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

AI|LS

Παραγωγικά Μοντέλα (II)
Diffusion Models

Ηλίας Μήτσουρας

Διάρθρωση της Παρουσίασης

- ***Εισαγωγή***
- Diffusion Models
  - Ιστορική Εξέλιξη
  - Διαδικασία Διάχυσης (Forward Diffusion)
  - Διαδικασία Αντίστροφης Διάχυσης (Reverse Diffusion)
  - Δίκτυο Αποθορυβοποίησης
  - Καθοδηγούμενη Σύνθεση Εικόνων
  - Stable Diffusion
  - Denoising Diffusion Implicit Models
  - Score-Based Generative Models
  - Noise Conditional Score Networks
  - Stochastic Differential Equations
  - Περιορισμοί
  - Generative Models Trilemma
- Εφαρμογές

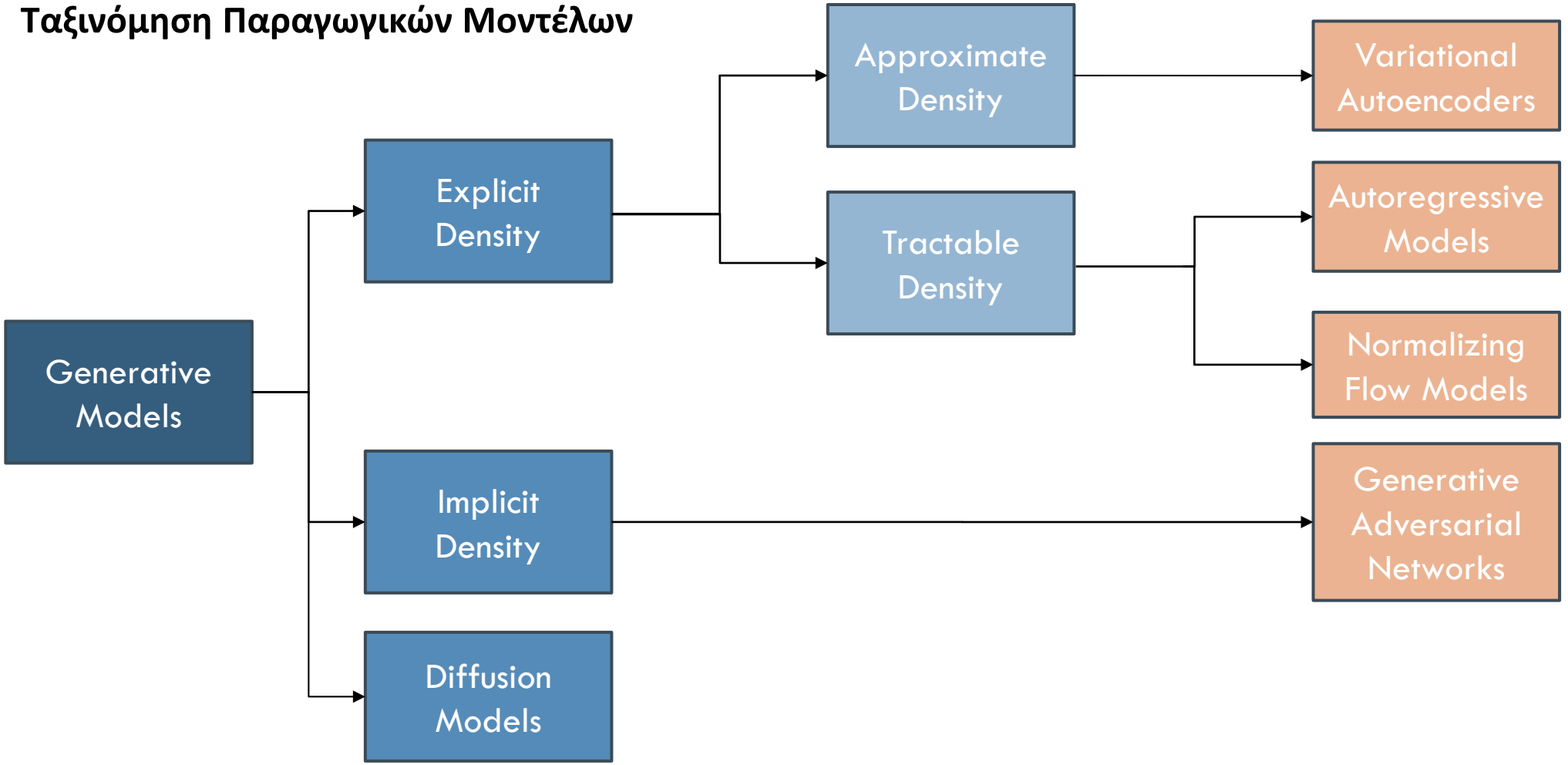
Εισαγωγή

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 $\boldsymbol{X} = \{\boldsymbol{x_1}, \boldsymbol{x_2}, \dots, \boldsymbol{x_n}\}$, και ενός συνόλου ετικετών $\boldsymbol{Y} = \{y_1, y_2, \dots, y_n\}$ (προαιρετικά), ένα παραγωγικό μοντέλο προσπαθεί να προσεγγίσει:
  - την από κοινού κατανομή $\boldsymbol{p}(\boldsymbol{X}, \boldsymbol{Y})$ ή
  - την κατανομή $\boldsymbol{p}(\boldsymbol{X})$, σε περίπτωση όπου δεν δίνεται το σύνολο $\boldsymbol{Y}$**.**
- Διαφέρουν από τα διακριτικά μοντέλα (*discriminative models*), τα οποία προσπαθούν να εκτιμήσουν την υπό συνθήκη κατανομή πιθανότητας $\boldsymbol{p}(\boldsymbol{Y} \mid \boldsymbol{X})$.
- Τα παραγωγικά μοντέλα ταξινομούνται βάσει του τρόπου με τον οποίο προσπαθούν να εκτιμήσουν τη συνάρτηση πυκνότητας πιθανότητας $\boldsymbol{p}_{\text{model}}(\boldsymbol{X})$.

Εισαγωγή

Ταξινόμηση Παραγωγικών Μοντέλων



Διάρθρωση της Παρουσίασης

- Εισαγωγή
- ***Diffusion Models***
  - Ιστορική Εξέλιξη
  - Διαδικασία Διάχυσης (Forward Diffusion)
  - Διαδικασία Αντίστροφης Διάχυσης (Reverse Diffusion)
  - Δίκτυο Αποθορυβοποίησης
  - Καθοδηγούμενη Σύνθεση Εικόνων
  - Stable Diffusion
  - Denoising Diffusion Implicit Models
  - Score-Based Generative Models
  - Noise Conditional Score Networks
  - Stochastic Differential Equations
  - Περιορισμοί
  - Generative Models Trilemma
- Εφαρμογές

Diffusion Models

Ιστορική Εξέλιξη

- **2015**
  - Πρώτες ιδέες & θεμελίωση των βασικών αρχών των μοντέλων διάχυσης
  - Περιορισμένη πρακτική εφαρμογή λόγω υπολογιστικών περιορισμών
- **2019**
  - Επέκταση των βασικών ιδεών και εισαγωγή της έννοιας της σταδιακής αποθορυβοποίησης
- **2021**
  - Εμφάνιση του *DALL-E* από την *OpenAI*
  - Πρώτη ολοκληρωμένη προσπάθεια αντιμετώπισης του προβλήματος της παραγωγής εικόνων από κειμενικές περιγραφές
- **2022**
  - Κυκλοφορία του *Stable Diffusion (open source)*
  - Εκρηκτική αύξηση της δημοτικότητας των μοντέλων διάχυσης
- **2023 – σήμερα**
  - Επέκταση σε νέους τομείς (βίντεο, 3D)
  - Ενσωμάτωση σε εμπορικές εφαρμογές

Diffusion Models

Εισαγωγή

- Τα μοντέλα διάχυσης πρωτοπαρουσιάστηκαν το 2015.
- Αποτελούν μία κλάση στατιστικών παραγωγικών μοντέλων.
- Εμπνέονται από τις αρχές της θερμοδυναμικής (non-equilibrium thermodynamics).
- **Λειτουργία:** διαδοχική έγχυση θορύβου στα διαθέσιμα δεδομένα (forward diffusion) και προσπάθεια εκμάθησης της αντίστροφης διαδικασίας αποθορυβοποίησης (reverse diffusion).

Σχήμα. Διαδικασίες forward και reverse diffusion.

Ho, J., Jain, A., & Abbeel, P. (2020). Denoising diffusion probabilistic models. Advances in neural information processing systems, 33, 6840-6851.

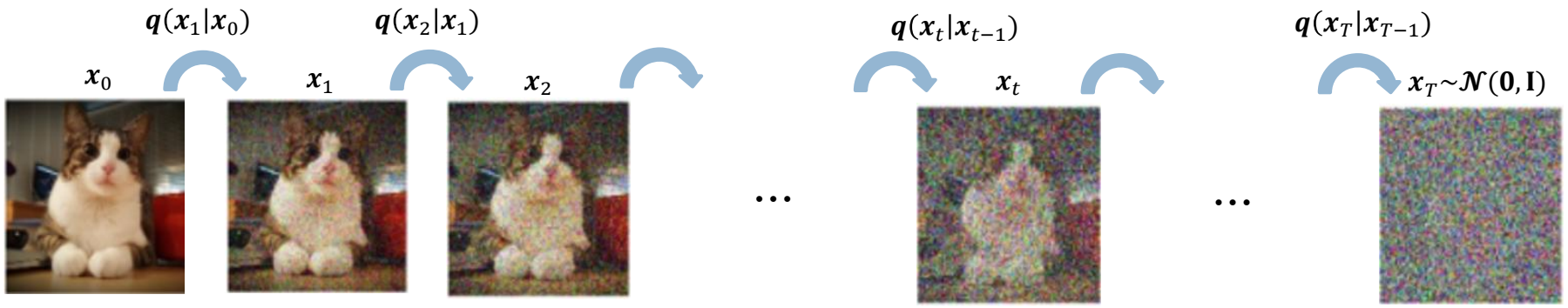
Diffusion Models

Διαδικασία Διάχυσης (Forward Diffusion)

- Διαδοχική προσθήκη θορύβου στα αρχικά δεδομένα μέχρι να καταστραφεί πλήρως η δομή τους → θόρυβος.
- Η forward diffusion μοντελοποιείται μέσω μιας Μαρκοβιανής αλυσίδας.
- Δοθέντος ενός δείγματος $\boldsymbol{x_0}$ από μια κατανομή $\boldsymbol{x_0} \sim \boldsymbol{q}(\boldsymbol{x}_0)$, κατά το forward diffusion προσθέτουμε διαδοχικά Γκαουσιανό θόρυβο στο δείγμα σε T βήματα, παράγοντας μια ακολουθία δειγμάτων θορύβου $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_T$.
- Χρησιμοποιείται ο πυρήνας μετάβασης $\boldsymbol{q}(\boldsymbol{x}_t|\boldsymbol{x}_{t-1})$, όπου,

$$\boldsymbol{q}(\boldsymbol{x}_t|\boldsymbol{x}_{t\text{-}1}) = \boldsymbol{\mathcal{N}}\left(\boldsymbol{x}_t; \sqrt{1-\beta_t}\boldsymbol{x}_{t\text{-}1}, \beta_t \mathbf{I}\right)$$

- Στην παραπάνω σχέση, το μέγεθος του βήματος ελέγχεται από μία παράμετρο, η οποία καλείται variance schedule $\{\beta_t \in (0,1)\}_{t=1}^{T}$.



Diffusion Models

Διαδικασία Διάχυσης (Forward Diffusion)

- Η από κοινού κατανομή των δειγμάτων $\boldsymbol{x}_1, \boldsymbol{x}_2, \dots, \boldsymbol{x}_T$, δεδομένου του αρχικού δείγματος $\boldsymbol{x}_0$, δίνεται από τη σχέση,

$$\boldsymbol{q}(\boldsymbol{x}_{1:T}|\boldsymbol{x}_0) = \prod_{t=1}^{T} \boldsymbol{q}(\boldsymbol{x}_t|\boldsymbol{x}_{t-1}),$$

  η οποία προκύπτει μέσω της Μαρκοβιανής ιδιότητας.

- Στην πράξη, ο πυρήνας $\boldsymbol{q}(\boldsymbol{x}_t|\boldsymbol{x}_{t-1})$ επιλέγεται ώστε να μετασχηματίζει την αρχική κατανομή των δεδομένων σε μία εκ των προτέρων γνωστή κατανομή → στην περίπτωσή μας κανονική κατανομή, $\boldsymbol{x}_T \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$.

