

IMAGE COMPLETION WITH DEEP LEARNING

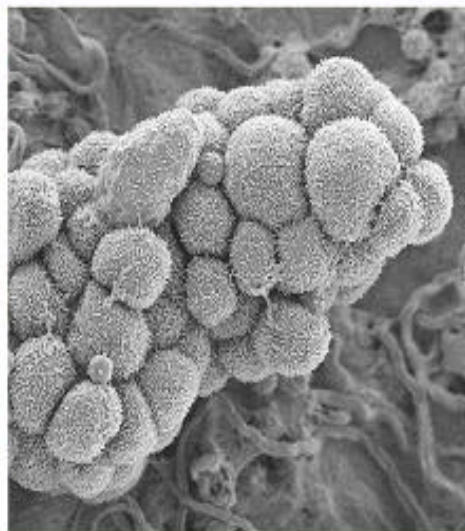
What is Deep Learning?

DEEP LEARNING

Deep learning (also known as deep structured learning or hierarchical learning) is part of a broader family of machine learning methods based on the layers used in artificial neural networks.

- use a cascade of multiple layers of nonlinear processing units for feature extraction and transformation. Each successive layer uses the output from the previous layer as input.
- learn in supervised (e.g., classification) and/or unsupervised (e.g., pattern analysis) manners.

DEEP LEARNING EVERYWHERE



INTERNET & CLOUD

Image Classification
Speech Recognition
Language Translation
Language Processing
Sentiment Analysis
Recommendation

MEDICINE & BIOLOGY

Cancer Cell Detection
Diabetic Grading
Drug Discovery

MEDIA & ENTERTAINMENT

Video Captioning
Video Search
Real Time Translation

SECURITY & DEFENSE

Face Detection
Video Surveillance
Satellite Imagery

AUTONOMOUS MACHINES

Pedestrian Detection
Lane Tracking
Recognize Traffic Sign

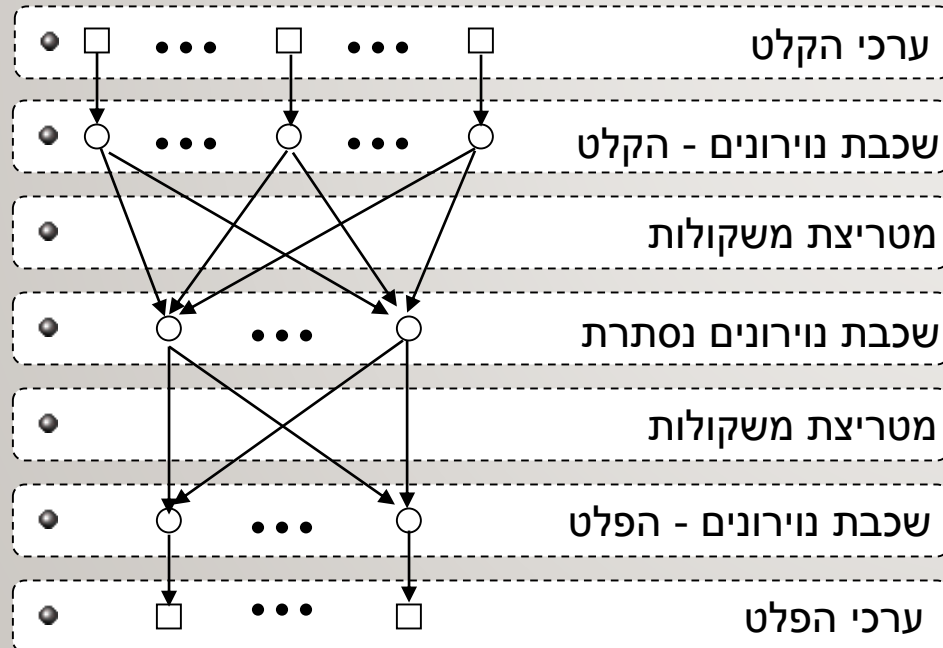
המודל

כדי להעביר את תכונות הנוירון הביולוגי לנוירון מלאכותי – ממוחשב, מתבצעת הפשטה של מונח הנוירון, ולא ממומשים במלואם כל המנגנונים הביולוגיים לפרטיהם.

הנוירון הממוחשב הוא יחידת עיבוד פשוטה, המקבלת קלט, מבצעת עיבוד ומספקת פלט.

הקלט הוא מערך של פלטים של נוירונים אחרים. הפלט הוא אות בינארי בעוצמה קבועה.

עוצמת הקשר בין פלט כלשהו לבין נקודת קלט הינה בעלת "משקל", ומתארת את מידת השפעת תוצאת החישוב של נוירון אחד על נוירון אחר המקבל קלט ממנו.

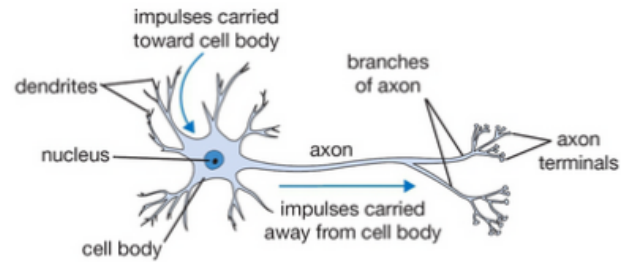


על מנת להפעיל את רשת הנוירונים, יש להגדיר את המשקולות בקשרים בין כל נוירון לשכנו.

המודל הפשוט ("המנוון"), מדבר על נוירון כמבצע פעולת סכימה פשוטה, המייצר קלט על פי נקודת סף (threshold) מוגדרת.

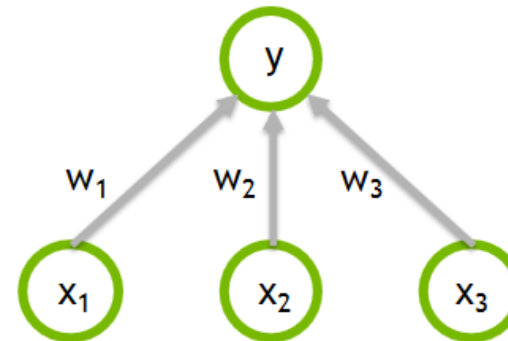
Artificial neurons

Biological neuron



From Stanford cs231n lecture notes

Artificial neuron

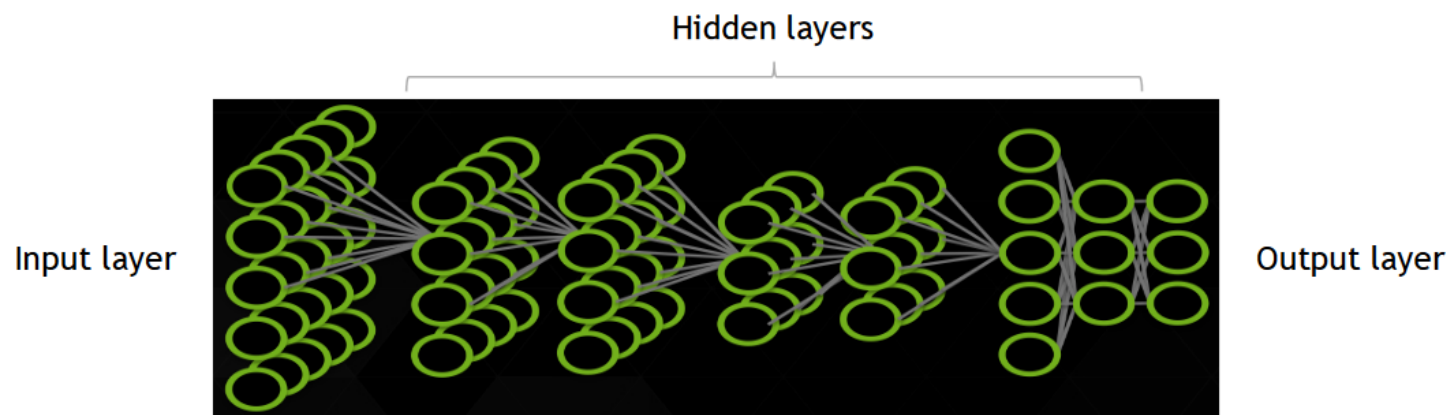


$$y = F(w_1x_1 + w_2x_2 + w_3x_3)$$

$$F(x) = \max(0, x)$$

Artificial neural network

A collection of simple, trainable mathematical units that collectively learn complex functions



