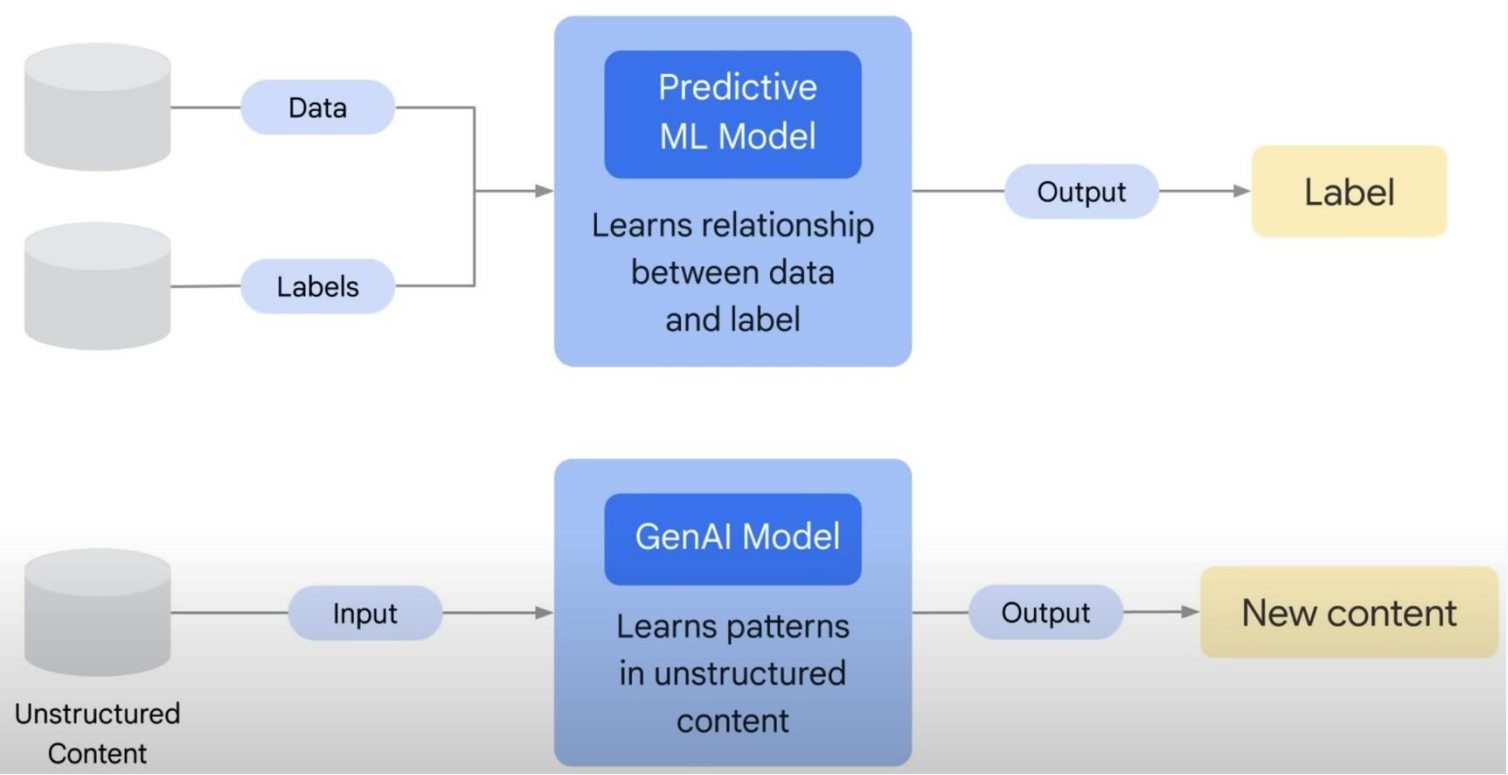
IA PER LA COMUNICAZIONE

Generative AI: L'intelligenza artificiale diventa creativa

Prof. ssa Romina Eramo
Dipartimento di Scienze della Comunicazione
Università degli Studi di Teramo
reramo@unite.it

Cos’è l’IA Generativa?

- Modelli che creano nuovi output: testo, immagini, video.
- Non producono etichette ma contenuti originali.
- Basati su reti neurali.



