



Εθνικό Μετσόβιο Πολυτεχνείο
Σχολή Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών
Εργαστήριο Συστημάτων Τεχνητής Νοημοσύνης και Μάθησης

AI|LS

Παραγωγικά Μοντέλα (I)

VAEs - GANs

Ηλίας Μήτσουρας

Διάρθρωση της Παρουσίασης

2

- ❑ **Εισαγωγή**
- ❑ Variational Autoencoders
- ❑ Generative Adversarial Networks (GANs)
- ❑ Εφαρμογές

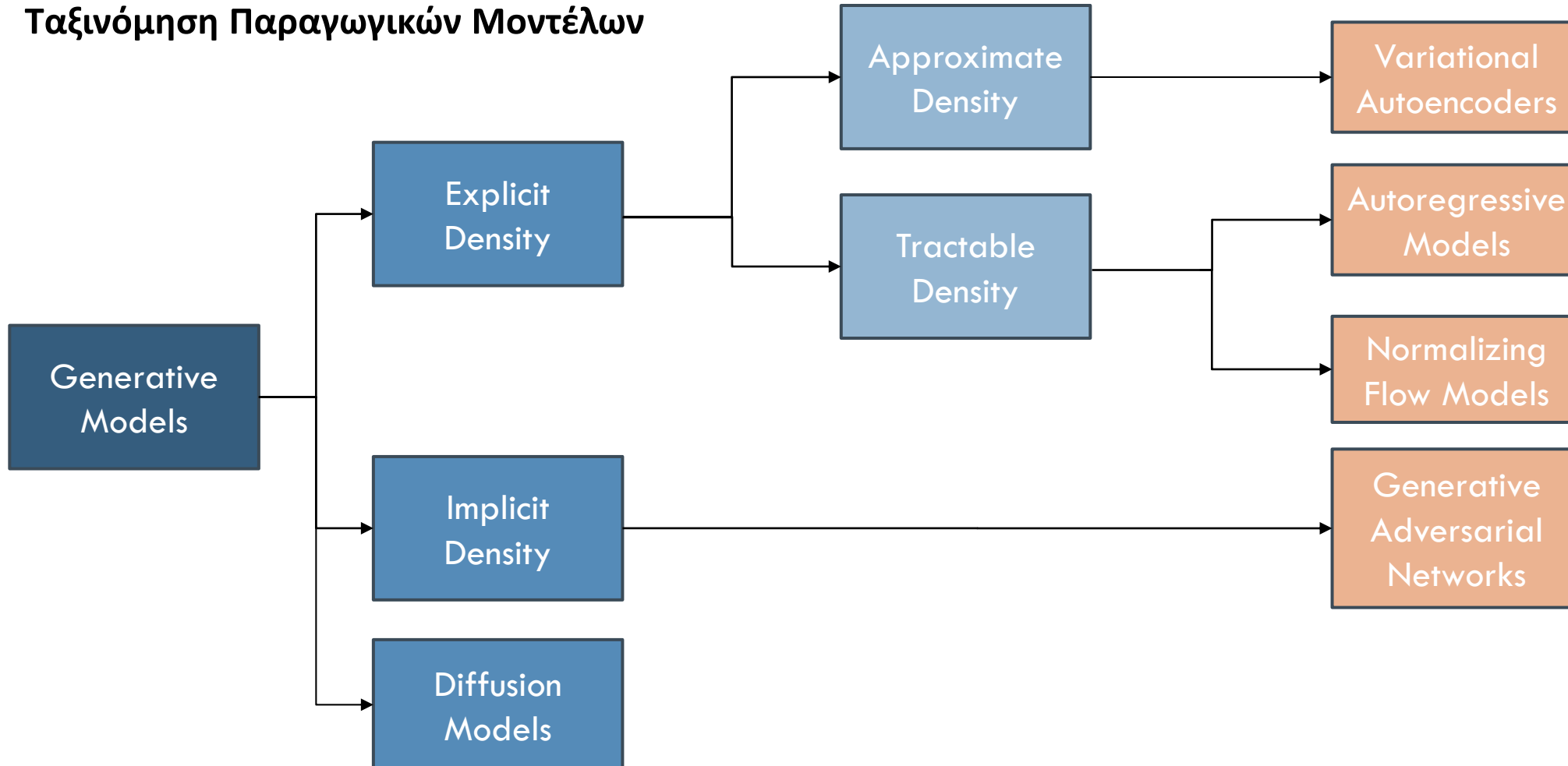
Εισαγωγή

3

Παραγωγικά Μοντέλα (*Generative Models*)

- Οικογένεια στατιστικών μοντέλων
- Σκοπός τους είναι ο προσδιορισμός των υποκείμενων μοτίβων και κατανομών των διαθέσιμων δεδομένων
→ **παραγωγή νέων δειγμάτων**, τα οποία μοιράζονται παρόμοια χαρακτηριστικά με τα αρχικά δεδομένα.
- Πιο επίσημα, δοθέντος ενός συνόλου δεδομένων $X = \{x_1, x_2, \dots, x_n\}$, και ενός συνόλου ετικετών $Y = \{y_1, y_2, \dots, y_n\}$ (προαιρετικά), ένα παραγωγικό μοντέλο προσπαθεί να προσεγγίσει:
 - την από κοινού κατανομή $p(X, Y)$ ή
 - την κατανομή $p(X)$, σε περίπτωση όπου δεν δίνεται το σύνολο Y .
- Διαφέρουν από τα διακριτικά μοντέλα (*discriminative models*), τα οποία προσπαθούν να εκτιμήσουν την υπό συνθήκη κατανομή πιθανότητας $p(Y | X)$.
- Τα παραγωγικά μοντέλα ταξινομούνται βάσει του τρόπου με τον οποίο προσπαθούν να εκτιμήσουν τη συνάρτηση πυκνότητας πιθανότητας $p_{\text{model}}(X)$.

Ταξινόμηση Παραγωγικών Μοντέλων



Διάρθρωση της Παρουσίασης

5

- Εισαγωγή
- ***Variational Autoencoders***
- Generative Adversarial Networks (Gans)
- Εφαρμογές

Variational Autoencoders

6

Μείωση Διαστάσεων (*Dimensionality Reduction*)

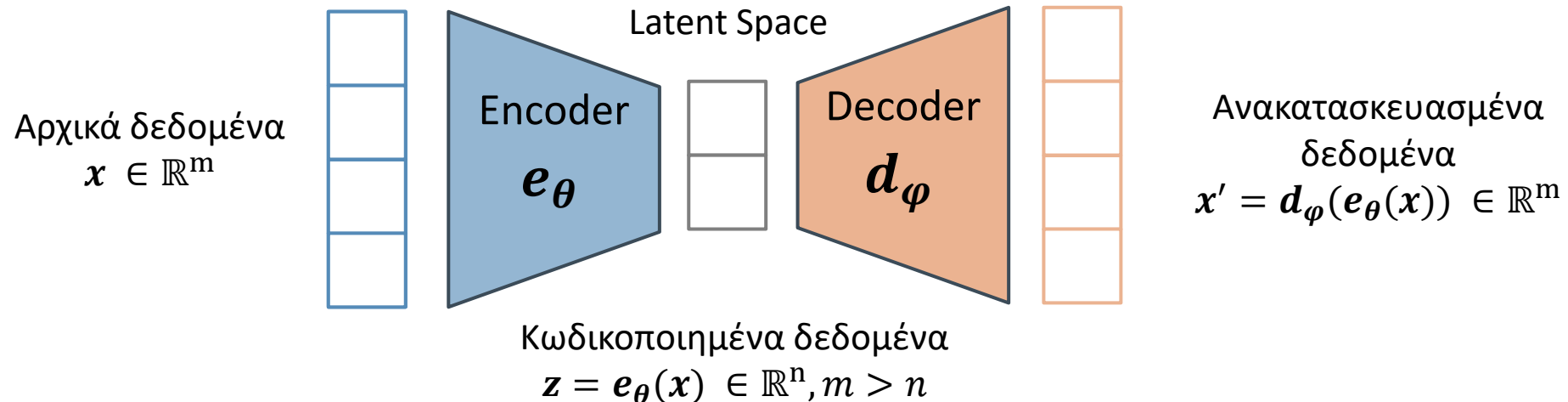
- ▣ Η μείωση Διαστάσεων αποσκοπεί στη **μείωση του πλήθους των χαρακτηριστικών** που χρησιμοποιούνται για να περιγράψουν τα δεδομένα.
- ▣ Είναι χρήσιμη σε διάφορες εφαρμογές οι οποίες απαιτούν δεδομένα χαμηλής διάστασης, όπως η οπτικοποίηση, η αποθήκευση κ.ά.
- ▣ Η μείωση της διάστασης των χαρακτηριστικών μπορεί να επιτευχθεί μέσω:
 - μεθόδων **επιλογής χαρακτηριστικών** (*feature selection*), στις οποίες επιλέγεται ένα υποσύνολο των αρχικών χαρακτηριστικών,
 - μεθόδων **εξαγωγής χαρακτηριστικών** (*feature extraction*), στις οποίες δημιουργείται ένα νέο σύνολο χαρακτηριστικών μικρότερου μεγέθους από το αρχικό, με τη μικρότερη δυνατή απώλεια πληροφορίας.

Variational Autoencoders

7

Autoencoders

- ▣ Νευρωνικό δίκτυο που μαθαίνει να μετασχηματίζει τα δεδομένα σε μία συμπιεσμένη μορφή και να ανακατασκευάζει τα αρχικά δεδομένα, με όσο το δυνατόν μικρότερες απώλειες.
- ▣ Ο *autoencoder* αποτελείται από δύο βασικά μέρη:
 - **Encoder:** χρησιμοποιείται για την παραγωγή μίας νέας, συμπιεσμένης αναπαράστασης των αρχικών δεδομένων.
 - **Decoder:** ανακατασκευάζει τα αρχικά δεδομένα βάσει της νέας συμπιεσμένης αναπαράστασης, με όσο το δυνατόν μικρότερες απώλειες.

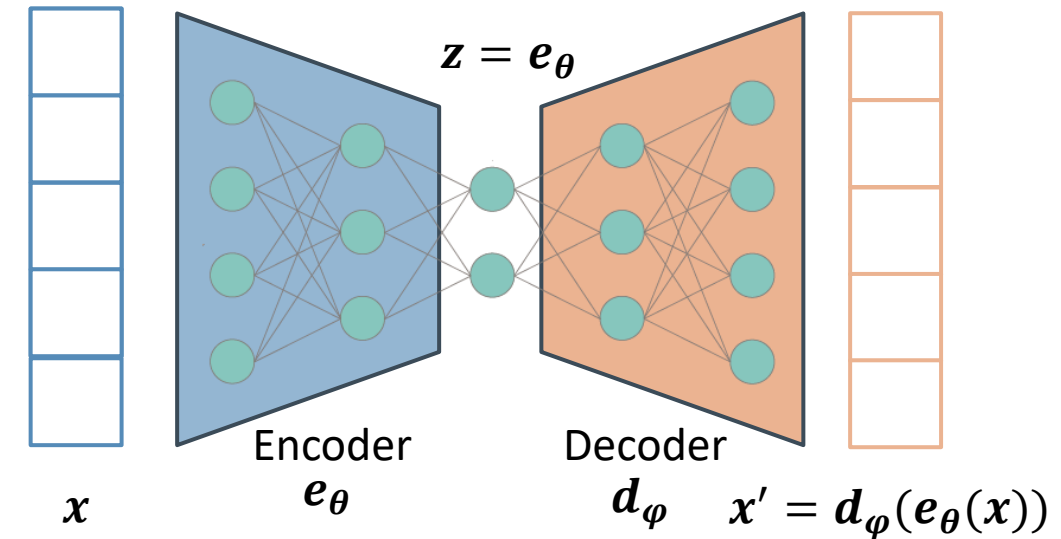


Variational Autoencoders

8

Autoencoders – Αρχιτεκτονική και Εκπαίδευση

- Τα δίκτυα του encoder και του decoder υλοποιούνται μέσω νευρωνικών δικτύων, τα οποία επιλέγονται βάσει του τύπου των διαθέσιμων δεδομένων (εικόνες → συνελκτικά δίκτυα, διανύσματα → fully connected layers).
- Ο autoencoder εκπαιδεύεται μέσω μη επιβλεπόμενης μάθησης, καθώς έχουμε μη επισημασμένα δεδομένα.
- Σε κάθε επανάληψη:
 - Ο encoder κωδικοποιεί τα δεδομένα εισόδου $\mathbf{z} = \mathbf{e}_{\theta}(\mathbf{x})$.
 - Ο decoder ανακατασκευάζει τα αρχικά δεδομένα βάσει της κωδικοποίησης του encoder, $\mathbf{x}' = \mathbf{d}_{\varphi}(\mathbf{e}_{\theta}(\mathbf{x}))$.
 - Η ανακατασκευή \mathbf{x}' συγκρίνεται με τα αρχικά δεδομένα \mathbf{x} και βάσει της ομοιότητάς τους ανανεώνονται τα βάρη των δικτύων.