- Αποδεικνύεται ότι:

Δειγματοληψία κατά το forward diffusion

Έστω $\boldsymbol{x}_t$ ένα δείγμα το οποίο δειγματοληπτείται από την κατανομή $\boldsymbol{q}(\boldsymbol{x}_t|\boldsymbol{x}_{t-1})$ στο βήμα t. Τότε το $\boldsymbol{x}_t$ μπορεί να εκφραστεί απευθείας, ως γραμμικός συνδυασμός του αρχικού δείγματος $\boldsymbol{x}_0$ και ενός διανύσματος θορύβου $\boldsymbol{\varepsilon}_t \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$, σύμφωνα με τη σχέση,

$$\boldsymbol{x}_t = \sqrt{\bar{\alpha}_t}\boldsymbol{x}_0 + \sqrt{1-\bar{\alpha}_t}\boldsymbol{\varepsilon}_t,$$

όπου, $\alpha_t = 1 - \beta_t$, $\bar{\alpha}_t = \prod_{i=1}^{t} \alpha_i$.

Diffusion Models

Διαδικασία Διάχυσης (Forward Diffusion)

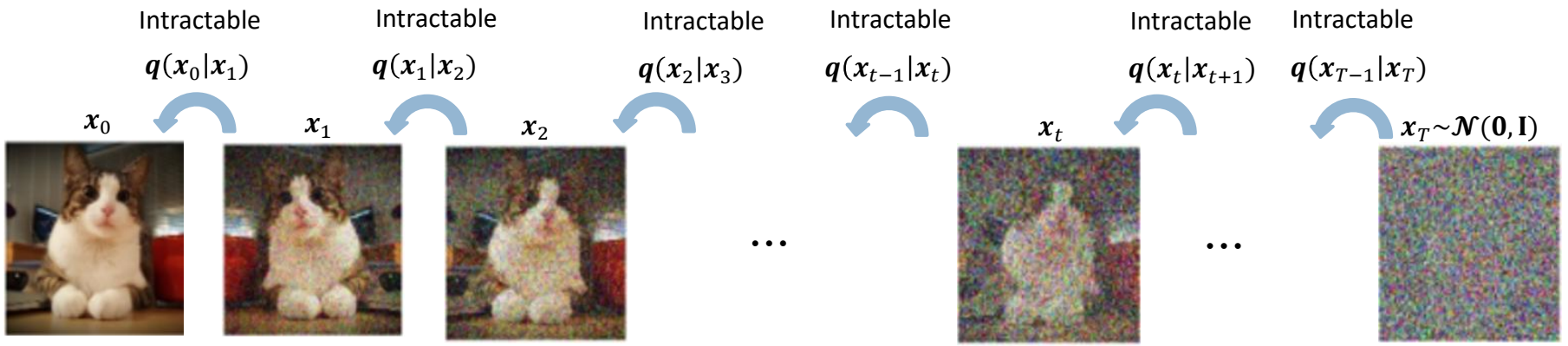
- Επομένως, οι κατανομές $\mathcal{N}(\boldsymbol{x}_t; \sqrt{\mathbf{1}-\beta_t}\boldsymbol{x}_{t-1}, \beta_t\mathbf{I})$ και $\mathcal{N}(\boldsymbol{x}_t; \sqrt{\bar{\alpha}_t}\boldsymbol{x}_0, (1-\bar{\alpha}_t)\mathbf{I})$ είναι ισοδύναμες.
- Αυτό επιτρέπει τη δειγματοληψία από οποιαδήποτε ενδιάμεση κατανομή κατά τη διάρκεια του forward diffusion, σε ένα μόνο βήμα.

Σχήμα. Η διάχυση καταστρέφει τη δομή.

Diffusion Models

Διαδικασία Αντίστροφης Διάχυσης (Reverse Diffusion)

- Αν αντιστρέψουμε τη διαδικασία του forward diffusion και κάνουμε δειγματοληψία από την κατανομή $\boldsymbol{q}(\boldsymbol{x}_{t-1}|\boldsymbol{x}_t)$, μπορούμε να συνθέσουμε νέα δείγματα, ξεκινώντας από Γκαουσιανό θόρυβο $\boldsymbol{x}_T \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$.



- **Το πρόβλημα:** η κατανομή $\boldsymbol{q}(\boldsymbol{x}_{t-1}|\boldsymbol{x}_t)$ δεν μπορεί **εκτιμηθεί** (intractable), γιατί απαιτεί υπολογισμούς με όλα τα διαθέσιμα δεδομένα. Πιο αναλυτικά,

$$\boldsymbol{q}(\boldsymbol{x}_{t-1}|\boldsymbol{x}_t) = \frac{\boldsymbol{q}(\boldsymbol{x}_t|\boldsymbol{x}_{t-1})\boldsymbol{q}(\boldsymbol{x}_{t-1})}{\boldsymbol{q}(\boldsymbol{x}_t)}, \text{ όπου } \boxed{\boldsymbol{q}(\boldsymbol{x}_t) = \int \boldsymbol{q}(\boldsymbol{x}_t|\boldsymbol{x}_0)\boldsymbol{q}(\boldsymbol{x}_0)\boldsymbol{d}\boldsymbol{x}_0} \text{.} \longrightarrow \text{Intractable}$$

Diffusion Models

Διαδικασία Αντίστροφης Διάχυσης (Reverse Diffusion)

- **Η λύση:** χρήση νευρωνικού δικτύου $\boldsymbol{p_\theta}(\boldsymbol{x}_{t-1}|\boldsymbol{x}_t)$ για την προσέγγιση των υπό συνθήκη κατανομών $\boldsymbol{q}(\boldsymbol{x}_{t-1}|\boldsymbol{x}_t)$, δηλαδή,

$$\boldsymbol{p_\theta}(\boldsymbol{x}_{t-1}|\boldsymbol{x}_t) = \boldsymbol{\mathcal{N}}\big(\boldsymbol{x}_{t-1}; \boldsymbol{\mu_\theta}(\boldsymbol{x}_t, t), \boldsymbol{\Sigma_\theta}(\boldsymbol{x}_t, t)\big),$$

  όπου η παράμετρος $\boldsymbol{\theta}$ υποδηλώνει τις παραμέτρους του δικτύου.
- **Πρόβλημα:** οι ποσότητες $\boldsymbol{\mu_\theta}(\boldsymbol{x}_t, t)$ και $\boldsymbol{\Sigma_\theta}(\boldsymbol{x}_t, t)$ δεν είναι γνωστές, ώστε να χρησιμοποιηθούν ως ground truth για την εκπαίδευση του δικτύου.
- **Υπενθύμιση:** η κατανομή $\boldsymbol{q}(\boldsymbol{x}_{t-1}|\boldsymbol{x}_t)$ δεν μπορεί **εκτιμηθεί** (intractable).
- Αποδεικνύεται, ότι εάν δεσμεύσουμε την κατανομή αυτή στο δείγμα $\boldsymbol{x_0}$, αυτή μπορεί πλέον να εκτιμηθεί. Δηλαδή,

$$\boldsymbol{q}(\boldsymbol{x}_{t-1}|\boldsymbol{x}_t, \boldsymbol{x}_0) = \boldsymbol{\mathcal{N}}\big(\boldsymbol{x}_{t-1}; \tilde{\boldsymbol{\mu}}_t(\boldsymbol{x}_t, \boldsymbol{x}_0), \tilde{\beta}_t \mathbf{I}\big).$$

- Εφαρμόζοντας τον κανόνα του Bayes για την κατανομή $\boldsymbol{q}(\boldsymbol{x}_{t-1}|\boldsymbol{x}_t, \boldsymbol{x}_0)$, λαμβάνουμε έπειτα από υπολογισμούς:

$$\tilde{\boldsymbol{\mu}}_t(\boldsymbol{x}_t, \boldsymbol{x}_0) = \frac{\sqrt{\alpha_t}(1-\bar{\alpha}_{t-1})}{1-\bar{\alpha}_t}\boldsymbol{x}_t + \frac{\sqrt{\bar{\alpha}_{t-1}}\beta_t}{1-\bar{\alpha}_t}\boldsymbol{x}_0 \text{ και } \tilde{\beta}_t = \frac{1-\bar{\alpha}_{t-1}}{1-\bar{\alpha}_t}\beta_t.$$

Diffusion Models

Διαδικασία Αντίστροφης Διάχυσης (Reverse Diffusion)

- Αν τώρα μετασχηματίσουμε την εξίσωση που χρησιμοποιούμε για τη δειγματοληψία κατά το forward diffusion, λαμβάνουμε,

$$\boldsymbol{x}_t = \sqrt{\bar{\alpha}_t}\boldsymbol{x}_0 + \sqrt{1-\bar{\alpha}_t}\boldsymbol{\varepsilon}_t \Rightarrow \boldsymbol{x}_0 = \frac{\boldsymbol{x}_t - \sqrt{1-\bar{\alpha}_t}\boldsymbol{\varepsilon}_t}{\sqrt{\bar{\alpha}_t}}.$$

- Αντικαθιστώντας την ποσότητα αυτή στη σχέση που υπολογίσαμε για τη μέση τιμή $\tilde{\boldsymbol{\mu}}_t(\boldsymbol{x}_t, \boldsymbol{x}_0)$ προκύπτει ότι,

$$\tilde{\boldsymbol{\mu}}_t(\boldsymbol{x}_t, \boldsymbol{x}_0) = \frac{1}{\sqrt{\alpha_t}}\left(\boldsymbol{x}_t - \frac{1-\alpha_t}{\sqrt{1-\bar{\alpha}_t}}\,\boldsymbol{\varepsilon}_t\right).$$

- Επομένως, αρκεί να εκπαιδεύσουμε το δίκτυό μας ώστε να προσεγγίζει τη μέση τιμή $\tilde{\boldsymbol{\mu}}_t(\boldsymbol{x}_t, \boldsymbol{x}_0)$.
- Ισοδύναμα, αρκεί να εκπαιδεύσουμε το δίκτυό μας ώστε να προσεγγίζει το θόρυβο $\boldsymbol{\varepsilon}_t$ για την εκάστοτε είσοδο $\boldsymbol{x}_t$, στο τυχόν βήμα t. Έπειτα, η μέση τιμή $\tilde{\boldsymbol{\mu}}_t(\boldsymbol{x}_t, \boldsymbol{x}_0)$ θα υπολογίζεται προσεγγιστικά μέσω της σχέσης,

$$\boldsymbol{\mu}_{\boldsymbol{\theta}}(\boldsymbol{x}_t, t) = \frac{1}{\sqrt{\alpha_t}}\left(\boldsymbol{x}_t - \frac{1-\alpha_t}{\sqrt{1-\bar{\alpha}_t}}\,\boldsymbol{\varepsilon}_{\boldsymbol{\theta}}(\boldsymbol{x}_t, t)\right),$$

όπου $\boldsymbol{\varepsilon}_{\boldsymbol{\theta}}(\boldsymbol{x}_t, t)$ η εκτίμηση του θορύβου $\boldsymbol{\varepsilon}_t$ στο βήμα t.

Diffusion Models

Δίκτυο Αποθορυβοποίησης

- Οι Ho et al. όρισαν τη συνάρτηση απώλειας $\boldsymbol{L}_t$ του δικτύου αποθορυβοποίησης, ώστε να ελαχιστοποιείται η διαφορά μεταξύ της εκτίμησης $\boldsymbol{\mu_\theta}(\boldsymbol{x}_t, t)$ και της μέσης τιμής $\tilde{\boldsymbol{\mu}}_t(\boldsymbol{x}_t, \boldsymbol{x}_0)$, σύμφωνα με τη σχέση,

$$\boldsymbol{L}_t = \mathbb{E}_{\mathbf{x}_0 \sim \mathbf{q}(\mathbf{x}_0), \boldsymbol{\varepsilon}_t \sim \boldsymbol{\mathcal{N}}(\mathbf{0}, \mathbf{I})} \left[\frac{1}{2\|\boldsymbol{\Sigma_\theta}(\boldsymbol{x}_t, t)\|_2^2} \|\tilde{\boldsymbol{\mu}}_t(\boldsymbol{x}_t, \boldsymbol{x}_0) - \boldsymbol{\mu_\theta}(\boldsymbol{x}_t, t)\|^2 \right]$$

$$= \mathbb{E}_{\mathbf{x}_0 \sim \mathbf{q}(\mathbf{x}_0), \boldsymbol{\varepsilon}_t \sim \boldsymbol{\mathcal{N}}(\mathbf{0}, \mathbf{I})} \left[\frac{(1 - a_t)^2}{2a_t(1 - \bar{a}_t)\|\boldsymbol{\Sigma_\theta}(\boldsymbol{x}_t, t)\|_2^2} \left\| \boldsymbol{\varepsilon}_t - \boldsymbol{\varepsilon_\theta}(\sqrt{\bar{\alpha}_t}\boldsymbol{x_0} + \sqrt{1 - \bar{\alpha}_t}\boldsymbol{\varepsilon}_t, t) \right\|^2 \right].$$

- Παρόλα αυτά, διαπίστωσαν ότι παραλείποντας τον πρώτο όρο στην παραπάνω σχέση, απλοποιείται σημαντικά η συνάρτηση απώλειας και βελτιώνεται η ποιότητα των παραγόμενων δειγμάτων:

$$\boldsymbol{L}_t^{simple} = \mathbb{E}_{t \sim \mathcal{U}[1,T], \mathbf{x}_0 \sim \mathbf{q}(\mathbf{x}_0), \boldsymbol{\varepsilon}_t \sim \boldsymbol{\mathcal{N}}(\mathbf{0}, \mathbf{I})} \left[\left\| \boldsymbol{\varepsilon}_t - \boldsymbol{\varepsilon_\theta}(\sqrt{\bar{\alpha}_t}\boldsymbol{x}_0 + \sqrt{1 - \bar{\alpha}_t}\boldsymbol{\varepsilon}_t, t) \right\|^2 \right].$$

Ho, J., Jain, A., & Abbeel, P. (2020). Denoising diffusion probabilistic models. Advances in neural information processing systems, 33, 6840-6851.