Tipologie di Modelli Generativi

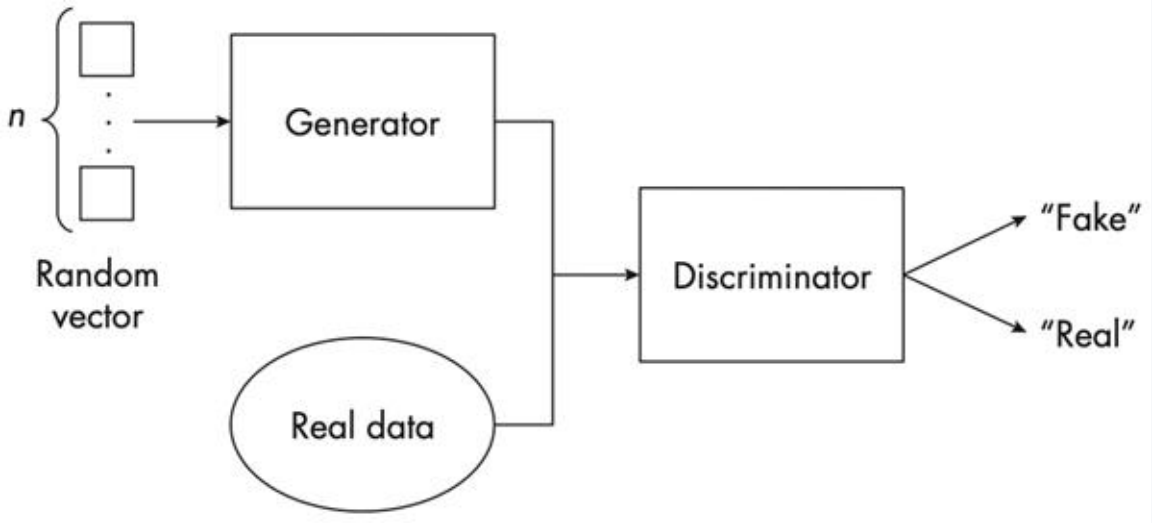
- **Reti Generative Avversarie (GAN)**: competizione tra generatore e discriminatore.
- **Modelli di Diffusione**: trasformano rumore in immagini realistiche.
- **Modelli Linguistici di Grandi Dimensioni**: trattati nella lezione successiva.

Reti Generative Avversarie (GAN)

- Struttura: Generatore e Discriminatore.
- Generatore: crea output falso.
- Discriminatore: distingue tra vero e falso.

Funzionamento di una GAN

- Generatore
  - Input: vettore casuale (rumore).
  - Output del generatore: esempio plausibile del set di training.
  - Obiettivo: confondere il discriminatore tra reale e falso.
- Discriminatore
  - Input: dati reali e output del generatore
  - Output: etichetta "Vero" o "Falso"