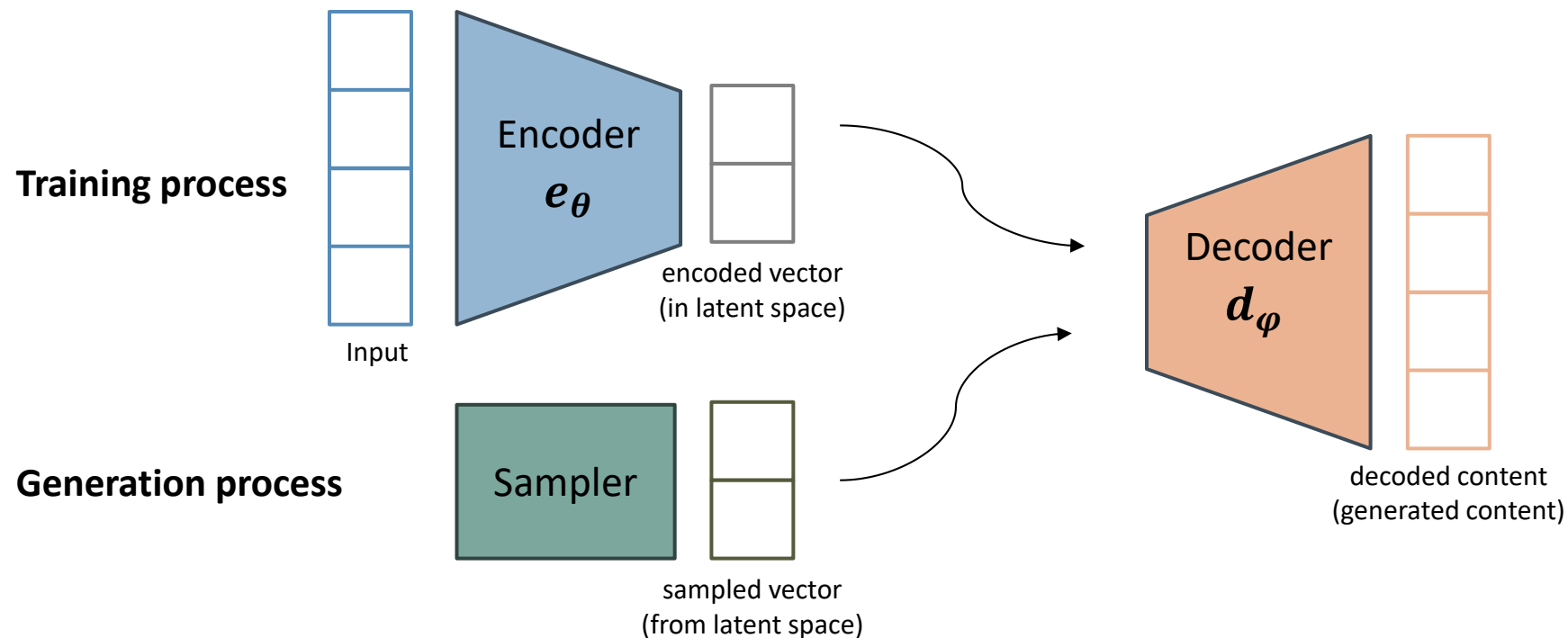
$$\text{Loss Function: } \mathcal{L} = \|\mathbf{x} - \mathbf{x}'\|_2 = \|\mathbf{x} - \mathbf{d}_{\varphi}(\mathbf{e}_{\theta}(\mathbf{x}))\|_2$$

Variational Autoencoders

9

Autoencoders και Παραγωγή νέων Δεδομένων

- Πώς συνδέεται ο autoencoder με την παραγωγή νέων δεδομένων;
- Θα μπορούσαμε να υποθέσουμε ότι, σε περίπτωση που ο latent χώρος είναι καλά δομημένος, τυχαία σημεία τα οποία δειγματοληπτούνται από αυτόν, μπορούν να αποκωδικοποιηθούν μέσω του decoder για την παραγωγή νέων συνθετικών δεδομένων.

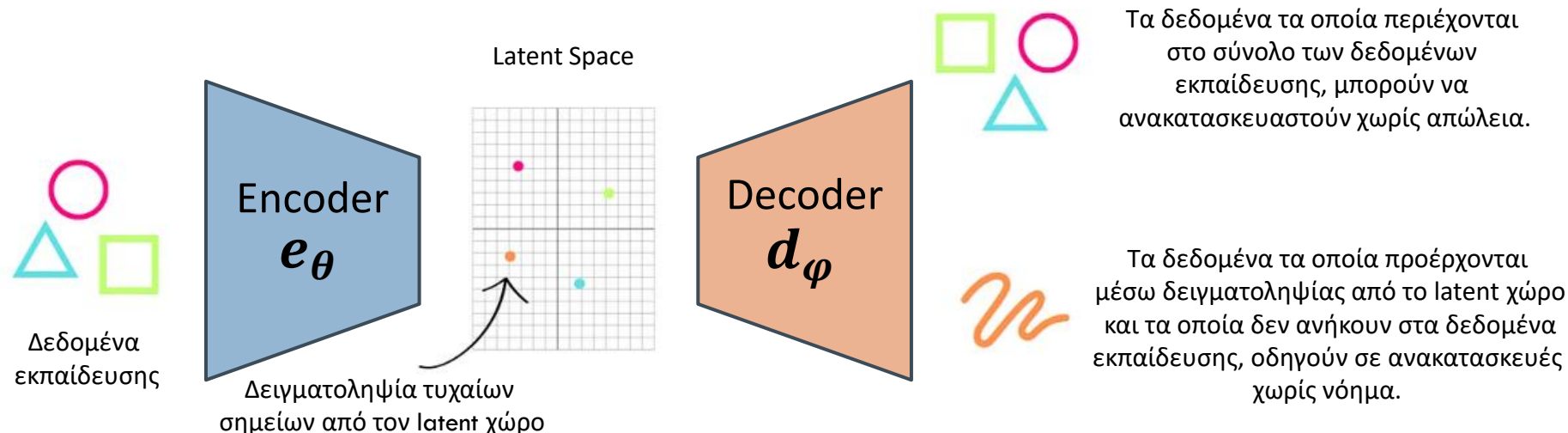


Variational Autoencoders

10

Autoencoders και Παραγωγή νέων Δεδομένων

- Στην πράξη δεν μπορούμε να εγγυηθούμε εκ των προτέρων ότι ο encoder θα οργανώσει τον latent χώρο με τρόπο ώστε να μπορούμε να δειγματοληπτήσουμε από αυτό και να παράγουμε νέα δείγματα.
- Ο υψηλός βαθμός ελευθερίας του autoencoder οδηγεί στην υπερπροσαρμογή, με αποτέλεσμα ο latent χώρος να είναι μη-κανονικοποιημένος (non-regularized).
- Αυτό έχει ως συνέπεια η αποκωδικοποίηση τυχαίων σημείων από τον latent χώρο να οδηγεί σε μη ρεαλιστικά ή χωρίς νόημα δεδομένα, τα οποία αποκλίνουν σημαντικά από τα πραγματικά δείγματα εκπαίδευσης.



Variational Autoencoders

11

Variational Autoencoders

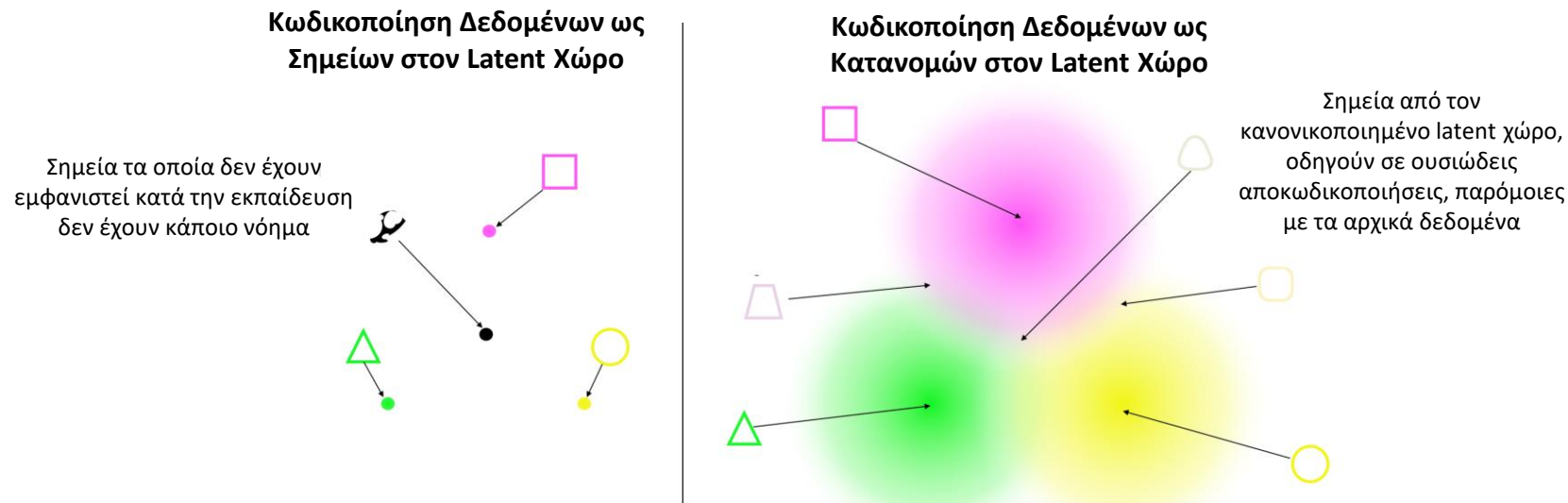
- ❑ Προκειμένου να χρησιμοποιήσουμε τον decoder του autoencoder για την παραγωγή νέων συνθετικών δειγμάτων, θα πρέπει να εξασφαλίσουμε ότι ο latent χώρος είναι **κατάλληλα κανονικοποιημένος**.
- ❑ Για το σκοπό αυτό χρησιμοποιούνται οι **Variational Autoencoders**.
- ❑ **Variational Autoencoder**
 - Πρόκειται για έναν autoencoder με μηχανισμό κανονικοποίησης, ώστε να αποφευχθεί το φαινόμενο της υπερπροσαρμογής και να εξασφαλιστούν χρήσιμες ιδιότητες στον latent χώρο.
 - Παρόμοια με τον autoencoder, περιλαμβάνει έναν encoder και έναν decoder, οι οποίοι εκπαιδεύονται ώστε να ελαχιστοποιούν το σφάλμα ανακατασκευής μεταξύ των αρχικών δεδομένων και των encoded-decoded δεδομένων.
 - Η βασική διαφορά έγκειται στο ότι, αντί η είσοδος x να κωδικοποιείται ως ένα διάνυσμα, κωδικοποιείται πλέον ως **μία κατανομή $p(z | x)$ στον latent χώρο**.
 - Επομένως, ο encoder επιστρέφει δύο διανύσματα, ένα διάνυσμα μέσων τιμών μ και ένα διάνυσμα τυπικών αποκλίσεων σ .

Variational Autoencoders

12

Variational Autoencoders

- Προκειμένου ο Variational Autoencoder να μπορεί να παράγει νέα συνθετικά δεδομένα, θα πρέπει ο latent χώρος να είναι:
 - **Συνεχής (Continuous):** Παρόμοιες εισοδοι θα πρέπει να έχουν παρόμοιες λανθάνουσες αναπαραστάσεις, διασφαλίζοντας ομαλές μεταβάσεις στον latent χώρο.
 - **Πλήρης (Complete):** Το μοντέλο θα πρέπει να αξιοποιεί αποδοτικά ολόκληρο τον latent χώρο.

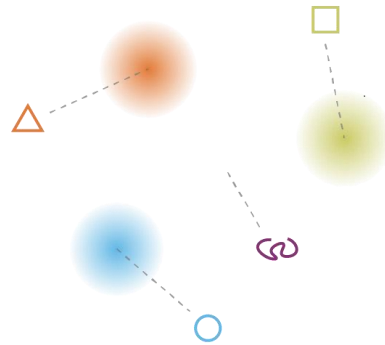


Variational Autoencoders

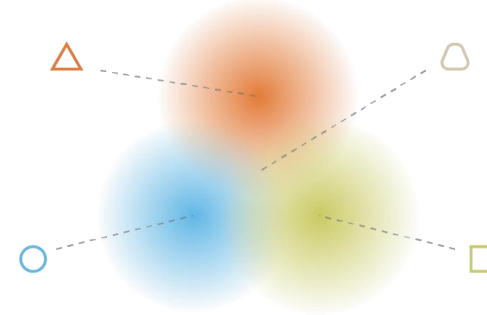
13

Variational Autoencoders

- Η τροποποίηση του encoder του VAE ώστε να μοντελοποιεί κατανομές πιθανότητες αντί για σημεία, δεν εγγυάται τις δύο ανωτέρω προδιαγραφές, καθώς αυτός μπορεί:
 - Να οδηγεί σε **πολύ μικρές διασπορές** στις λανθάνουσες αναπαραστάσεις, με αποτέλεσμα να συμπεριφέρονται σαν σταθερά σημεία.
 - Να **απομακρύνει αρκετά τις μέσες τιμές** των κατανομών, οδηγώντας σε κενά και ασυνέχεια στον latent χώρο.
- Για να αποφύγουμε τα προβλήματα αυτά:
 - **Κανονικοποίηση μέσης τιμής και τη διασποράς** των κατανομών των λανθανουσών αναπαραστάσεων
→ Κανονική κατανομή.