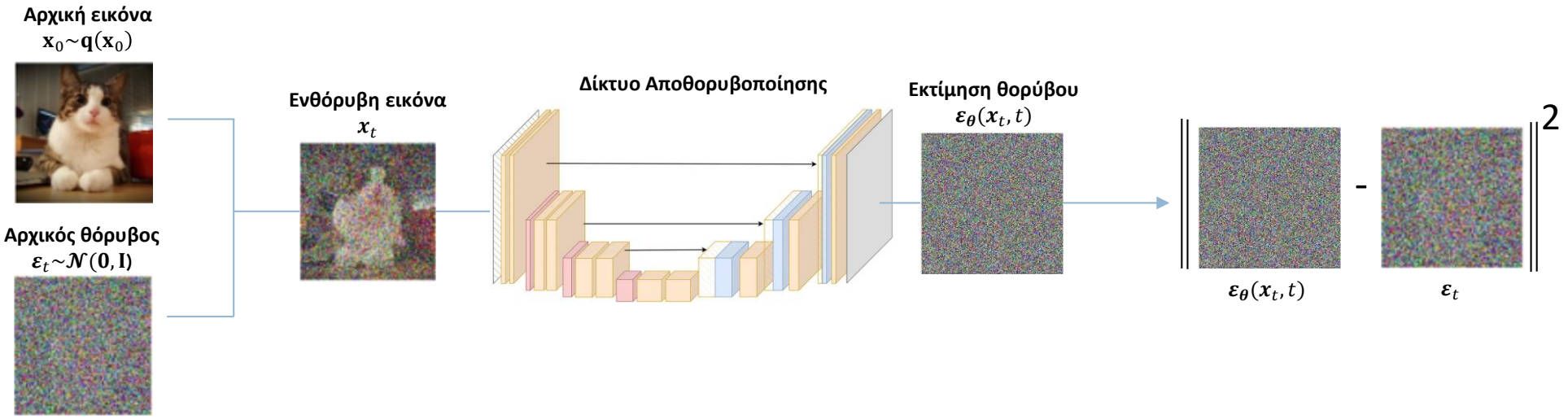
Diffusion Models

Δίκτυο Αποθορυβοποίησης

- **Διαδικασία εκπαίδευσης**

Algorithm 1 Training

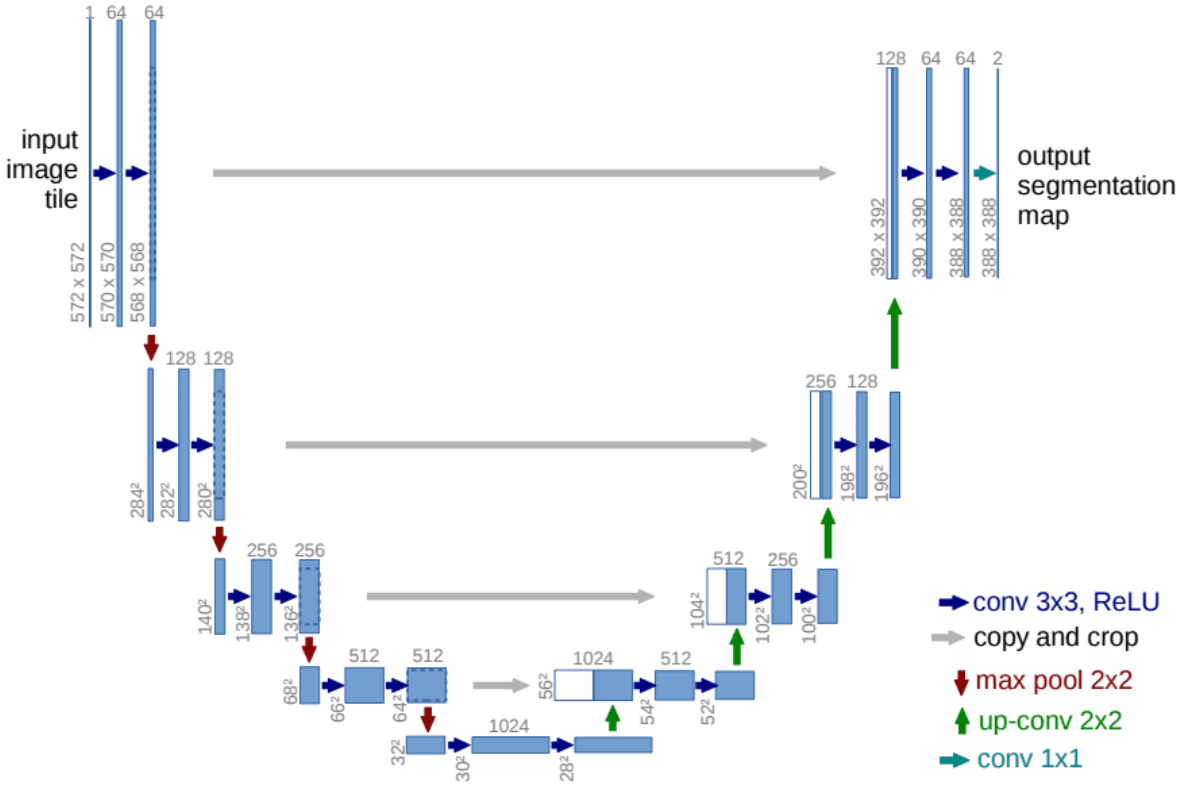
1: **repeat**
2: $\mathbf{x}_0 \sim q(\mathbf{x}_0)$
3: $t \sim \text{Uniform}(\{1, \dots, T\})$
4: $\boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$
5: Take gradient descent step on
$\nabla_\theta \left\| \boldsymbol{\epsilon} - \boldsymbol{\epsilon}_\theta(\sqrt{\bar{\alpha}_t}\mathbf{x}_0 + \sqrt{1-\bar{\alpha}_t}\boldsymbol{\epsilon}, t) \right\|^2$
6: **until** converged



Diffusion Models

Δίκτυο Αποθορυβοποίησης

- **Αρχιτεκτονική**
  - Η πιο συνηθισμένη αρχιτεκτονική του δικτύου αποθορυβοποίησης είναι αυτή του δικτύου U-Net, η οποία περιλαμβάνει εκτός από συνελικτικά layers και attention layers.
  - Οι πιο καινούριες αρχιτεκτονικές βασίζονται εξ ολοκλήρου σε δομές transformers, οι οποίες είναι ειδικά σχεδιασμένες για τις ανάγκες των diffusion μοντέλων (Diffusion transformers).



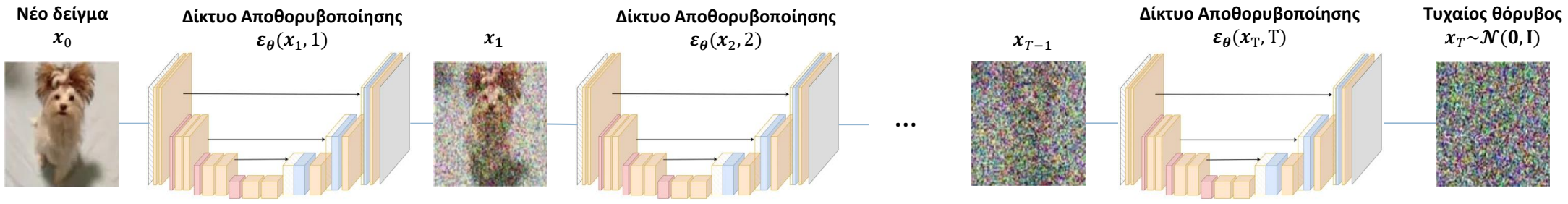
Diffusion Models

Δίκτυο Αποθορυβοποίησης

- **Σύνθεση νέων δειγμάτων**
  - Αξιοποιώντας το εκπαιδευμένο δίκτυο αποθορυβοποίησης $\boldsymbol{\varepsilon_\theta}(\boldsymbol{x}_t, t)$, μπορούμε να ακολουθήσουμε τη reverse diffusion διαδικασία βήμα προς βήμα και να παράξουμε νέα δεδομένα.

Algorithm 2 Sampling

1: $\mathbf{x}_T \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$
2: **for** $t = T, \ldots, 1$ **do**
3: $\mathbf{z} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ if $t > 1$, else $\mathbf{z} = \mathbf{0}$
4: $\mathbf{x}_{t-1} = \frac{1}{\sqrt{\alpha_t}} \left(\mathbf{x}_t - \frac{1-\alpha_t}{\sqrt{1-\bar{\alpha}_t}} \boldsymbol{\epsilon}_\theta(\mathbf{x}_t, t) \right) + \sigma_t \mathbf{z}$
5: **end for**
6: **return** $\mathbf{x}_0$



Diffusion Models

Καθοδηγούμενη Σύνθεση Εικόνων

- Καθοδήγηση της διαδικασίας σύνθεσης των εικόνων βάσει εξωτερικών συνθηκών (π.χ. κειμενικές περιγραφές).
- Ας υποθέσουμε ότι θέλουμε να παράξουμε νέα δείγματα βάσει της συνθήκης $\boldsymbol{y}$ → ετικέτα της κλάσης της εικόνας.
- Οι Sohl-Dickstein et al. και αργότερα οι Dhariwal και Nichol έδειξαν ότι μπορούμε να χρησιμοποιήσουμε ένα μοντέλο ταξινόμησης $\boldsymbol{p_\varphi}(\boldsymbol{y}|\boldsymbol{x}_t, t)$, για να καθοδηγήσουμε τη διαδικασία του reverse diffusion προς την επιθυμητή κλάση $\boldsymbol{y}$.
- Για να το επιτύχουμε αυτό, εκπαιδεύουμε τον ταξινομητή $\boldsymbol{p_\varphi}(\boldsymbol{y}|\boldsymbol{x}_t, t)$ στις θορυβώδεις εικόνες $\boldsymbol{x}_t$, ώστε να προβλέπει τις κλάσεις τους $\boldsymbol{y}$.
- Έπειτα, αντί να λαμβάνουμε δείγματα από την κατανομή $\mathcal{N}\left(\boldsymbol{\mu_\theta}(\boldsymbol{x}_t, t), \boldsymbol{\Sigma_\theta}(\boldsymbol{x}_t, t)\right)$, λαμβάνουμε δείγματα από την κατανομή $\mathcal{N}\left(\hat{\boldsymbol{\mu}}(\boldsymbol{x}_t|\boldsymbol{y}, t), \boldsymbol{\Sigma_\theta}(\boldsymbol{x}_t, t)\right)$, όπου,

$$\hat{\boldsymbol{\mu}}(\boldsymbol{x}_t|\boldsymbol{y}, t) = \boldsymbol{\mu_\theta}(\boldsymbol{x}_t, t) + s\boldsymbol{\Sigma_\theta}(\boldsymbol{x}_t, t)\boldsymbol{\nabla}_{\boldsymbol{x}_t}\log \boldsymbol{p_\varphi}(\boldsymbol{y}|\boldsymbol{x}_t, t).$$

- Η μέθοδος αυτή καλείται **Classifier Guidance**.
- Αν και αποτελεσματική, απαιτεί την εκπαίδευση ενός επιπλέον μοντέλου ταξινόμησης στο ίδιο σύνολο δεδομένων → Πολύ υψηλό υπολογιστικό κόστος.

