

Παραγωγικά Δίκτυα Μάθησης με Αντιπαλότητα (Generative Adversarial Networks)

Βαθιά Μηχανική Μάθηση

ΔΠΜΣ Επιστήμης Δεδομένων & Μηχανικής Μάθησης

Εθνικό Μετσόβιο Πολυτεχνείο

Γιώργος Αλεξανδρίδης

Εισαγωγικές Έννοιες

Διακριτικά και Παραγωγικά Μοντέλα

- **Διακριτικά** (*discriminative*) μοντέλα

- Λειτουργία ταξινόμησης

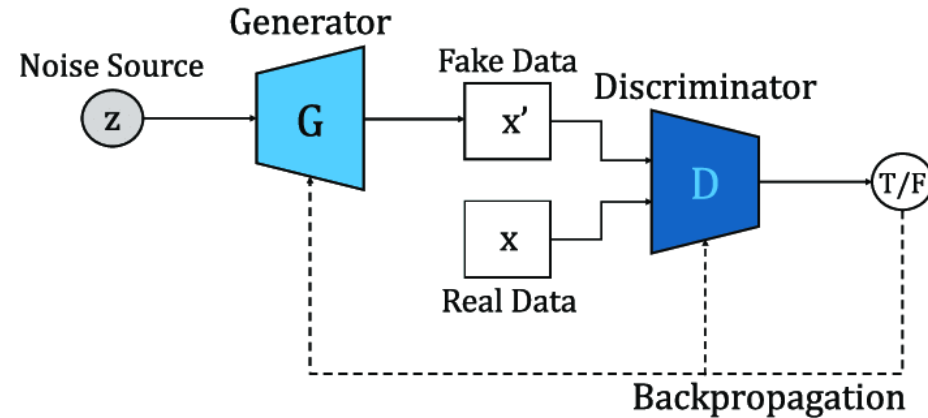
- Δεδομένων των χαρακτηριστικών ενός δείγματος $x \in X$ του χώρου της εισόδου, προέβλεψε την κλάση (ετικέτα) του y
- Εκμάθηση της κατανομής $p(Y|X = x) \Rightarrow$ όρια μεταξύ των κλάσεων

- Παραδείγματα: MLP, SVM, RBF, ...

- **Παραγωγικά** (*generative*) μοντέλα

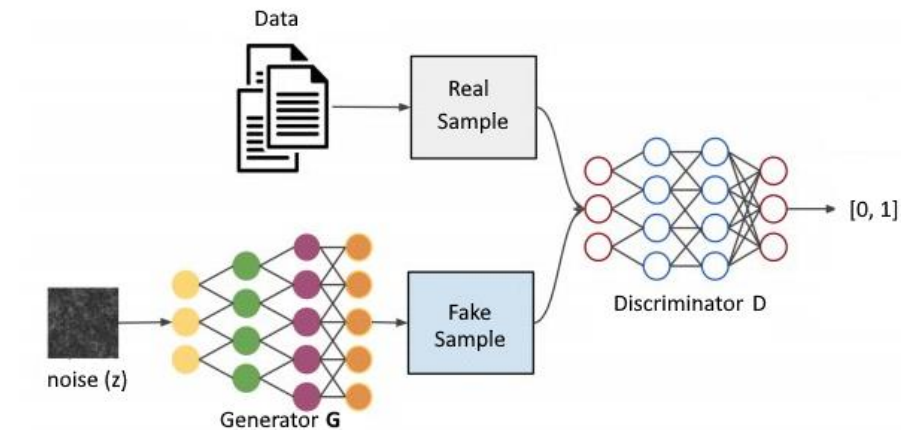
- Δεδομένης μια κλάσης $y \in Y$, προέβλεψε τα χαρακτηριστικά των δειγμάτων X που ανήκουν σε αυτή
- Εκμάθηση της κατανομής $p(X|Y = y)$ των επιμέρους κλάσεων των δεδομένων
- Παραδείγματα: Απλός Μπεϋζιανός ταξινομητής, Γκαουσιανά μοντέλα μίξης, ...