Given sufficient training data an artificial neural network can approximate very complex functions mapping raw data to output decisions

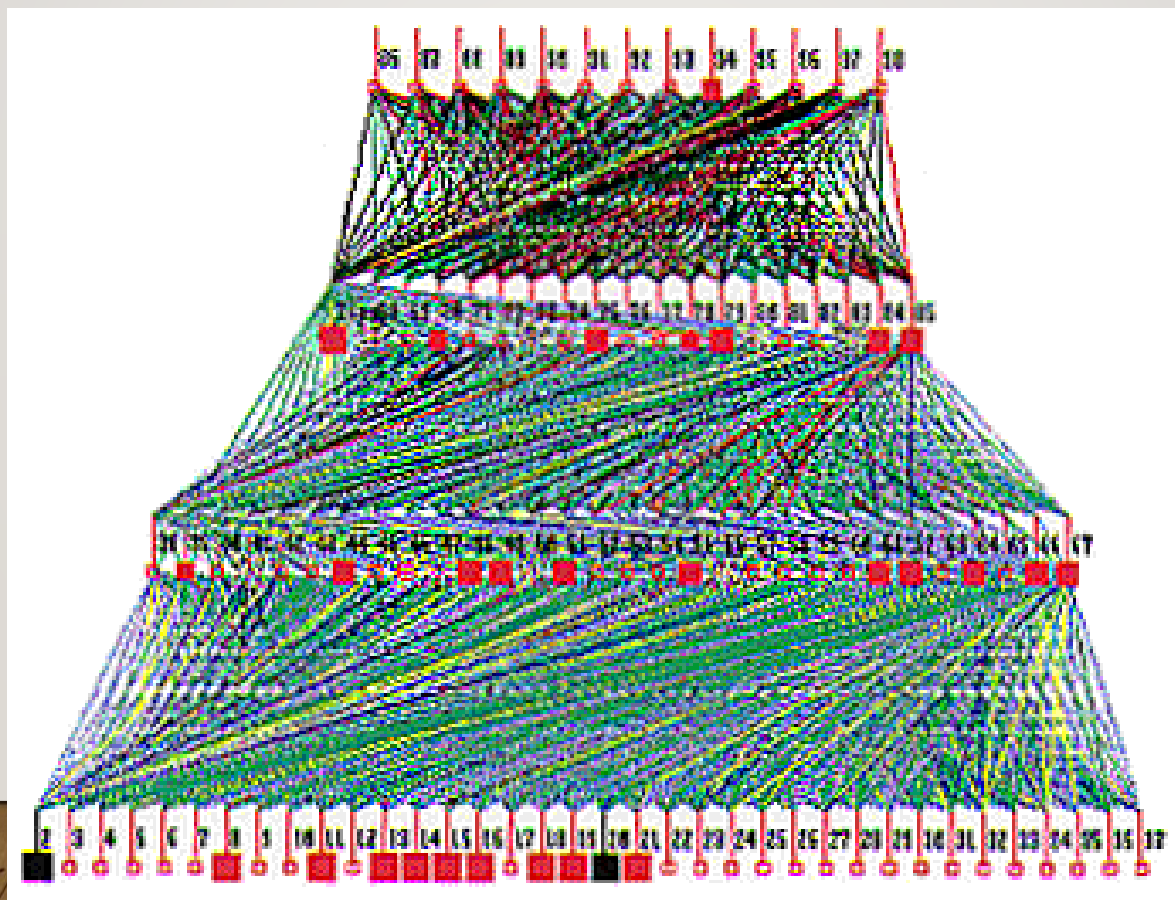
Question



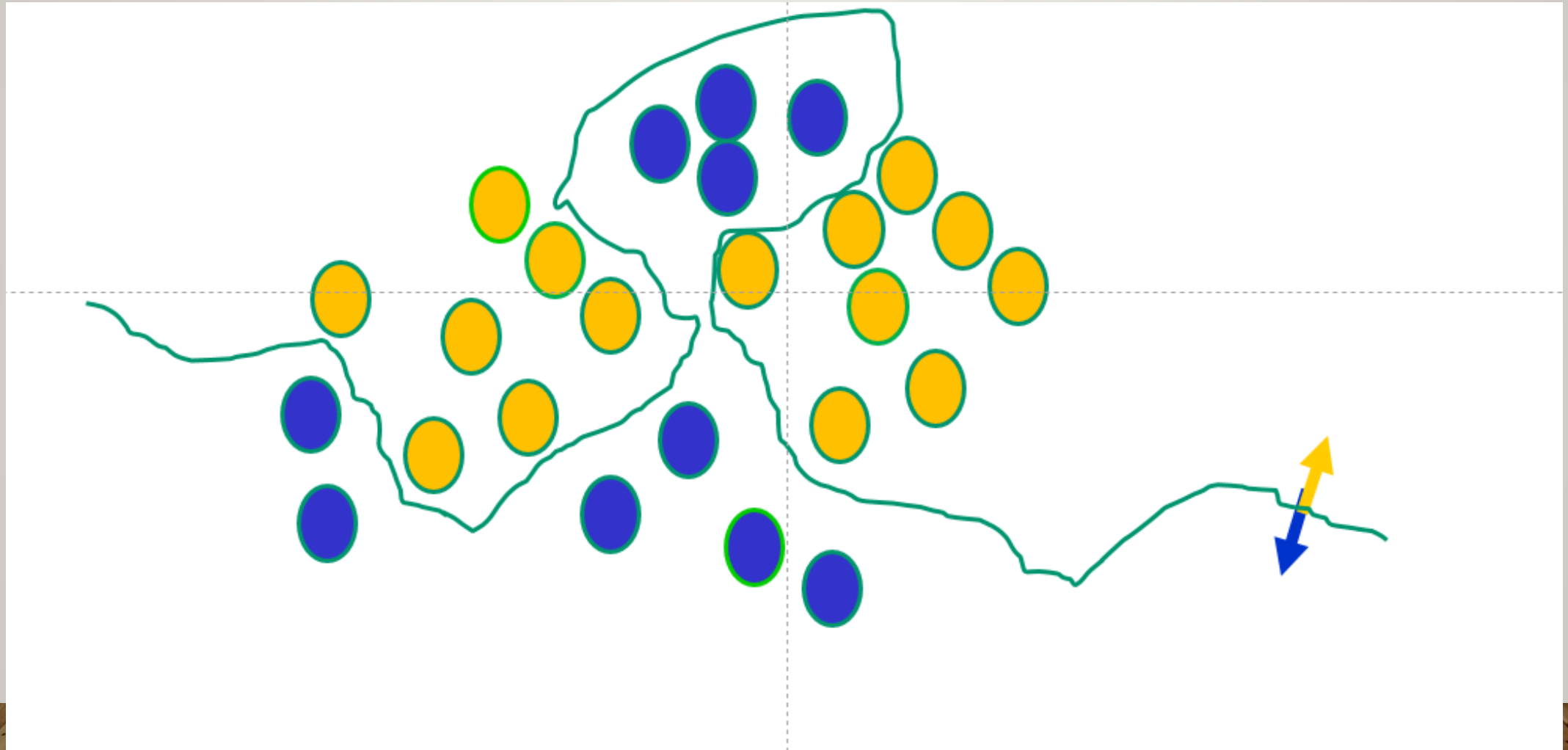
QUESTION

How do we
find this
line?