Τι μπορεί να συμβεί χωρίς κανονικοποίηση ❌

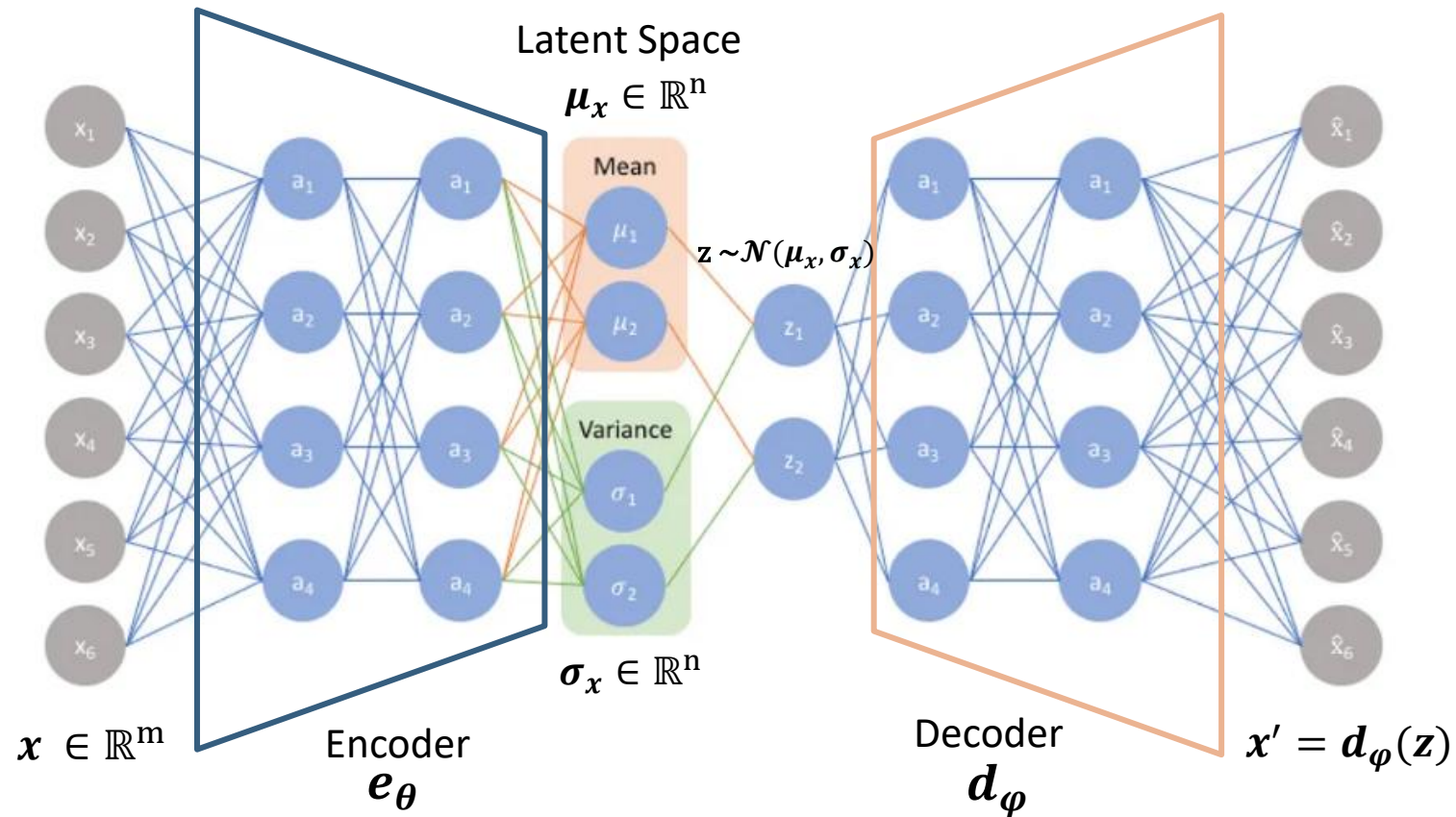


✓ Τι θέλουμε να επιτύχουμε μέσω κανονικοποίησης

Variational Autoencoders

14

Variational Autoencoders



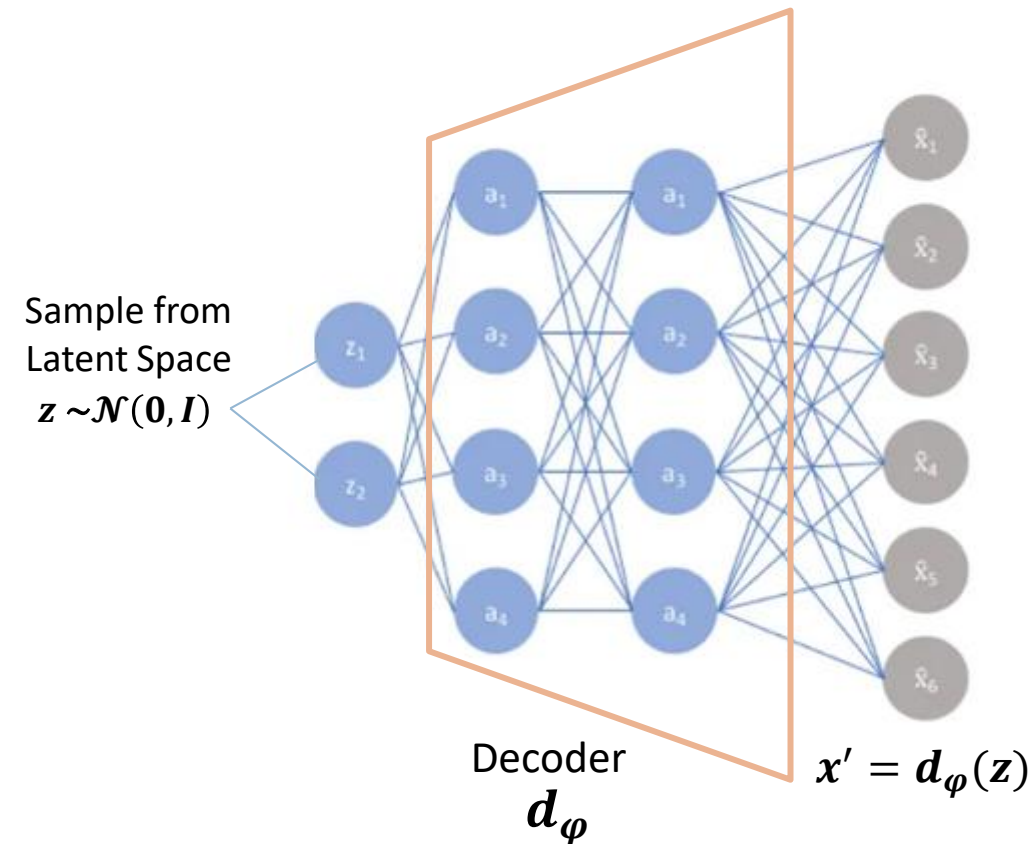
$$\text{Loss Function: } \mathcal{L} = \|x' - d_\phi(e_\theta(x))\|_2 + D_{KL}(\mathcal{N}(\mu_x, \sigma_x) \parallel \mathcal{N}(0, I))$$

Variational Autoencoders

15

Variational Autoencoders – Σύνθεση Νέων Δειγμάτων

- Μπορούμε να χρησιμοποιήσουμε μόνο τον Decoder ενός VAE για να παράξουμε νέα δείγματα.
- **Διαδικασία**
 1. Δειγματοληψία στον latent χώρο, θεωρώντας μια κανονική κατανομή, $\mathbf{z} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$.
 2. Αποκωδικοποίηση της λανθάνουσας αναπαράστασης \mathbf{z} μέσω του Decoder και παραγωγή νέου δείγματος $\mathbf{x}' = \mathbf{d}_\varphi(\mathbf{z})$.



Variational Autoencoders

16



Δείγματα από VAE.

Διάρθρωση της Παρουσίασης

17

- Εισαγωγή
- Variational Autoencoders
- ***Generative Adversarial Networks (GANs)***
- Εφαρμογές

Generative Adversarial Networks (GANs)

18

Generative Adversarial Networks (GANs)

- Πρωτοπαρουσιάστηκαν από τον Ian Goodfellow και τους συνεργάτες του το 2014.
- **Σκοπός:** η παραγωγή νέων δειγμάτων παρόμοιων με τα αρχικά διαθέσιμα δεδομένα.
- Η **αρχιτεκτονική** των GANs αποτελείται από 2 νευρωνικά δίκτυα που ανταγωνίζονται μεταξύ τους:
 - **Generator**
Δημιουργεί **νέα συνθετικά δεδομένα** από **τυχαίο θόρυβο**, με στόχο να είναι όσο το δυνατόν πιο ρεαλιστικά, ούτως ώστε ο Discriminator να μην μπορεί να τα διακρίνει από τα αρχικά δεδομένα.
 - **Discriminator**
Λειτουργεί ως **κριτής**, προσπαθώντας να διακρίνει εάν τα δεδομένα τα οποία δέχεται ως είσοδο είναι πραγματικά (real) ή ψεύτικα (fake).
- Τα δύο δίκτυα ανταγωνίζονται συνεχώς, με τρόπο ώστε:
 - Ο Generator να βελτιώνεται στην παραγωγή ρεαλιστικών συνθετικών δειγμάτων και
 - Ο Discriminator να βελτιώνεται στην ανίχνευση των ψεύτικων δειγμάτων.

Generative Adversarial Networks (GANs)

19

Αρχιτεκτονική GANs

□ Generator

- Πρόκειται για βαθύ νευρωνικό δίκτυο, το οποίο λαμβάνει ως είσοδο τυχαίο θόρυβο (latent χώρος) και παράγει ρεαλιστικά δεδομένα, παρόμοια με τα αρχικά (π.χ. εικόνες).
- Μαθαίνει την υποκείμενη κατανομή των δεδομένων.

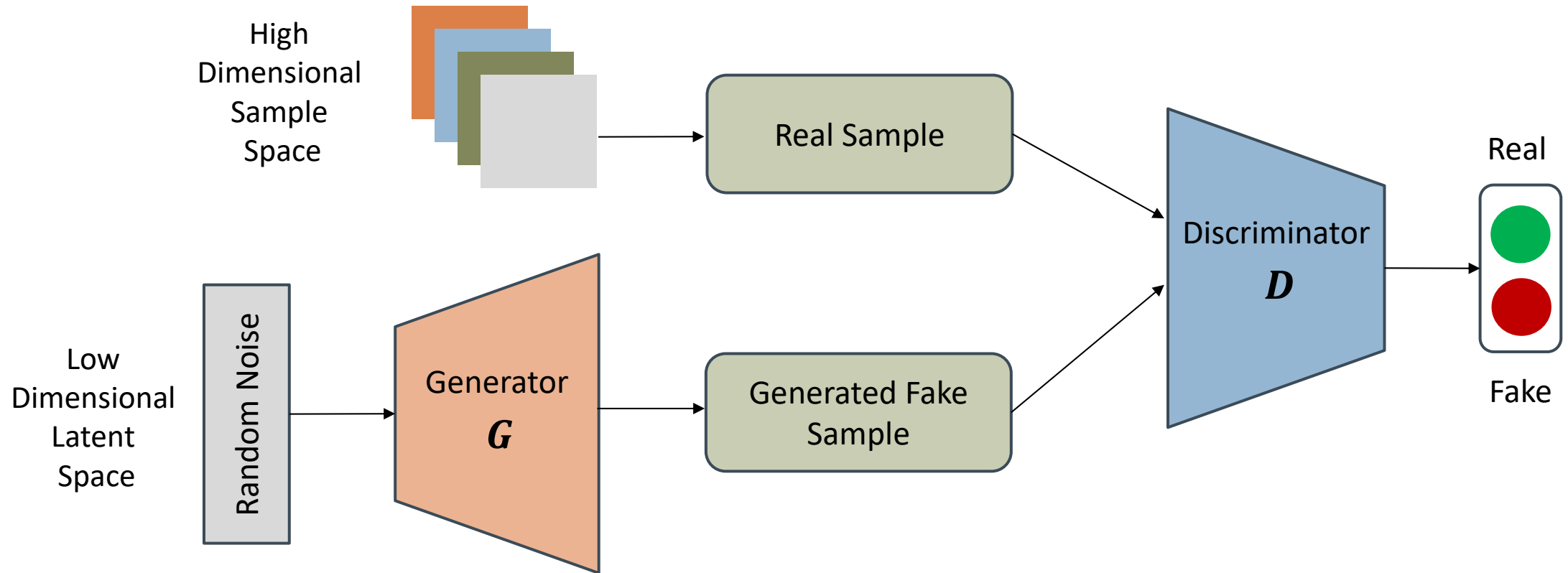
□ Discriminator

- Νευρωνικό δίκτυο το οποίο χρησιμοποιείται για τη δυαδική ταξινόμηση των δεδομένων τα οποία λαμβάνει στην είσοδό του (αληθινά ή ψεύτικα).
- Λαμβάνει δεδομένα από δύο πηγές:
 - **Αληθινά δεδομένα** από το σύνολο των διαθέσιμων δεδομένων, τα οποία χρησιμοποιούνται ως **positive examples** κατά την εκπαίδευση.
 - **Συνθετικά δεδομένα** από τον Generator, τα οποία χρησιμοποιούνται ως **negative examples** κατά την εκπαίδευση.
- Βελτιώνει την ικανότητά του να διακρίνει αληθινά και ψεύτικα δεδομένα μέσω εκπαίδευσης.
- Η αρχιτεκτονική του εξαρτάται από τον τύπο των δεδομένων εισόδου (εικόνες → συνελικτικά δίκτυα).