Παραγωγικά Δίκτυα Μάθησης με Αντιπαλότητα



- Generative Adversarial Networks (GANs)
- Σύστημα Μηχανικής Μάθησης που προτάθηκε το **2014** από τον **Goodfellow**, το οποίο αποτελείται από 2 διακριτά νευρωνικά δίκτυα
 - **Διευκρινιστής** (*discriminator*)
 - Αποφασίζει αν τα δεδομένα εισόδου του είναι αληθινά
 - **Γεννήτορας** (*generator*)
 - Προσπαθεί να δημιουργήσει δεδομένα τα οποία ο διευκρινιστής θα θεωρήσει ως αληθινά

Εκπαίδευση GAN



- Στην απλούστερη περίπτωση, ο γεννήτορας **δεν «βλέπει» ποτέ** τα πραγματικά δεδομένα
 - Πρέπει να **«μάθει»** να τα **δημιουργεί**
- Τεχνική Μάθησης: Απώλεια Μάθησης με Αντιπαλότητα (*Adversarial loss*)
 - Αρχικά ο γεννήτορας παράγει τυχαίο θόρυβο
 - Με βάση την απόκριση που λαμβάνει από τον διευκρινιστή, μαθαίνει τελικά την κατανομή των χαρακτηριστικών της εισόδου

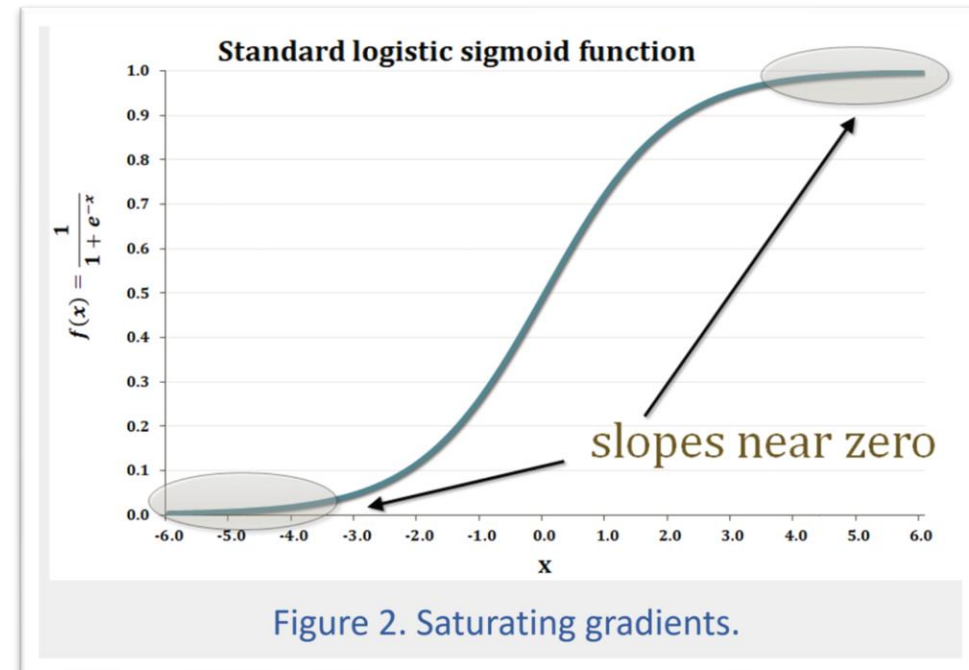
Εκπαίδευση GAN

- Για την **αποφυγή υπερπροσαρμογής**, μπορούν να χρησιμοποιηθούν **τεχνικές ομαλοποίησης** κατά τη διαδικασία της μάθησης
 - λ.χ. DropOut
- **Συνεχές «παιχνίδι αντιπαλότητας»**
 - Διευκρινιστής μαθαίνει να εντοπίζει *έκτοπες τιμές, σφάλματα* και γενικότερα κάθε τι εκτός της *«κανονικότητας»* των δεδομένων
 - Ο γεννήτορας μαθαίνει την κατανομή των δεδομένων εισόδου



Πρόβλημα της Κυριαρχίας

- Αν το ένα δίκτυο **κυριαρχήσει** επί του άλλου, η εκπαίδευση **σταματά** και για τα δύο
- Περίπτωση δυαδικής ταξινόμησης
 - Ο διευκρινιστής έχει σιγμοειδή συνάρτηση ενεργοποίησης
 - Αν για τον *οποιοδήποτε λόγο*, το ανταγωνιστικό παιχνίδι **«κολλήσει»** σε μια από τις δύο **ακριανές περιοχές**, η εκπαίδευση σταματά



