A convolutional neural network (CNN) is a class of deep neural networks applied to analyzing visual imagery. CNNs are regularized versions of multilayer perceptrons (input layer, hidden layer, output layer) where neurons in one layer are connected to neurons in the next. Inter-connectedness makes models prone to data overfitting. CNN enables scientists to assemble complex patterns into smaller simpler patterns. Inspired by animal species visual cortex research in 1980s.*

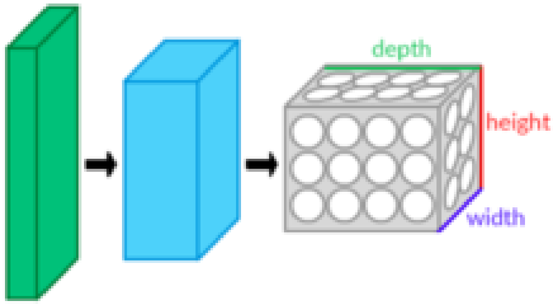
(a (process below)) Convolution layer: tensor inputs (i.e. a math array object) ==> (image) x (width x height) x (depth). Layers have learning filters called kernels that compute dot producing 2-d maps (neurons) who iterate and learn features of spatial position inputs.

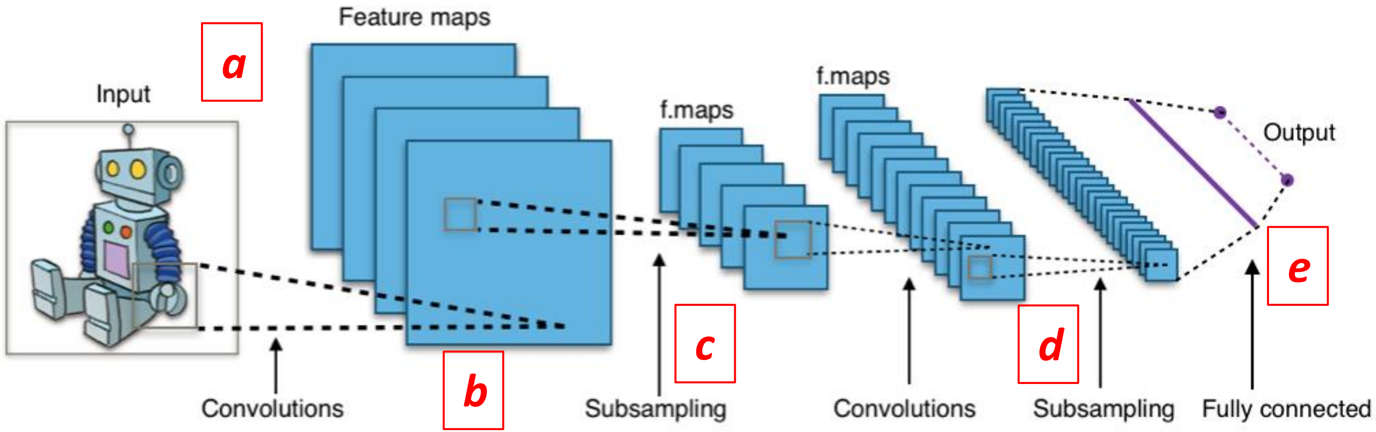
(b) Neurons (layer outputs) are filters along a depth dimension of a small input used to connect between tensor input feature maps. Neuron connectivity between layers becomes a hyperparameter (receptive field) whose (width x height) extend depth wise. The algorithm's "…architecture ensures learnt filters produce a strong response to spatially local input pattern(s)."

(c) Algorithm iterates through feature maps & convolutions generating *depth*, *stride*, and *zero-padding*.
Depth controls # layer neurons connecting a region based on learned edges, such as blobs of color. (d) Stride controls how depth columns of (width x height) are allocated by adjusting pixels until resulting output volume has smaller spatial dimensions.
Zero-padding are zero(0) input values applied to 'input volume borders' to help control the output of spatial size volume.

(e) Fully-connected states are neurons across layers in a flat matrix adjusted with weight & bias vectors from learning iterations.

(f) Neuron fit formula is a function of input volume size (W), kernel size in convolution (K), stride applied (S) + zero-padding (P).