Dhariwal, P., & Nichol, A. (2021). Diffusion models beat GANs on image synthesis. arXiv. https://arxiv.org/abs/2105.05233

Diffusion Models

Καθοδηγούμενη Σύνθεση Εικόνων

- Για να αντιμετωπίσουν το πρόβλημα αυτό, οι Ho et al. απέδειξαν ότι η καθοδήγηση μέσω της συνθήκης $\boldsymbol{y}$ μπορεί να υλοποιηθεί και χωρίς τη χρήση του εξωτερικού ταξινομητή $\boldsymbol{p_\varphi}(\boldsymbol{y}|\boldsymbol{x}_t, t)$.
- **Λύση:** από κοινού εκπαίδευση ενός **υπό συνθήκη** δικτύου αποθορυβοποίησης $\boldsymbol{\varepsilon_\theta}(\boldsymbol{x}_t, \boldsymbol{y}, t)$ και ενός **απλού** δικτύου αποθορυβοποίησης $\boldsymbol{\varepsilon_\theta}(\boldsymbol{x}_t, \mathbf{y} = \boldsymbol{\emptyset}, t)$. Στην πραγματικότητα πρόκειται για **το ίδιο δίκτυο.**
- Η εκπαίδευση πραγματοποιείται με μικτό τρόπο → τυχαία επιλογή μεταξύ ετικετών $\boldsymbol{y}$ και $\boldsymbol{y} = \boldsymbol{\emptyset}$, ούτως ώστε το μοντέλο να εκτεθεί τόσο στην υπό συνθήκη όσο και στην χωρίς συνθήκη παραγωγή.
- Για τη παραγωγή νέων δειγμάτων ακολουθείται και πάλι η ίδια διαδικασία (Αλγόριθμος 2), με τη διαφορά ότι:

$$\tilde{\boldsymbol{\varepsilon}}_{\boldsymbol{\theta}}(\boldsymbol{x}_t, \boldsymbol{y}, t) = \boldsymbol{\varepsilon}_{\boldsymbol{\theta}}(\boldsymbol{x}_t, \mathbf{y} = \boldsymbol{\emptyset}, t) + w\big(\boldsymbol{\varepsilon}_{\boldsymbol{\theta}}(\boldsymbol{x}_t, \boldsymbol{y}, t) - \boldsymbol{\varepsilon}_{\boldsymbol{\theta}}(\boldsymbol{x}_t, \mathbf{y} = \boldsymbol{\emptyset}, t)\big),$$

όπου w μία παράμετρος η οποία καλείται **guidance scale** και η οποία ορίζει το βαθμό της επίδρασης της συνθήκης $\boldsymbol{y}$ στη διαδικασία της σύνθεσης.

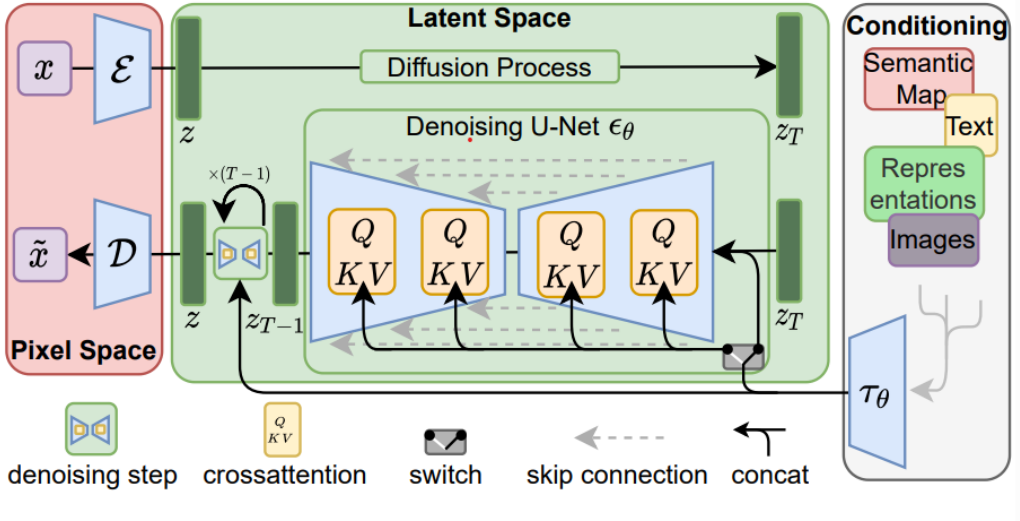
- **Ερώτημα:** πώς η συνθήκη $\boldsymbol{y}$ εισέρχεται στο δίκτυο αποθορυβοποίησης $\boldsymbol{\varepsilon_\theta}$;

Ho, J., & Salimans, T. (2022). Classifier-free diffusion guidance. arXiv. https://arxiv.org/abs/2207.12598

Diffusion Models

Stable Diffusion

- Πρόκειται για ένα latent μοντέλο διάχυσης → διαδικασία διάχυσης και αντίστροφης διάχυσης σε latent χώρο.
- Χρησιμοποιεί ένα δίκτυο αποθορυβοποίησης $\boldsymbol{\varepsilon_\theta(z_t, y, t)}$ τύπου U-Net, όπου $\boldsymbol{z_t = \mathcal{E}(x)}$ και $\boldsymbol{x \sim q(x_0)}$.
- Κάθε block του U-Net αποτελείται από συνελικτικά layers ακολουθούμενα από self-attention και cross-attention layers.
- Η κειμενική καθοδήγηση $\boldsymbol{y}$ εισέρχεται στη διαδικασία της διάχυσης μέσω των cross-attention layers.



Rombach, R., Blattmann, A., Lorenz, D., Esser, P., & Ommer, B. (2022). High-resolution image synthesis with latent diffusion models. arXiv. https://arxiv.org/abs/2112.10752

Diffusion Models

Denoising Diffusion Implicit Models

- Στα παραδοσιακά μοντέλα διάχυσης (DDPMs) η διαδικασία του forward diffusion μπορεί να περιλαμβάνει αρκετές χιλιάδες βήματα.
- Για να παράξουμε νέα δείγματα, θα πρέπει να ακολουθήσουμε αντίστροφα όλα τα βήματα αυτά → υψηλοί χρόνοι εκτέλεσης.
- **Λύση:** τα Denoising Diffusion Implicit Models, τα οποία επεκτείνουν την ιδέα των παραδοσιακών μοντέλων διάχυσης σε μη-Μαρκοβιανές περιπτώσεις.

Εκπαίδευση	Δειγματοληψία
Μεγάλος αριθμός χρονικών βημάτων T (συνήθως μερικές χιλιάδες)	Υποσύνολο των βημάτων που χρησιμοποιούνται κατά την εκπαίδευση

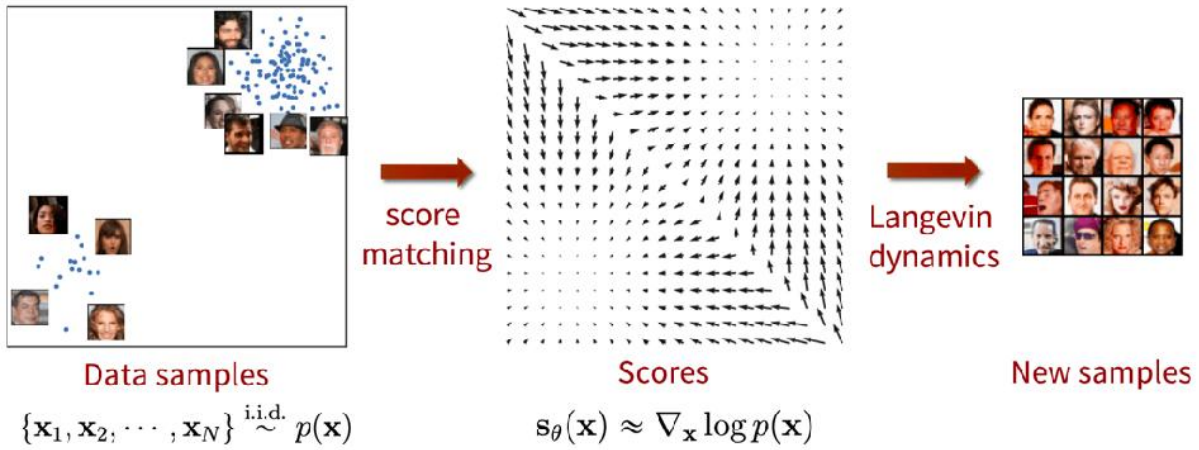
Song, J., Meng, C., & Ermon, S. (2022). Denoising diffusion implicit models. arXiv. https://arxiv.org/abs/2010.02502

Diffusion Models

Score-Based Generative Models

- Η βασική ιδέα των μοντέλων αυτών είναι η χρήση ενός νευρωνικού δικτύου $\boldsymbol{s_\theta}$ για την εκτίμηση της **score function** και η δημιουργία νέων δειγμάτων με χρήση score-based μεθόδων δειγματοληψίας.
- Η score function μιας πυκνότητας πιθανότητας $\boldsymbol{p(x)}$ ορίζεται ως $\boldsymbol{\nabla_x \log p(x)}$.
- Πρόκειται ουσιαστικά για ένα διανυσματικό πεδίο, το οποίο δείχνει προς την κατεύθυνση εκείνη στην οποία μεγιστοποιείται η λογαριθμική πυκνότητα των δεδομένων.
- Το νευρωνικό δίκτυο εκπαιδεύεται ώστε να προσεγγίζει την ποσότητα αυτή, δηλαδή $\boldsymbol{s_\theta(x) \approx \nabla_x \log p(x)}$, σύμφωνα με τη συνάρτηση απώλειας,

$$\mathbb{E}_{\boldsymbol{p(x)}}\left[\|\boldsymbol{\nabla_x \log p(x) - s_\theta(x)}\|_2^2\right]$$



Diffusion Models

Score-Based Generative Models

- **Δειγματοληψία με Langevin Dynamics**
  - Αφού έχουμε εκπαιδεύσει το δίκτυο $\boldsymbol{s_\theta}$, μπορούμε να παράξουμε νέα δείγματα χρησιμοποιώντας μια Markov Chain Monte Carlo μέθοδο δειγματοληψίας, τα **Langevin Dynamics.**
  - Η μέθοδος αυτή, επιτρέπει τη λήψη δειγμάτων από μία κατανομή $\boldsymbol{p(x)}$, χρησιμοποιώντας αποκλειστικά τη score function $\boldsymbol{\nabla_x \log p(x)}$.
  - Πιο συγκεκριμένα, ξεκινώντας από ένα δείγμα $\boldsymbol{x_0 \sim \pi(x)}$, όπου $\boldsymbol{\pi(x)}$ μια prior κατανομή και με βήμα $\epsilon > 0$, η μέθοδος των Langevin Dynamics υπολογίζει αναδρομικά, για *Τ βήματα,* το ακόλουθο:

$$\tilde{\boldsymbol{x}}_t = \tilde{\boldsymbol{x}}_{t-1} + \frac{\boldsymbol{\epsilon}}{2}\boldsymbol{\nabla_x \log p}(\tilde{\boldsymbol{x}}_{t-1}) + \sqrt{\boldsymbol{\epsilon}}\boldsymbol{z}_t$$

όπου $\boldsymbol{z_t \sim \mathcal{N}(0, \mathrm{I})}$.

  - Όταν $\epsilon \to 0$ και $T \to \infty$, η κατανομή των δειγμάτων $\tilde{\boldsymbol{x}}_T$ προσεγγίζει την κατανομή $\boldsymbol{p(x)}$.
  - Στην πράξη, επιλέγουμε ένα αρκετά μικρό ϵ και ένα αρκετά μεγάλο T.

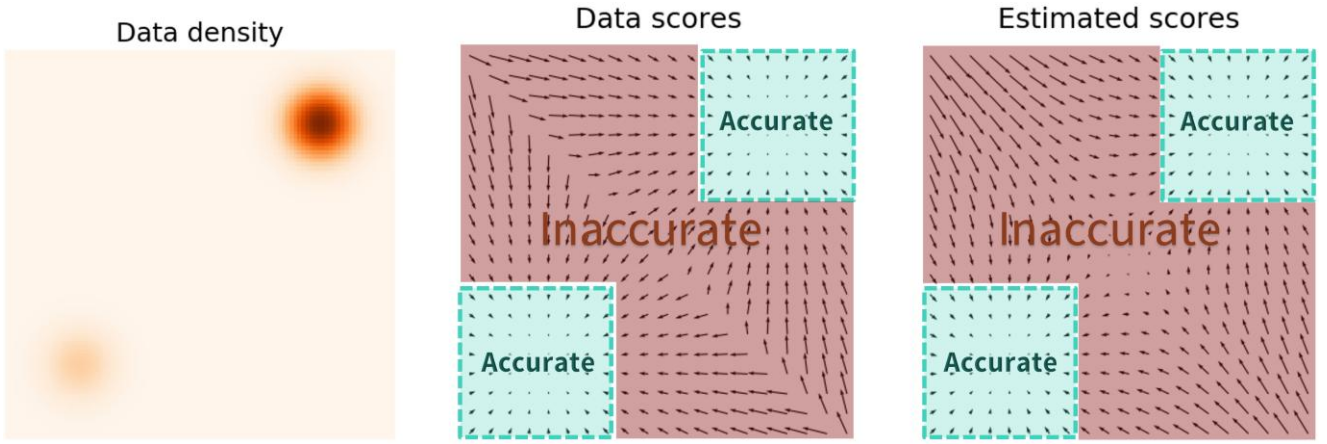
Diffusion Models

Score-Based Generative Models

- **Περιοχές Χαμηλής Πυκνότητας Δεδομένων (Low Data Density Regions)**
  - Το βασικό πρόβλημα των score-based generative μοντέλων είναι ότι η προσέγγιση της score function δεν είναι ακριβής σε περιοχές χαμηλής συγκέντρωσης των δεδομένων.
  - Τα score-based generative μοντέλα εκπαιδεύονται βάσει της συνάρτησης απώλειας,

$$\mathbb{E}_{\boldsymbol{p}(\mathbf{x})}\left[\|\boldsymbol{\nabla}_{\boldsymbol{x}}\log \boldsymbol{p}(\boldsymbol{x})-\boldsymbol{s}_{\boldsymbol{\theta}}(\boldsymbol{x})\|_2^2\right] = \int \boldsymbol{p}(\boldsymbol{x})\ \|\boldsymbol{\nabla}_{\boldsymbol{x}}\log \boldsymbol{p}(\boldsymbol{x})-\boldsymbol{s}_{\boldsymbol{\theta}}(\boldsymbol{x})\|_2^2 \boldsymbol{dx}.$$