Generative Adversarial Networks (GANs)

20

Αρχιτεκτονική GANs

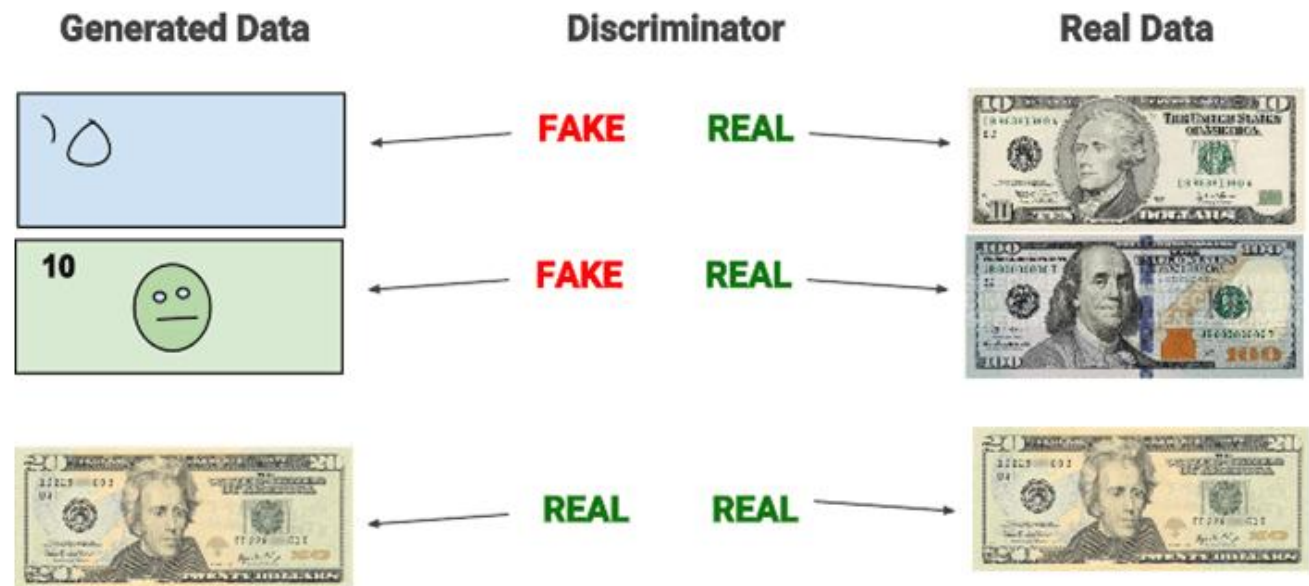


Generative Adversarial Networks (GANs)

21

Εκπαίδευση GANs

- Η εκπαίδευση των δικτύων πραγματοποιείται με **μη επιβλεπόμενο** (unsupervised) και **ανταγωνιστικό** (adversarial) τρόπο.
- Στην αρχή της εκπαίδευσης, ο Generator παράγει ψεύτικα δεδομένα, τα οποία ο Discriminator εύκολα διακρίνει ως ψεύτικα.
- Όσο η εκπαίδευση συνεχίζεται, ο Generator παράγει δεδομένα όλο και πιο όμοια με τα αρχικά, με αποτέλεσμα να δυσκολεύει τον Discriminator.
- Τελικά, καθώς η ικανότητα του Generator να παράγει αληθοφανή δεδομένα βελτιώνεται, η ικανότητα του Discriminator να ταξινομεί τα δεδομένα ως αληθινά ή ψεύτικα μειώνεται.



Generative Adversarial Networks (GANs)

22

Εκπαίδευση GANs

- Η βασική ιδέα κατά την εκπαίδευση των GANs είναι:
 - Εάν ο Discriminator ταξινομεί ορθά τα δεδομένα που λαμβάνει στην είσοδό του ως αληθινά ή ψεύτικα, ενισχύει την ικανότητά του.
 - Εάν ο Generator ξεγελά επιτυχημένα τον Discriminator μέσω των δεδομένων που παράγει, λαμβάνει κάποιο reward, ενώ ο Discriminator κάποιο penalty.
- Τα δίκτυα του Generator και του Discriminator εκπαιδεύονται σε **ξεχωριστές εναλλασσόμενες** περιόδους. Πιο συγκεκριμένα:
 1. Ο Discriminator εκπαιδεύεται για κάποιο αριθμό βημάτων.
 2. Έπειτα εκπαιδεύεται ο Generator.
 3. Τα βήματα 1. και 2. επαναλαμβάνονται μέχρι τη σύγκλιση.
- Κατά την εκπαίδευση του Discriminator (Βήμα 1.) ο Generator παραμένει σταθερός (frozen).
- Κατά την εκπαίδευση του Generator (Βήμα 2.) ο Discriminator παραμένει σταθερός (frozen).

Generative Adversarial Networks (GANs)

23

Εκπαίδευση GANs – Εκπαίδευση Discriminator

1. Δειγματοληψία ενός συνόλου **αληθινών δεδομένων** από το σύνολο των δεδομένων εκπαίδευσης.
2. Δημιουργία ενός συνόλου **ψεύτικων δεδομένων** μέσω του Generator.
3. Τροφοδότηση **αληθινών και ψεύτικων** δεδομένων στον Discriminator.
4. Δυναμική ταξινόμηση των δεδομένων σε αληθινά ή ψεύτικα.
5. Ανανέωση των βαρών του Discriminator, ώστε να **μεγιστοποιηθεί** η συνάρτηση,

$$\mathcal{L}_D = \mathbb{E}_{x \sim p_{data}} [\log D(x)] + \mathbb{E}_{z \sim p_z} \left[\log (1 - D(G(z))) \right].$$

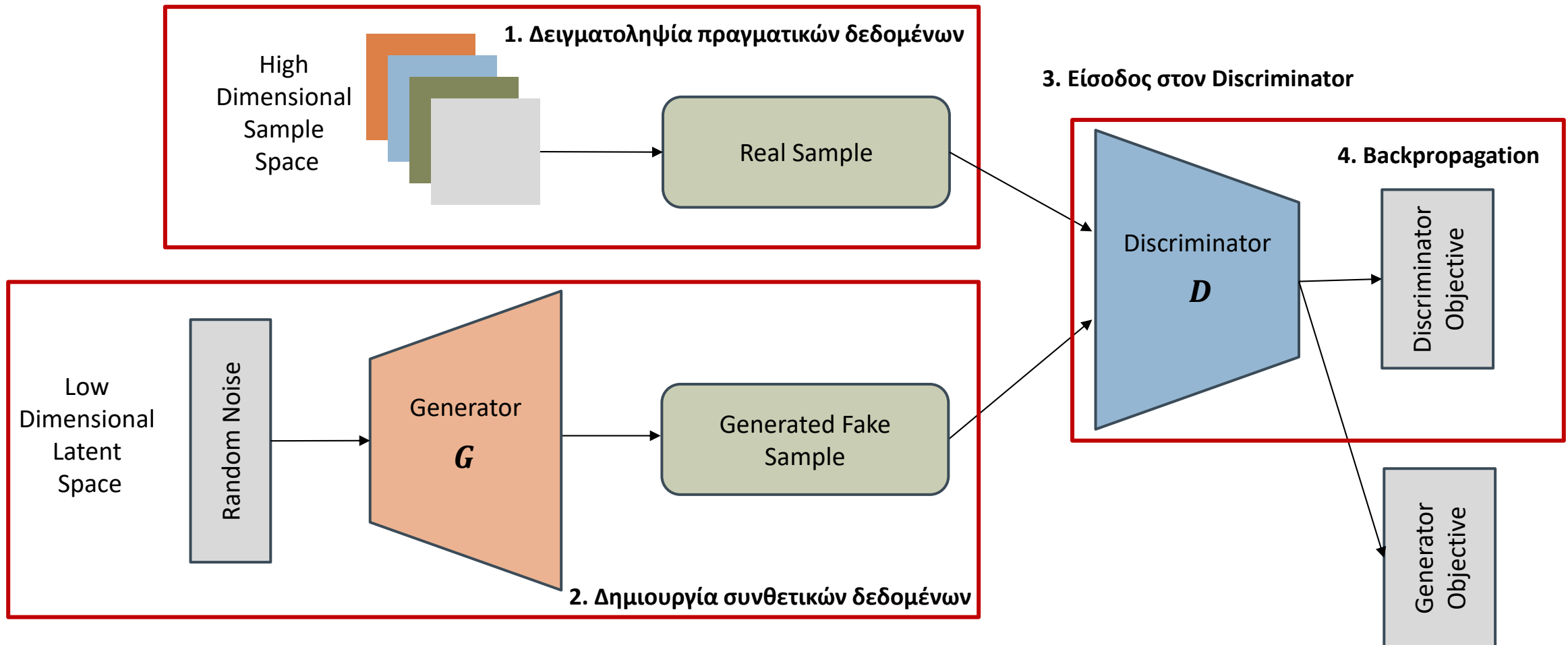
Στην ανωτέρω συνάρτηση:

- $D(x)$ η πιθανότητα το δεδομένο x να είναι αληθινό, όπως αυτή εκτιμάται από τον Discriminator.
- $G(z)$ η έξοδος του Generator, δοθέντος τυχαίου θορύβου z .
- $D(G(z))$ η πιθανότητα το ψεύτικο δείγμα $G(z)$ να είναι αληθινό, όπως αυτή εκτιμάται από τον Discriminator.
- Ο τύπος προκύπτει από την εφαρμογή της binary cross-entropy loss αθροιστικά για τα αληθινά και τα συνθετικά δεδομένα: $\mathcal{L}_D = \text{BCE}(1, D(x)) + \text{BCE}(0, D(G(z)))$.