Wasserstein GAN

Wasserstein GAN (WGAN)

- Προτάθηκε το **2017** από τον **Wasserstein**
- Η **συνάρτηση αποτίμησης** του διευκρινιστή δεν είναι πλέον η **σιγμοειδής** αλλά η **γραμμική**
 - Σε περίπτωση κυριαρχίας του ενός δικτύου επί του άλλου, ο γεννήτορας δε θα λαμβάνει συνεχώς τις ίδιες τιμές (0 ή 1)
- Επειδή ο διευκρινιστής δεν επιστρέφει τιμές στο $[0, 1]$ αλλά σε ένα ευρύτερο σύνολο, ονομάζεται πλέον **κριτής** (*critic*)

Wasserstein GAN (WGAN)

- Συνάρτηση απώλειας κριτή
 - D_{REAL} : Εκτίμηση κριτή κατά πόσο το δείγμα είναι πραγματικό
 - D_{FAKE} : Εκτίμηση κριτή κατά πόσο το δείγμα είναι ψεύτικο
- Μεταβολή προς την **αντίθετη κατεύθυνση** της κλίσης
 - $L_C = -(D_{REAL} - D_{FAKE})$
 - $L_G = -D_{FAKE}$
- Ένας καλός κριτής επιστρέφει **ψηλές τιμές** για τα **πραγματικά δείγματα** και **χαμηλές** για τα **ψεύτικα**
- Στόχος γεννήτορα είναι ο **ανάποδος**
 - Θέλει να κάνει τον κριτή να επιστρέψει υψηλές τιμές για τα δείγματα που ο ίδιος (ο γεννήτορας) παράγει

Περιορισμένο εύρος αποτελεσμάτων

- Συχνά ο γεννήτορας καταλήγει να επαναδημιουργεί ένα **συγκεκριμένο μόνο εύρος** των δεδομένων εισόδου
 - Αν το **μοναδικό** του κριτήριο είναι να «**κοροϊδέψει**» τον κριτή, τότε μπορεί να το πετύχει δημιουργώντας **δεδομένα** από τις πιο *πυκνές περιοχές* της κατανομής τους

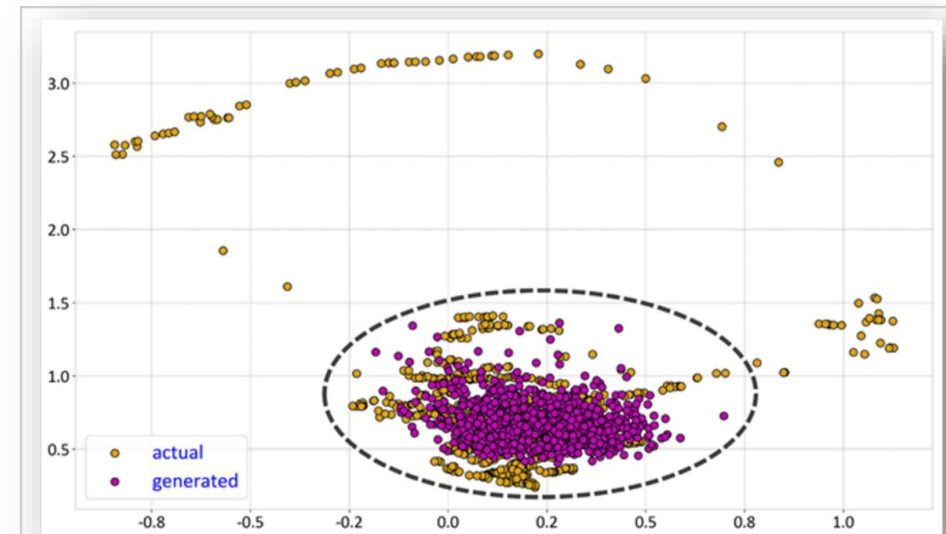


Figure 5. Generator maximizes success by staying in the bulk of the data.