רשת המשתמשת בשכבה אחת , תוכל לפתור בעיות ליניאריות בלבד, לכן
נדרוש רשת בעלת מס. שכבות $1 <$ וכל זה בשביל פתרון בעיות לא
ליניאריות.



דוגמא למודל המבוסס מספר שכבות



Traditional machine perception

Hand crafted feature extractors

Raw data



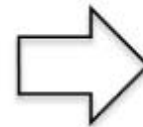
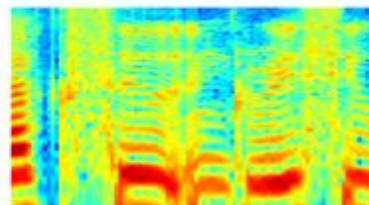
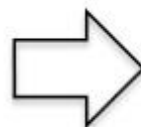
Feature extraction



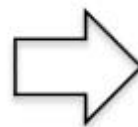
Classifier/
detector

SVM,
shallow neural net,
...

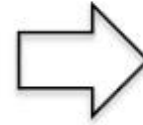
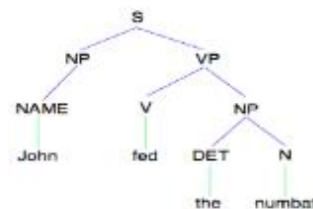
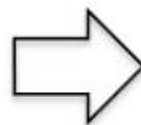
Result



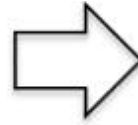
HMM,
shallow neural net,
...



Speaker ID,
speech transcription, ...



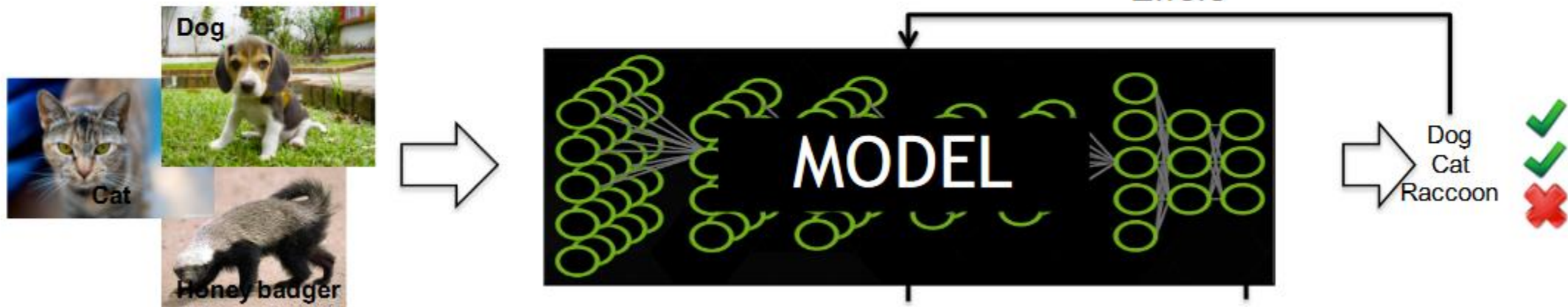
Clustering, HMM,
LDA, LSA
...



Topic classification,
machine translation,
sentiment analysis...

Deep learning approach

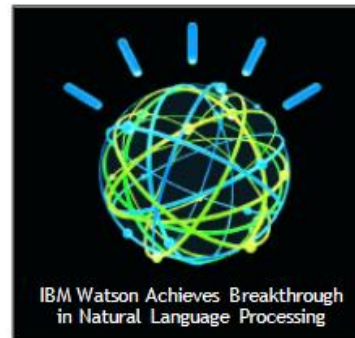
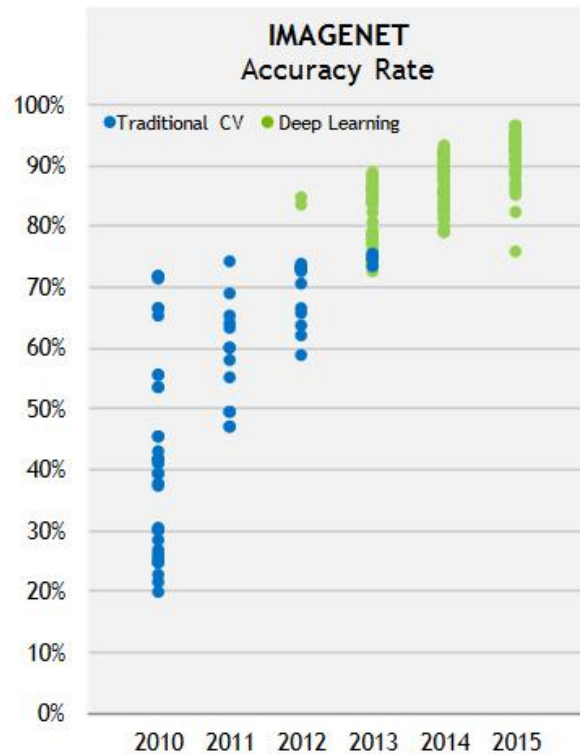
Train:



Deploy:



