Computational Social Science with Images and Audio

Elliott Ash, **Philine Widmer**

06 October 2023

CNN are the current computer vision workhorse.

But not the only horse in the stable!

- Convolutional neural nets (abbreviated as CNNs or CovNets) are specialized neural networks for CV tasks
  - For various tasks, such as object detection, facial recognition, image segmentation, sentiment analysis
- Much of the recent image-based social science research uses CNN
- Other approaches are used, too:
  - Classic techniques: in the pre-deep learning and pre-CNN era, CV relied on hand-crafted features and algorithms
    - Edge detection, color histograms, template matching, ...
  - Other neural nets: recurrent neural networks (RNN), generative adversarial networks (GAN), transformers, stable diffusion models
    - RNN are often used for video analysis, GAN for image generation
  - Hybrid approaches: traditional approaches and deep learning
  - Unsupervised and semi-supervised learning: e.g., clustering

One promise of neural nets is a more end-to-end approach.

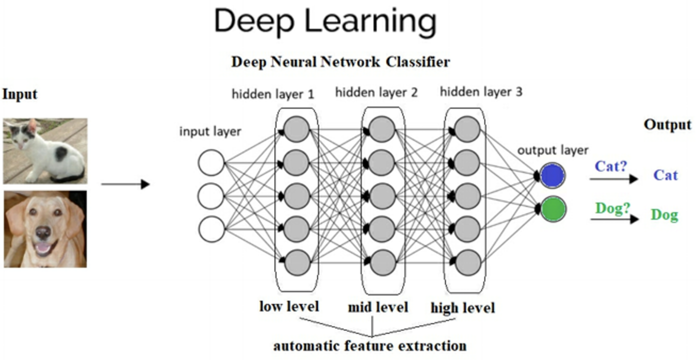


Figure: Dey (2018)[1]

[1] Dey, S. (2018). Hands-On Image Processing with Python: Expert Techniques for Advanced Image analysis and Effective Interpretation of Image Data. Packt Publishing Ltd.

CNN typically contain an input layer, hidden layers, and an output layer.

- Typically: input layer $\rightarrow$ hidden layers $\rightarrow$ output layer
- What do different convolutional layers do?
  - Layers close to the input layer learn low-level features (e.g., lines)
  - Middle layers learn complex abstract features (combining lower-level features)
  - Layers closer to the output interpret the extracted features in the light of the classification task

A typical application for a CNN in social science could be a visual sentiment analyzer.

- Two ways to think about emotions and images: expressed and evoked
  - What emotions are expressed? → Typically facial analysis/emotion annotation, objective-ish
  - What emotions are evoked? → For example, image sentiment, subjective-ish
- CNNs do a moderately good job of predicting evoked emotions (60–70% accuracy),[2] depending on...
  - The dataset being used
  - The architecture of the CNN
  - The training process
  - Maybe most importantly: How "evoked emotions" are defined and labeled

[2] Peng, K.-c., et al. (2015). A Mixed Bag of Emotions: Model, Predict, and Transfer Emotion Distributions. In: 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) pp. 1–9.

What emotions do these images evoke?

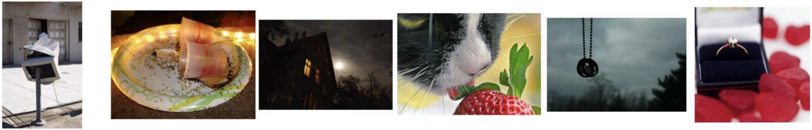


Figure: Source: Peng, K.-c., et al. (2015)

What emotion does each of these images evoke?

"Ground truth" in the Emotion6 Database (focusing on Ekman's 6 basic emotions plus neutral)

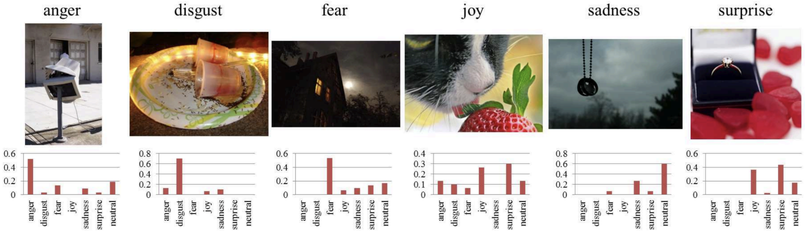


Figure: Source: Peng, K.-c., et al. (2015)

CNN are traditionally employed in supervised settings.

- CNNs are a specialized type of deep learning algorithm working with raw pixel values
- CNNs use many prelabeled training images to "learn" which pixel combinations are associated with the desired labels
- Once the performance is satisfactory, the algorithm can then be used to label large numbers of additional images quickly and at low cost
- Existing trained CNNs can be borrowed and fine-tuned ("transfer learning") for (rather) accurate results using a smaller training set of images
  - When is this a good/bad idea?
- Can you think of social science-relevant examples other than sentiment?

