Advances in Pattern Recognition with Convolutional Neural Networks

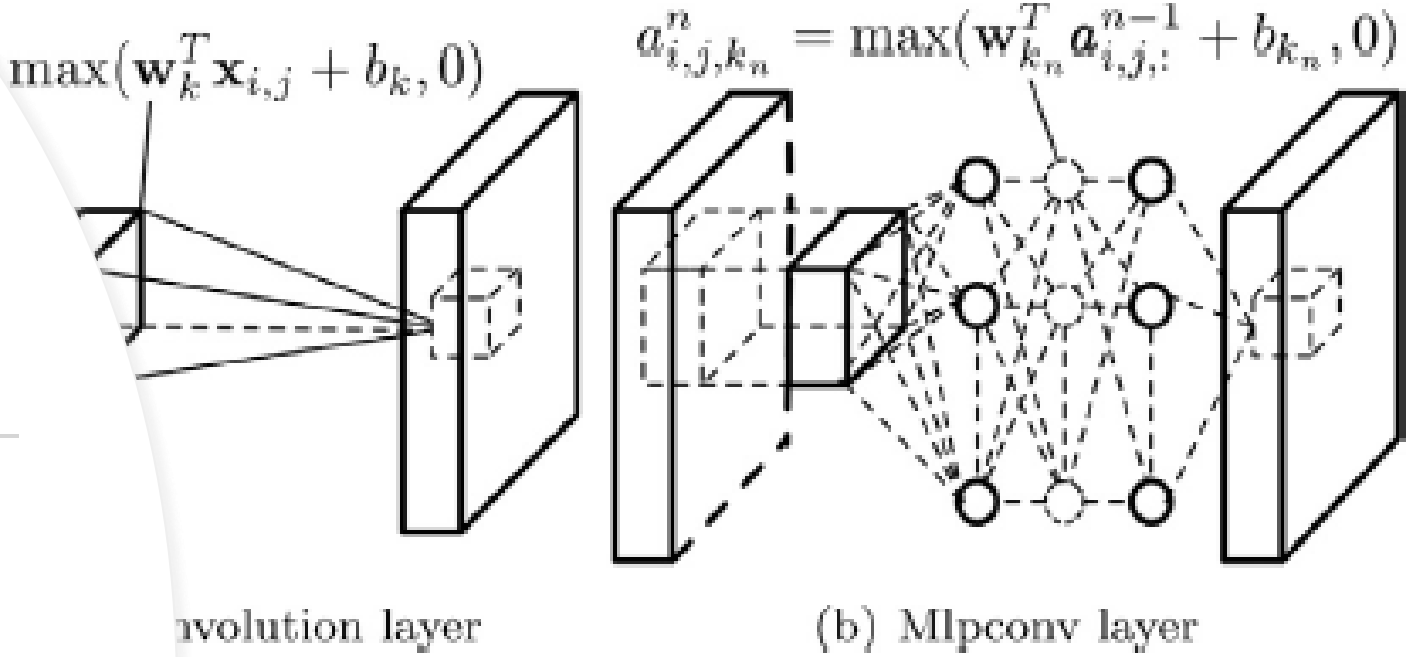


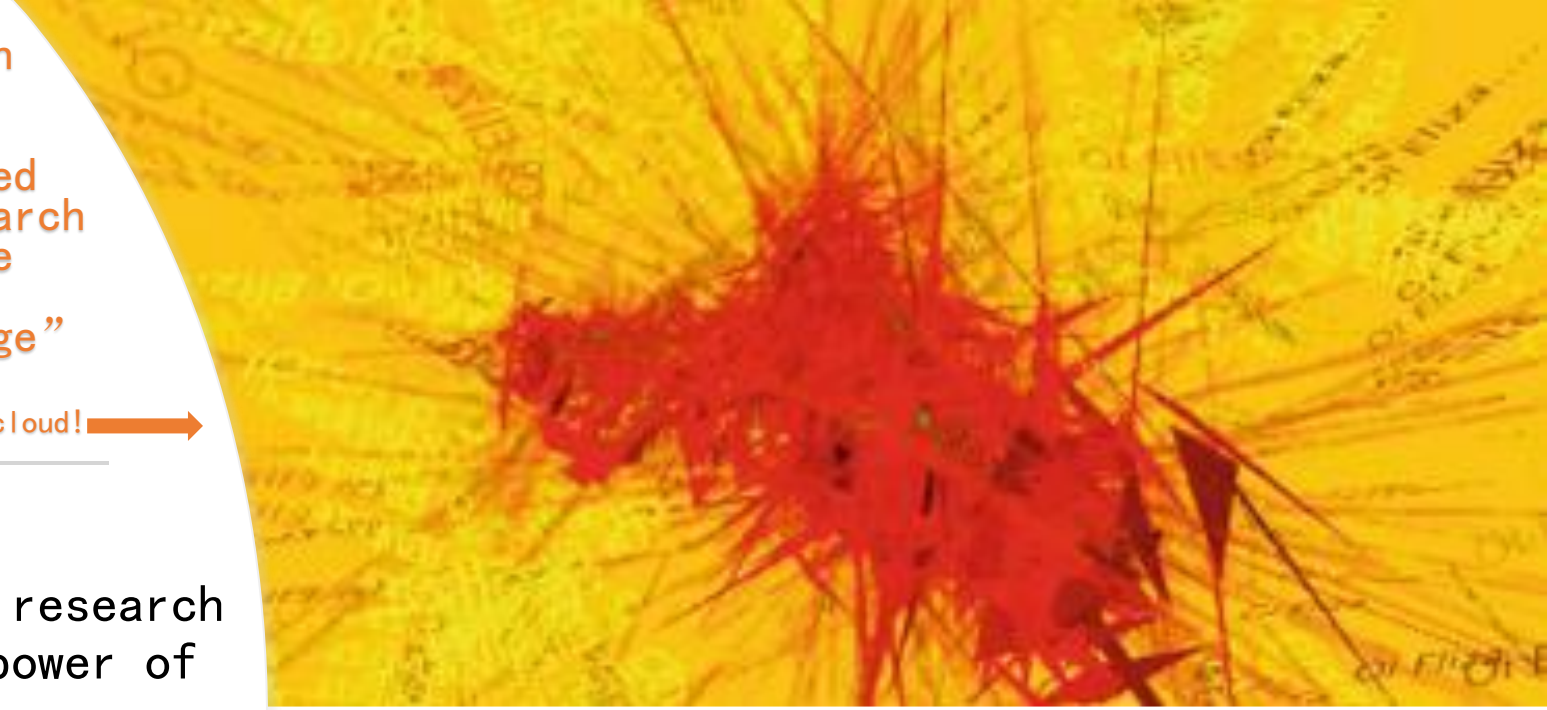
Image to Text, Text Classification & Interactive Natural Language Processing

By: Brian Hogan
Professor McCracken
ist644 - NLP

NLP ELIZA "Interlace"
Artwork by Tania Fraga

NLP artwork using an ELIZA system is an interesting form bridging human speech with a dynamic artform. Used as presentation background as images filled with curious language and shapes. Research did not have any algorithm or technique advancements. Paper was interested in computers generating a "poetic language" from real-time NLP interactions.*