Generative Adversarial Networks (GANs)

24

Εκπαίδευση GANs – Εκπαίδευση Discriminator



Generative Adversarial Networks (GANs)

25

Εκπαίδευση GANs – Εκπαίδευση Generator

1. Δημιουργία ενός συνόλου **ψεύτικων δεδομένων** μέσω του Generator.
2. Τροφοδότηση **μόνο ψεύτικων** δεδομένων στον Discriminator.
3. Δυναμική ταξινόμηση των δεδομένων σε αληθινά ή ψεύτικα.
4. Ανανέωση των βαρών του Generator, ώστε να **ελαχιστοποιηθεί** η συνάρτηση,

$$\mathcal{L}_G = \mathbb{E}_{z \sim p_z} \left[\log \left(1 - D(G(z)) \right) \right].$$

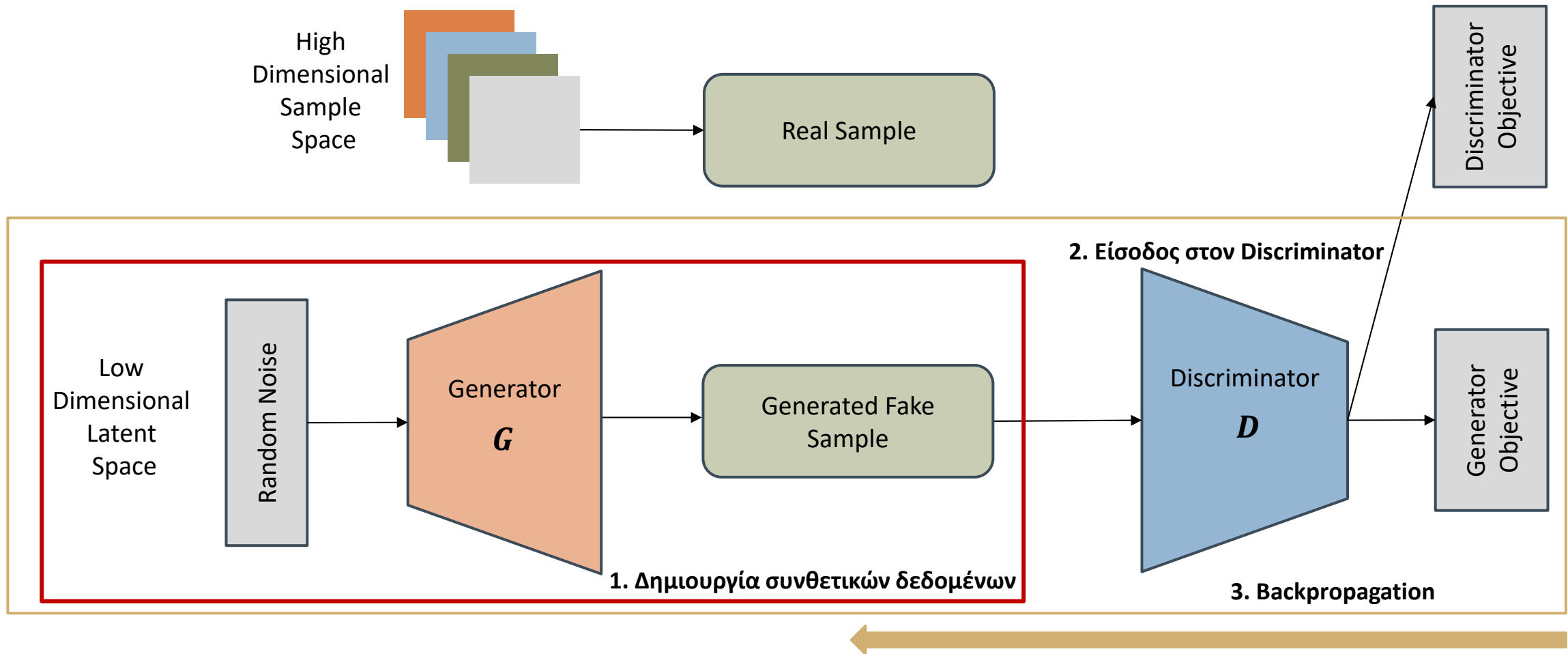
Στην ανωτέρω συνάρτηση:

- $G(z)$ η έξοδος του Generator, δοθέντος τυχαίου θορύβου z .
- $D(G(z))$ η πιθανότητα το ψεύτικο δείγμα $G(z)$ να είναι αληθινό, όπως αυτή εκτιμάται από τον Discriminator.

Generative Adversarial Networks (GANs)

26

Εκπαίδευση GANs – Εκπαίδευση Generator



Generative Adversarial Networks (GANs)

27

Εκπαίδευση GANs – Αλγόριθμος Εκπαίδευσης

Algorithm 1 Minibatch stochastic gradient descent training of generative adversarial nets. The number of steps to apply to the discriminator, k , is a hyperparameter. We used $k = 1$, the least expensive option, in our experiments.

for number of training iterations **do**

for k steps **do**

- Sample minibatch of m noise samples $\{z^{(1)}, \dots, z^{(m)}\}$ from noise prior $p_g(z)$.
- Sample minibatch of m examples $\{x^{(1)}, \dots, x^{(m)}\}$ from data generating distribution $p_{\text{data}}(x)$.
- Update the discriminator by ascending its stochastic gradient:

$$\nabla_{\theta_d} \frac{1}{m} \sum_{i=1}^m \left[\log D(x^{(i)}) + \log (1 - D(G(z^{(i)}))) \right].$$

end for

- Sample minibatch of m noise samples $\{z^{(1)}, \dots, z^{(m)}\}$ from noise prior $p_g(z)$.
- Update the generator by descending its stochastic gradient:

$$\nabla_{\theta_g} \frac{1}{m} \sum_{i=1}^m \log (1 - D(G(z^{(i)}))).$$

end for

The gradient-based updates can use any standard gradient-based learning rule. We used momentum in our experiments.

Training the
Discriminator

Training the
Generator

Generative Adversarial Networks (GANs)

28

Εκπαίδευση GANs – Προβλήματα Σύγκλισης

▣ Vanishing (ή Exploding) Gradients

- Σε περίπτωση που ο Discriminator είναι «πολύ καλός» κριτής, η πιθανότητα $D(G(z)) \rightarrow 0$ για κάθε z και έτσι η παράγωγος $\nabla \{\log(1 - D(G(z)))\}$ «εξαφανίζεται», με αποτέλεσμα ο Generator να σταματάει να μαθαίνει.
- Αντίστοιχα, σε περίπτωση που ο Discriminator είναι «πολύ αδύναμος» κριτής παρουσιάζεται το φαινόμενο του exploding gradient.

▣ Ταλαντώσεις και Αδυναμία Σύγκλισης

- Εμφάνιση ταλαντώσεων στις εξόδους και στις τιμές της συνάρτησης απώλειας.
- **Αιτία:** ανισορροπία στις ανανεώσεις των βαρών των δικτύων του Generator και του Discriminator και έλλειψη θεωρητικού υποβάθρου το οποίο να εξασφαλίζει τη σύγκλιση.

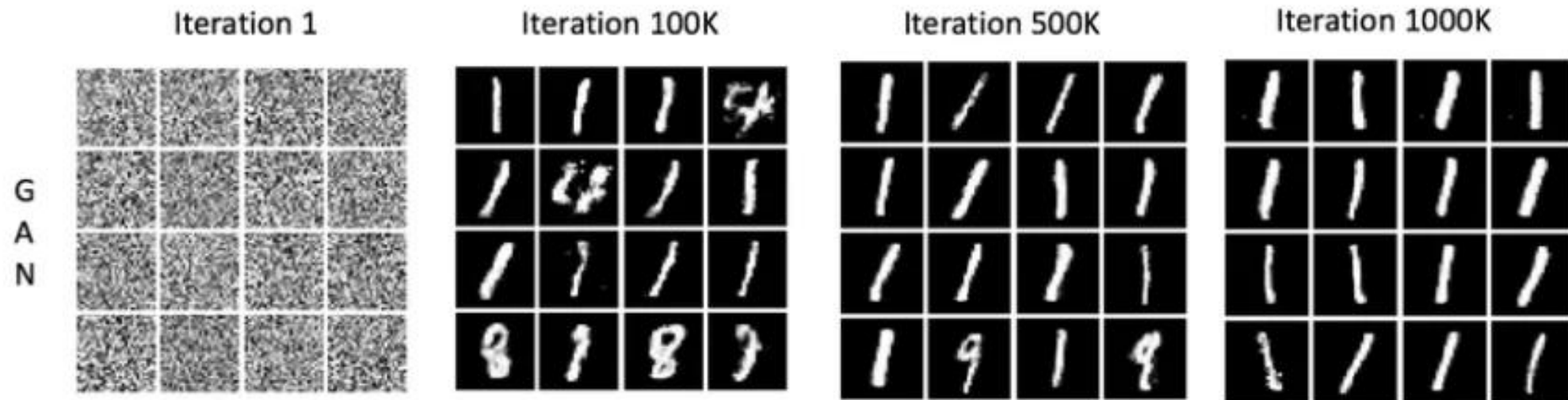
Generative Adversarial Networks (GANs)

29

Εκπαίδευση GANs – Προβλήματα Σύγκλισης

□ Mode Collapse

- Φαινόμενο κατά το οποίο ο Generator παράγει δείγματα περιορισμένων μοτίβων και ποικιλίας.
- **Αιτία:** ο Generator συγκλίνει κατά την εκπαίδευση σε ένα περιορισμένο σύνολο δεδομένων, το οποίο ξεγελά σταθερά τον Discriminator και προσκολλάται σε αυτό (zero gradients).



Generative Adversarial Networks (GANs)

30

Εκπαίδευση GANs – Minimax game

- Πρακτικά, η εκπαίδευση των GANs προσομοιάζει ένα **παίγνιο μηδενικού και σταθερού αθροίσματος** (two-player zero-sum game), στο οποίο το κέρδος του ενός παίκτη (Generator) ισούται με τη ζημιά του άλλου (Discriminator).
- Ακολουθείται η στρατηγική **Minimax** για την εκπαίδευση των μοντέλων, βάσει της συνολικής αντικειμενικής συνάρτησης:

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

Generative Adversarial Networks (GANs)

31

Conditional GANs (CGANs)

- Βελτιωμένη εκδοχή των GANs στην οποία τόσο ο Generator όσο και ο Discriminator γίνονται conditioned σε κάποια επιπλέον πληροφορία (π.χ. ετικέτες, κείμενο).
- Αυτό καθιστά τη διαδικασία της σύνθεσης πιο ελεγχόμενη και στοχευμένη.
- **Generator στα CGANs:** χρησιμοποιείται για την παραγωγή συνθετικών δεδομένων. Δέχεται ως είσοδο:
 - **Τυχαίο θόρυβο (z)**
 - **Επιπλέον συνθήκη (y):** πρόκειται για επιπλέον δεδομένα π.χ. ετικέτες, τα οποία ελέγχουν τη διαδικασία σύνθεσης δειγμάτων του Generator.
 - Ο Generator συνδυάζει και τις δύο εισόδους για να παράξει ρεαλιστικά συνθετικά δεδομένα. Για παράδειγμα, εάν $y \rightarrow \text{'cat'}$, θα παράξει την εικόνα μιας γάτας.
- **Discriminator στα CGANs:** δυαδικός ταξινομητής των δεδομένων εισόδου. Δέχεται ως είσοδο:
 - **Αληθινά και ψεύτικα δεδομένα**
 - **Επιπλέον συνθήκη (y):** πρόκειται για επιπλέον δεδομένα π.χ. ετικέτες, τα ίδια με αυτά που λαμβάνει ο Generator.

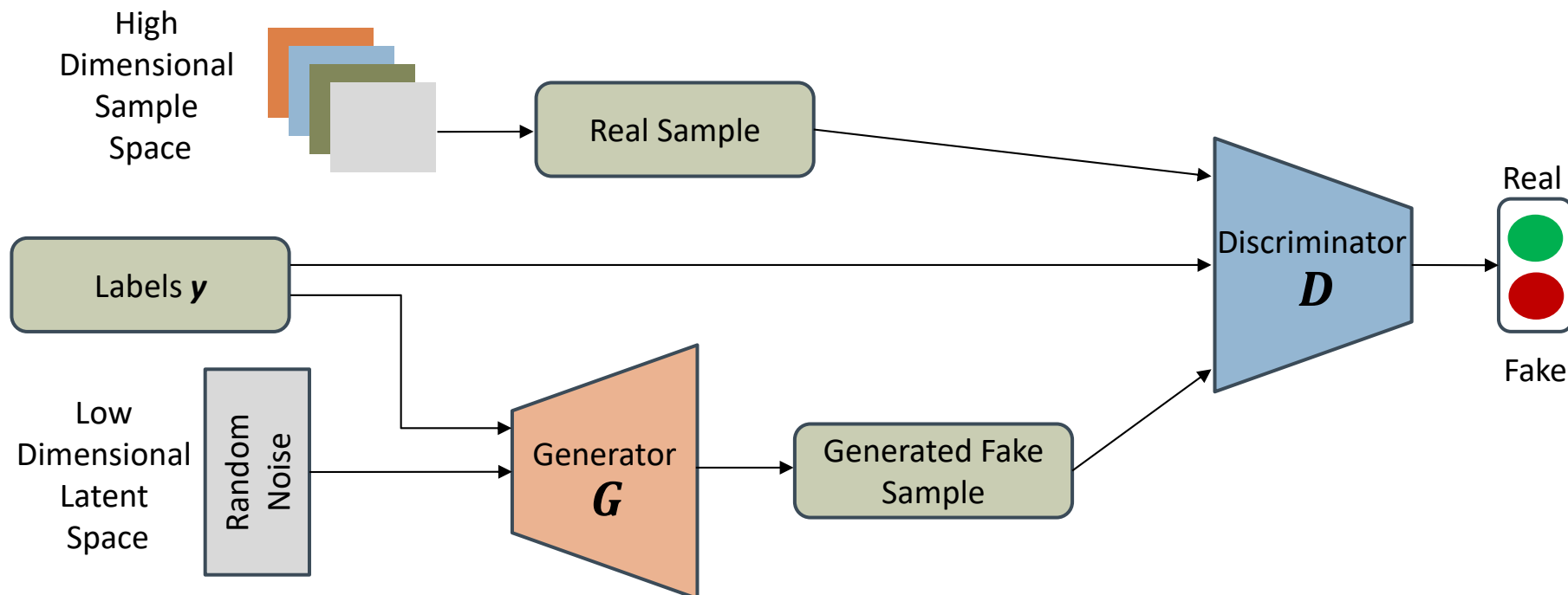
Generative Adversarial Networks (GANs)

