Computational Social Science with Images and Audio

Elliott Ash, **Philine Widmer**

27 October 2023

CNN are helpful beyond classification.

- So far, we have discussed classical machine learning and CNNs
- When discussing CNNs, we've primarily considered (supervised) image analysis
- As you recall, a classification model typically proceeds as follows:
  - Feed the raw image pixel data to the model
  - Convolve the pixel values (by color channel) to produce feature maps
  - Apply non-linear activation functions
  - Apply additional layers like pooling to down-sample and possibly dense (fully connected) layers
  - Use an output layer to map the results to probabilities
- Now, what if we want to *generate* instead of analyse images?

A typical architecture used for image generation is GAN.

- GAN stands for Generative Adversarial Network
- Start with a random noise vector
  - Usually a 1D array of random numbers
- Feed this noise vector into a generator network
  - The generator transforms the noise into an image
    - It uses deconvolution layers and non-linear activations
- The generated image is passed to a discriminator network
  - The discriminator, a typical CNN classifier, predicts if the image is real or generated
- Both networks are trained together:
  - Generator aims to deceive the discriminator
  - Discriminator aims to identify real vs. generated correctly

What is deconvolution?

It is also called transposed convolution.

- Given a matrix z,

$$z = \begin{bmatrix} 1 & 3 & 4 \\ 6 & 1 & 2 \\ 1 & 2 & 1 \end{bmatrix}$$

- we can introduce zeroes between the elements...

$$z' = \begin{bmatrix} 1 & 0 & 3 & 0 & 4 \\ 0 & 0 & 0 & 0 & 0 \\ 6 & 0 & 1 & 0 & 2 \\ 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 2 & 0 & 1 \end{bmatrix}$$

- and then “regularly” convolve the expanded matrix x' with a filter
- This “upscales” the image
- In practice, there might be more efficient algorithms!

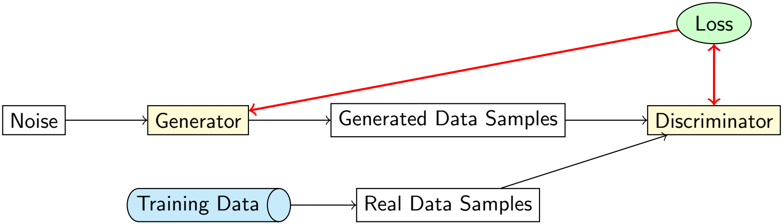
GAN are "adversarial" because the generator and the discriminator play against each other.

- z: The random noise vector
- $G(z)$: The generator's function transforming z into an image
- $D(x)$: The discriminator's function estimating the probability that an image x is real
- A potential training objective:

$$\min_G \max_D V(D, G) = \mathbb{E}_{x \sim p_{\text{data}}(x)}[\log D(x)] + \mathbb{E}_{z \sim p_z(z)}[\log(1 - D(G(z)))]$$

  - $p_{\text{data}}(x)$ is the true data distribution
  - $p_z(z)$ is the noise distribution
  - The generator minimizes V while the discriminator maximizes it

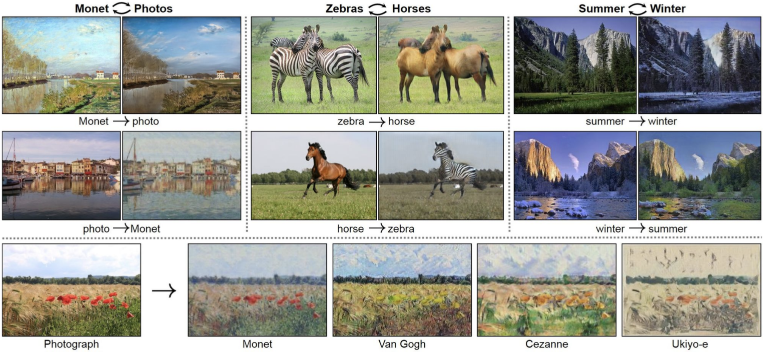
The loss function connects discriminator and generator.



How do we want the “game” to end?

- Who should “win” the game?
- Typically, convergence would imply the following:
  - Nash equilibrium: neither player can reduce their cost without the other player changing their strategy
  - The generator’s images are indistinguishable from real ones
  - The discriminator outputs 50% (real vs. fake) for every image
- In practice, training a GAN is challenging
  - How could it go sideways?
  - Researchers have proposed many fixes involving architectural choices, modified loss functions, regularization, monitoring, learning curricula, ...

Typically, we don't just want *any* realistic image.



Use cases of CycleGAN. Source: https://github.com/hyunbo9/yonsei

Figure: CycleGAN

cGAN or CycleGAN are examples of GAN with more specific requirements.

- Conditional GAN (cGAN):
  - In a cGAN, the generator receives random noise *and* some additional information (a condition)
  - This condition could be a label, another image, or any instruction about the output type
  - The discriminator receives the generated image and the condition
- CycleGAN:
  - Learns mappings between two domains (A and B) without paired training examples
  - Uses two GANs: G: A $\rightarrow$ B and F: B $\rightarrow$ A
  - Cyclic consistency losses ensure that the domain transformations make sense

Evaluating the performance of GAN is not straightforward.