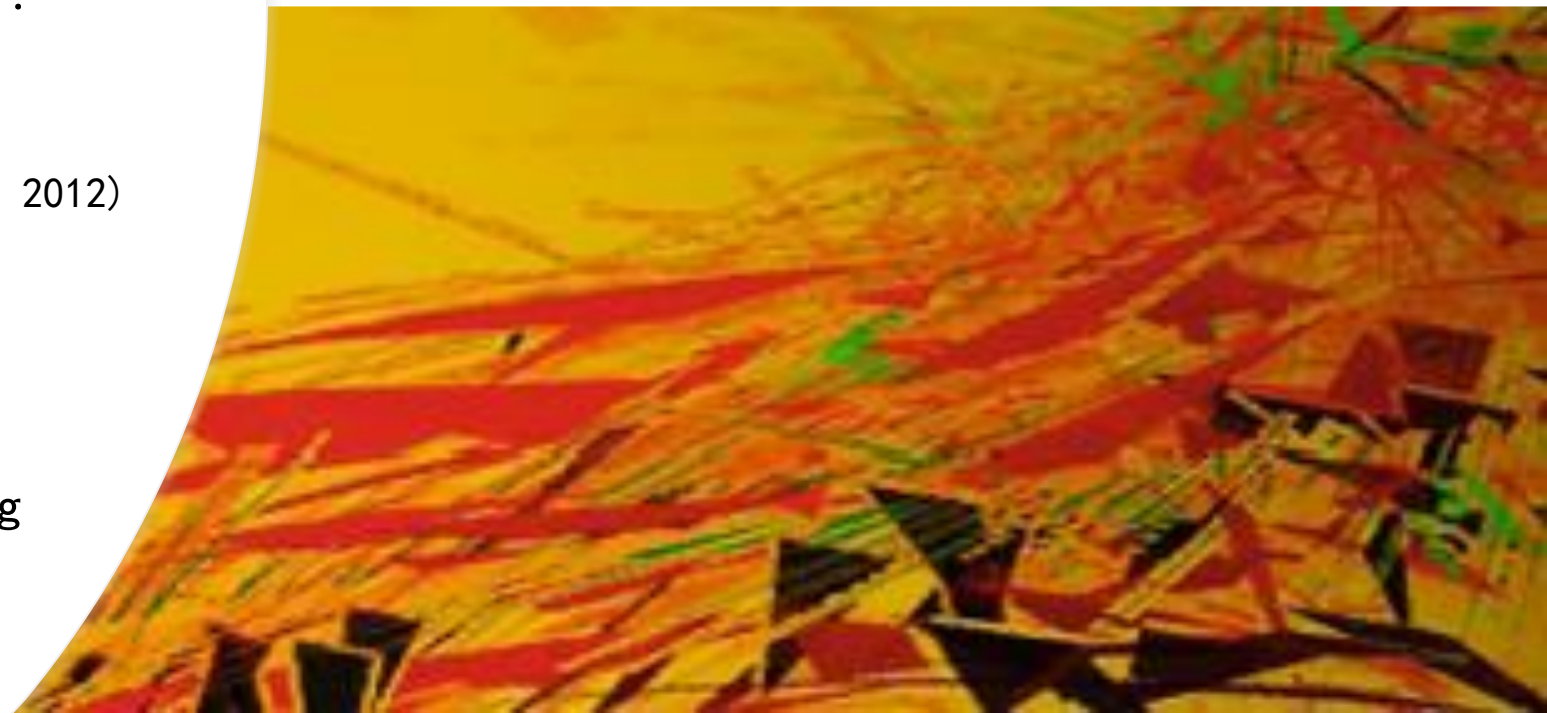
Wordsmash vs. Wordcloud!



Overview

The following slides blend current research and news stories on the scope and power of convolutional neural networks (CNN).

- CNN: how they work
- CNN: algorithm improvements (since 2012)
- Application area successes:
  - Facial pattern recognition
    - pros & cons
  - Natural Language Understanding
    - text-to-picture systems



*Fraga T. (2018) Designing Behaviors to Interactively Interlace Natural Language Processing, Text to Speech Procedures and Algorithmic Images.

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.*

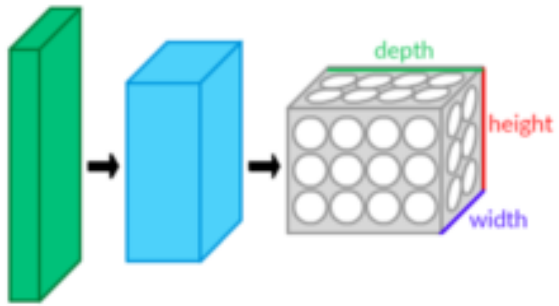
(a (process below)) Convolution layer: tensor inputs (i.e. a math array object) ==> (image) x (width x height) x (depth). Layers have learning filters called kernels that compute dot producing 2-d maps (neurons) who iterate and learn features of spatial position inputs.