32

Conditional GANs – Εκπαίδευση

- Τα CGANs εκπαιδεύονται προσομοιάζοντας και αυτά το **παίγνιο μηδενικού και σταθερού αθροίσματος** με Minimax στρατηγική, βάσει της αντικειμενική συνάρτησης:

$$\min_G \max_D \mathbb{E}_{x \sim p_{data}} [\log D(x|y)] + \mathbb{E}_{z \sim p_z} [\log (1 - D(G(z|y)))] .$$



Generative Adversarial Networks (GANs)

33

Deep Convolutional GANs (DCGANs)

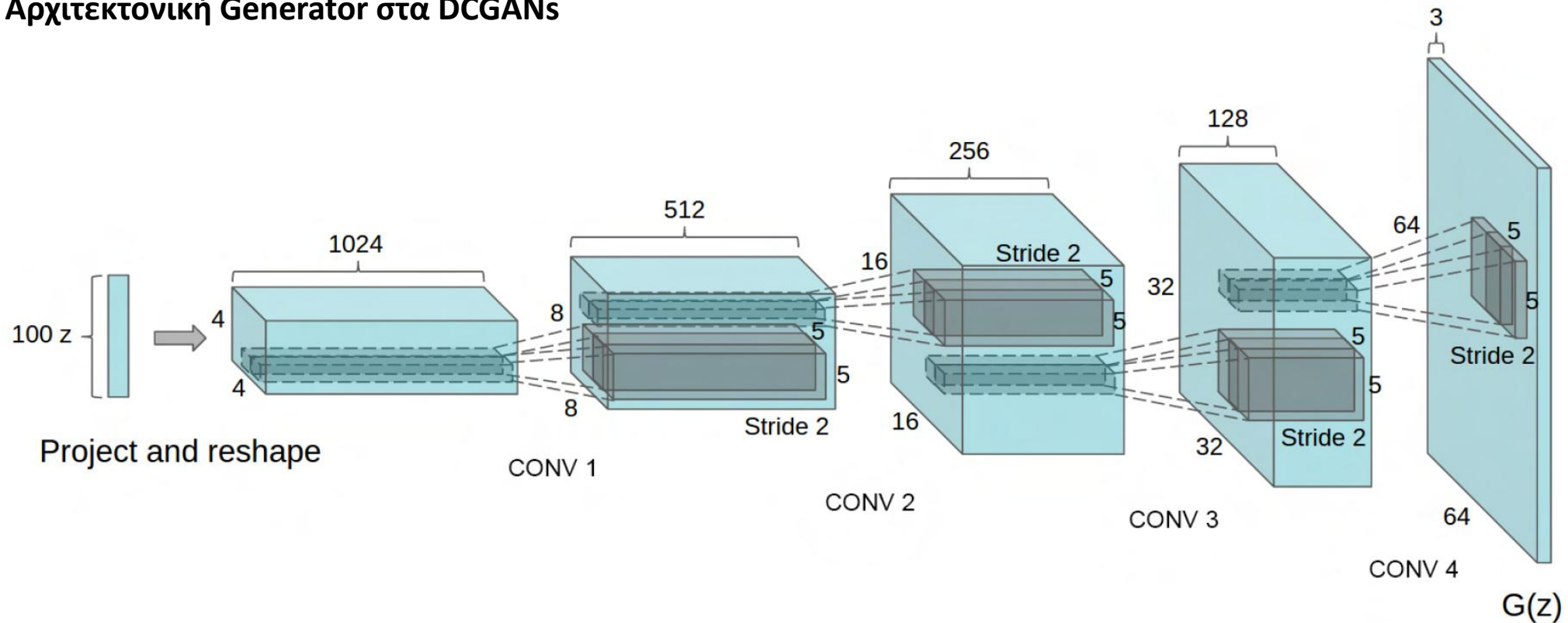
- Τα Deep Convolutional GANs (DCGANs) αναπτύχθηκαν προκειμένου να περιοριστεί το φαινόμενο του mode collapse.
- Χρησιμοποιούν συνελκτικά layers στα δίκτυα των Generator και Discriminator (αντί για fully connected)
→ εξαγωγή **χωρικών εξαρτήσεων** και παραγωγή πιο ποικίλων δειγμάτων.
- **Generator στα DCGANs**
 - Δέχεται ως είσοδο λανθάνουσες αναπαραστάσεις θορύβου 100 διαστάσεων, $\mathbf{z} \in \mathbb{R}^{100}$.
 - Μέσω ενός fully connected layer μετασχηματίζει την είσοδο σε ένα διάνυσμα διαστάσεων $4 * 4 * 1024$, το οποίο γίνεται reshape σε διαστάσεις $4 \times 4 \times 1024$.
 - Εφαρμόζει 4 διαδοχικά layers με Transposed Convolutions (διπλασιασμός των χωρικών διαστάσεων) χρησιμοποιώντας διαφορετικό πλήθος φίλτρων .
 - Η τελική έξοδος έχει διαστάσεις $64 \times 64 \times 3$.
 - Χρησιμοποιεί συναρτήσεις ενεργοποίησης ReLU στα ενδιάμεσα layers και Tanh στην έξοδο.

Generative Adversarial Networks (GANs)

34

Deep Convolutional GANs (DCGANs)

□ Αρχιτεκτονική Generator στα DCGANs



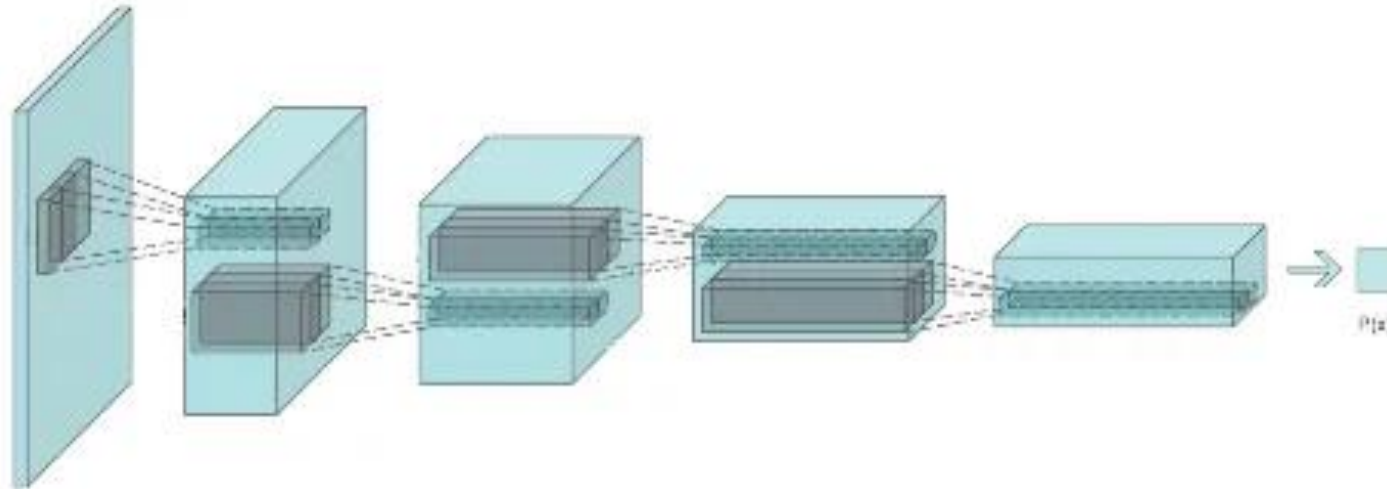
Generative Adversarial Networks (GANs)

35

Deep Convolutional GANs (DCGANs)

■ Discriminator στα DCGANs

- Δέχεται ως είσοδο εικόνες διαστάσεων $64 \times 64 \times 3$.
- Αποτελείται από συνελκτικά layers με strided-convolutions.
- Χρησιμοποιεί συνάρτηση ενεργοποίησης LeakyReLU στα ενδιάμεσα layers και Sigmoid στην έξοδο.



Generative Adversarial Networks (GANs)

36

Energy-based GANs (EBGANs)

■ Energy-Based Μοντέλα

- Είδος μοντέλων μηχανικής μάθησης τα οποία μαθαίνουν να αποδίδουν μία τιμή ενέργειας στις εισόδους.
- Είσοδος με επιθυμητά χαρακτηριστικά (ρεαλιστική εικόνα) → Χαμηλή ενέργεια.
- Είσοδος με μη επιθυμητά χαρακτηριστικά (ψεύτικη εικόνα) → Υψηλή ενέργεια.

- Τα EBGANs αποτελούν μια παραλλαγή των βασικών Generative Adversarial Networks.

■ Discriminator στα EBGANs

- Χρησιμοποιεί μια αρχιτεκτονική autoencoder.
- Αντί για πιθανότητα ταξινόμησης, υπολογίζει πόσο καλά ανακατασκευάζεται η είσοδός του:

$$D(x) = ||x - \text{AutoEncoder}(x)||^2$$

- Οι αληθινές εικόνες θα πρέπει να ανακατασκευάζονται καλά → Χαμηλή ενέργεια.
- Οι ψεύτικες εικόνες θα πρέπει να ανακατασκευάζονται με απώλειες → Υψηλή ενέργεια.

Generative Adversarial Networks (GANs)

37

Energy-based GANs (EBGANs)

Generator στα EBGANs

- Προσπαθεί να παράξει εικόνες οι οποίες ξεγελούν τον Discriminator, ούτως ώστε να τους αποδίδει χαμηλή ενέργεια.
- Εκπαιδεύεται ώστε να ελαχιστοποιεί την ενέργεια των εικόνων που παράγει.

Εκπαίδευση EBGAN

- Δοθέντος ενός αληθινού δείγματος x , ενός συνθετικού δείγματος $G(z)$ και ενός θετικού margin m , οι αντικειμενικές συναρτήσεις \mathcal{L}_G και \mathcal{L}_D του Generator και Discriminator αντίστοιχα, ορίζονται ως:

$$\mathcal{L}_D(x, z) = D(x) + \max\{0, [m - D(G(z))]\}$$

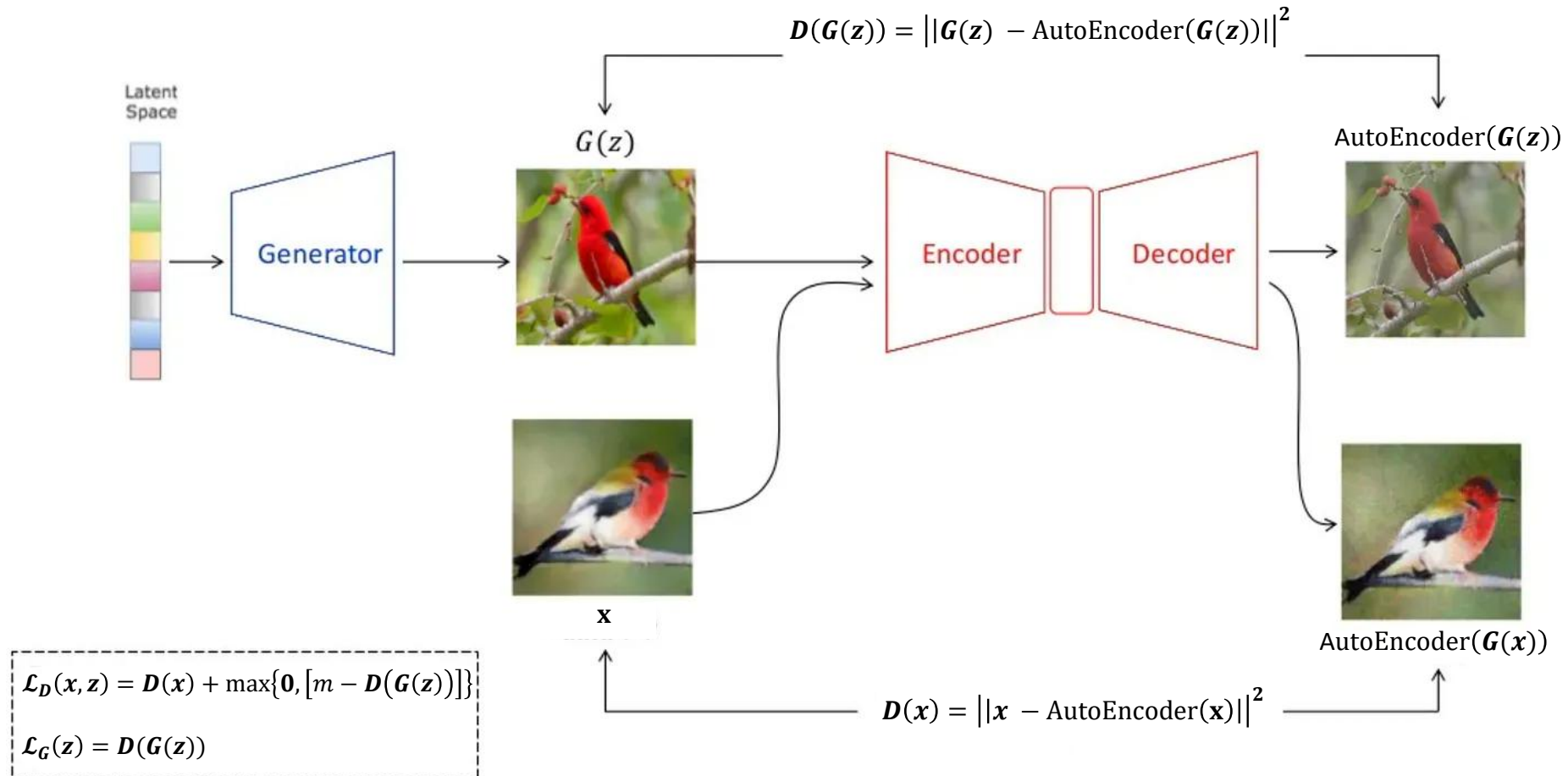
$$\mathcal{L}_G(z) = D(G(z))$$

- Ο Discriminator προσπαθεί να ελαχιστοποιήσει την \mathcal{L}_D .
- Ο παράμετρος m περιορίζει το penalty που δέχεται ο Generator λόγω της επιτυχούς ταξινόμησης του Discriminator \rightarrow φράσει την τιμή του $D(G(z))$ και ενισχύει την ευστάθεια κατά την εκπαίδευση.
- Εάν $D(G(z)) < m$:
 - Η συνθετική εικόνα μοιάζει με αληθινή.
 - Η συνάρτηση απώλειας αυξάνεται κατά $m - D(G(z))$.
- Εάν $D(G(z)) \geq m$:
 - Η συνθετική εικόνα αναγνωρίζεται ήδη ως ψεύτικη.
 - Ο Discriminator δεν αλλάζει τη λειτουργία του.
- Ο Generator προσπαθεί να ελαχιστοποιήσει την \mathcal{L}_G .