Also applies to other generative methods.

- It is challenging because of its unsupervised nature
- Quantitative evaluation:
  - The inception score (IS) measures the quality and diversity of generated images ($\uparrow$)
  - The Frechet inception distance (FID) measures the distance between feature vectors of real and generated samples ($\downarrow$)
  - Precision, recall, and the F1 score can be used with pre-trained classifiers
- Qualitative evaluation:
  - Visual assessment of generated images
  - Comparisons with real data samples (e.g., human validation $\rightarrow$ might require expertise!)
- All metrics come with limitations; no definitive metric

An example for quantitative evaluation is the IS.

- The inception score (IS) evaluates the quality and diversity of generated images
- Intuition: the generated images should be diverse and easily classifiable

$$IS(G) = \exp\left(\mathbb{E}_{x\sim p_g}[D_{KL}(p(y|x) \parallel p(y))]\right)$$

- Where
  - x is a generated image
  - $p(y|x)$ is the conditional class distribution given image x (obtained using a classifier)
  - $p(y)$ is the marginal class distribution over all generated images (basically the average of $p(y|x)$)
  - D_{KL} is the Kullback-Leibler divergence
- GANs can "game" the IS (e.g., unrealistic, high-scoring images)

How can generated images be used in social science?

- Experimentation
  - For example, how does changing one aspect of an image influence its appeal?
- Simulation
  - For example, when tracking and analyzing the evolution of culture, one could use generated images to simulate potential future trends
  - Experimental simulation of social interactions (to study human behavior in controlled but realistic settings)
- Studies of perceptions and biases
  - Generated image could reflect public stereotypes
- Other ideas?

To conclude, let us briefly discuss again misinformation...

... and the role of AI

- In the past Swiss elections (10/2023), five parties have released a joint "AI Codex"
- It acknowledges that technological advancements, in particular Generative AI, influence political communication
  - Quick production of content in large quantities and increasing quality
- The parties highlight some risks, such as falsely attributing statements or facts to political actors
- Currently, there are no rules on the use of AI-generated content
- The parties commit to preventing deception of the public through AI
- They promise to adhere to these principles:
  - Declare the AI authorship for auditory and visual campaign elements
  - No use of such AI-generated content for negative campaigns

Five Swiss parties committed to “responsible” use of Generative AI.

KI-Kodex

gemeinsame Erklärung von GRÜNE, GLP, SP, Die Mitte und EVP zum verantwortungsvollen Umgang mit KI- generierten Inhalten

Die Technologien unterliegen einem rasanten Wandel und beeinflussen immer stärker auch die politische Kommunikation. Generative künstliche Intelligenz (KI) bietet neue Möglichkeiten, automatisiert Text, Ton, Bilder und Videos zu generieren – einfach, rasch, in grosser Menge und in immer besserer Qualität.

KI ist eine Chance, birgt aber auch Gefahren. Denn mit KI lassen sich verfängliche auditive und/oder visuelle Inhalte erstellen. Also Bilder, Tondokumente oder Videos, die täuschend echt sind, aber für Negativ-Kampagnen missbraucht werden, indem sie politischen Akteurinnen und Akteuren falsche Tatsachen oder Aussagen unterstellen.

Da bisher Regeln zum Umgang mit KI- generierten Inhalten fehlen, verpflichten wir uns dem Ziel, eine absichtliche Täuschung der Öffentlichkeit mithilfe von KI zu verhindern, damit nicht das Vertrauen in die Demokratie untergraben wird.

Zu diesem Zweck halten wir uns in der politischen Kommunikation an die folgenden beiden Grundsätze:

(1) Wir deklarieren die Urheberschaft von KI bei der Erstellung von auditiven und/oder visuellen Kampagnenelementen.

(2) Wir nutzen keine KI-erzeugten auditiven und/oder visuellen Inhalte für Negativ-Kampagnen.

Dadurch tragen wir dazu bei, einen Umgang mit KI zu finden, der unserem politischen System gerecht wird. Denn die Demokratie baut auf Vertrauen auf und auf dem fairen Wettbewerb zwischen unterschiedlichen politischen Lösungsvorschlägen und Werten.

Die Erklärung gilt für die nationalen Parteien. Alle Generalsekretariate verpflichten sich ihre Kantonalsektionen und die Kandidierenden für die Gesamterneuerungswahlen auf diese Erklärung hinzuweisen.

Unterzeichnet von:

Balthasar Glättli, Präsident GRÜNE
Jürg Grossen, Präsident GLP
Mattea Meyer & Cédric Wermuth, Co-Präsidium SP
Gerhard Pfister, Präsident Die Mitte
Lilian Studer, Präsidentin EVP

Figure: The five parties’ codex in the original (2023).