 # neurons to fit a volume

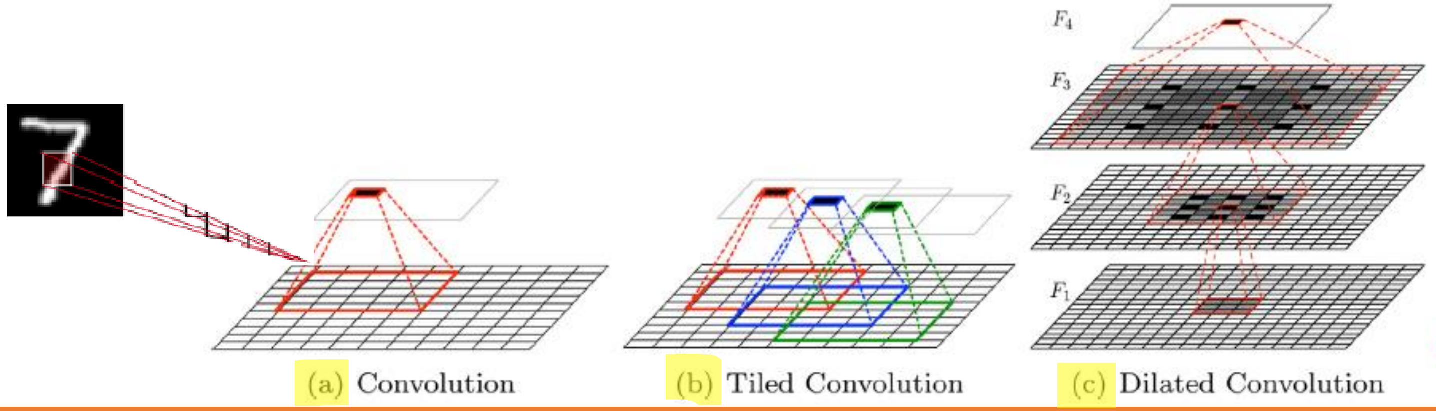
$$\frac{W-K+2P}{S}+1.$$

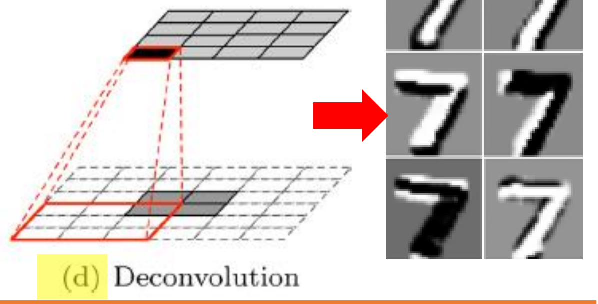
f

*References: Wikipedia (see reference slide)
By: Brian Hogan | Syracuse Univ | 2019 (664)

By: Brian Hogan | Syracuse Univ | 2019 (664)

CNN: algorithm improvements (since 2012)





According to Gu, et. al., by increasing depth, i.e. # layer neurons connecting a region based on learned edges, a network can better approximate the target function with increased nonlinearity generating better feature representations but must balance network complexity, overfitting, and computational efficiency. There are **27** major CNN improvements since 2012 including: hyper-planes (*dilation*), reducing nodes between layers (neuron *pooling*), adjusting image loss functions (weight adjustments), and improved methods for *regularization,* i.e. adjusting image overfitting by decaying weight scores by rewarding *invariance* (p 361).

(a) Convolution: basic view of image or text parsing, a.k.a. an image patch. Layers attached by neurons across image landscape.

(b) Tiled Convolution: CNN tiles and multiplies feature maps to learn rotational and scale invariant features. Separate kernels are learned within the same layer, and the complex invariances can be learned implicitly by square-root pooling over neighboring units (p. 356). A user sets tile size quantity effecting distance over which weight scores are shared improving image capture.

(c) Dilated Convolution: introduces a hyper-parameter to convolutional layer by *inserting zeros* between filter elements. Increasing the network's receptive field size widens relevant information and improves performance outcomes for scene segmentation, machine translation, speech synthesis, and speech recognition. Figure C is zooming in, or widening, pixel grouping.

(d) Deconvolution: essentially convolution run backwards. Results in multiple image outputs, i.e. a larger pixel area, of a single activation a condensed image subset. Convolutions are run within "each" single activation (subset) improving the dilation factor for each input feature map. Groupings are tied together with neurons across subsets. Method improves visualization, semantic segmentation, and visual question answering. Reference: Gu, J, et. al., (2018). Recent advances in convolutional neural networks. *Pattern Recognition*.

References: By: Brian Hogan |bphogan@syr.edu | 2019	Reference Type:
Chapellent-Lanier, T. (2018). Facial recognition algorithms are getting a lot better, NIST study finds. Retrieved from: https://www.fedscoop.com/facial-recognition-algorithms-getting-lot-better-nist-study-finds	Internet News Article
Editor (2019). Files, Not Faces. The Economist Magazine. May 25th 2019, pg27.	Magazine
Fraga T. (2018) Designing Behaviors to Interactively Interlace Natural Language Processing, Text to Speech Procedures and Algorithmic Images. In: Marcus A., Wang W. (eds) Design, User Experience, and Usability: Designing Interactions. DUXU 2018. Lecture Notes in Computer Science, vol 10919. Springer, Cham. Retrieved from: https://doi-org.libezproxy2.syr.edu/10.1007/978-3-319-91803-7_33	Research Article
Grother, P., Ngan, M., Hanaoka, K. Ongoing Face Recognition Vendor Test (FRVT) Part 2: Identification. National Institute of Standards and Technology. *U.S. Department of Commerce*. Retrieved from: https://nvlpubs.nist.gov/nistpubs/ir/2018/NIST.IR.8238.pdf	U.S. Gov. Whitepaper
Gu, J., Wang, Z., Kuen, J., Ma, L., Shahroudy, A., Shuai, B., Liu, T., Wang, X., Wang, G., Cai, J., Chen, T. (2018). Recent advances in convolutional neural networks, *Pattern Recognition*, pgs. 354-377. Retrieved from: https://doi.org/10.1016/j.patcog.2017.10.013.	Research Article
Wikipedia (2019). Convolutional Neural Network. https://en.wikipedia.org/wiki/Convolutional_neural_network.	Internet Encyclopedia
Zakraoui, J., Saleh, M. & Al Ja'am, J. Text-to-picture tools, systems, and approaches: a survey. *Multimedia Tools Appl* (2019) 78: 22833. https://doi-org.libezproxy2.syr.edu/10.1007/s11042-019-7541-4	Research Article
Editors, (2019).www.wordseye.com	website

Research articles sourced from Syracuse University Research Library Portal = SUMMON