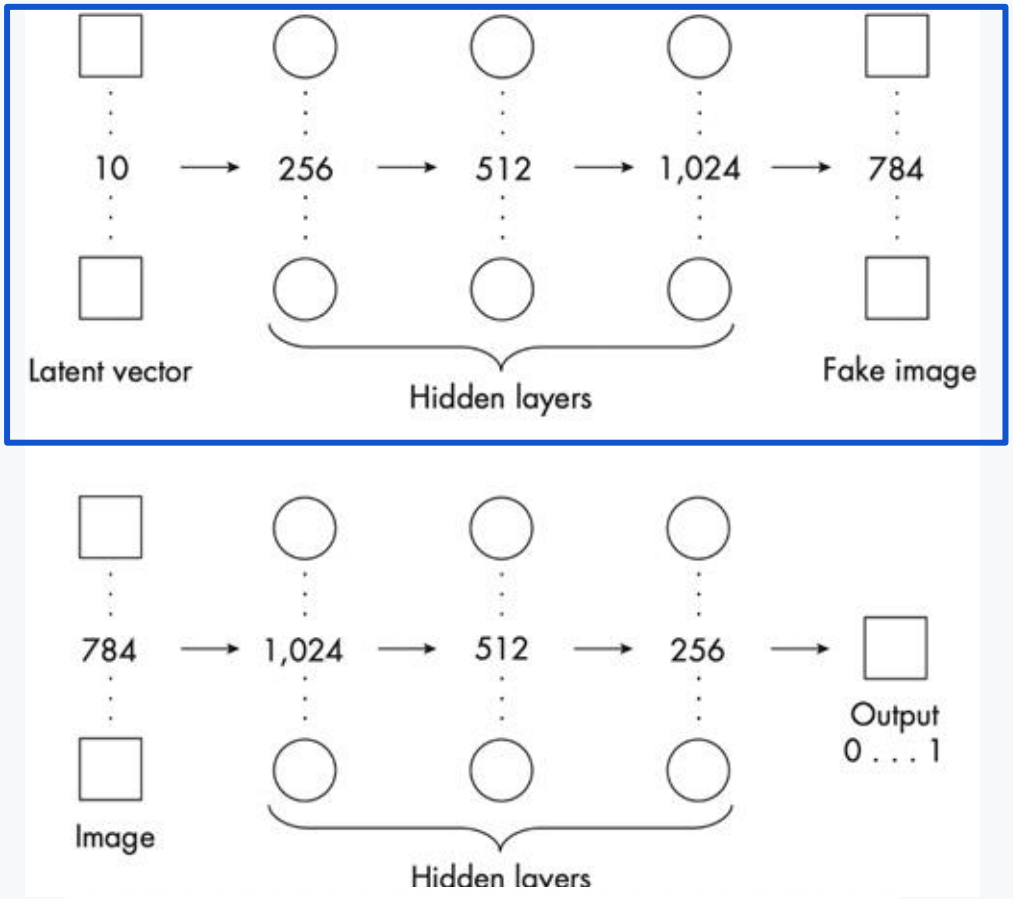


Esempio con MNIST

- **Generatore**: trasforma un vettore casuale di 10 numeri (n=10) in un'immagine numerica
  - Una volta addestrato, possiamo fornire al generatore qualsiasi raccolta di 10 valori intorno allo zero, e il generatore produrrà una nuova immagine di cifre come output, imitando così il processo che ha creato il set di dati MNIST
- **Risultati dopo il training del generatore**: immagini chiare e distinte (dopo epoche avanzate)
- **Discriminatore**: valuta la somiglianza con i dati reali e produce valori da 0 a 1
  - Convinzione del discriminatore che l'input sia reale (output vicino a 1) o falso (output vicino a 0)

Esempio con MNIST

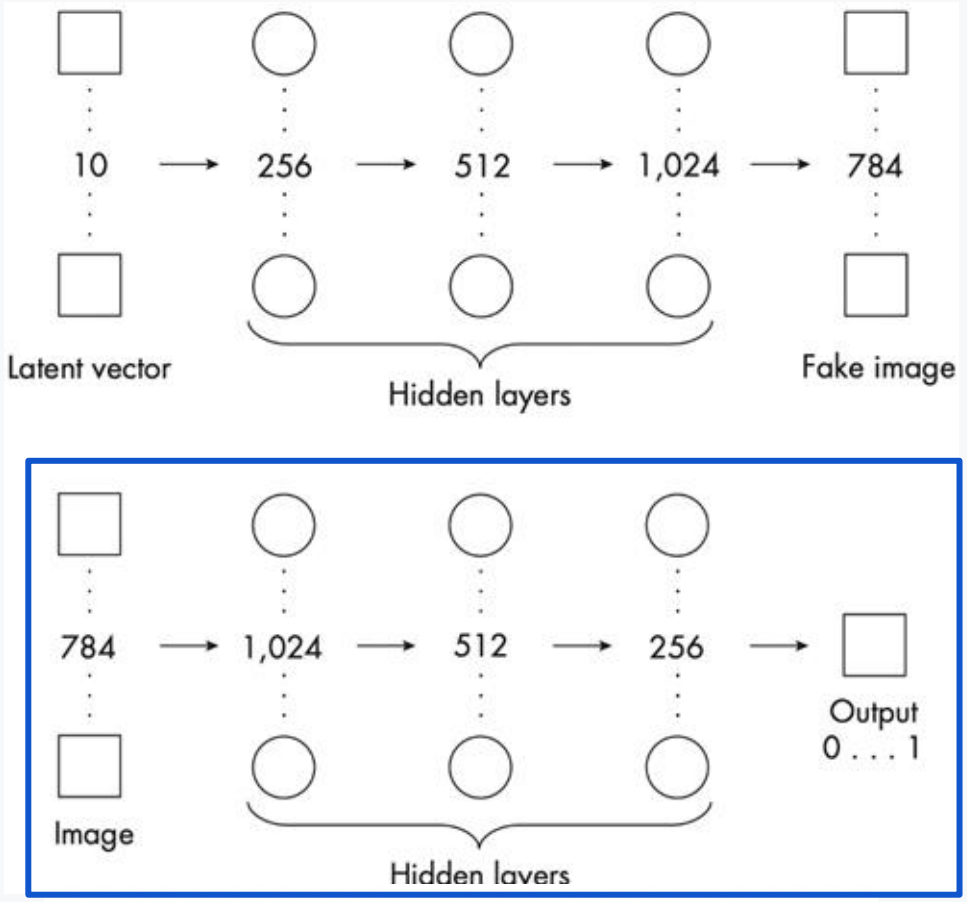
- **Generatore**: semplice GAN basata su reti neurali tradizionali
  - il modello ha tre strati nascosti, con *256*, *512* e *1.024* nodi, e uno strato di output di *784* nodi per produrre l'immagine
  - i nodi del livello nascosto utilizzano una versione modificata dell'unità lineare rettificata chiamata *ReLU leaky*



