

Χρήση των Generative Adversarial Networks για επιλογή πόζας και στιλ σε εφαρμογές σχεδιασμού μόδας

Επιβλέποντες: καθ. κ. Μήτκας Περικλής - κ. Αντώνης Χρυσόπουλος

Αθανάσιος Χαρισούδης
achariso@ece.auth.gr

Τμήμα Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών
Τομέας Ηλεκτρονικής & Υπολογιστών
Αριστοτέλειο Πανεπιστήμιο Θεσσαλονίκης

13 Ιουλίου 2021

- 1 Σύνοψη της Εργασίας
- 2 Παραγωγική Μοντελοποίηση & GANs
- 3 Εφαρμογή σε Εικόνες Μόδας
- 4 Αποτελέσματα & Αξιολόγηση

- 1 Σύνοψη της Εργασίας
- 2 Παραγωγική Μοντελοποίηση & GANs
- 3 Εφαρμογή σε Εικόνες Μόδας
- 4 Αποτελέσματα & Αξιολόγηση

Σύνοψη της Εργασίας

Σκοπός

Δημιουργία ενός πολυ-εργαλείου μηχανικής μάθησης με δυνατότητες παραγωγής ρεαλιστικών εικόνων μόδας και εφαρμογής ρεαλιστικών μετασχηματισμών σε αυτές.

Αναμενόμενα Αποτελέσματα

Το εργαλείο αυτό το οπλίσαμε με τις ακόλουθες δυνατότητες:

- ✓ **Αλλαγή Πόζας**
- ✓ **Εξαγωγή Ρούχου**
- ✓ **Ταίριασμα Στιλ**
- ✓ **Παραγωγή ρεαλιστικών εικόνων μόδας**

Βήματα Υλοποίησης

- 1 Σχεδίαση μοντέλων GAN (βιβλιογραφία)
- 2 Υλοποίηση μοντέλων (PyTorch)
- 3 Εκπαίδευσή μοντέλων σε σύνολα δεδομένων εικόνων μόδας (Colab, Kaggle, vast.ai)

- 1 Σύνοψη της Εργασίας
- 2 Παραγωγική Μοντελοποίηση & GANs**
- 3 Εφαρμογή σε Εικόνες Μόδας
- 4 Αποτελέσματα & Αξιολόγηση

Παραγωγική Μοντελοποίηση

«What I understand, I can create»

Παραγωγική Μοντελοποίηση: παραγωγή τεχνητών πολυμεσικών δεδομένων από αλγόριθμους ή μοντέλα μηχανικής μάθησης.

Παραγωγική Μοντελοποίηση Εικόνων: έγκειται στην παραγωγή εικόνων από στατιστικά μοντέλα που επιδεικνύουν *ρεαλισμό* και *ποικιλομορφία*. Ιδανικά δεν διακρίνονται από αυτές του συνόλου εκπαίδευσης.

(Στατιστικά) Βαθιά Παραγωγικά Μοντέλα: εκπαιδεύονται ώστε να αιχμαλωτίσουν την πιθανοτική κατανομή που ακολουθούν τα δεδομένα εκπαίδευσης (απαιτείται πρότερη γνώση - *No Free Lunch*). Κατόπιν, καλούνται να παράξουν δείγματα μη-διαρκίσιμα από τα πρώτα.

Generative Adversarial Networks (GANs)

«Generative Adversarial Networks», Goodfellow et al., 2014



Δίκτυο Discriminator

- ✓ Διαδική ταξινόμηση εικόνων σε πραγματικές/τεχνητές
- ✓ Αποτελείται από συνελκτικές στρώσεις σε σειρά
- ✓ Πιο απλό έργο από Generator

Δίκτυο Generator

- ✓ Παραγωγή ρεαλιστικών εικόνων που μπερδεύουν τον Discriminator
- ✓ Αποτελείται από αναστροφές συνελκτικές στρώσεις
- ✓ Είσοδος: Τυχαίος θόρυβος (χωρίς συνθήκη) - Εικόνα (υπό-συνθήκη)

Αντιπαραθητική Εκπαίδευση

- Το κάθε δίκτυο μεταβάλλει μόνο τις δικές του παραμέτρους

$$J_D(\bar{\theta}_D, \bar{\theta}_G) \approx -\mathbb{E}_{x \sim p_{\text{data}}} \log [D(x)] - \mathbb{E}_{z \sim p_{\text{prior}}} \log [1 - D(G(z))]$$

$$J_G(\bar{\theta}_G, \bar{\theta}_D) \approx \mathbb{E}_{z \sim p_{\text{prior}}} \log [1 - D(G(z))]$$