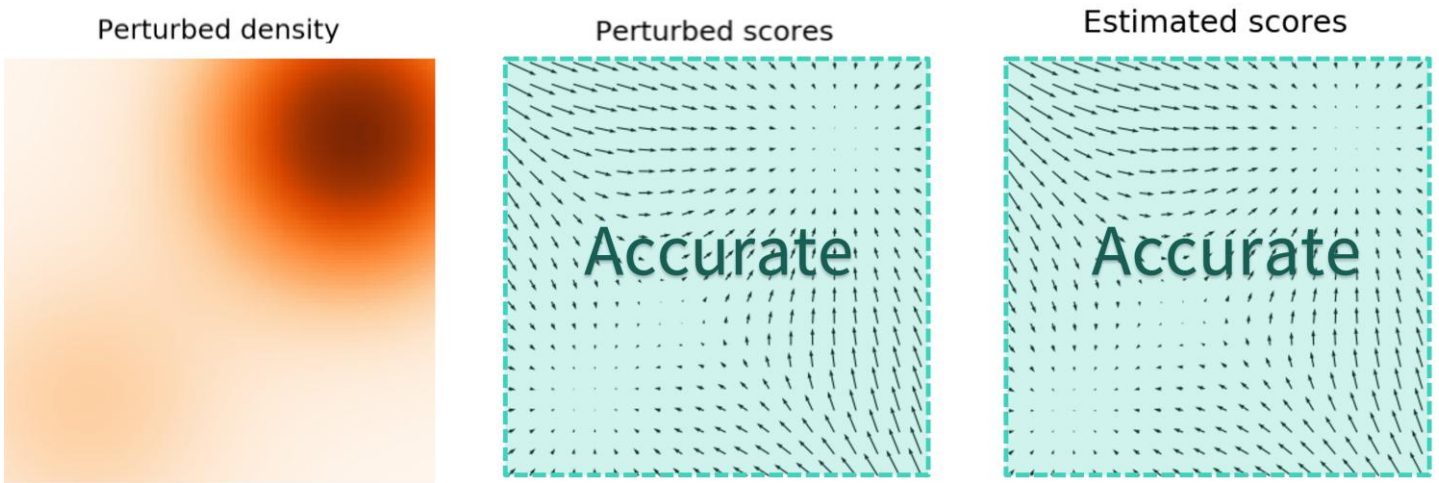
  - Επομένως, οι περιοχές χαμηλής συγκέντρωσης αμελούνται κατά την εκπαίδευση, λόγω του όρου $\boldsymbol{p}(\boldsymbol{x})$.



Diffusion Models

Noise Conditional Score Networks

- Αποτελούν είδος score-based generative μοντέλων, τα οποία αναπτύχθηκαν για να αντιμετωπιστεί το πρόβλημα των low data density regions.
- Χρησιμοποιούν ένα νευρωνικό δίκτυο για την εκτίμηση της score function της κατανομής των δεδομένων **έπειτα από την προσθήκη θορύβου.**
- Με την προσθήκη του θορύβου ομαλοποιείται η πυκνότητα των περιοχών και αποφεύγονται οι περιοχές χαμηλής πυκνότητας.



Diffusion Models

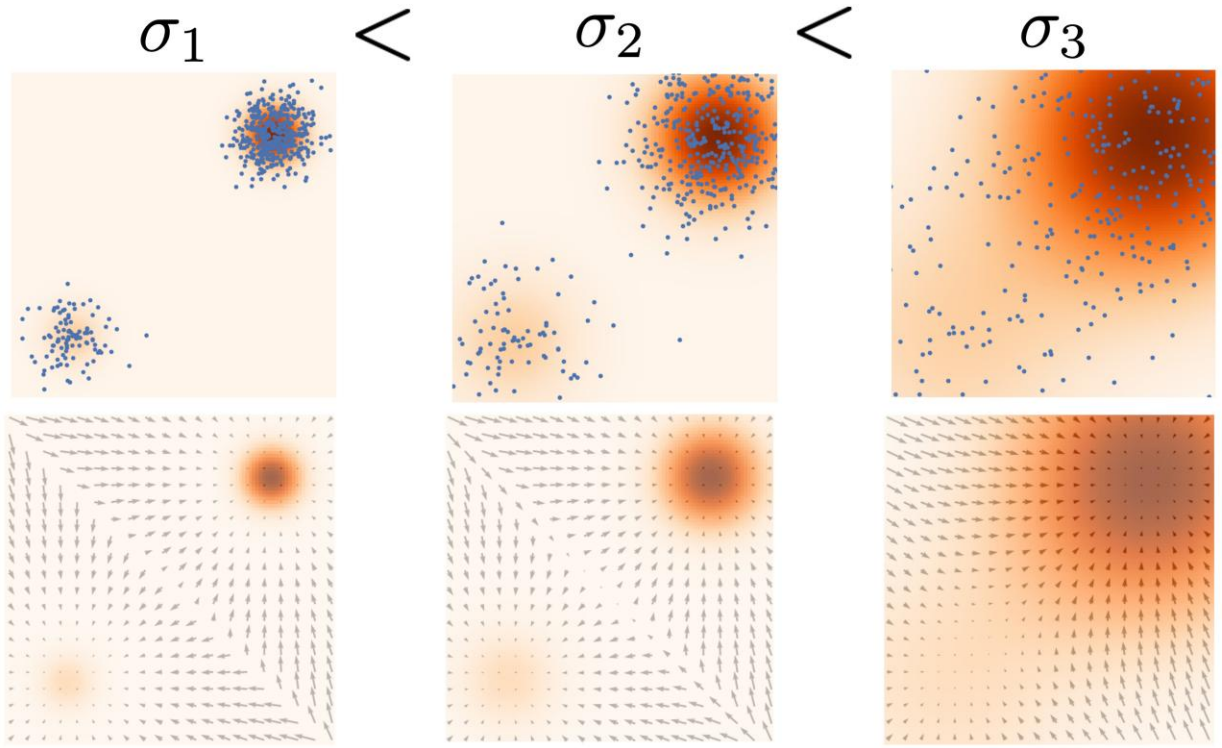
Noise Conditional Score Networks

- Έστω $\sigma_1 < \sigma_2 < \cdots < \sigma_L$ μια ακολουθία επιπέδων θορύβου.
- Προσθέτουμε Γκαουσιανό θόρυβο $\boldsymbol{\mathcal{N}}\left(\mathbf{0}, \sigma_i^2 \mathbf{I}\right), i = 1,2,\ldots,L$ στα δεδομένα και λαμβάνουμε την ενθόρυβη κατανομή,

$$\boldsymbol{p}_{\boldsymbol{\sigma}_i}(\boldsymbol{x}) = \int \boldsymbol{p}(\boldsymbol{y})\boldsymbol{p}(\boldsymbol{x}|\boldsymbol{y})\mathrm{d}\boldsymbol{y}$$

$$= \int \boldsymbol{p}(\boldsymbol{y})\boldsymbol{\mathcal{N}}\left(\boldsymbol{x}; \boldsymbol{y}, \sigma_i^2 \mathbf{I}\right)\mathrm{d}\boldsymbol{y}.$$

- Εκπαιδεύουμε ένα Noise Conditional score-based νευρωνικό δίκτυο $\boldsymbol{s}_{\boldsymbol{\theta}}(\boldsymbol{x}, \sigma_i)$, το προσεγγίζει αυτή τη score function, δηλαδή $\boldsymbol{s}_{\boldsymbol{\theta}}(\boldsymbol{x}, \sigma_i) \approx \boldsymbol{\nabla}_{\boldsymbol{x}} \log \boldsymbol{p}_{\sigma_i}(\boldsymbol{x})$, για κάθε $i = 1,2,\ldots,L$.



Diffusion Models

Noise Conditional Score Networks

- Σε κάθε επίπεδο θορύβου, η αντικειμενική συνάρτηση ισούται με,

$$\ell(\boldsymbol{\theta};\sigma) = \frac{1}{2}\mathbb{E}_{\boldsymbol{x}\sim\boldsymbol{p}(\boldsymbol{x})}\mathbb{E}_{\tilde{\boldsymbol{x}}\sim\mathcal{N}(\boldsymbol{x},\sigma^2\mathbf{I})}\left[\left\|\boldsymbol{s}_{\boldsymbol{\theta}}(\tilde{\boldsymbol{x}},\sigma) + \frac{\tilde{\boldsymbol{x}}-\boldsymbol{x}}{\sigma^2}\right\|_2^2\right]$$

- Συνδυάζοντας την παραπάνω συνάρτηση για κάθε $\sigma \in \{\sigma_i\}_{i=1}^{L}$, λαμβάνουμε την ολική αντικειμενική συνάρτηση εκπαίδευσης του δικτύου $\boldsymbol{s}_{\boldsymbol{\theta}}(\boldsymbol{x},\sigma)$:

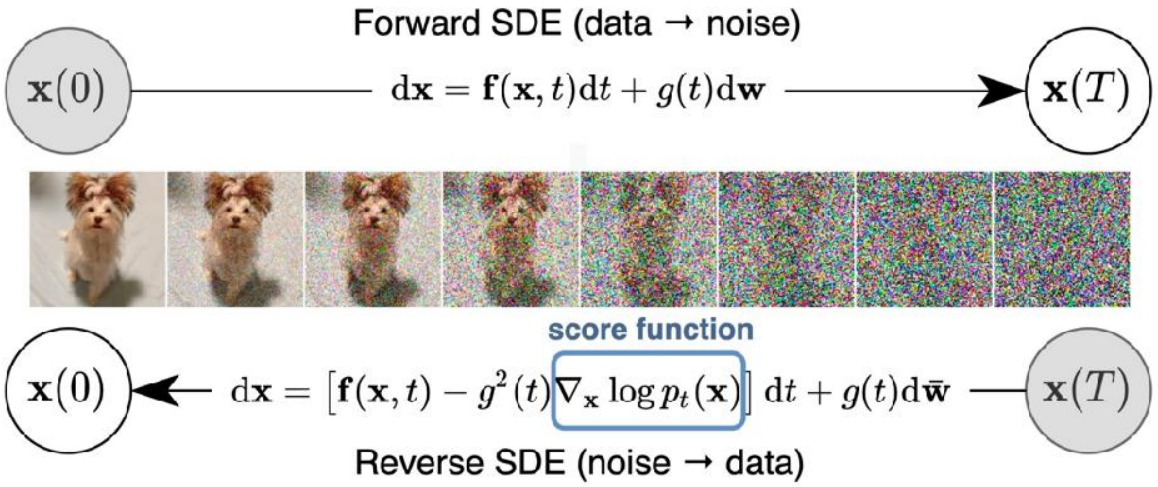
$$\mathcal{L}\left(\boldsymbol{\theta},\{\sigma_i\}_{i=1}^{L}\right) = \frac{1}{L}\sum_{i=1}^{L}\lambda(\sigma_i)\ell(\boldsymbol{\theta};\sigma_i),$$

όπου $\lambda(\cdot)$ είναι ένας συντελεστής, ο οποίος συχνά επιλέγεται ίσος με $\lambda(\sigma) = \sigma^2$.