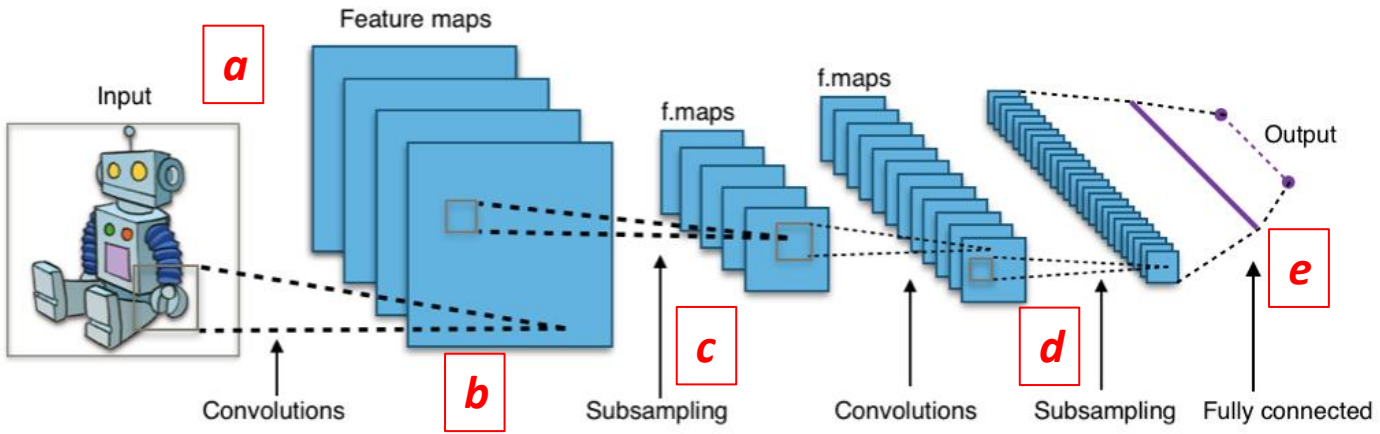
(b) Neurons (layer outputs) are filters along a depth dimension of a small input used to connect between tensor input feature maps. Neuron connectivity between layers becomes a hyperparameter (receptive field) whose (width x height) extend depth wise. The algorithm's "…architecture ensures learnt filters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





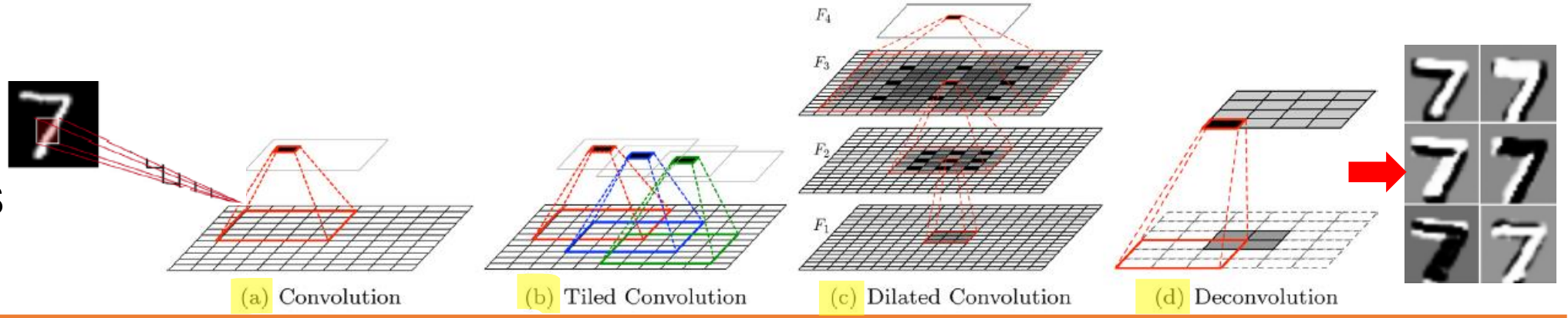
 # neurons to fit a volume

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

f

*References: Wikipedia (see reference slide)

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.