Generative Adversarial Networks (GANs)

38

Energy-based GANs (EBGANs)



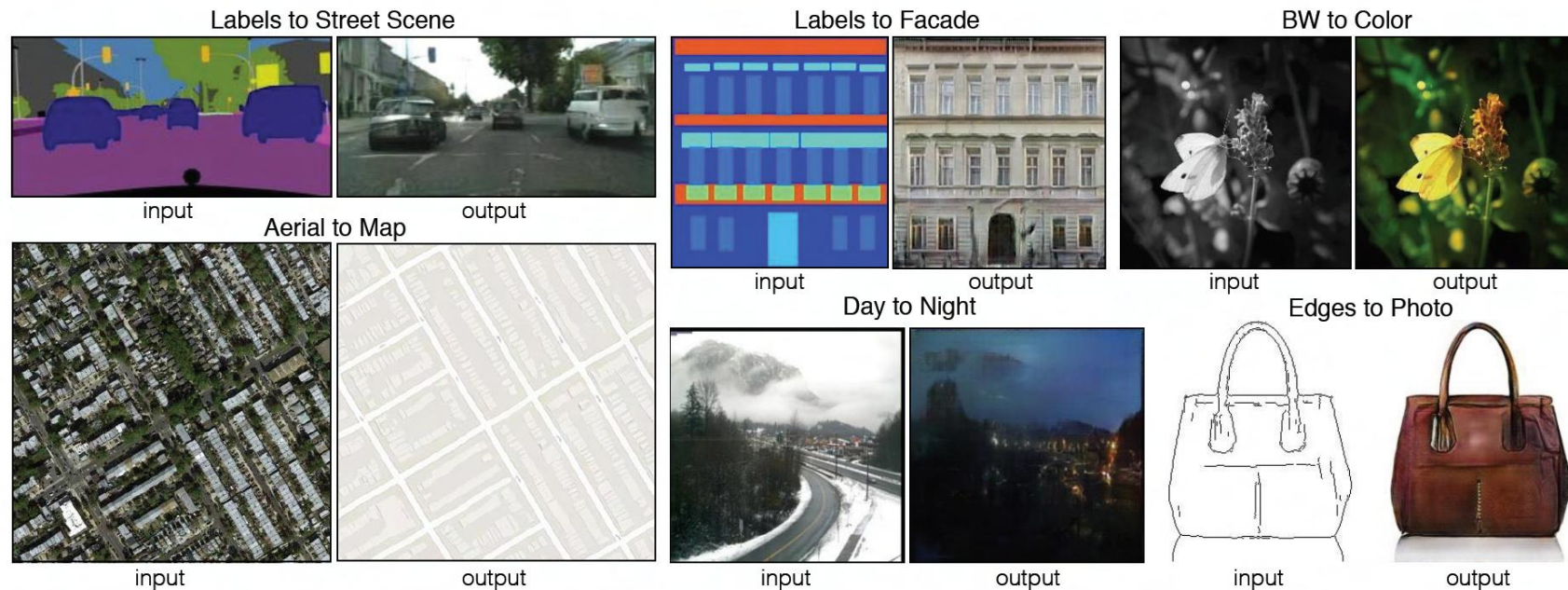
Διάρθρωση της Παρουσίασης

39

- Εισαγωγή
- Variational Autoencoders
- Generative Adversarial Networks (Gans)
- **Εφαρμογές**

Image-to-Image Translation

- Σκοπός του Image-to-Image Translation είναι η μεταφορά εικόνων από ένα αρχικό πεδίο (source domain) σε ένα άλλο (target domain), διατηρώντας την αναπαράσταση του αρχικού περιεχομένου.
- Πληθώρα εφαρμογών

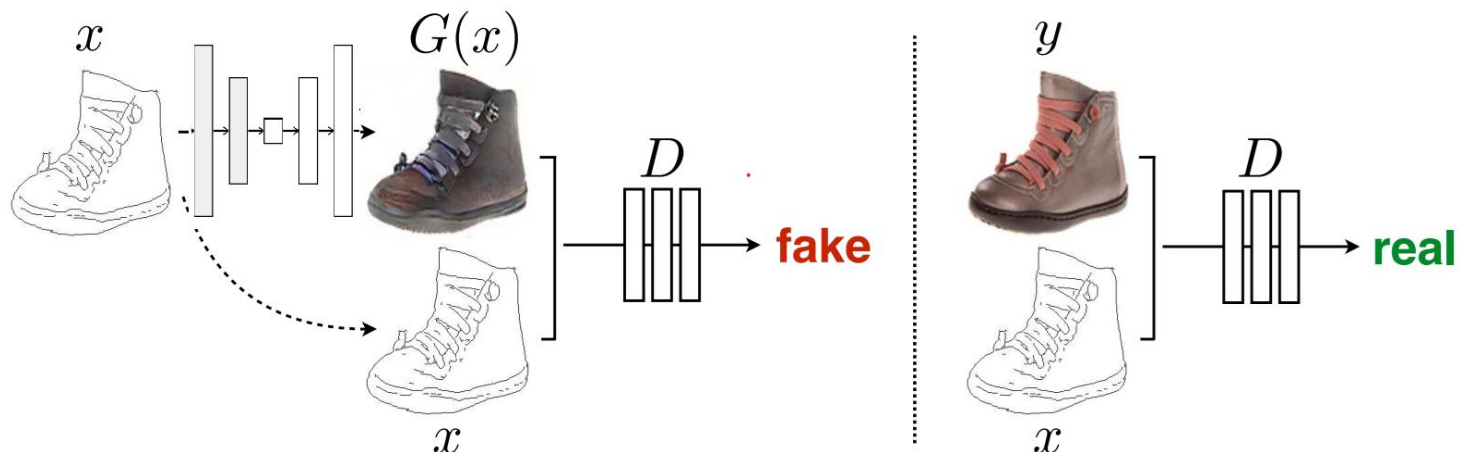


Εφαρμογές

41

Image-to-Image Translation

- Ο Generator παράγει συνθετικά δείγματα βάσει της συνθήκης y .
- Τόσο ο Generator όσο και ο Discriminator έχουν πρόσβαση στη συνθήκη y .
- Ο Discriminator μαθαίνει να ταξινομεί τα ζεύγη της μορφής {συνθήκη, εικόνα} ως αληθινά ή ψεύτικα.

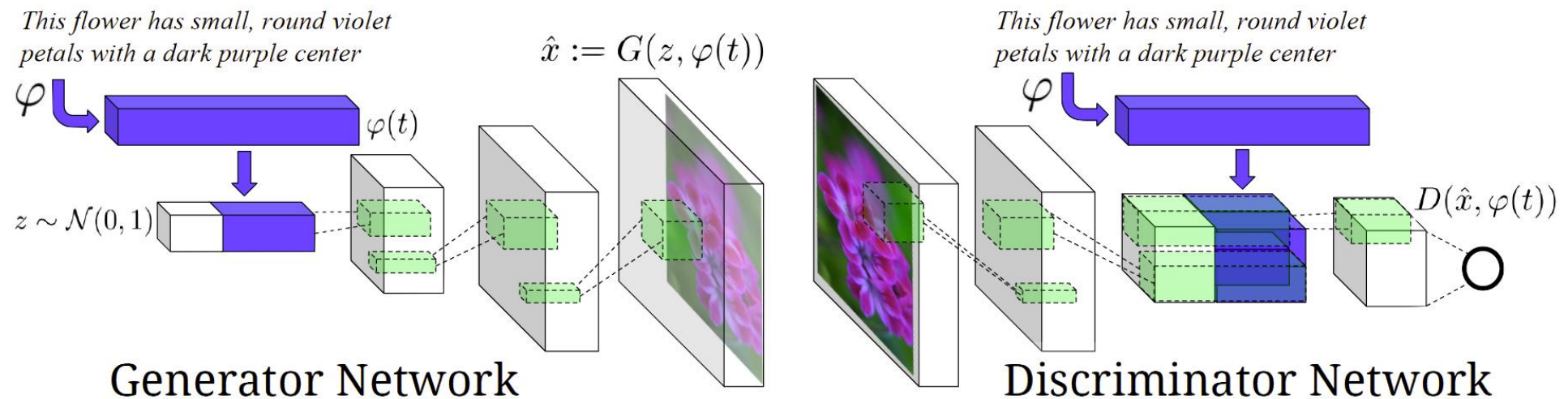


Εφαρμογές

42

Text-to-Image Synthesis

- Εισαγωγή του embedding του κειμένου τόσο στον Generator όσο και στον Discriminator.
- Positive Examples \rightarrow {Right Image, Right Text}
- Negative examples \rightarrow {Fake Image, Right Text} και {Right Image, Wrong Text}



Super Resolution

- Στο task του Image Super-Resolution θέλουμε να αυξήσουμε την ανάλυση μιας εικόνας χαμηλής ανάλυσης.

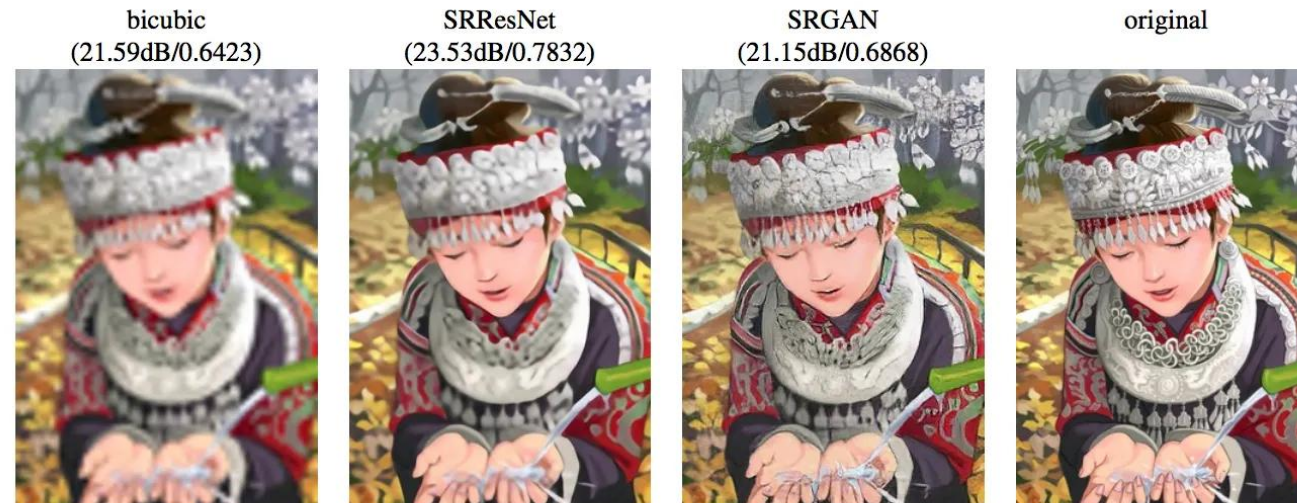


Figure 2: From left to right: bicubic interpolation, deep residual network optimized for MSE, deep residual generative adversarial network optimized for a loss more sensitive to human perception, original HR image. Corresponding PSNR and SSIM are shown in brackets. [4× upscaling]