GAN generator (top) and discriminator (bottom)

Esempio con MNIST

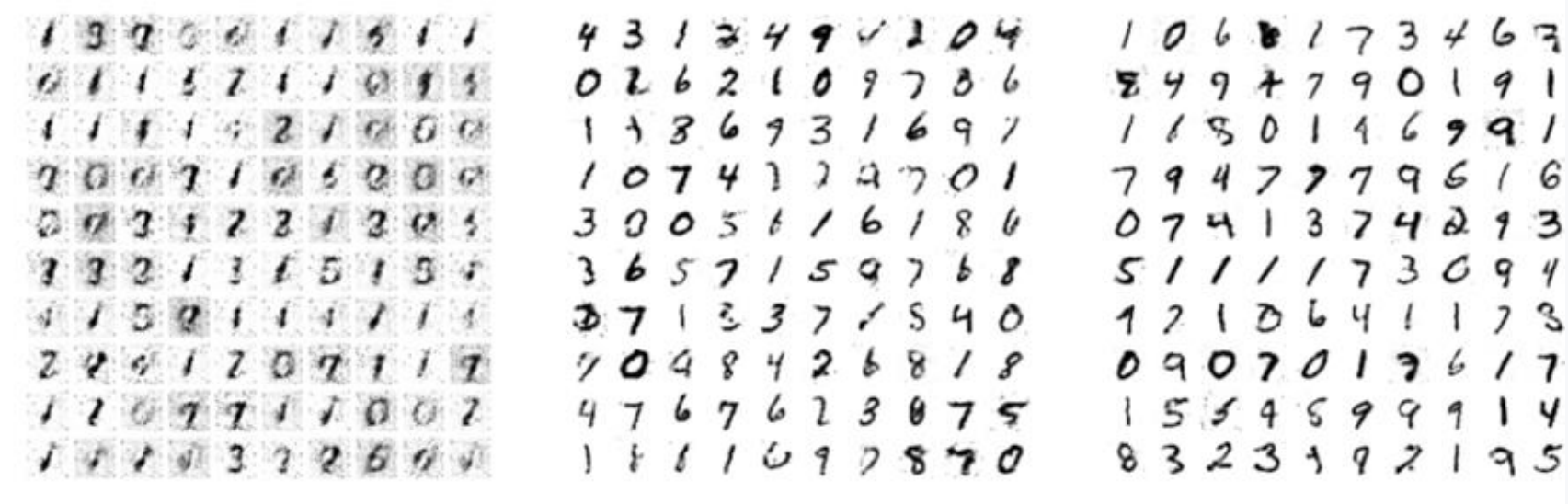
- **Generatore**: semplice GAN basata su reti neurali tradizionali
  - il modello ha tre strati nascosti, con 256, 512 e 1.024 nodi, e uno strato di output di 784 nodi per produrre l'immagine
  - i nodi del livello nascosto utilizzano una versione modificata dell'unità lineare rettificata chiamata ReLU leaky



GAN generator (top) and discriminator (bottom)

Esempio con MNIST

- Il generatore si comporta male dopo un singolo passaggio attraverso i dati di addestramento
- Dopo 60 epoche, il generatore produce una gamma completa di cifre, alcuni sono ancora confusi
- Dopo 200 epoche, la maggior parte delle cifre sono distinte e ben definite. Il generatore è addestrato.