THE AI RACE IS ON

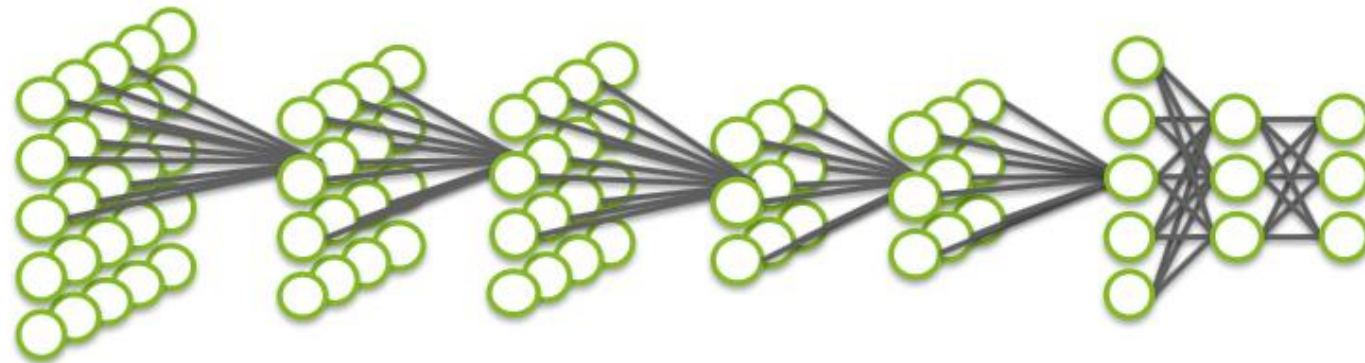


MOST COMMON NETWORKS

DNN - all fully connected layers

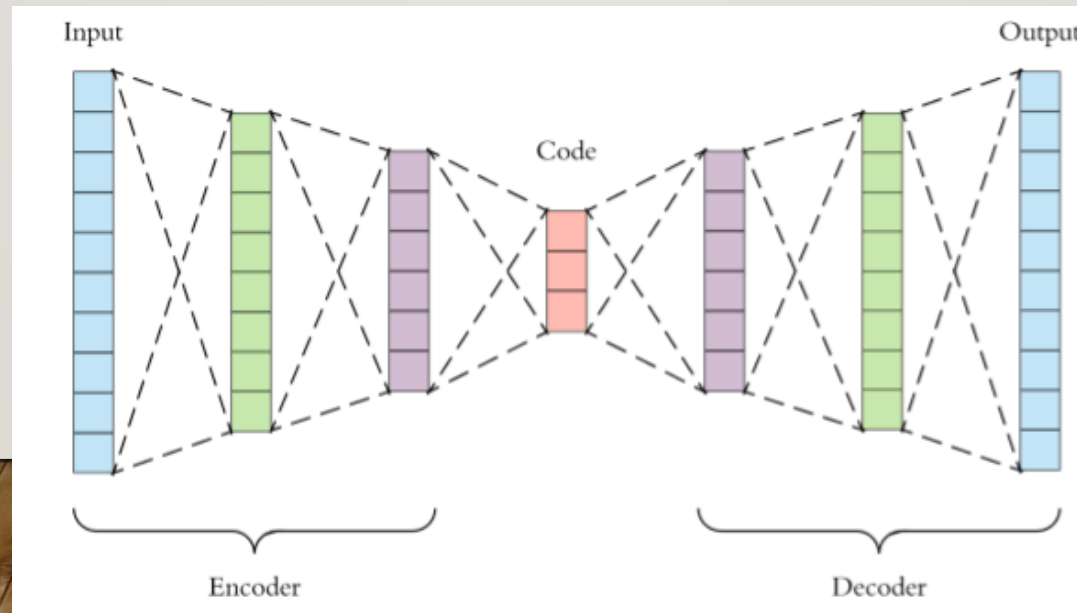
CNN - some convolutional layers

RNN - recurrent neural network, LSTM



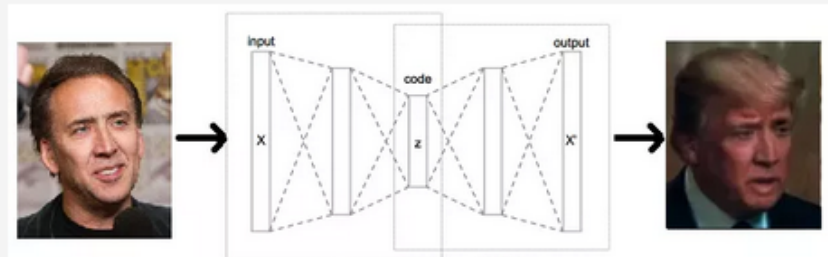
AUTOENCODER קלאסי

AutoEncoder היא רשת נוירונים שאומנה להעתיק את הקלט שלה לפלט שלה, זאת במטרה להפחתת מימד. היא כוללת פונקצית קידוד שיוצרת שכבות נסתרות המכילות קוד המתאר את הקלט. בהמשך הרשת יש מפענח אשר יוצר שחזור של הקלט מהשכבה מוסתרת. במצב שכזה AutoEncoder יכול להיות שימושי, על ידי הורדת שכבה נסתרת אחת, הוא מסוגל לאלץ את הרשת ליצור ייצוג דחוס של הנתונים בשכבה מוסתרים על ידי למידה קורלציות בנתונים.



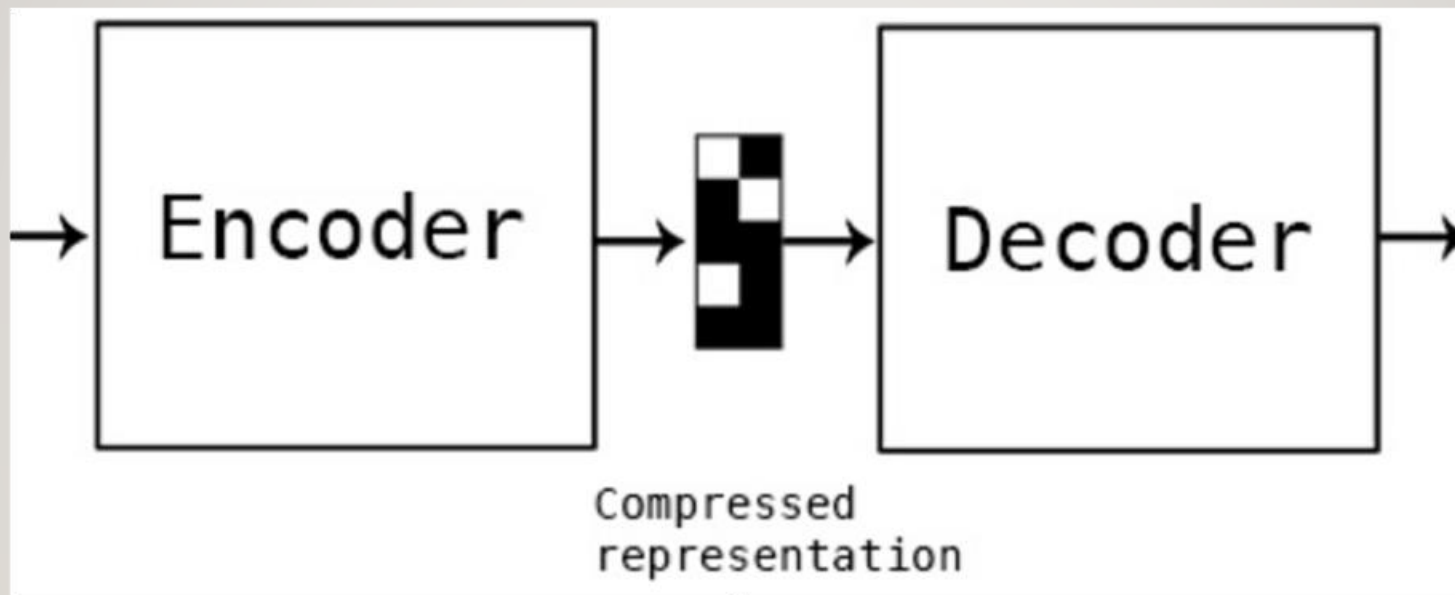
ה Encoder מקבל קלט כלשהוא (במקרה שלנו תמונת פנים), מעבד אותו ומחזיר פיסת נתונים קטנה יותר (z בתמונה). אם למשל תמונת פנים מיוצגת ע"י עשרות אלפי ערכים מספריים, אזי z במוצא ה Encoder (נקרא גם הנתונים נסתרים latent variables) יכול להיות בסך הכל עשרות מספרים. ז"א ה Encoder **מעבד ומאבד** מידע, אבל השאיפה לאמן אותו כך שהמידע שיוציא ייצג את הקלט באופן כזה שיהיה ניתן לשחזר את הקלט המקורי.