Super Resolution

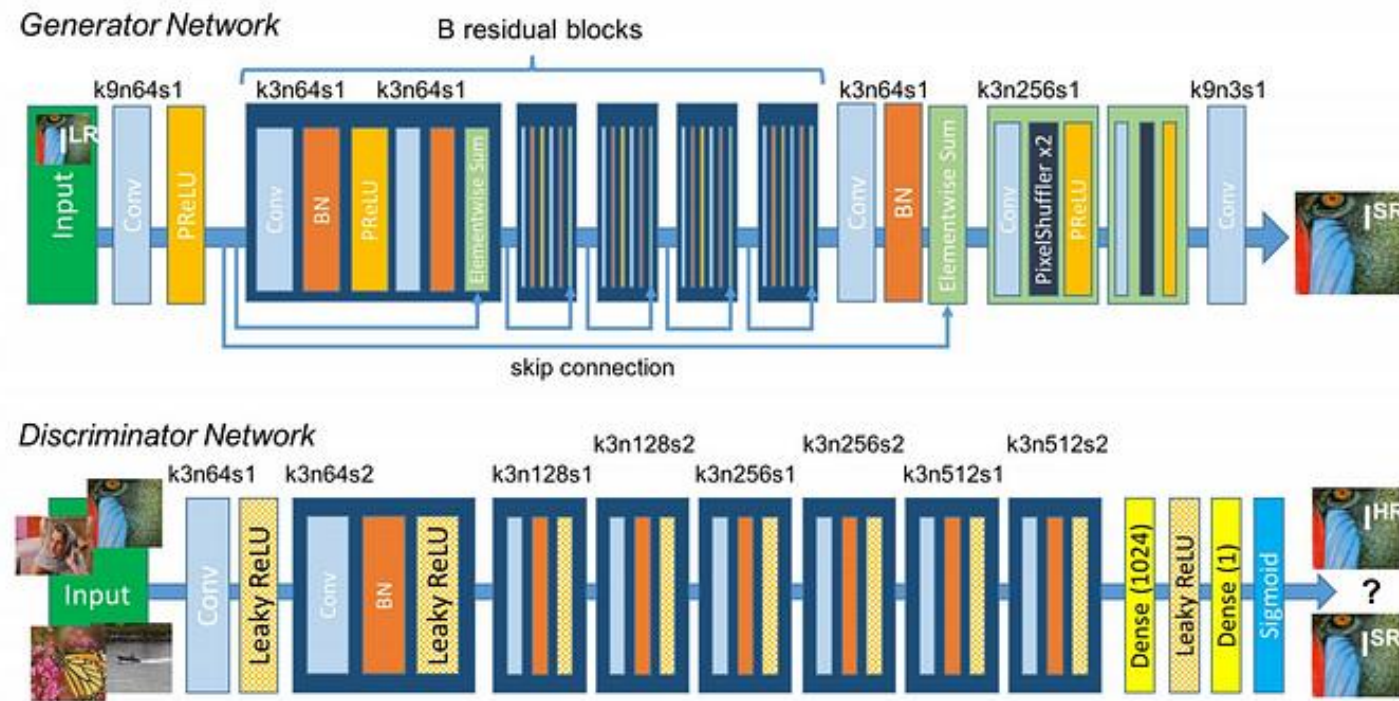
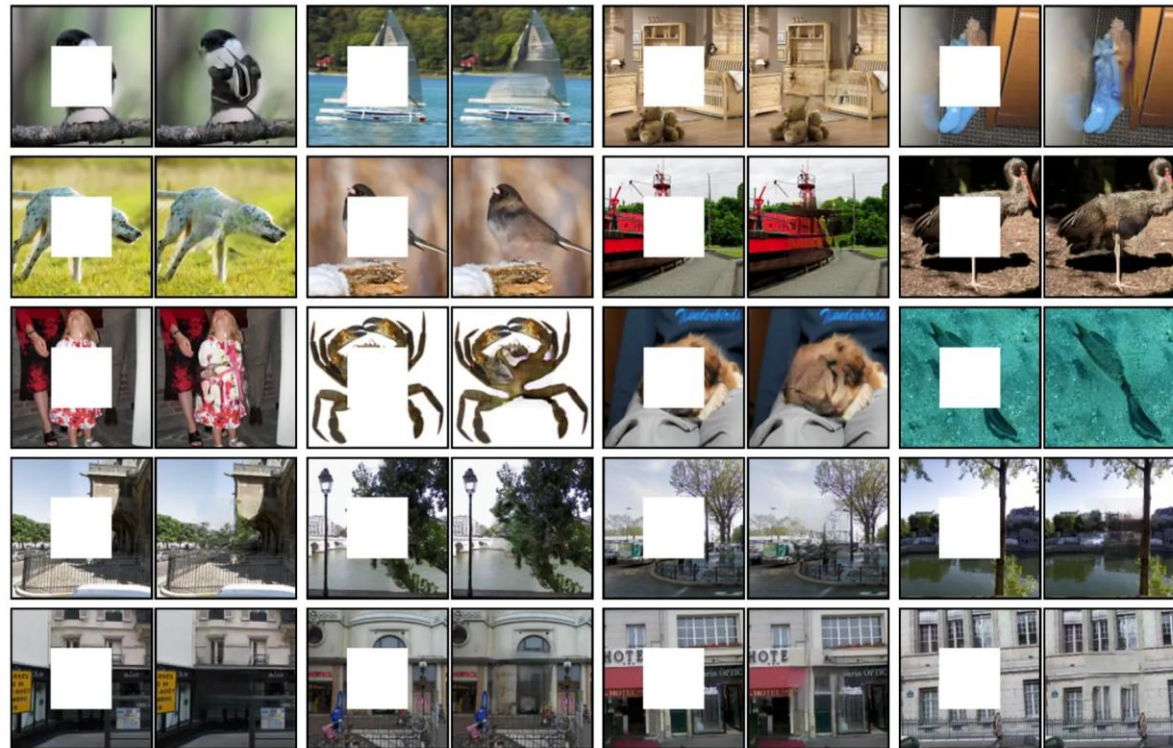


Image Inpainting

- Στο task του Image Inpainting προσπαθούμε να ανακατασκευάσουμε και να συμπληρώσουμε κενές περιοχές μιας εικόνας.

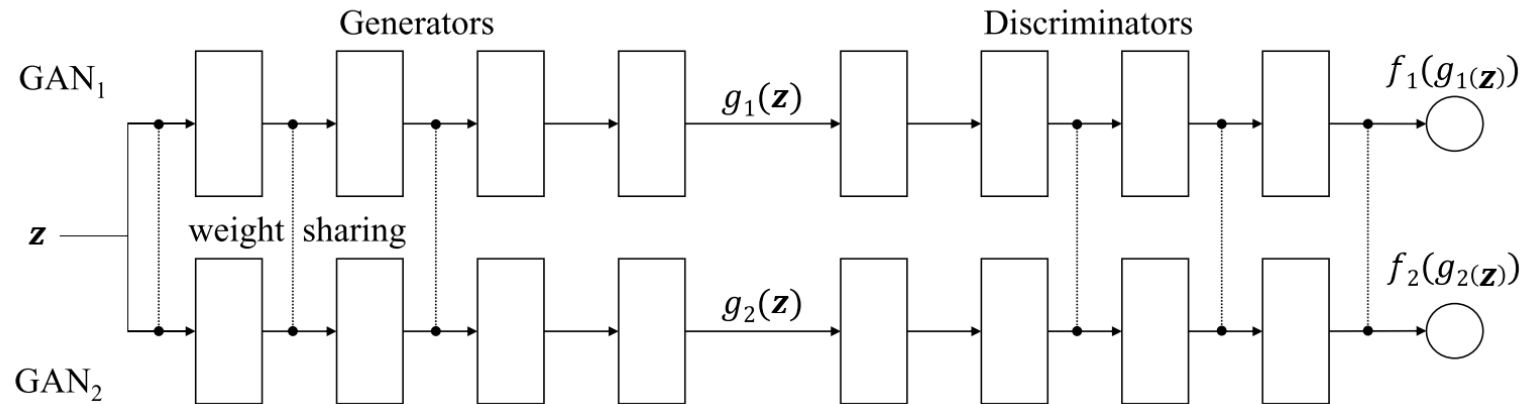


Εφαρμογές

46

Coupled Gans

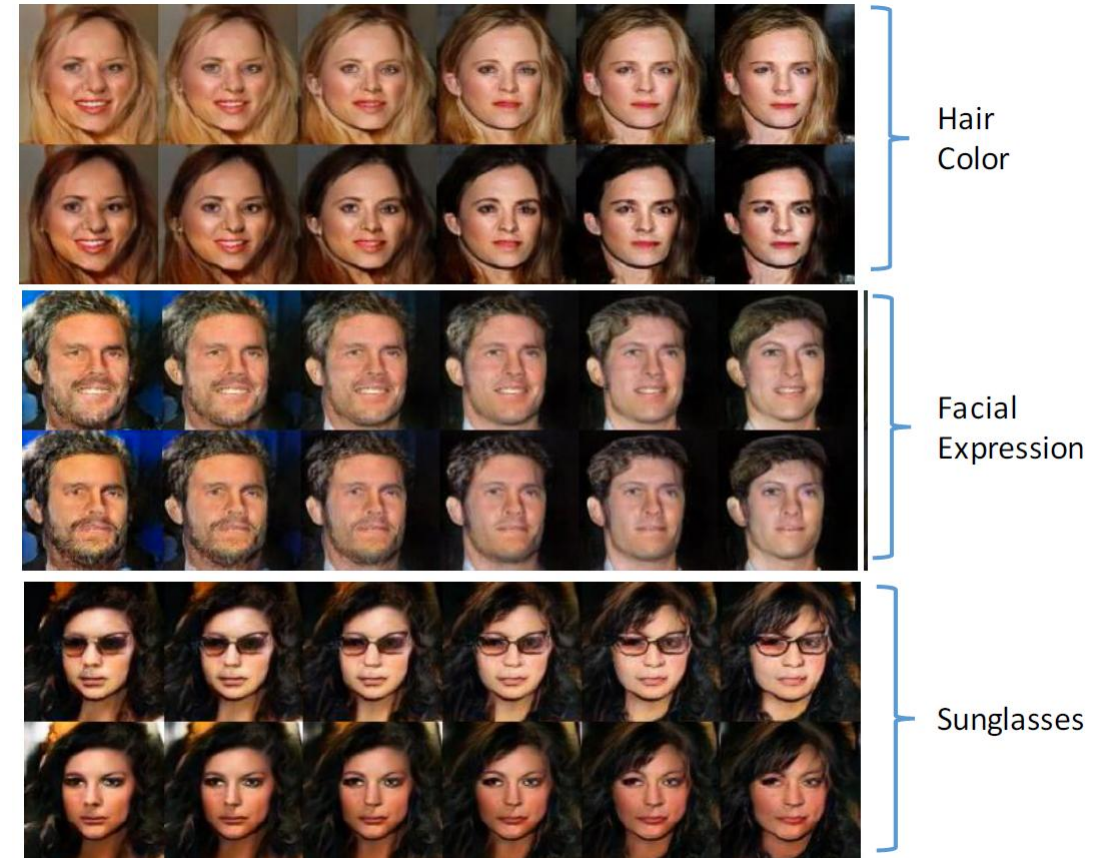
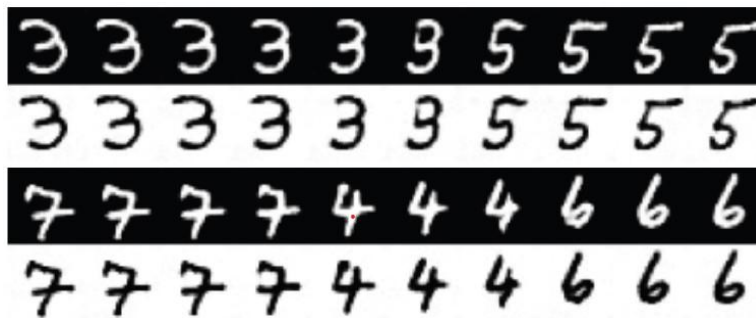
- Χρησιμοποιούνται για την εκμάθηση της από κοινού κατανομής εικόνων που προέρχονται από διαφορετικά πεδία (domains).
- Χρησιμοποιούν δύο GANs τα οποία έχουν μοιραζόμενα βάρη.
 - Οι Generator μοιράζονται τα βάρη τους στα πρώτα επίπεδα.
 - Οι Discriminators μοιράζονται τα βάρη τους στα τελευταία επίπεδα.
- Οι τελικές εικόνες παράγονται από την ίδια αρχική λανθάνουσα αναπαράσταση \mathbf{z} .



Εφαρμογές

47

Coupled Gans



Generative Adversarial Networks (GANs)

48



Δείγματα από GAN.



Παραγωγή με χρήση του Stable Diffusion