Diffusion Models

Stochastic Differential Equations

- Τόσο τα παραδοσιακά μοντέλα διάχυσης, όσο και τα score-based generative μοντέλα, μπορούν να θεωρηθούν ως διακριτοποιήσεις στοχαστικών διαφορικών εξισώσεων.
- Στις έως τώρα προσεγγίσεις, χρησιμοποιείται ένα πεπερασμένο πλήθος επιπέδων θορύβου.
- Μπορούμε να επεκτείνουμε τη λογική αυτή και να χρησιμοποιήσουμε ένα συνεχές φάσμα κατανομών θορύβου για να προσθέσουμε θόρυβο στα δεδομένα.
- Οι κατανομές των ενθόρυβων δεδομένων, εξελίσσονται χρονικά σύμφωνα με μία Στοχαστική Διαφορική Εξίσωση:



Diffusion Models

Stochastic Differential Equations

- Παρατηρούμε ότι υπάρχουν αρκετές αναλογίες μεταξύ των παραδοσιακών μοντέλων διάχυσης, των Score-based Generative μοντέλων και των SDEs.
  - Τα διακριτά βήματα $t = 1, 2, \dots, T$, γενικεύονται σε βήματα στο συνεχές φάσμα t ϵ [0, T].
  - Τα διακριτά επίπεδα θορύβου που ελέγχονται μέσω της παραμέτρου β_t, αντιστοιχούν στο συντελεστή διάχυσης $g(t)$ της SDE.
- Προκειμένου τώρα να παράξουμε νέα δείγματα $\boldsymbol{x(0)} \sim \boldsymbol{p_0}$, μπορούμε, ξεκινώντας από ένα δείγμα $\boldsymbol{x}(T) \sim \boldsymbol{p}_T$ να ακολουθήσουμε την αντίστροφη SDE,

$$\boldsymbol{dx} = [\boldsymbol{f}(\boldsymbol{x}, t) - \boldsymbol{g}(t)^2 \boldsymbol{\nabla}_{\boldsymbol{x}} \log_t \boldsymbol{p}(\boldsymbol{x})]\boldsymbol{dt} + \boldsymbol{g}(t) d\overline{\boldsymbol{w}}.$$

- Για να κάνουμε δειγματοληψία από την κατανομή $\boldsymbol{p}_0$, θα πρέπει να προσομοιώσουμε την αντίστροφη SDE, η οποία προϋποθέτει την εκτίμηση της score function $\boldsymbol{\nabla}_{\boldsymbol{x}} \log_t \boldsymbol{p}(\boldsymbol{x})$.
- Η score function αυτή είναι προφανώς η ίδια με αυτή της περίπτωσης των Score-based Generative Μοντέλων.

Diffusion Models

Περιορισμοί

- Υψηλό Υπολογιστικό Κόστος
  - Ανάγκη για πληθώρα υπολογιστικών πόρων → δύσκολη υλοποίηση σε πραγματικό χρόνο
  - Αργή και ιδιαίτερα απαιτητική σε hardware εκπαίδευση
- Δυσκολία Γενίκευσης σε Άγνωστα Δεδομένα
  - Ανάγκη επανεκπαίδευσης και προσαρμογής
- Υψηλό Επίπεδο Πολυπλοκότητας
  - Αδυναμία ενσωμάτωσης και χρήσης σε εφαρμογές οι οποίες απαιτούν γνώση του ακριβούς τρόπου λειτουργίας τους.
- Ελλιπής ικανότητα ερμηνείας των αποτελεσμάτων
  - Προκλήσεις σε εφαρμογές ερμηνεύσιμης τεχνητής νοημοσύνης (Explainable AI)
- Πολωμένα Αποτελέσματα
  - Οι παραγόμενες εικόνες μπορεί να παρουσιάζουν απόκλιση από τις κειμενικές περιγραφές, λόγω έλλειψης ισορροπίας στα δεδομένα εκπαίδευσης.

"A group of watches showing 5 minutes past 12"

Παραγωγή με χρήση του Stable Diffusion

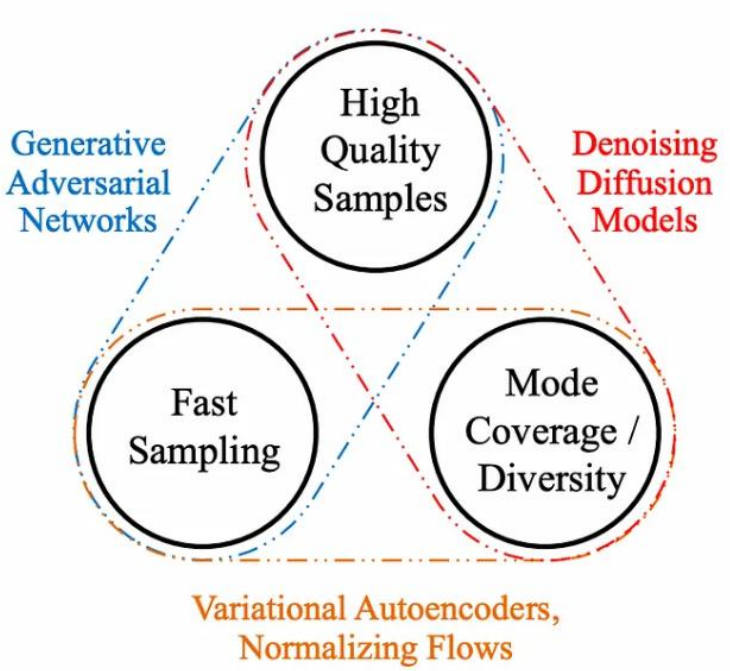
"A left-handed person writing down on a notebook"

Παραγωγή με χρήση του Stable Diffusion

Diffusion Models

Generative Models Trilemma

- Βασικοί στόχοι ενός παραγωγικού μοντέλου:
  1) Παραγωγή υψηλής ποιότητας δειγμάτων,
  2) Παραγωγή diverse δειγμάτων και
  3) Γρήγορη παραγωγή δειγμάτων με χαμηλό υπολογιστικό κόστος.
- Δεν υπάρχει κάποιο παραγωγικό framework που να ικανοποιεί και τα 3 κριτήρια.
- Variational Autoencoders (VAEs) → χαμηλή ποιότητα παραγόμενων δειγμάτων.
- Generative Adversarial Networks (GANs) → περιορισμένη ποικιλία στα παραγόμενα δείγματα.
- Diffusion Models → υψηλό υπολογιστικό κόστος (μεγάλος αριθμός βημάτων για παραγωγή νέων δειγμάτων).



Διάρθρωση της Παρουσίασης

- Εισαγωγή
- Diffusion Models
  - Ιστορική Εξέλιξη
  - Διαδικασία Διάχυσης (Forward Diffusion)
  - Διαδικασία Αντίστροφης Διάχυσης (Reverse Diffusion)
  - Δίκτυο Αποθορυβοποίησης
  - Καθοδηγούμενη Σύνθεση Εικόνων
  - Stable Diffusion
  - Denoising Diffusion Implicit Models
  - Score-Based Generative Models
  - Noise Conditional Score Networks
  - Stochastic Differential Equations
  - Περιορισμοί
  - Generative Models Trilemma
- ***Εφαρμογές***

Εφαρμογές

Text-to-Image Synthesis

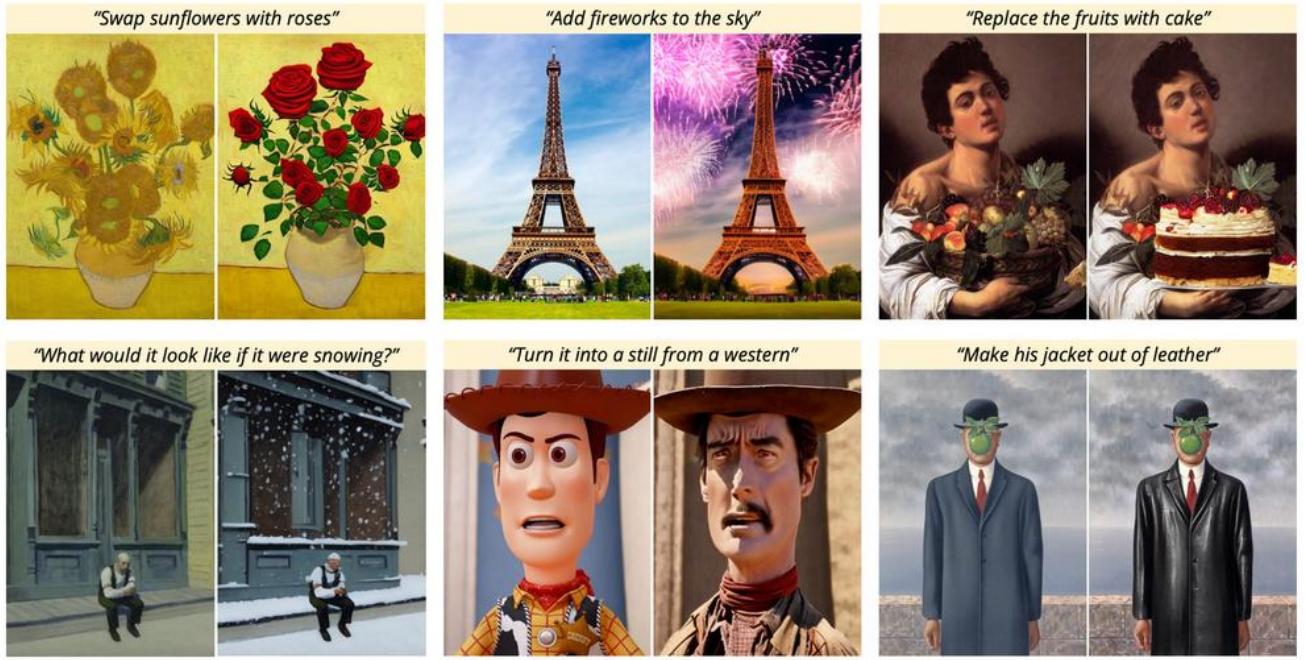
- Τα μοντέλα διάχυσης μπορούν να χρησιμοποιηθούν για την **παραγωγή εικόνων** από κειμενικές περιγραφές.
- **Γνωστά μοντέλα:**
  - *DALL-E 2, DALL-E 3*
  - *Stable Diffusion*
  - *MidJourney*
  - *Imagen*

"a hedgehog using a calculator" | "a corgi wearing a red bowtie and a purple party hat" | "robots meditating in a vipassana retreat" | "a fall landscape with a small cottage next to a lake"

Εφαρμογές

Image Editing

- Τα μοντέλα διάχυσης μπορούν να χρησιμοποιηθούν για την **επεξεργασία εικόνων**, διατηρώντας το φωτορεαλισμό και την ευκρίνειά τους.
  - Αλλαγή εκφράσεων προσώπου
  - Προσθήκη/αφαίρεση αντικειμένων
  - Αλλαγή χρωμάτων κ.ά.
- **Γνωστά μοντέλα:**
  - *DreamBooth*
  - *Stable Diffusion (+ ControlNet)*
  - *Instruct Pix2Pix*

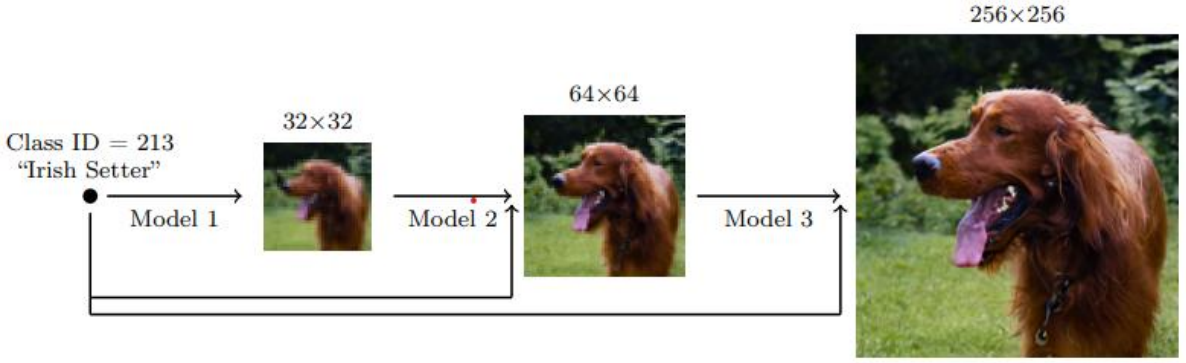


Brooks, T., Holynski, A., & Efros, A. A. (2023). *InstructPix2Pix: Learning to follow image editing instructions.*

Εφαρμογές

Image Super-Resolution

- Στο task του Image Super-Resolution θέλουμε να αυξήσουμε την ανάλυση μιας εικόνας χαμηλής ανάλυσης.
- Ιδιαίτερα χρήσιμο για τη βελτίωση της ανάλυσης παλιών φωτογραφιών και φωτογραφιών από πλάνα ασφαλείας.
- Οι Ho et al. 2022 πρότειναν μια αρχιτεκτονική πολλαπλών σταδίων, στην οποία μια σειρά από μοντέλα διάχυσης χρησιμοποιούνται για να αυξήσουν διαδοχικά την ανάλυση των εικόνων.
- Ξεκινώντας με ένα βασικό μοντέλο, το οποίο παράγει μια εικόνα σε χαμηλή ανάλυση, ακολουθούν upsampling μοντέλα, τα οποία αυξάνουν προοδευτικά την ανάλυση και το επίπεδο των λεπτομερειών.



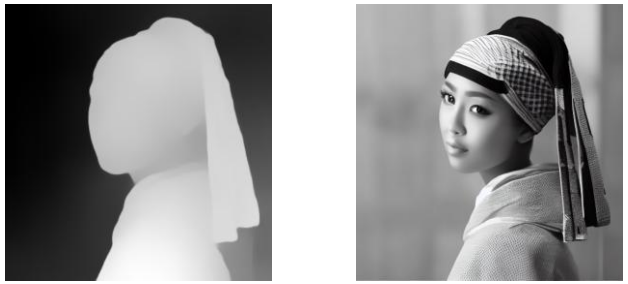
Ho, J., Saharia, C., Chan, W., Fleet, D. J., Norouzi, M., & Salimans, T. (2022). Cascaded diffusion models for high fidelity image generation. *Journal of Machine Learning Research, 23*(47), 1–33.