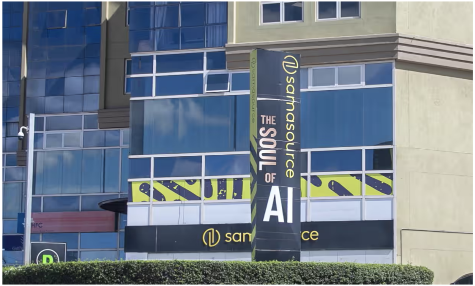
Training CNN requires substantial computing power and a lot of data.

- Access to graphics processing unit (GPU) *greatly* increases efficiency
- Supervised learning algorithms like CNNs require labeled images:
  - Options for generating original labels include hiring coders, using commercial services like LabelBox, Mechanical Turk
  - Researchers can also use benchmark image datasets with well-validated labels
  - Or rely on "creative" ways of labeling
    - For instance, annotate partisanship based on the data point's source
    - Intuition: a desired label might strongly correlate with an observable characteristic of the data ("proxy variable")
  - What are some ethical challenges of image annotation?
- A lot of data requires a lot of (robust) storage

Data labeling still represents *a lot* of human work.

‘It’s destroyed me completely’: Kenyan moderators decry toll of training of AI models

Employees describe the psychological trauma of reading and viewing graphic content, low pay and abrupt dismissals



Office premises of Sama in Nairobi, Kenya. Photograph: Tony Karumba/AFP/Getty Images

Figure: Source: The Guardian (2023)[3]

[3] “It’s destroyed me completely”: Kenyan moderators decry toll of training of AI models. The Guardian. (2023). https://www.theguardian.com/technology/2023/aug/02/ai-chatbot-training-human-toll-content-moderator-meta-openai

Von Kempelen's "automaton": a technology masterpiece?

- In 1769, inventor Wolfgang von Kempelen introduced a chess-playing automaton[4]
- It was a robot/"dummy" in Turkish attire, seated next to a large cabinet with machinery and a chessboard on top
- Visitors could inspect the cabinet and its complex machinery
- The Turk defeated notable figures such as Napoleon Bonaparte and Benjamin Franklin
- The term "automaton" gained popularity from this narrative
- Despite von Kempelen's renowned inventor status, skepticism persisted due to the machine's unbelievable capability
- Have you heard the story? What do you think?

[4] Story as recounted in: Shiller, R. J. (2019). Narratives about technology-induced job degradations then and now (No. w25536). National Bureau of Economic Research.

Von Kempelen's "automaton": an illusion masterpiece!

- There was a complex mechanism that obscured the visitors':
- They did not see that there was a man, a chess master, hiding inside
- A magnet system allowed the master to see the positions of the pieces on the board
- A system of controls enabled him to move the arms of the automaton to move the pieces

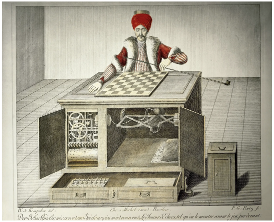


Figure: Source: Encyclopaedia Britannica, Photo credits: INTERFOTO/Alamy

Getting back to the technicalities: CNN are ANN.

- Artificial Neural Networks (ANN) are a broad category of statistical models
- They use **intermediate abstract representations** (hidden layers) of the input data to learn and predict outcomes
- CNN are an ANN subtype for processing grid-like data structures such as images, where spatial hierarchies and local patterns (edges, textures, etc.) are important
- These intermediate representations result from complex interactions and nonlinear transformations of the input matrix
- Coefficients are estimated for these intermediate representations
- *Deep* learning refers to the presence of *multiple* hidden layers
- Less standardized notions of “breadth” (e.g., # of neurons in a layer)

Convolution is fundamental for the abstract representations in CNN.

- The primary operation in CNNs is convolution, not just simple matrix multiplication
- Convolution involves applying a filter (or kernel) over the input data to produce a feature map
- This helps detect local patterns, such as edges or textures, in different regions of the input data
- Along with convolution, non-linear transformations are applied using an activation function (e.g., RELU) to introduce non-linearity into the model
  - This helps the network capture complex relationships in the data
- These feature maps (or convolved outputs) form new abstract representations that can be fed into subsequent layers

Nonlinear transformations are commonly applied to the hidden layers in deep learning models.

- Nonlinear transformations are also known as activation functions
- Common activation functions include RELU, Sigmoid, and Tanh

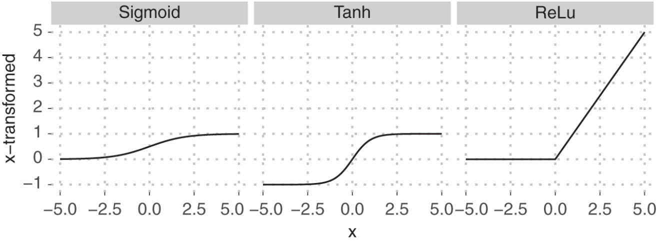


Figure: Three common nonlinear transformations (activation functions) in deep learning. Source: Webb Williams et al. (2020), cf. first slide set.