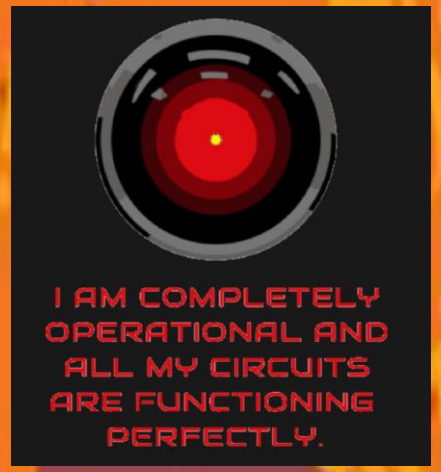
(c) Dilated Convolution: introduces a hyper-parameter to convolutional layer by *inserting zeros* between filter elements. Increasing the network's receptive field size widens relevant information and improves performance outcomes for scene segmentation, machine translation, speech synthesis, and speech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

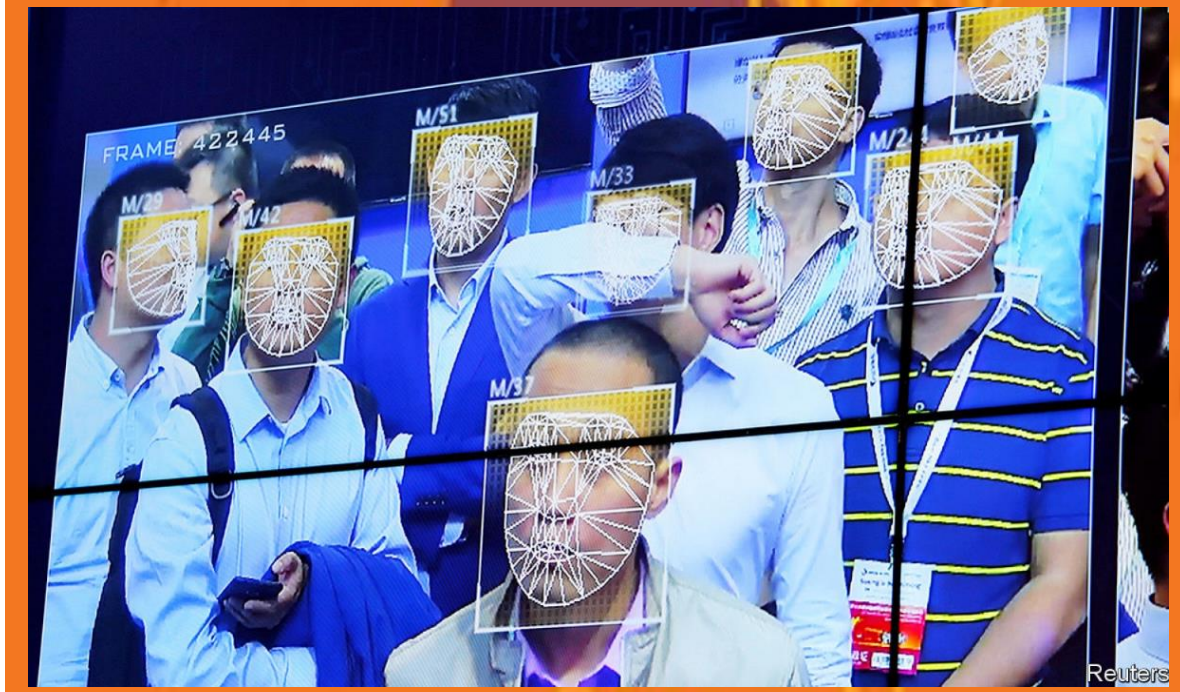
(d) Deconvolution: essentially convolution run backwards. Results in multiple image outputs, i.e. a larger pixel area, of a single activation a condensed image subset. Convolutions are run within "each" single activation (subset) improving the dilation factor for each input feature map. Groupings are 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*.