ה Decoder מקבל את אותו הקוד המקוצר z ומטרתו לעבד אותו ולהרחיב אותו לגודלו מקורי (במקרה שלנו שוב תמונת פנים). האימון של שני המרכיבים Encoder, Decoder נעשה במשותף והתגמול לכל אחד על הצלחתו היא כאשר התמונה במוצא ה Decoder זהה לתמונה בכניסת ה Encoder. אימון מוצלח של AutoEncoder משמעותו סוג של דחיסת נתונים או קידוד של הרבה מידע במעט מידע. יש לכך שימושים רבים אך במקרה שלנו השימוש הוא לייצג בתמציתיות מבנה פנים של אדם א כדי לשחזר (באמצעות Autoencoder שאומן אחרת) תמונה של פנים של אדם ב.



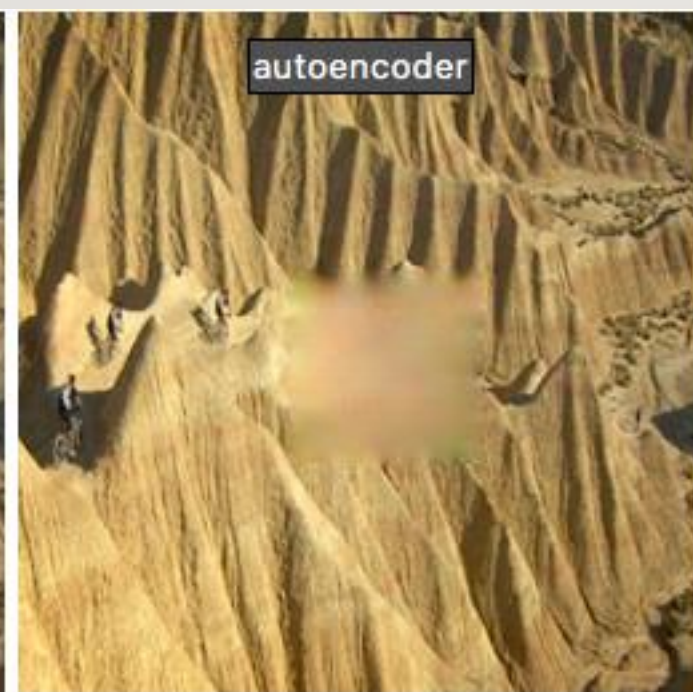
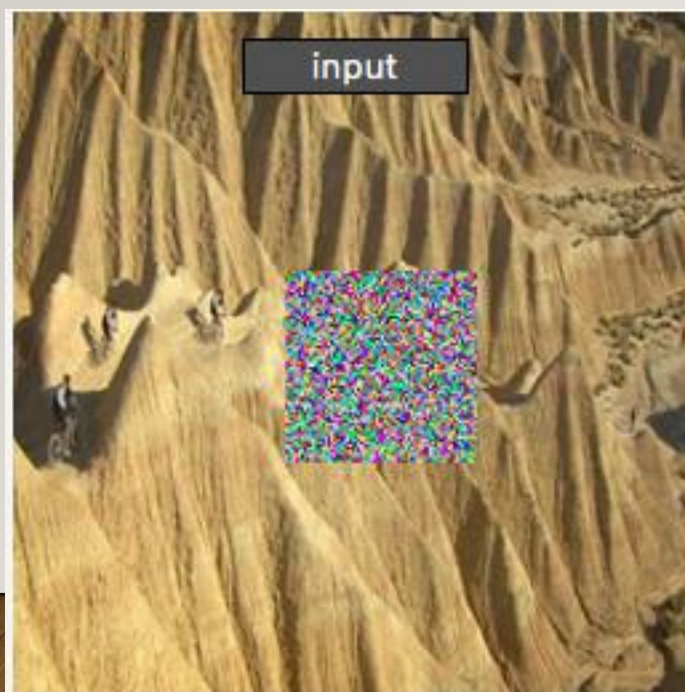
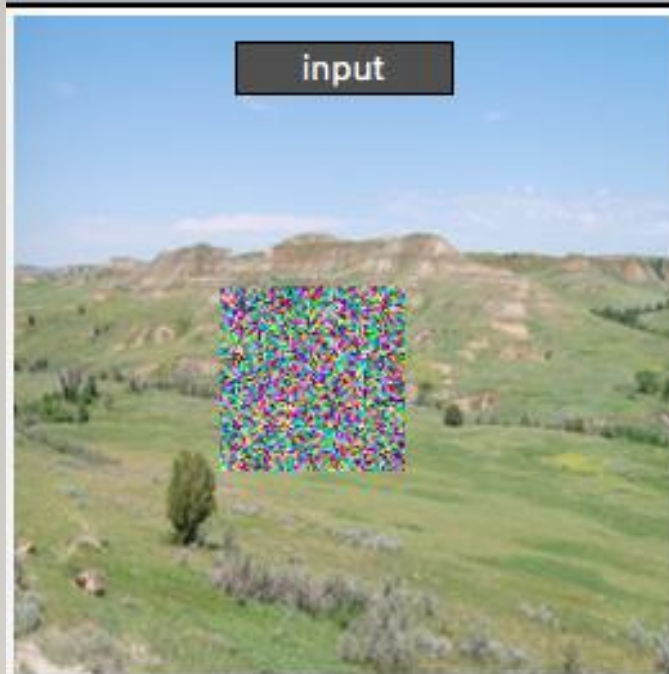


Noisy input



Denoised image

The feature we want to
extract from the image



רשת נוירונים גנרטיבית – GAN

רשת גנרטיבית אדברסרית מורכבת משני מודלים של רשתות נוירונים, Generator ו-Discriminator. הרשתות מתחרות זו בזו כדי להשיג מטרה מסוימת.

מטרתו של ה-Discriminator היא לזהות אם הקלט "אמיתי" – כלומר, שייכים למערך המקורי – או אם הוא "מזויף" – שנוצר על ידי הזיפן. אפשר לדמות את ה-Discriminator למשטרה שמנסה לאתר הונאה.

ה-Generator שואף ליצור נתונים חדשים שדומים כמה שיותר למקור. אפשר לדמות את ה-Generator לשודד, היוצר תמונות מזויפות כדי להערים על המשטרה.

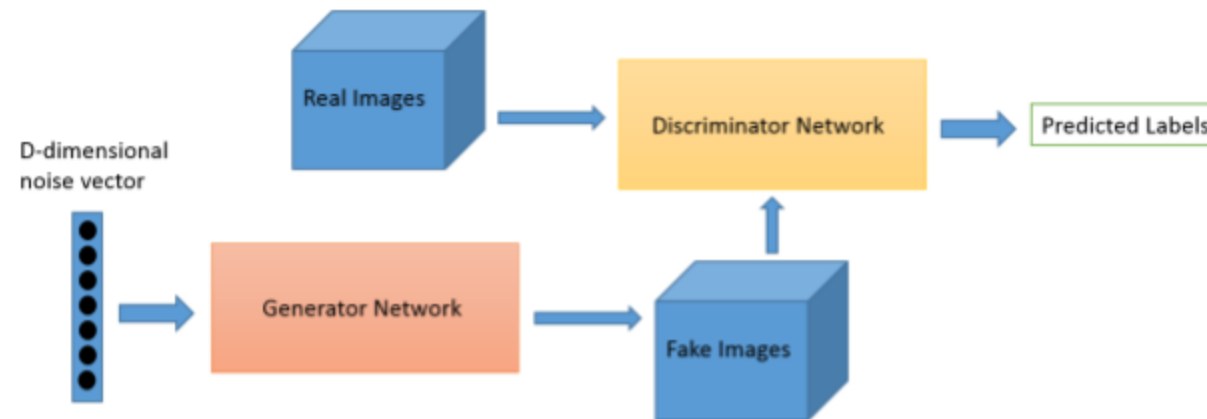


Figure 1: GAN general flow



Generator

Discriminator



Real Money



Fake Money



Counterfeiter prints fake money. It is labeled as fake for police training. Sometimes, the counterfeiter attempts to fool the police by labeling the fake money as real.