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
- **Depth-to-Image Generation**
  - Μετατροπή χάρτη βάθους (depth map) σε εικόνα.
  - Χάρτης βάθους → Εικόνα σε γκρίζα κλίμακα. Τα μαύρα αναπαριστούν περιοχές μεγάλου και τα άσπρα περιοχές χαμηλού βάθους.

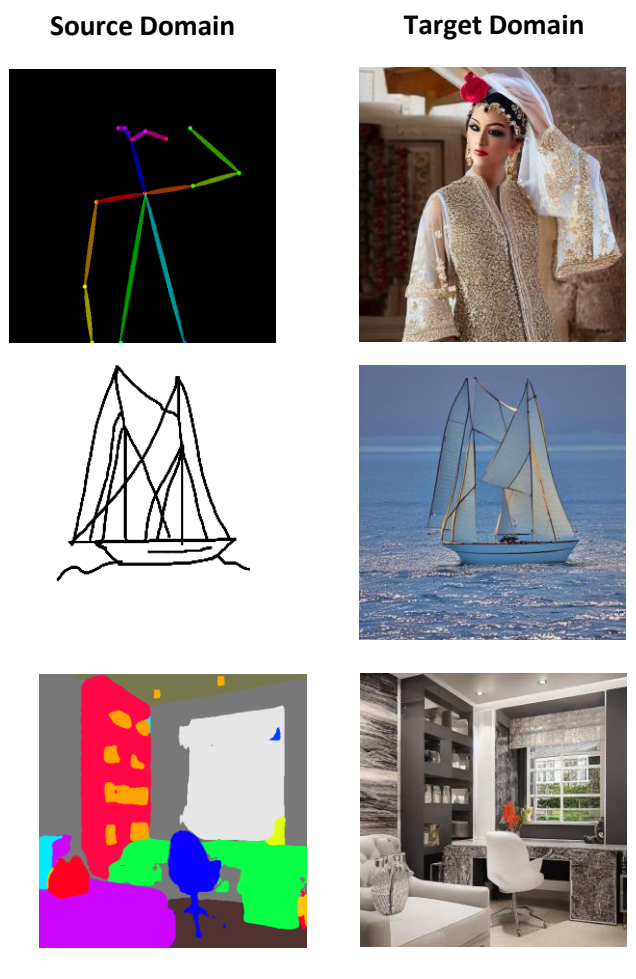
Source Domain **Target Domain**



Εφαρμογές

Image-to-Image Translation

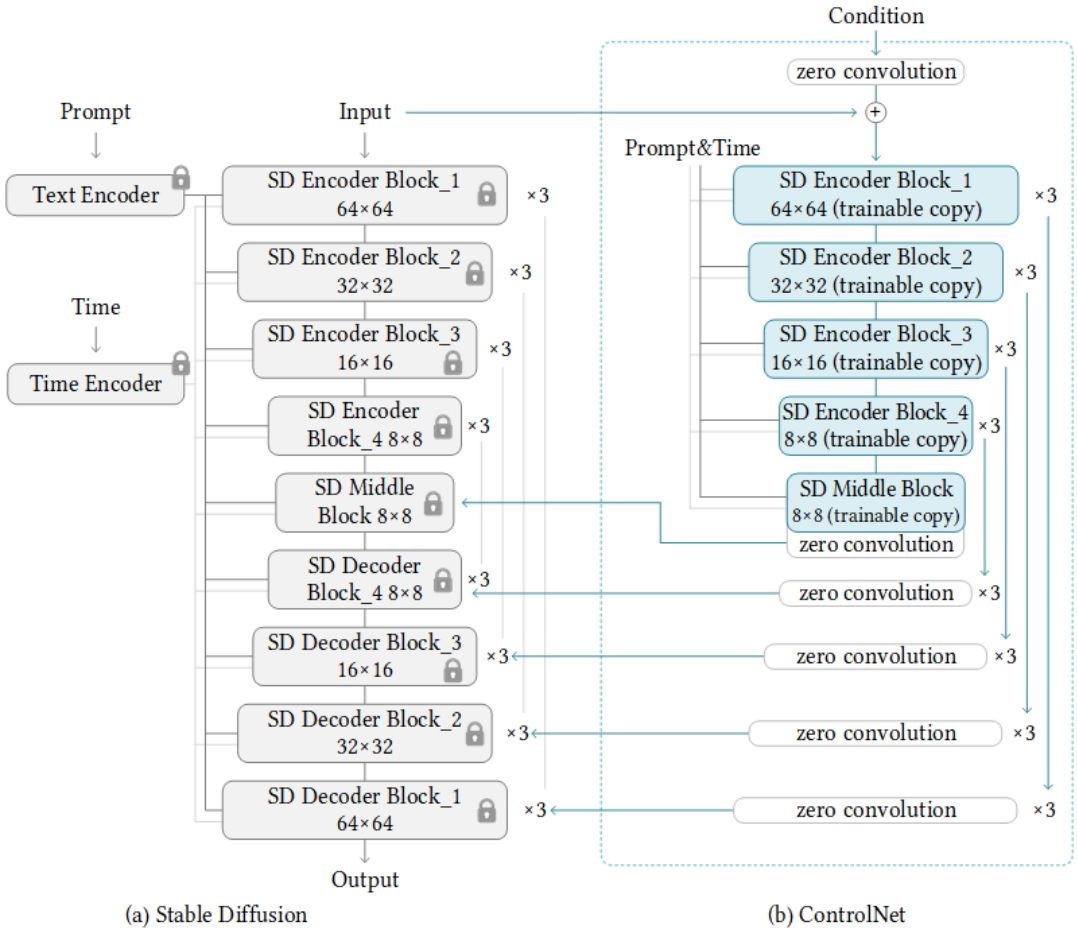
- **Pose-to-Image Generation**
  - Μετατροπή εικόνας σκελετού (skeleton image) σε εικόνα.
- **Sketch-to-Image Generation**
  - Μετατροπή εικόνας σκίτσου (sketch image) σε εικόνα.
- **Semantic Segmentation-to-Image Generation**
  - Παραγωγή εικόνων από χάρτες σημασιολογικής κατάτμησης (segmentation maps).



Εφαρμογές

Image-to-Image Translation - ControlNet

- Adapter-like δίκτυο, το οποίο λειτουργεί παράλληλα με τη βασική αρχιτεκτονική του Stable Diffusion.
- Επιτρέπει την καθοδήγηση της σύνθεσης βάσει εξωτερικών συνθηκών.
- Κατά την εκπαίδευση το Stable Diffusion παραμένει frozen και ανανεώνονται μόνο τα βάρη του ControlNet.

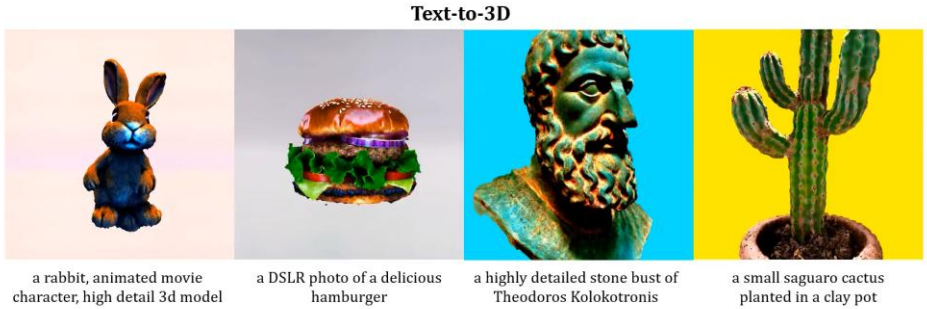


Zhang, L., Rao, A., & Agrawala, M. (2023). Adding conditional control to text-to-image diffusion models. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*.

Εφαρμογές

3D Content Generation

- **Text-to-3D Shape Generation**



- **Image-to-3D Scene Generation**

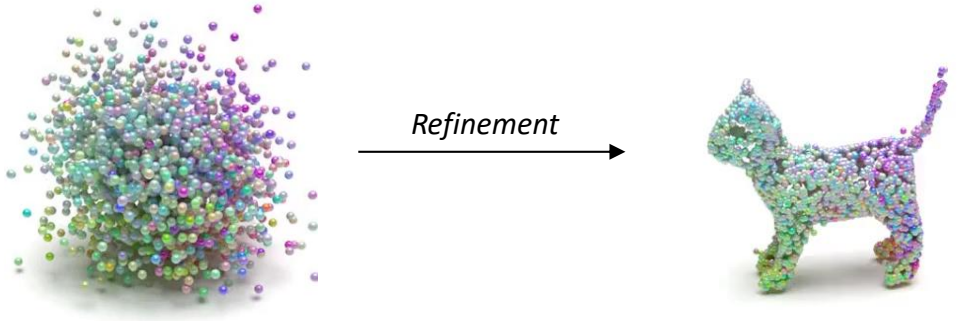
Poole, B., Jain, A., Barron, J. T., & Mildenhall, B. (2022). *DreamFusion: Text-to-3D using 2D Diffusion*.

Huang, Z., Guo, Y.-C., An, X., Yang, Y., Li, Y., Zou, Z.-X., Liang, D., Liu, X., Cao, Y.-P., & Sheng, L. (2024). *MIDI: Multi-Instance Diffusion for Single Image to 3D Scene Generation*.

Εφαρμογές

3D Content Generation

- **Text-to-Point Cloud Generation**
  - **Point Cloud:** σύνολο από σημεία στον 3Δ χώρο, όπου κάθε σημείο έχει συντεταγμένες x, y, z και αναπαριστά ένα συγκεκριμένο σημείο στην επιφάνεια ενός αντικειμένου ή ενός περιβάλλοντος.
  - Η αρχιτεκτονική LION της NVIDIA χρησιμοποιεί diffusion models για να παράξει το αρχικό point cloud από τυχαίο θόρυβο και έπειτα κάνει refinement για την παραγωγή του τελικού point cloud.



Coarse Point Cloud

Final Point Cloud

Zeng, X., Vahdat, A., Williams, F., Gojcic, Z., Litany, O., Fidler, S., & Kreis, K. (2022). LION: Latent point diffusion models for 3D shape generation. In *Advances in Neural Information Processing Systems (NeurIPS)*.

Εφαρμογές

Text or Image-to-Video Generation

- Τα μοντέλα διάχυσης μπορούν να χρησιμοποιηθούν για την παραγωγή βίντεο από κειμενικές περιγραφές ή εικόνες.
- **Γνωστά μοντέλα:**
  - *Stable Video Diffusion*
  - *Sora*
  - *ModelScopeT2V*
  - *CogVideoX*

Source Image

Generated Video

Παραγωγή με χρήση του Stable Video Diffusion.

Text Prompt

“Confident teddy bear surfer rides the wave in the tropics”

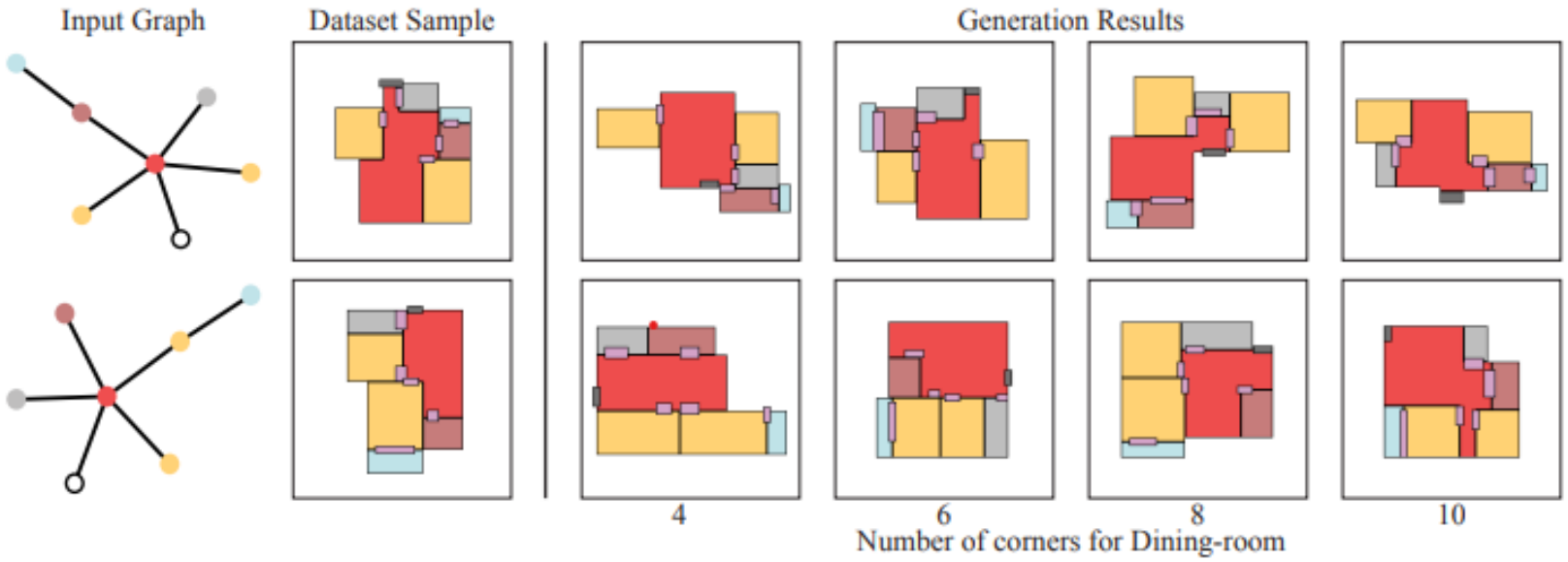
Generated Video

Παραγωγή με χρήση του ModelScopeT2V.