Αύξηση εύρους αποτελεσμάτων

- Λειτουργία του κριτή σε **μικρο-δέσμες εισόδων** (*minibatch discrimination*)
 - Ο κριτής συλλέγει στατιστικές πληροφορίες για τα δείγματα που του παρέχει ο γεννήτορας
 - Αν η κατανομή των παραγόμενων δειγμάτων είναι πολύ διαφορετική από αυτή των πραγματικών, ο κριτής μπορεί να ζητήσει ο γεννήτορας να μεταβάλλει το εύρος του
- Παροχή και στον κριτή και στον γεννήτορα **επιπλέον μεταβλητών**
 - Έχουν τον ρόλο της **υπό συνθήκη πληροφορίας** (*conditional information*)
 - Ενημερώνουν τα δίκτυα για το περιβάλλον στο οποίο λειτουργούν, κατευθύνοντας τις εξόδους τους

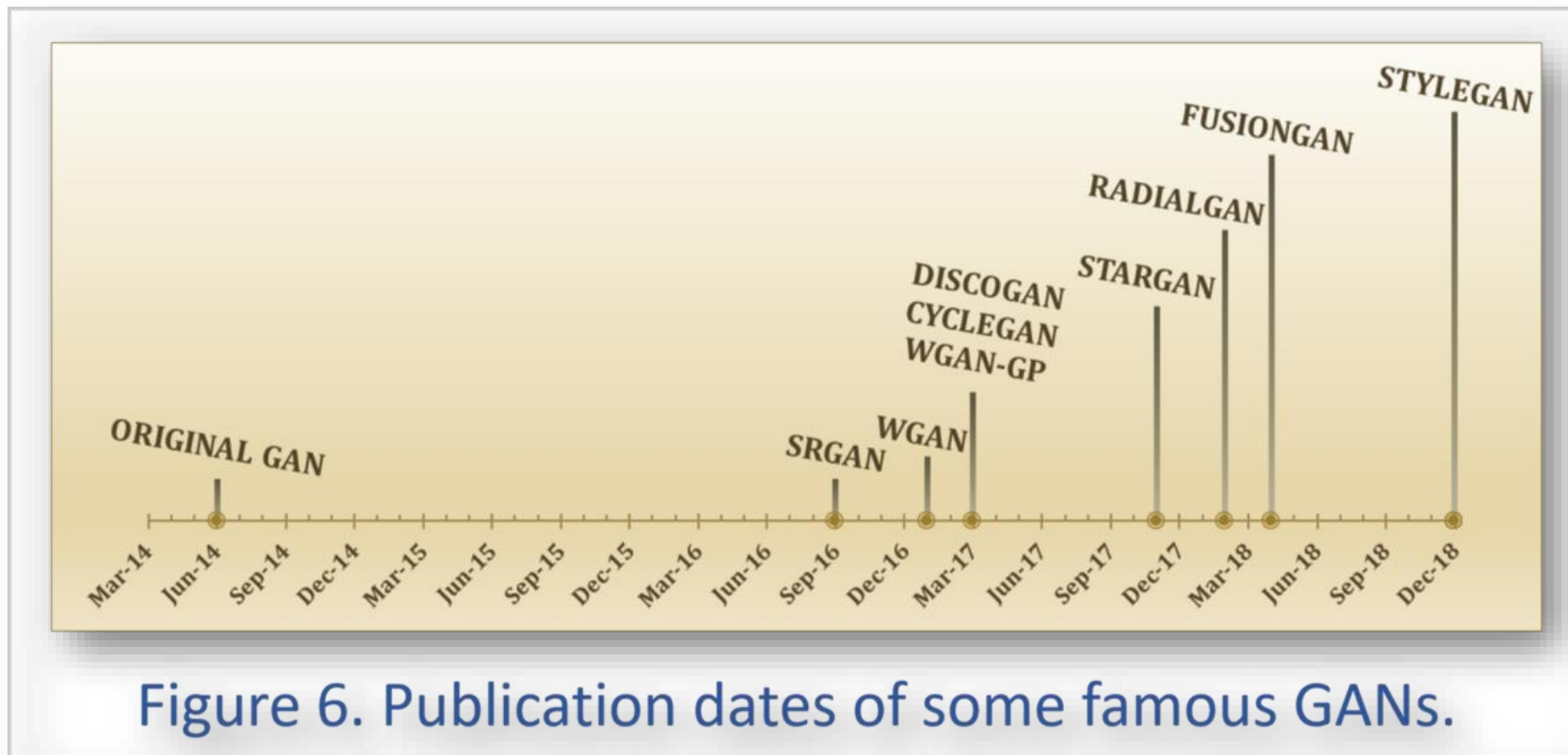
Βελτίωση εκπαίδευσης των WGANs: Βελτιστοποιητές (optimizers)