Αξιολόγηση Παραγόμενων Εικόνων

Μετράμε δύο επιθυμητές ιδιότητες των GANs: ποιότητα (fidelity) και ποικιλομορφία (diversity). Τέσσερις μετρικές αξιολόγησης:

- 1 **Inception Score (IS)**: χρήση προ-εκπαιδευμένου ταξινομητή εικόνας και σύγκριση προβλέψεων ταξινόμησης. **(το υψηλότερο το καλύτερο)**
- 2 **Fréchet Inception Distance (FID)**: σύγκριση μεταξύ πραγματικών και τεχνητών εικόνων από προσαρμογή Gaussians στις εμβυθίσεις και απόσταση αυτών κατά Fréchet **(το χαμηλότερο το καλύτερο)**
- 3 **Precision-Recall-F₁-Score**: σύγκριση εμβυθίσεων μεταξύ παραγόμενων και πραγματικών εικόνων με προσέγγιση των αντίστοιχων supports των κατανομών τους **(το υψηλότερο το καλύτερο)**
- 4 **Structural Similarity Index (SSIM)**: σύγκριση φωτεινότητας, αντίθεσης και δομής για εύρεση ομοιότητας κάθε ζεύγους πραγματικής - τεχνητής εικόνας **(το υψηλότερο το καλύτερο)**

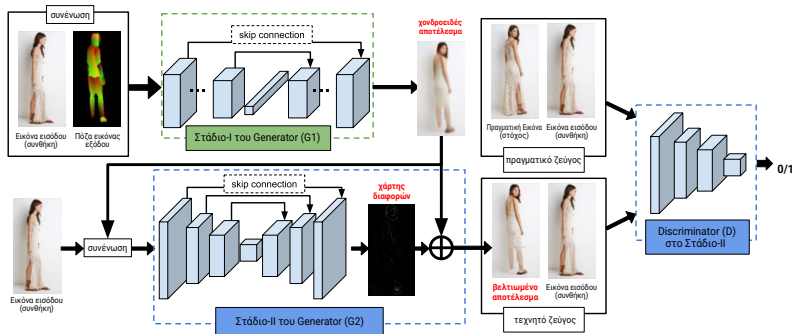
Ευρέως αποδεκτό: FID και F1-Score οι πιο αξιόπιστες.

- 1 Σύνοψη της Εργασίας
- 2 Παραγωγική Μοντελοποίηση & GANs
- 3 Εφαρμογή σε Εικόνες Μόδας**
 - Αλλαγή Πόζας (PGPG - *PoseGAN*)
 - Εξαγωγή Ρούχου (PixelDTGAN)
 - Ταίριασμα Στιλ (DiscoGAN - *CycleGAN*)
 - Παραγωγή Ρεαλιστικών Εικόνων (*StyleGAN*)
- 4 Αποτελέσματα & Αξιολόγηση

- 1 Σύνοψη της Εργασίας
- 2 Παραγωγική Μοντελοποίηση & GANs
- 3 Εφαρμογή σε Εικόνες Μόδας**
 - Αλλαγή Πόζας (PGPG - *PoseGAN*)
 - Εξαγωγή Ρούχου (PixelDTGAN)
 - Ταίριασμα Στιλ (DiscoGAN - CycleGAN)
 - Παραγωγή Ρεαλιστικών Εικόνων (StyleGAN)
- 4 Αποτελέσματα & Αξιολόγηση

Αλλαγή Πόζας (PGPG - PoseGAN)

«Pose Guided Person Image Generation», Ma et al., 2017



Σχήμα 1: Συνολική αρχιτεκτονική του μοντέλου PGPG (PoseGAN). Φαίνονται τα δύο στάδια του Generator, ο Discriminator καθώς και οι εισοδοί/έξοδοι όλων των υποδι-κτύπων του μοντέλου.