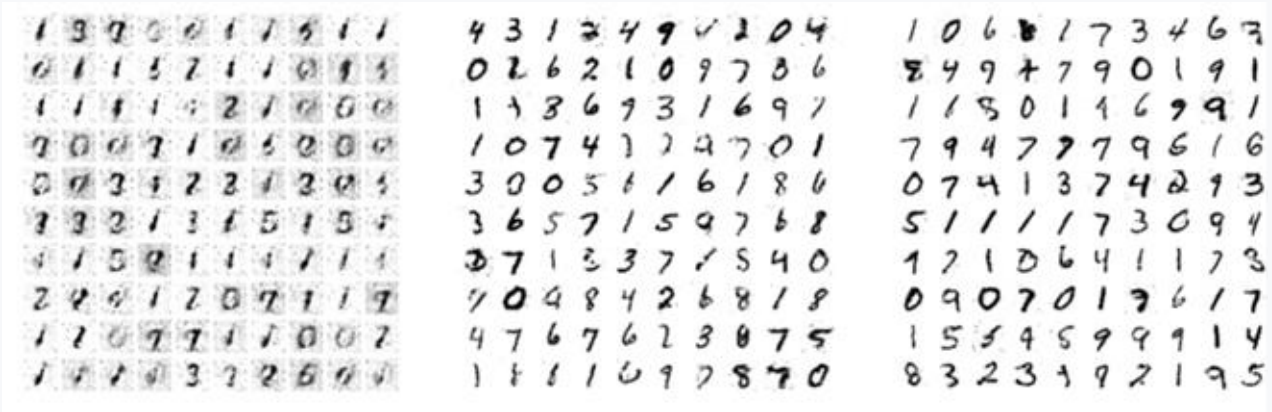


Problemi nelle GAN (incontrollabili)

- Collasso della modalità: output monotono.
  - Il generatore impara come creare una serie di esempi buoni e rimane intrappolato nel produrre solo quell'output e non la diversità di immagini desiderata
    - La predilezione del GAN per le cifre uno era evidente fin dall'inizio dell'addestramento

Digit	Percentage
0	10.3
1	21.4
2	4.4
3	7.6
4	9.5
5	6.0
6	9.1
7	14.4
8	4.4
9	12.9

The Actual Digit Distribution



GAN condizionali

- Le GAN condizionali aggiungono un'informazione specifica (un condizionamento) come input al generatore e al discriminatore.
- Il condizionamento è spesso un vettore one-hot o un'altra forma di informazione di classe che specifica una categoria (esempio: "0", "1", "2" per le cifre MNIST).
  - Durante l'addestramento, la rete impara a generare dati che appartengono a una specifica classe quando viene fornito il condizionamento corrispondente.
- Applicazioni:
  - Generare immagini specifiche per una classe desiderata, come cifre da 0 a 9 o tipi di oggetti (ad esempio, "cani" o "gatti").

GAN condizionali

- GAN condizionale addestrata sul dataset MNIST
  - è possibile generare cifre “4” specificando l'etichetta di classe come input.
  - oppure è possibile produrre 10 campioni di ogni cifra

Le GAN condizionali ci permettono di selezionare la classe di output desiderata, cosa che le GAN non condizionate non possono fare, ma cosa succede se vogliamo regolare caratteristiche specifiche dell'immagine di output?



GAN controllabili

- Le GAN controllabili permettono di modificare caratteristiche specifiche degli output generati attraverso movimenti in uno spazio latente appreso.
- Lo spazio latente (generato dai vettori casuali) viene analizzato per individuare direzioni semantiche significative (ad esempio, barba o colore dei capelli nei volti umani).
- Spostarsi lungo queste direzioni modifica le caratteristiche dell'immagine in modo controllato.
- Applicazioni:
  - Modifica di attributi visivi, come l'aggiunta di occhiali o la regolazione dell'età in immagini di volti.
  - Adattamento progressivo delle caratteristiche visive per design, moda, o personalizzazioni.
  - Esempio pratico: Usare una GAN controllabile per modificare il colore dei capelli in un ritratto generato, passando da nero a biondo.