The police are trained to distinguish between. Sometimes, the police give feedback to the counterfeiter about why the money is fake.

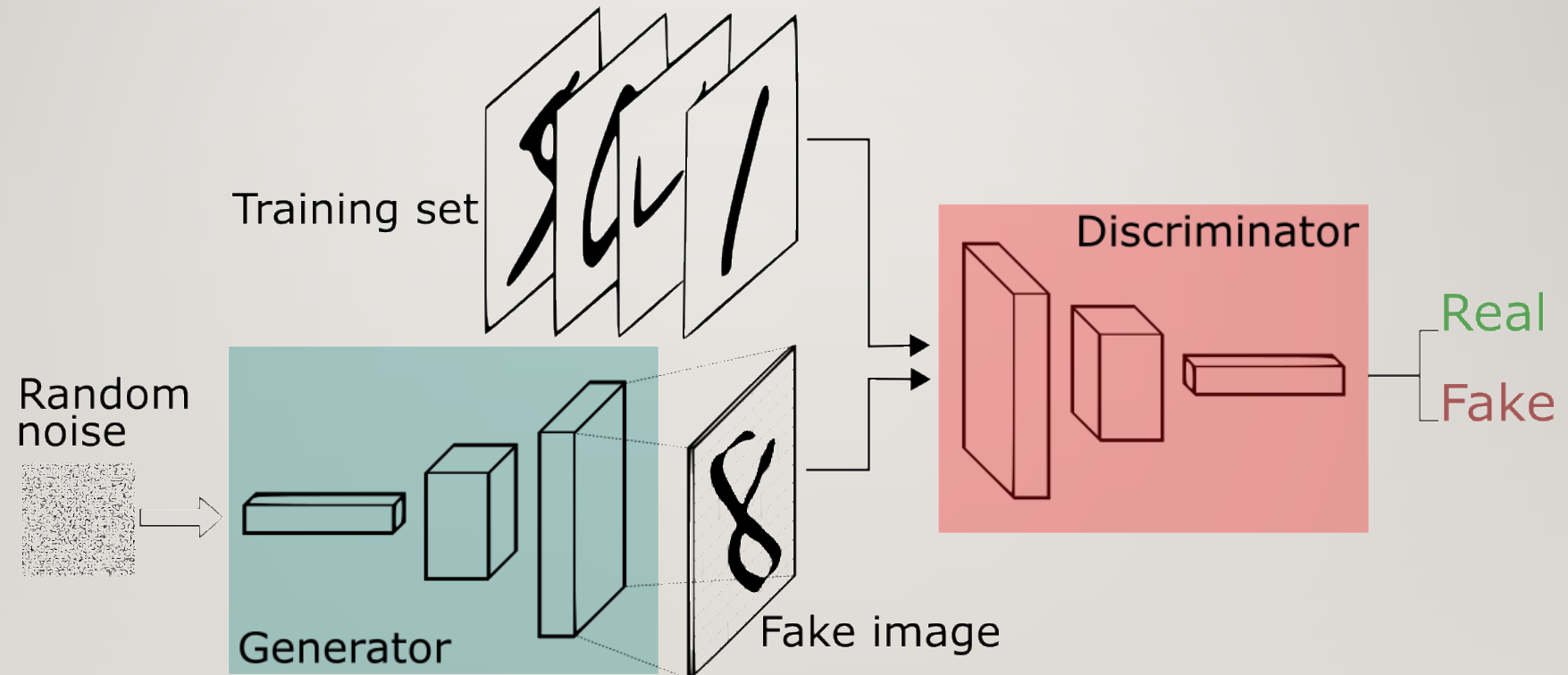
איך מודלים אלה מתקשרים?

מצד אחד, ה- Generator (הזייפן) צריך ללמוד איך ליצור תמונות בצורה מספיק טובה, שה- Discriminator לא יהיה מסוגל להבחין שזה מזויף.

מצד שני, ה- Discriminator מתבונן כיצד שיטת הזיוף של ה- Generator משתפרת ולומד עצמו להבחין טוב יותר.

המשחק הזה נמשך עד שה- Generator מצליח ליצור דימויים מציאותיים כל כך, שה- Discriminator לא יכול היה לומר שהם מזויפים.