Facial Pattern Recognition

Who are you standing next to ?

- Facial recognition in China is a common feature crossing city sidewalks, subway stations, and tourist parks.
- Many office workers in Beijing's financial district clock in and out of work by scanning faces.
- "There is a feeling that everything you say and do is being monitored. It is terrifying.*"
- YouTube uploads 500 videos per hour and has sophisticated tools monitoring their content.**
- Convolutional neural networks assist in scanning for undesired content, such as gun violence.
- As their speed and proficiency increases real-time matching of a person's name with their image will become increasingly commonplace.





*Editors, (2019, November 9). The first face-off. *The Economist*, 71.

**Editors, (2019, May 19). Now playing, everywhere. *The Economist*, 17.

HAL Movie Image Courtesy of Microsoft Bing Search

Facial Pattern Recognition Pros & Cons

Situation: Deep convolutional neural networks contribute meaningfully in both facial and image recognition technology and *"the technology with the greatest potential to change policing is also the least visible to the public"* (Economist).*

Overall Advantages**	Overall Disadvantages*
An *Industrial revolution* in CNN accuracy has been achieved	Crime rarer than 1990s; technology skepticism is growing as facial recognition spreads in America
CNN algorithm is 20 times faster searching databases & finding matches	Algorithm reliability not confirmed across all race & age groups
0.2% error rate 12.3 million individual portraits	Weariness of using technology to monitor citizens
Quality algorithms across multiple applications	Possess unique threat to civil rights & liberties
New algorithms outperforming 2014 technology	"Technology exists...best thing...is to help shape it...*"

Miscellaneous

- Active development in private industry for sale, such as Amazon's "Rekognition"*
- Active evaluation by U.S. government for standards, concerns, and cause & effect social research***
- Facebook using technology to label people in uploaded photographs^^.

*References: Editor (2019, May 26). Files, Not Faces. *The Economist*, 27.

**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

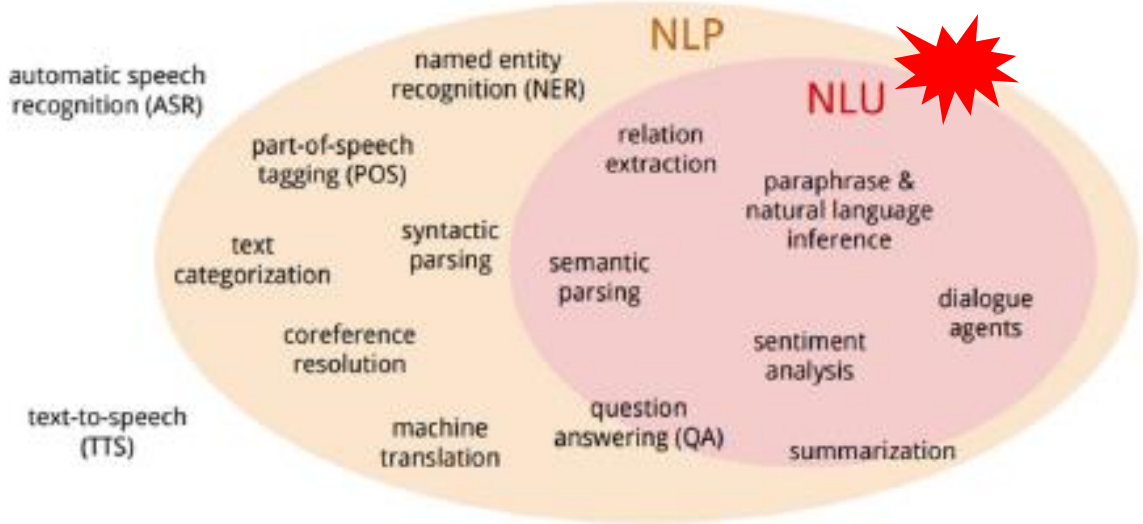
***Grother, P., Ngan, M., Hanaoka, K. Ongoing Face Recognition Vendor Test (FRVT) Part 2: Identification. *National Institute of Standards and Technology*.