GAN controllabili

- Muovendosi attraverso lo spazio astratto rappresentato dal vettore di rumore si produce un'immagine di output dopo l'altra.
- Le GAN controllabili scoprono direzioni significative nello spazio del rumore.
- Muovendosi lungo una di queste direzioni si altera la caratteristica relativa alla direzione.

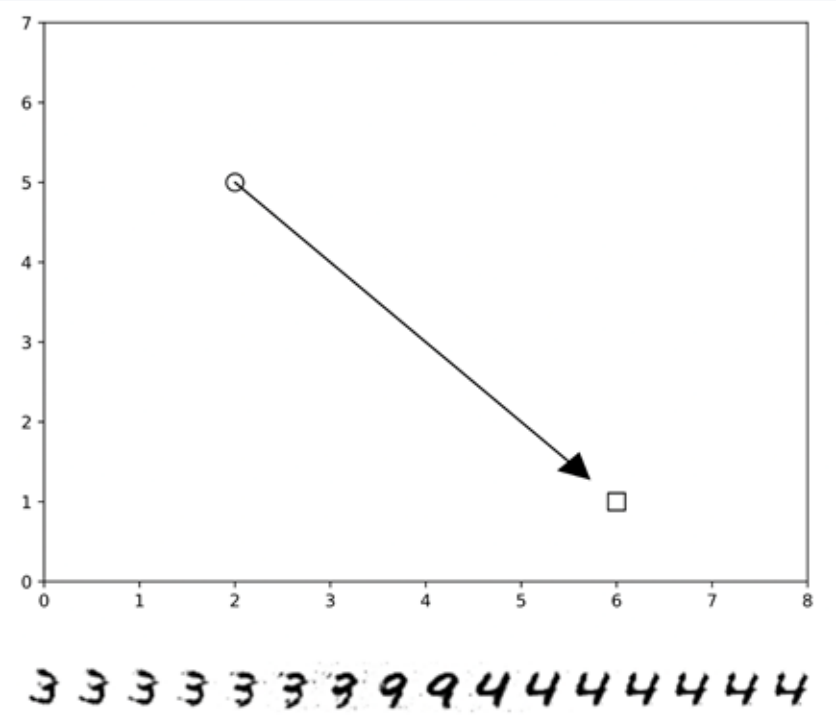
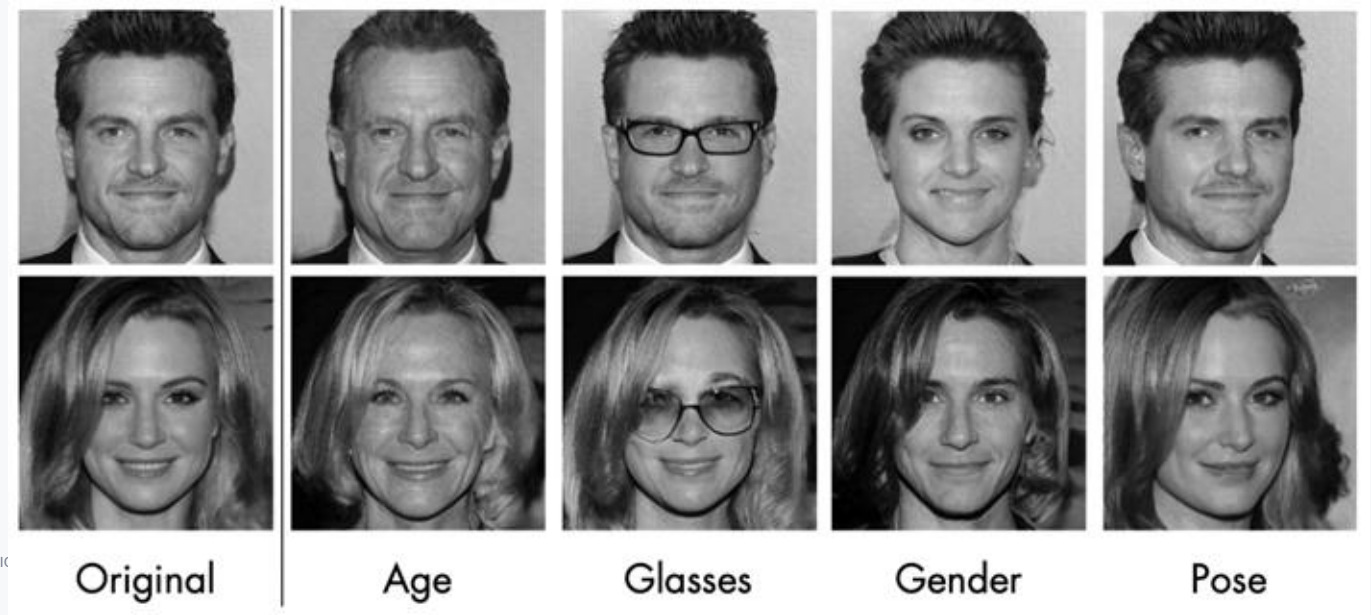