Gan Architecture

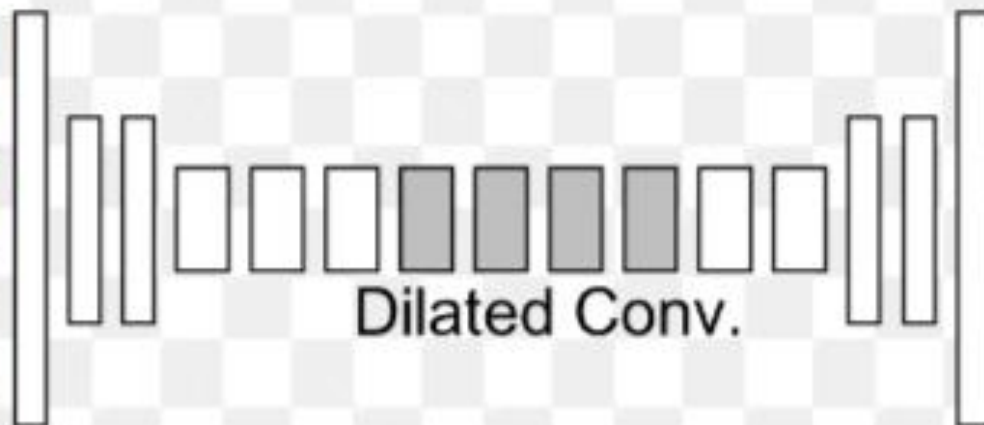




Raw



Input and Mask



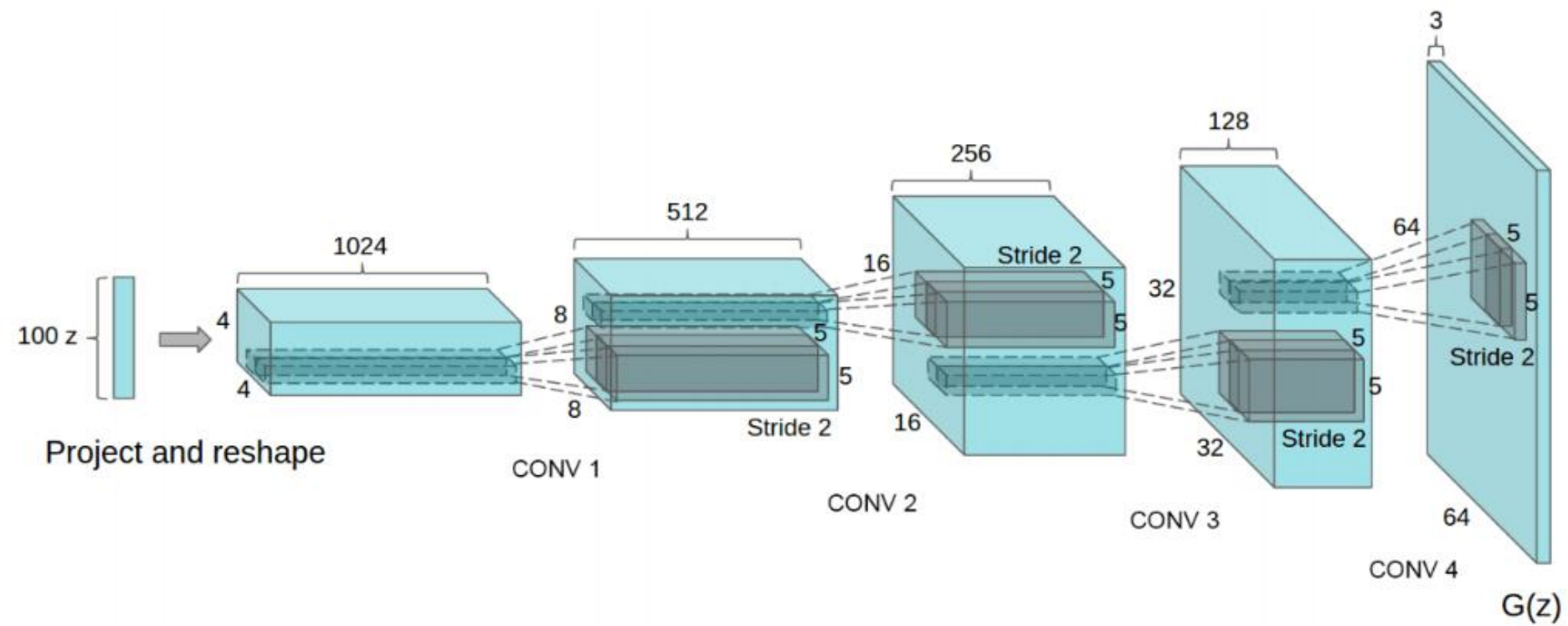
Dilated Conv.

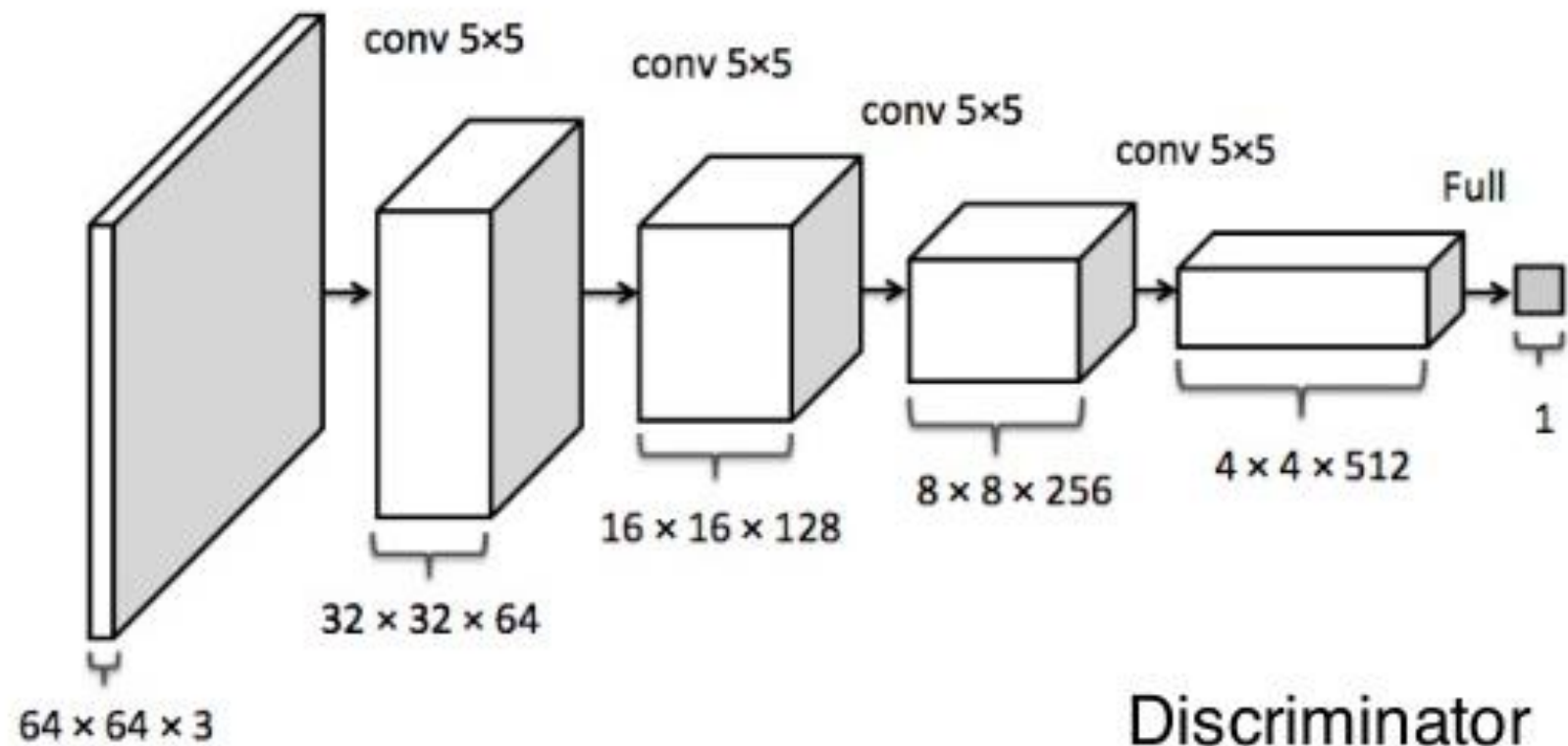


Coarse Result

GAN VS AUTOENCODER

פרמטרים	Autoencoder	GAN
מתאים ל-	עיבוד דטה קיים	משימות גנראטיביות
דרך דגימת הנתונים	מהתפלגויות רציפות וידועות מראש	וקטורים רנדומליים
תוצאות לעומת אימונים	קל יותר לאמן רשת VAE אבל היא מייצרת תמונות באיכות נמוכה יותר.	רשת GAN נוטה לייצר תמונות חדות יותר, אבל קשה יותר לאמן אותה.





$$\min_{\theta_g} \max_{\theta_d} \left[\mathbb{E}_{x \sim p_{data}} \log \underbrace{D_{\theta_d}(x)}_{\substack{\text{Discriminator output} \\ \text{for real data } x}} + \mathbb{E}_{z \sim p(z)} \log(1 - \underbrace{D_{\theta_d}(G_{\theta_g}(z))}_{\substack{\text{Discriminator output for} \\ \text{generated fake data } G(z)}}) \right]$$

Objective Function

הגדרה מתמטית

נגדיר Z בתור וקטור רנדומלי ואת x כתמונת קלט.

G ו- D הם פונקציות המייצגות את ה-Generator וה-Discriminator בהתאמה.

הפלט של ה-Generator הוא תמונה מזויפת ($G(z)$).

עכשיו בוא נגדיר כי עבור קלט כלשהו, הפלט של ה-Discriminator (D) הוא ההסתברות ($[0,1]$) כי הקלט הוא תמונה אמיתית.

