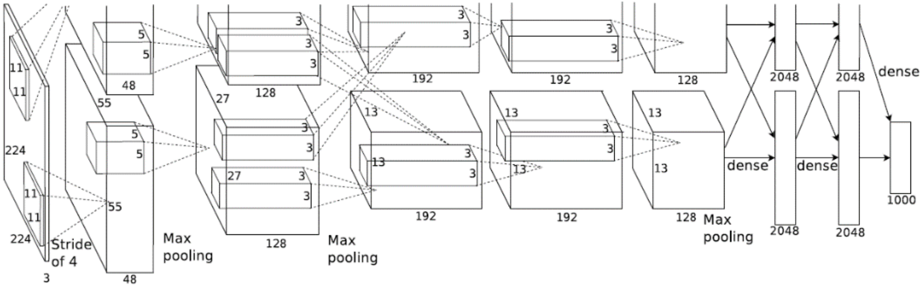
# Max Pooling



# Fully Connected Layers



# AlexNet Architecture




*Feature Extractor*
*via*
*Convolution*

*Classifier*
*via*
*Neural Network*

Source: Krizhevsky et al., 2012

# References

# References I


[1] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015, pp. 1–9.

[2] M. D. Zeiler and R. Fergus, "Visualizing and understanding convolutional networks," in *Proceedings of the ECCV International Workshop on Statistical Learning in Computer Vision*. Springer, 2014, pp. 818– 833.

[3] R. Wu, S. Yan, Y. Shan, Q. Dang, and G. Sun, "Deep Image: Scaling up Image Recognition," Arxiv, p. 12, 2015.

[4] F. Perronnin, Y. Liu, J. S´anchez, and H. Poirier, "Large-scale image retrieval with compressed fisher vectors," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2010, pp. 3384–3391.

[5] G. Csurka, C. R. Dance, L. Fan, J. Willamowski, and C. Bray, "Visual categorization with bags of keypoints," *Proceedings of the ECCV International Workshop on Statistical Learning in Computer Vision*, pp. 59–74, 2004.

# References II


[6] Y. Lin, F. Lv, S. Zhu, M. Yang, T. Cour, K. Yu, L. Cao, and T. Huang, “Large-scale image classification: Fast feature extraction and SVM training,” in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2011, pp. 1689–1696.

[7] S. Lazebnik, C. Schmid, and J. Ponce, “Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories,” in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 2, 2006, pp. 2169–2178.

[8] K. Simonyan and A. Zisserman, “Very Deep Convolutional Networks for Large-Scale Image Recognition,” Proceedings of the ICLR, pp. 1–14, 2015.

[9] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “ImageNet Classification with Deep Convolutional Neural Networks,” *Advances In Neural Information Processing Systems*, pp. 1–9, 2012.

[10] Y. Li, J. Yosinski, J. Clune, H. Lipson, and J. Hopcroft, “Convergent Learning: Do different neural networks learn the same representations?” in ICLR, 2016, pp. 1–21.

# References III


[11] G, Hinton, S. Osindero, and Y. Teh, “A Fast Learning Algorithm for Deep Belief Nets,” *Neural Computation*, vol. 18, no. 7, pp. 1527–54, 2006.

[12] C. P. Marc’ Aurelio, S. Chopra, and T. LeCun, “Efficient Learning of Sparse Representations with an Energy-based Model,” *Advances in Neural Information Processing Systems*, vol. 19, pp. 1137-1144, 2006.

[13] Y, Bengio, P. Lamblin, D. Popovici, and H. Larochelle, “Greedy Layer-Wise Training of Deep Networks,” *Advances in Neural Information Processing Systems*, vol 19., no. 1, pp. 153, 2007.

[14] K. He, X. Zhang, S. Ren, and J. Sun, “Spatial Pyramid Pooling in Deep Convolutional Networks for Visual Recognition,” *Proceedings of the ECCV International Workshop on Statistical Learning in Computer Vision*, pp. 153-361, 2014.

# Questions