Αλλαγή Πόζας (PGPG - PoseGAN)

Σύνοψη Μοντέλου - Παράμετροι Εκπαίδευσης

Σύνοψη του μοντέλου PoseGAN

Όνομα Μοντέλου	PoseGAN ^{GT}
Κωδικός Configuration	128_MSE_256_4_6_5_none_1e4_true_false_false
Εφαρμογή	Αλλαγή πόζας σε ανθρώπους-μοντέλα
Κατηγορία Εφαρμογής	Συζευγμένη Μετατροπή Εικόνας-σε-Εικόνα
Κατηγορία Παραγωγής	Υπο-συνθήκη (εικόνα)
Σχετικό Μοντέλο	PGPG
Σχετικό Άρθρο	«Pose Guided Person Image Generation» [81]
Αριθμός GANs	ένα (1)
Αριθμός Generators	έναν (1) - αποτελούμενος από δύο υποδίκτυα, G1 και G2
Αριθμός Discriminators	έναν (1)
Τύπος Generator(s)	G1: U-Net με skip connections (4 blocks ανά κατεύθυνση + πλήρως συνδεδεμένη στρώση στη στένωση) G2: U-Net με skip connections (6 blocks ανά κατεύθυνση + Dropout στον encoder)
Τύπος Discriminator(s)	PatchGAN Discriminator (5 blocks, 32x32 receptive field, Κανονικοποίηση Φάσματος)



Ενδεικτικές εικόνες που δίνονται στο μοντέλο από τον φορτωτή δεδομένων
Σύνολο Δεδομένων: In-Shop Clothes Retrieval Benchmark (DeepFashion)

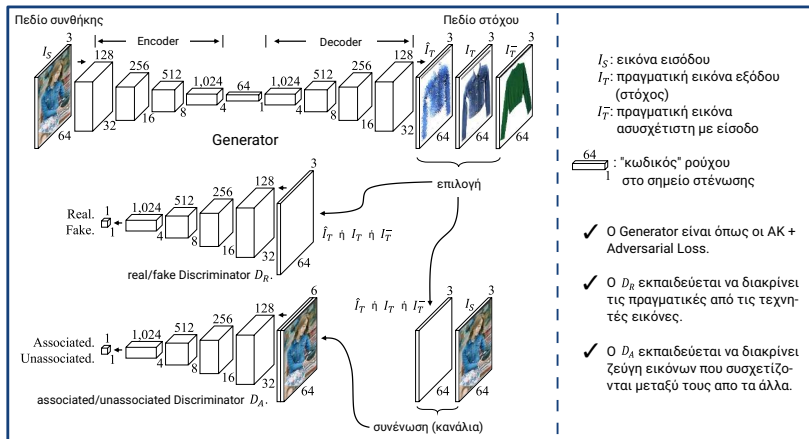
Παράμετροι εκπαίδευσης του μοντέλου PoseGAN

Όνομα Μοντέλου	PoseGAN ^{GT}
Κωδικός Configuration	128_MSE_256_4_6_5_none_1e4_true_false_false
Συνάρτ. Κόστους Generator	G1: L ₁ Loss (ανακατασκευής) G2: L ₁ Loss (ανακατασκευής, βάρος: $\lambda_{recos} = 1 \rightarrow 5 \rightarrow 10$) + Ελαχίστων Τετραγώνων (MSE) (αντιπαραθετική)
Συνάρτ. Κόστους Discriminator(s)	Ελαχίστων Τετραγώνων (MSE)
Αριθμός βελτ/τών για Generator(s)	έναν (1) (από κοινού εκπαίδευση των υποδικτύων G1 και G2 του Generator)
Αριθμός βελτ/τών για Discriminator(s)	έναν (1)
Τύπος βελτ/τών για Generator(s)	Adam, με παραμέτρους ($lr=0.0001, \beta_1=0.9, \beta_2=0.999$)
Τύπος βελτ/τών για Discriminator(s)	Adam, με παραμέτρους ($lr=0.0001, \beta_1=0.9, \beta_2=0.999$)
Σύνολο δεδομένων εκπαίδευσης	ICRB (DeepFashion)
Μέγεθος συνόλου δεδομένων	46.4K εικόνες με πόζα 92.5K ζεύγη εικόνων αλλαγής πόζας
Ανάλυση εικόνων	128x128px
Μέγεθος ομάδας	48 εικόνες/batch
Αριθμός epochs	93 epochs (162.533 επαναλήψεις)
Χρόνος Εκπαίδευσης	περίπου έξι (6) μέρες σε 16GB GPUs
# Παραμέτρων Generator(s)	393.9M εκπαιδευόμενες παράμετροι (G1: 276M, G2: 117.4M)
# Παραμέτρων Discriminator(s)	1.6M εκπαιδευόμενες παράμετροι
# Παραμέτρων	395.1M εκπαιδευόμενες παράμετροι συνολικά