- Η έξοδος του κριτή είναι πλέον **μη-σταθερή** (*non-stationary*)
 - Δεν είναι στο $[0,1]$
- Χρήση βελτιστοποιητών που δεν περιέχουν **παράγοντα ορμής** (*momentum*)
 - RMSProp, Adam/SGD χωρίς ορμή
 - Ο παράγοντας ορμής «συμπιέζει» (squashes) την έξοδο στο $[0,1]$

Βελτίωση εκπαίδευσης των WGANs: Ποινή κλίσης (Gradient penalty)

- **Αποτρέπει** τις κλίσεις της γραμμικής εξόδου από το να γίνουν πολύ «απότομες»
- **Διαδικασία**
 1. Λήψη δειγμάτων από τα πραγματικά δεδομένα και από τα «ψεύτικα» που παράγει ο γεννήτορας
 2. Υπολογισμός της κλίσης, μετρώντας την μεταβολή της εξόδου του κριτή
 3. Υπολογισμός της νόρμας της κλίσης και δημιουργία παράγοντα ποινής, ανάλογα με το πόσο αυτή απέχει από το 1
 4. Προσθήκη του όρου ποινής του προηγούμενου βήματος στη συνάρτηση απώλειας του κριτή
- **Πλεονέκτημα**
 - Αύξηση της ευστάθειας των GANs
- **Μειονέκτημα**
 - Αύξηση χρόνου εκπαίδευσης για τον υπολογισμό του όρου ποινής