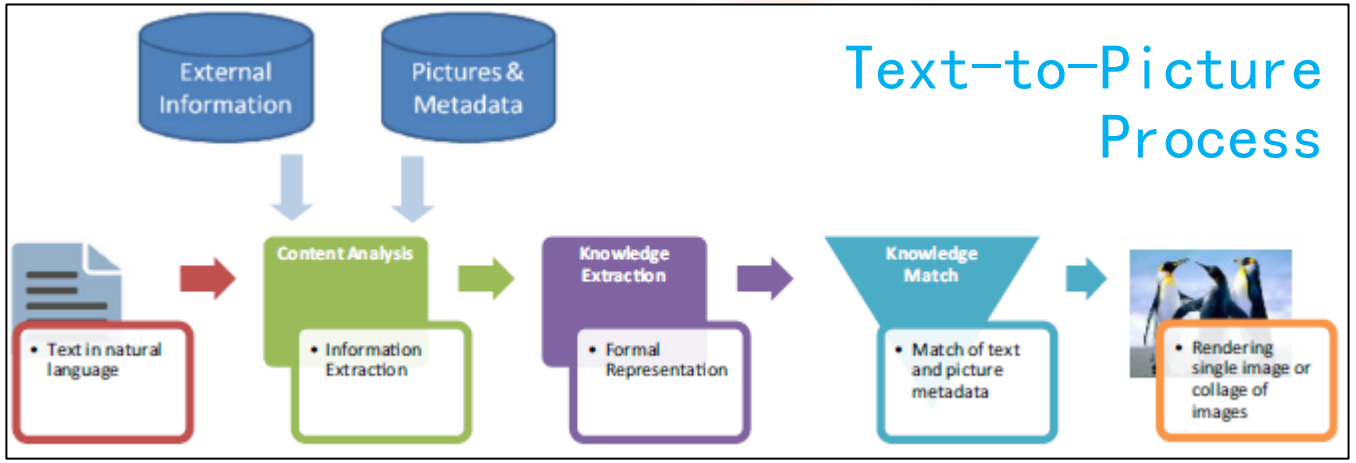
^^ Editor (2019, November 9). The first face-off. *The Economist*, 70.

Natural Language Understanding: Text-to-Pictures

Natural Language Understanding results in transforming natural languages from one representation into another by 'discovering hidden semantics' and converting them into a semantic text-to-picture representation.*

- Text-to-pictures positioned for language learning, education, rehab. from cerebral injuries, language disorders, and communication with autistic children.
- (many)systems lack strong NLP semantic analysis for capturing dependencies while matching text data to image metadata.
- NLU speaks to deeper semantic reasoning & relation extraction (ist644 \ wk_10_lecture)
- Presents as a growing field.