משימתו של ה-Generator היא לרמות את ה-Discriminator מהבחנה בין תמונות מזויפות לאמיתיות. כלומר, **מיקסום של $D(G(z))$** .

המשימה של ה-Discriminator תהיה **למקסם את $D(x)$ ולמזער את $D(G(z))$** .

באמצעות פונקציית \log נקצין את העונשים.

שים לב כי שתי רשתות הנוירוניים מנסות לייעל את אותה פונקציה (המונח השני), אבל אחת מהן מנסה למקסם אותו והשנייה מנסה למזער את זה של בעיה נקרא משחק Minimax.

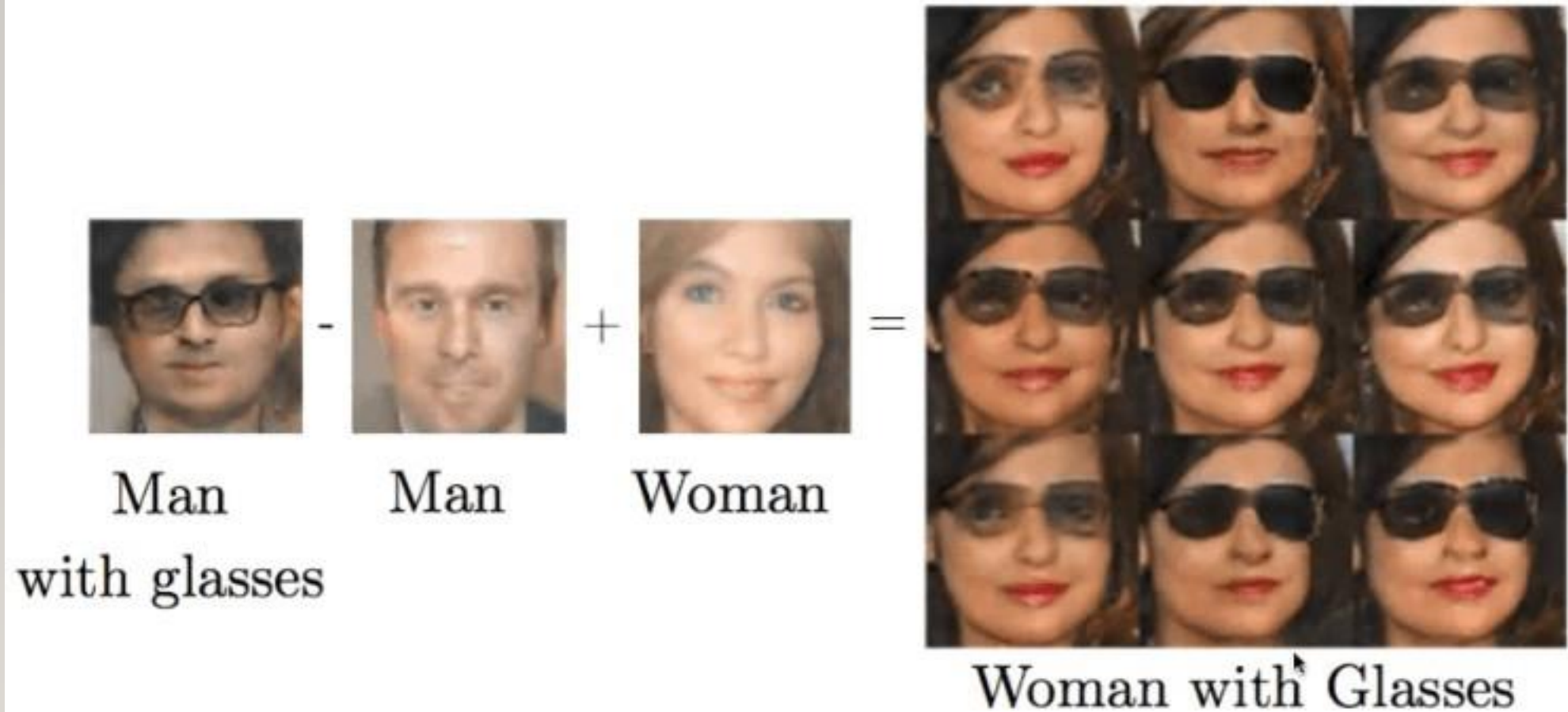


As you probably know, this is former
US President Barack Obama.





חיבור וחיסור פיצרים



The image displays a visual equation for face manipulation. On the left, three individual face images are shown in a row: a man with glasses, a man, and a woman. Below these images are the labels "Man with glasses", "Man", and "Woman" respectively. In the center, there are mathematical symbols: a minus sign, a plus sign, and an equals sign. To the right of the equals sign is a 3x3 grid of nine face images, all showing the woman's face with glasses. Below this grid is the label "Woman with Glasses".

$$\begin{array}{c} \text{Man} \\ \text{with glasses} \end{array} - \begin{array}{c} \text{Man} \end{array} + \begin{array}{c} \text{Woman} \end{array} = \begin{array}{c} \text{Woman with Glasses} \end{array}$$

שימושים נוספים

Labels to Street Scene



input



output

Aerial to Map

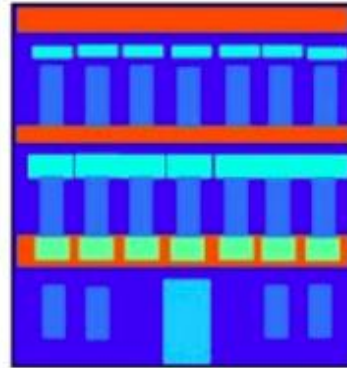


input



output

Labels to Facade



input



output

BW to Color



input



output

Day to Night



input



output

Edges to Photo



input



output

input



original



autoencoder



gan



input



original



autoencoder



gan



input



original



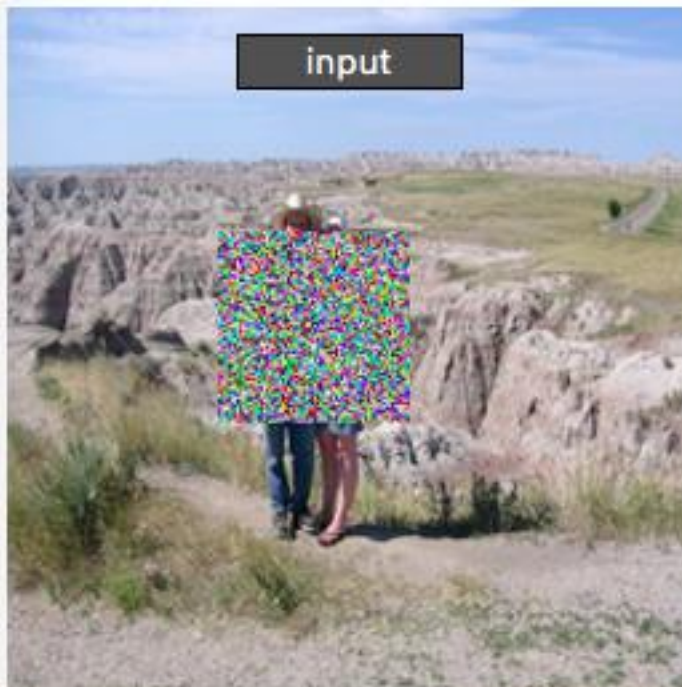
autoencoder



gan



input



original



autoencoder



gan