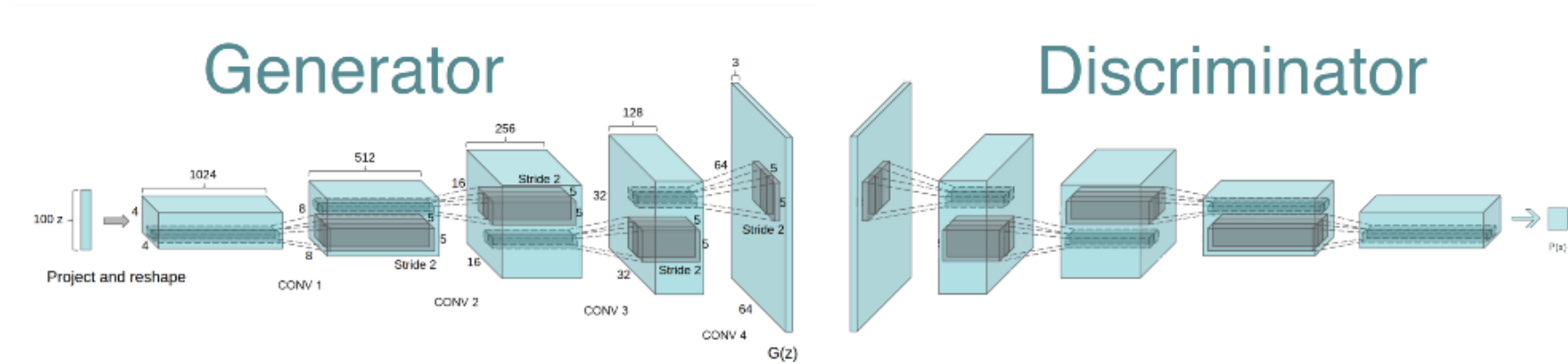
Εξέλιξη των GANs



Αύξηση των ερευνητικών εργασιών σε GANs μετά την εμφάνιση του WGAN

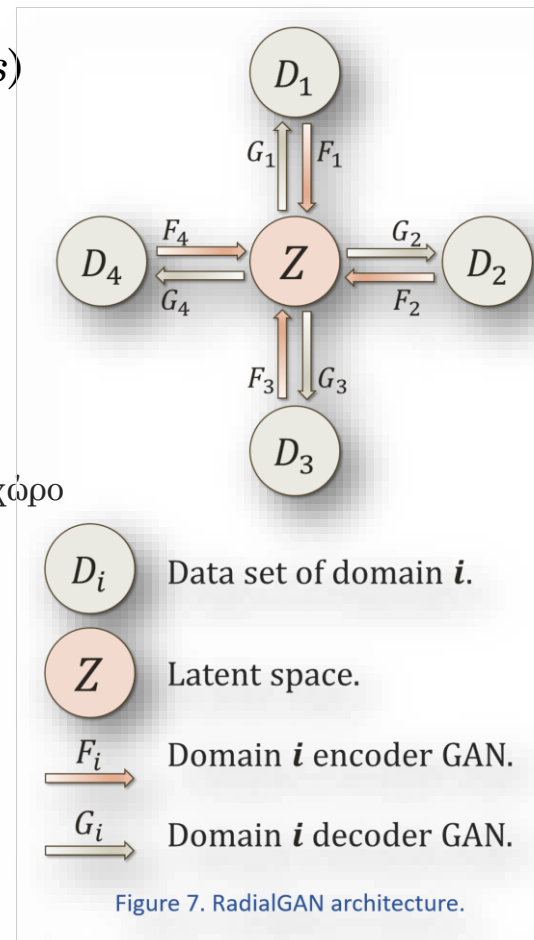
Επεκτάσεις

Deep Convolutional GAN (DC-GAN)



RadialGAN

- Κατάλληλο για συνδυασμό δεδομένων από πολλαπλά **πεδία** (*domains*)
 - **Μεταφορά πεδίου** (*domain transfer*)
- Μετασχηματισμός σε **λανθάνοντα χώρο** (*latent space*)
- Κάθε σύνολο δεδομένων έχει δικό του:
 1. Δίκτυο **κωδικοποίησης** (*encoder network*)
 2. Δίκτυο **αποκωδικοποίησης** (*decoder network*)
 - Ο διευκρινιστής επιβεβαιώνει ότι η πληροφορία που έρχεται από το λανθάνοντα χώρο ταιριάζει με τις ιδιότητες του συγκεκριμένου πεδίου
- **Κυκλική συνέπεια** (*cycle consistency*)
 - Μεταφορά πληροφορίας μεταξύ του λανθάνοντα χώρου και των πεδίων
 - Και προς τις δύο κατευθύνσεις
- Παράλληλη εκπαίδευση όλων των δικτύων οδηγεί στη δημιουργία **επαυξημένου συνόλου δεδομένων**



StyleGAN

- Συνδυασμός **Progressive GAN (ProGAN)** και **μεταφοράς στυλ μέσω νευρώνων (neural style transfer)**
- **ProGAN**
 - Αναπτύσσει αρχική εικόνα μικρής διάστασης (πχ 4×4 ή 8×8 pixel) μέχρι αυτή να θεωρηθεί ρεαλιστική από τον διευκρινιστή
 - Στην επόμενη φάση, προστίθεται ένα επίπεδο υψηλότερης ανάλυσης και επαναλαμβάνεται η εκπαίδευση
 - Η διαδικασία ολοκληρώνεται όταν φτάσουμε στο επιθυμητό επίπεδο ανάλυσης (πχ 1024×1024 pixel)
- **StyleGAN**
 - Μεταφορά στυλ μέσω της προσαρμογής των βαρών του επιπέδου του στυλ στο επίπεδο της εικόνας