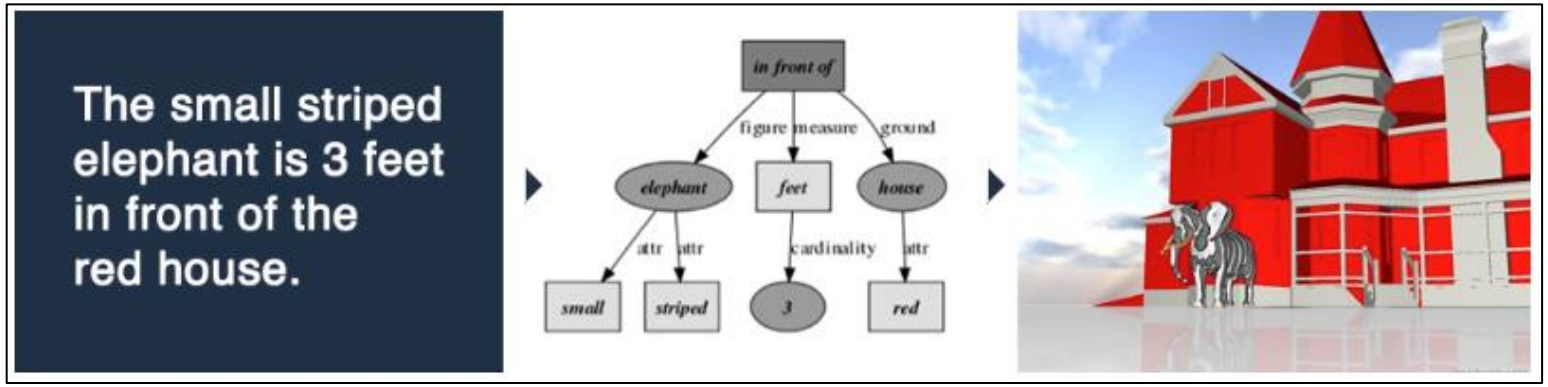


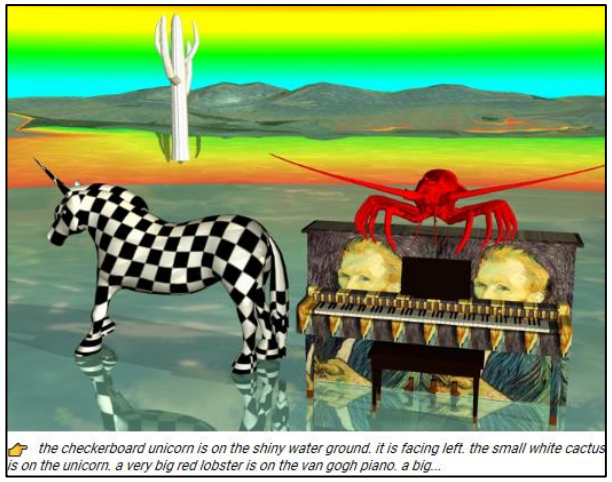
*Zakraoui, J., Saleh, M., Ja'am, J., (2019). Text-to-picture tools, systems, and approaches: a survey. *Multimedia Tools and Applications*.

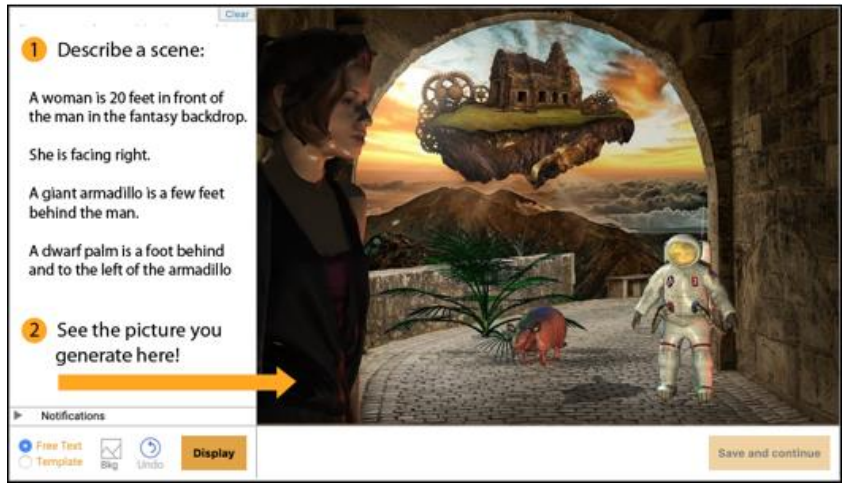
NLU text-to-picture systems: WordsEye (www.wordseye.com)

- Works by parsing text input into a semantic representation rendered as a 3D image. Relies on a large database for linguistic and world knowledge about objects, their parts, and properties.
- "...by using [simple] language, anyone can describe 3d scenes very quickly and easily."
- Promoted as educational tool for ESL, special needs, art, creative writing, and literacy skills education.
- Mobile app for image creation and social media sharing. Other users can refigure a received image.

~~~Semantic Rendering Process~~~







*Zakraoui, J., Saleh, M., Ja'am, J., (2019). Text-to-picture tools, systems, and approaches: a survey. *Multimedia Tools and Applications*.
~~~

References:	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