Figure 6-6: Moving through a two-dimensional noise space and interpolated MNIST digits

GAN controllabili

- La figura mostra due volti falsi generati a sinistra e volti aggiustati a destra
- Le regolazioni corrispondono al movimento attraverso lo spazio del rumore dalla posizione originale dell'immagine lungo le direzioni apprese che rappresentano l'età, gli occhiali, il sesso e la posa.



Modelli di Diffusione

- Creano dati nuovi (come immagini) partendo da rumore casuale
- A differenza delle GAN, non utilizzano una competizione tra reti, ma si basano su un processo probabilistico per apprendere come trasformare il rumore in dati realistici
- Esempio di Processo di Diffusione:
  - Input: Un'immagine (es. un volto).
  - Fase di Diffusione: Viene aggiunto rumore gradualmente fino a trasformare il volto in puro rumore casuale.
  - Fase Inversa: Il modello rimuove gradualmente il rumore, ricostruendo il volto originale.

Diffusione Condizionale

- Input: incorporamento testuale o immagine di base.
- Controllo: specifica dettagli dell'immagine finale.

Esempi di Applicazioni di Modelli di Diffusione

- Stable Diffusion: Generazione di immagini realistiche guidate da un prompt testuale.
- DALL-E 2: Generazione di immagini basate su descrizioni testuali.
- Immagini condizionate: Creazione di immagini con specifiche caratteristiche o elementi stilistici.
- Super-risoluzione: Aumentare la qualità e il dettaglio di immagini a bassa risoluzione.

Esempio di Modello di Diffusione

- Gioconda di Leonardo da Vinci e cinque varianti
- Le variazioni sono prodotte in risposta all'immagine originale e a un prompt di testo:
- *“Ritratto di una donna che indossa un abito marrone nello stile di Da Vinci, colori tenui e di terra”*



Errori dei Modelli di Diffusione

- La generazione di immagini AI non è perfetta
- La forma esatta del prompt di testo influenza fortemente il processo di generazione dell'immagine



Vasi generati da un modello di diffusione

- Immagini generate da permutazioni delle parole ornato, verde e vaso.
- L'immagine iniziale del rumore era la stessa ogni volta; variava solo l'ordine delle tre parole.
- Tre dei vasi sono simili, ma il quarto è molto diverso
- Tutti e quattro sono validi esemplari di vasi verdi decorati



Vasi generati da un modello di diffusione

- Il prompt "ornato, verde, vaso" è stato modificato cambiando "verde" con altri colori (rosso, malva, giallo e blue) utilizzando la stessa immagine di rumore iniziale di prima
- Il vaso blu (l'ultimo) differisce in modo significativo



Conclusione

- Confronto GAN e Diffusione
  - GAN: basate sulla competizione, necessitano di training accurato
  - Diffusione: addestramento senza competizione, più adatte a condizionamenti testuali
- L'IA generativa è rivoluzionaria: arte, ricerca e industria.
- Prossimo passo: modelli linguistici di grandi dimensioni.

Esempio

Prompt:

Vista mozzafiato del Serengeti con fauna selvatica errante, vasto e dettagliato paesaggio naturale

GPT 3



GPT 4 (1)



GPT 4 (2)



GPT 6