Εφαρμογές

Floorplan Synthesis

- Δημιουργία κατόψεων εσωτερικού χώρου.
- Η διασύνδεση των χώρων αναπαρίσταται μέσω γράφων.
- Ικανότητα ορισμού γωνιών ανά χώρο.



Shabani, M. A., Hosseini, S., & Furukawa, Y. (2022). *HouseDiffusion: Vector Floorplan Generation via a Diffusion Model with Discrete and Continuous Denoising*.

Εφαρμογές

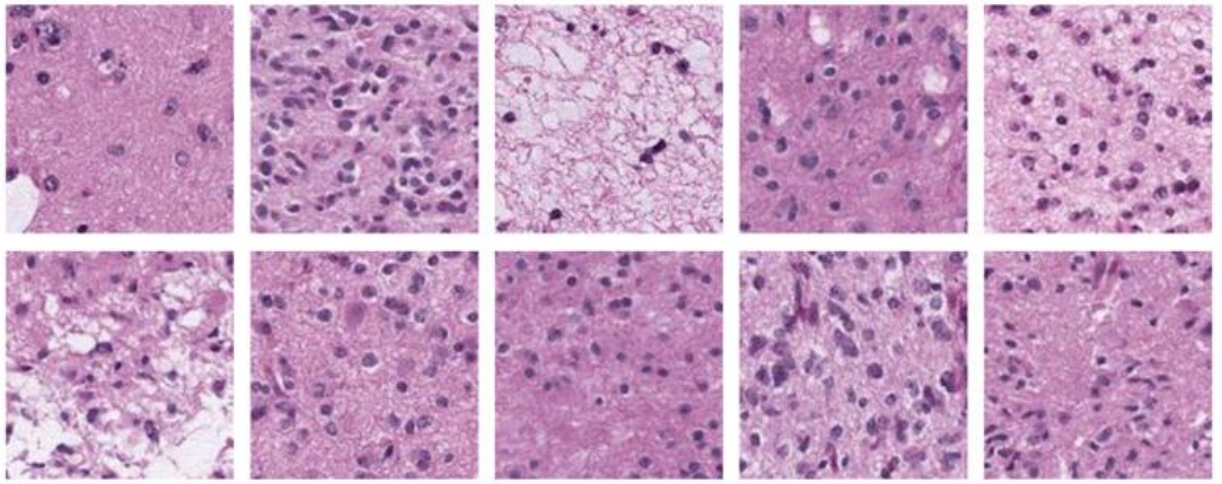
Biomedical Domain

- Τα παραγωγικά μοντέλα έχουν παρουσιάζουν σημαντική επίδραση στον τομέα της ιατρικής απεικόνισης (medical imaging), παρέχοντας εργαλεία που διευκολύνουν το έργο των ιατρών και των ασθενών.
- Η διαδικασία συλλογής απεικονιστικών δεδομένων είναι συχνά πολύπλοκη και χρονοβόρα.
  - Εξειδικευμένα πρωτόκολλα και περιορισμοί.
  - Έλλειψη εξειδικευμένων επαγγελματιών για την επισήμανση (annotation) των δεδομένων.
  - Ζητήματα ιδιωτικότητας και ανάγκη για ρητή συναίνεση των ασθενών για διάθεση των προσωπικών τους δεδομένων.
  - Έντονη ανισορροπία στις κλάσεις των δεδομένων, λόγω της φύσης ορισμένων παθολογιών.
- Τα παραγωγικά μοντέλα προσφέρουν λύσεις στα προβλήματα αυτά καθώς:
  - Δημιουργούν ρεαλιστικά συνθετικά δεδομένα ιατρικών απεικονίσεων.
  - Αίρουν το ζήτημα της προστασίας της ιδιωτικότητας, μέσω δεδομένων που δεν βασίζονται σε πραγματικούς ασθενείς.

Εφαρμογές

Biomedical Domain

- **Synthetic Data Generation**
  - Έχει αποδειχθεί (Chen et al., 2021) ότι η συνδυαστική χρήση **συνθετικών και πραγματικών δεδομένων** για την εκπαίδευση ταξινομητών σε εικόνες ιστολογίας, **βελτιώνει την απόδοση**.
  - Μελέτη η οποία διεξήχθη για την αξιολόγηση των μορφολογικών χαρακτηριστικών συνθετικών και πραγματικών εικόνων σε παθολόγους διαφόρων επιπέδων εμπειρίας (Moghadam et al., 2023), έδειξε ότι οι παθολόγοι δεν μπορούσαν να διακρίνουν τις συνθετικές εικόνες από τις πραγματικές.



Συνθετικές ιστοπαθολογικές εικόνες που δημιουργήθηκαν από το μοντέλο MFDPM (Moghadam et al., 2023).

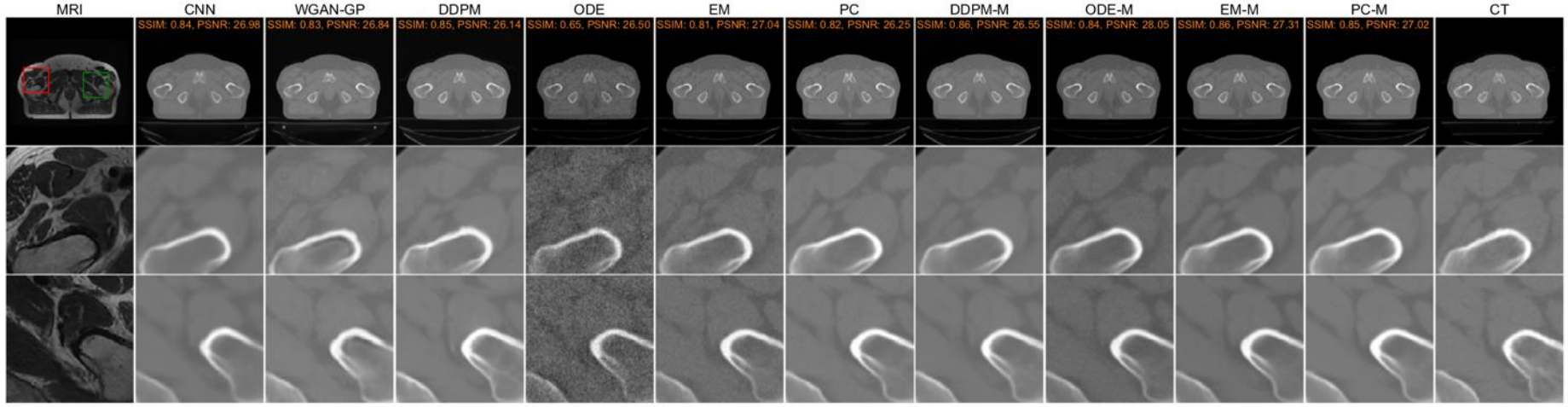
Chen, Richard J., Ming Y. Lu, Tiffany Y. Chen, Drew FK Williamson, and Faisal Mahmood. "Synthetic data in machine learning for medicine and healthcare." Nature Biomedical Engineering 5, no. 6 (2021): 493-497.

Moghadam, Puria Azadi, Sanne Van Dalen, Karina C. Martin, Jochen Lennerz, Stephen Yip, Hossein Farahani, and Ali Bashashati. "A morphology focused diffusion probabilistic model for synthesis of histopathology images." In Proceedings of the IEEE/CVF winter conference on applications of computer vision, pp. 2000-2009. 2023.

Εφαρμογές

Biomedical Domain

- **Image-to-Image Translation**
  - Οι πολυτροπικές (multi-modality) ιατρικές εικόνες είναι καθοριστικές για τη διάγνωση και τη θεραπεία.
  - Σε αρκετές περιπτώσεις, λείπει κάποια απεικονιστική ακολουθία.
  - Τα παραγωγικά μοντέλα έχουν παρουσιάσει ενθαρρυντικά αποτελέσματα στην παραγωγή των απουσιαζουσών απεικονιστικών modalities, π.χ. μετατροπή από *MRI* σε Αξονική Τομογραφία (*CT*).



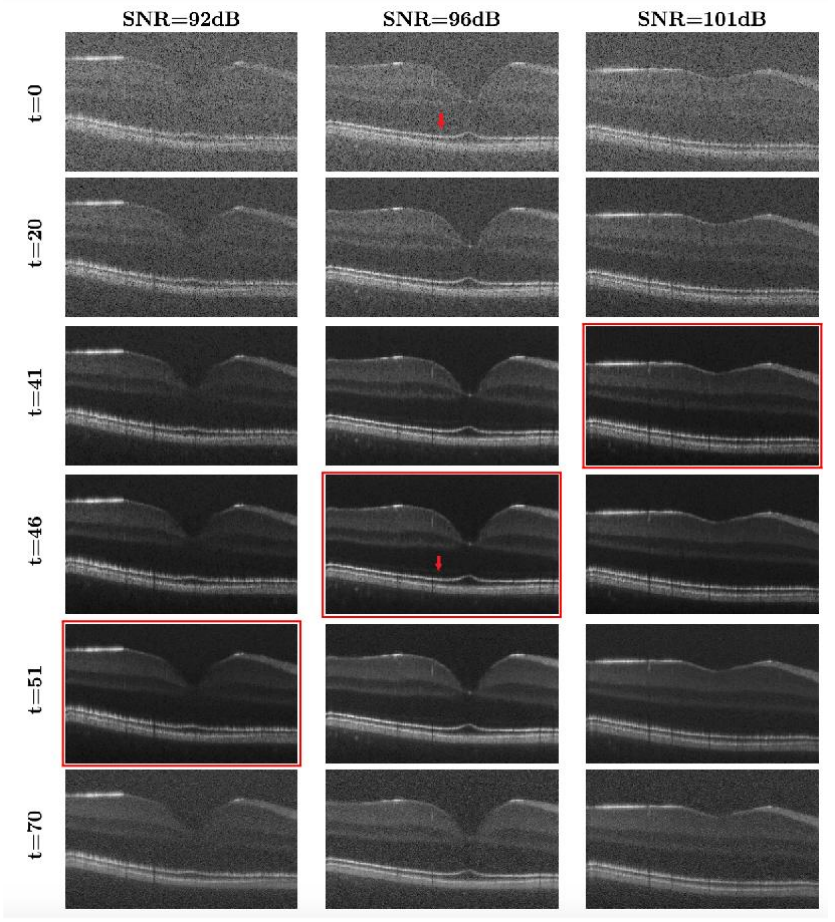
Διαφορετικές μέθοδοι για translation από MRI εικόνες σε CT Scans (Lyu and Wang, 2022).

Lyu, Qing, and Ge Wang. "Conversion between CT and MRI images using diffusion and score-matching models." arXiv preprint arXiv:2209.12104 (2022).

Εφαρμογές

Biomedical Domain

- **Image Denoising**
  - Στην ιατρική απεικόνιση είναι συχνό το φαινόμενο κάποιες εικόνες να επηρεάζονται από θόρυβο κατά τη διάρκεια της απόκτησής τους ή των περαιτέρω σταδίων επεξεργασίας.
  - Λόγω τη φύσης τους τα μοντέλα διάχυσης μπορούν να χρησιμοποιηθούν για αποθορυβοποίηση.
  - Οι Hu et al. (2022) χρησιμοποίησαν ένα DDPM για την αποθορυβοποίηση εικόνων του αμφιβληστροειδούς χιτώνα (retina) του ματιού.



Αποτελέσματα αποθορυβοποίησης για διάφορα επίπεδα SNR και για διαφορετικές τιμές t. Το καλύτερο για κάθε επίπεδο SNR επισημαίνεται με κόκκινο πλαίσιο (Hu et al., 2022).

Hu, Dewei, Yuankai K. Tao, and Ipek Oguz. "Unsupervised denoising of retinal OCT with diffusion probabilistic model." In Medical Imaging 2022: Image Processing, vol. 12032, pp. 25-34. SPIE, 2022.

Παραγωγή με χρήση του Stable Diffusion