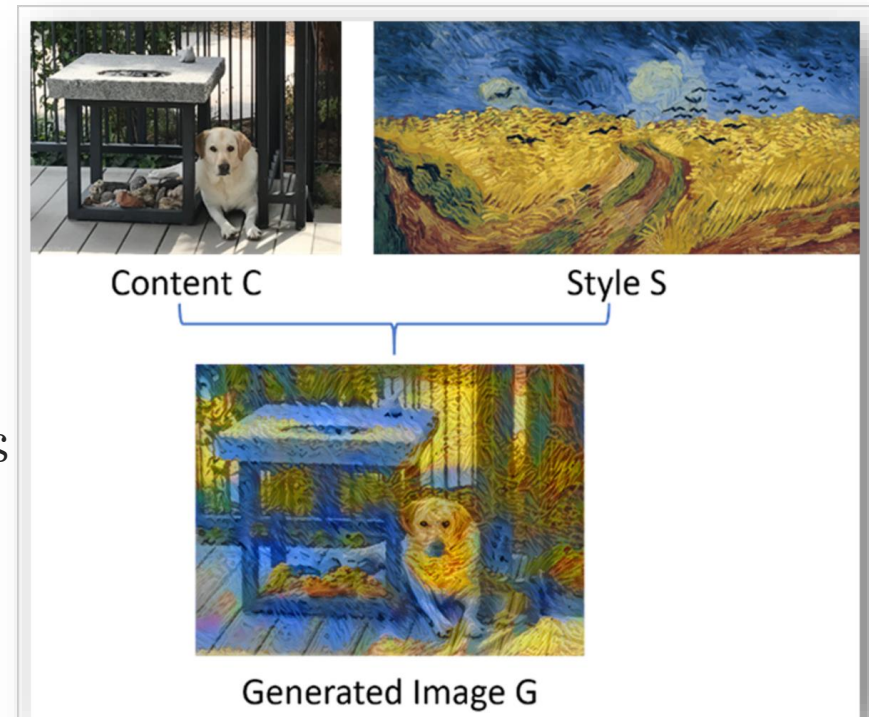


Figure 10. Neural style transfer redraws a content image in the design of a style image.

AdaIN: Adaptive Instance Normalization

- Τεχνική μεταφοράς στυλ που χρησιμοποιείται στο StyleGAN
- Δεν περιλαμβάνει βελτιστοποίηση και συνεπώς είναι μια γρήγορη τεχνική
- Εφαρμογή του AdaIN σε συνελκτικό επίπεδο εικόνας
 1. Κανονικοποίηση του επιπέδου
 - Αφαίρεση μέσης τιμής και διαίρεση με τυπική απόκλιση
 2. Κλιμάκωση του κανονικοποιημένου επιπέδου έτσι ώστε να προσαρμοστεί στην τυπική απόκλιση του επιπέδου του στυλ
 3. **Ολισθηση** (*shift*) των βαρών του κανονικοποιημένου και του κλιμακούμενου επιπέδου μέσω της προσθήκης της μέσης τιμής του επιπέδου του στυλ

Εφαρμογές

This Person Does Not Exist



<https://thispersondoesnotexist.com/>

Μετάφραση κειμένου σε εικόνα



Han Zhang, et al. (2016) - *StackGAN: Text to Photo-realistic Image Synthesis with Stacked Generative Adversarial Networks*

Σύνθεση μουσικής

- [MuseGAN](#)

- Παραγωγή «πολυφωνικής» μουσικής (πολλαπλών οργάνων)
- Είτε από την αρχή, είτε ως συνοδεία κομματιού που δίνεται από τον χρήστη
- Εκπαίδευση στο Lakh Pianoroll Dataset
 - Περιέχει pop μουσική αποτελούμενη από μπάσο, ντραμς, κιθάρα, πιάνο και έγχορδα

- [Pokemon Music](#)

- Συνδυασμός δικτύου LSTM και RNN-GAN
- Εκπαίδευση σε 5 ώρες μουσικής Pokemon, υπό τη μορφή αρχείων MIDI

Βιβλιογραφία

- Επιστημονικά άρθρα
 - Ian J. Goodfellow et al. [Generative Adversarial Networks](#)
 - Martin Arjovsky et al. [Wasserstein GAN](#)
 - Alec Radford et al. [Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks](#)
 - Tim Salimans et al. [Improved Techniques for Training GANs](#)
 - Ishaan Gulrajani et al. [Improved Training of Wasserstein GANs](#)
 - Lars Mescheder et al. [Which Training Methods for GANs do Actually Converge?](#)
 - Tero Karras et al. [Progressive Growing of GANs for Improved Quality, Stability, and Variation](#)
 - Leon A. Gatys et al. [A Neural Algorithm of Artistic Style](#)

Βιβλιογραφία

- Άρθρα στον Παγκόσμιο Ιστό
 - [A Beginner's Guide to Generative Adversarial Networks \(GANs\)](#)
 - [A Leap into the Future: Generative Adversarial Networks](#)
 - [Understanding Generative Adversarial Networks \(GANs\)](#)