- 1 Σύνοψη της Εργασίας
- 2 Παραγωγική Μοντελοποίηση & GANs
- 3 Εφαρμογή σε Εικόνες Μόδας**
 - Αλλαγή Πόζας (PGPG - *PoseGAN*)
 - Εξαγωγή Ρούχου (PixelDTGAN)**
 - Ταίριασμα Στιλ (DiscoGAN - *CycleGAN*)
 - Παραγωγή Ρεαλιστικών Εικόνων (*StyleGAN*)
- 4 Αποτελέσματα & Αξιολόγηση

Εξαγωγή Ρούχου (PixelDTGAN)

«Pixel-Level Domain Transfer», Yoo et al., 2016



Σχήμα 2: Αρχιτεκτονική Μοντέλου PixelDTGAN

Εξαγωγή Ρούχου (PixelDTGAN)

Σύνοψη Μοντέλου - Παράμετροι Εκπαίδευσης

Σύνοψη του μοντέλου PixelDTGAN

Όνομα Μοντέλου	PixelDTGAN ^{GT}
Κωδικός Configuration	px1dtgan_default
Εφαρμογή	Εξαγωγή ρούχου του ανθρώπου στην εικόνα εισόδου
Κατηγορία Εφαρμογής	Συζευγμένη Μετατροπή Εικόνας-σε-Εικόνα
Κατηγορία Παραγωγής	Υπο-συνθήκη (εικόνα)
Σχετικό Μοντέλο	PixelDTGAN
Σχετικό Άρθρο	«Pixel-Level Domain Transfer» [66]
Αριθμός GANs	ένα (1)
Αριθμός Generators	ένα (1)
Αριθμός Discriminators	δύο (2) - real/fake (D_R) και associated/unassociated (D_A)
Τύπος Generator(s)	U-Net χωρίς skip connections (5 blocks ανά κατεύθυνση, με δίνουσα 100 στοιχείων στο σημείο στένωσης)
Τύπος Discriminator(s)	2 × PatchGAN Discriminator (4 blocks, 16×16 receptive field, Κανονικοποίηση Φάσματος)
# Παραμέτρων Generator(s)	37.7M εκπαιδευσιμες παράμετροι
# Παραμέτρων Discriminator(s)	D_R : 6.2M εκπαιδευσιμες παράμετροι D_A : 6.2M εκπαιδευσιμες παράμετροι
# Παραμέτρων	50.1M εκπαιδευσιμες παράμετροι συνολικά

Ενδεικτικές εικόνες που δίνονται στο μοντέλο από τον φορτωτή δεδομένων
Σύνολο Δεδομένων: **LookBook + ICRB (DeepFashion)**

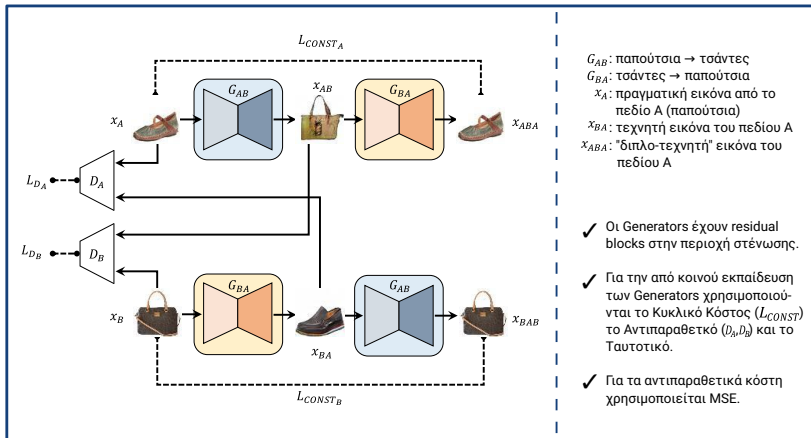
Παράμετροι εκπαίδευσης του μοντέλου PixelDTGAN

Όνομα Μοντέλου	PixelDTGAN ^{GT}
Κωδικός Configuration	px1dtgan_default
Συναρτ. Κόστους Generator	L_1 Loss (ανακατασκευής, βάρος: $\lambda_{recon} = 4$ εκπαιδευσιμο) + Ελαχίστων Τετραγώνων (MSE) (αντιπαραθετική)
Συναρτ. Κόστους Discriminator(s)	Ελαχίστων Τετραγώνων (MSE)
Αριθμός βελτ/τών για Generator(s)	ένας (1)
Αριθμός βελτ/τών για Discriminator(s)	δύο (2) (ξεχωριστή εκπαίδευση των D_R και D_A)
Τύπος βελτ/τών για Generator(s)	Adam, με παραμέτρους ($\text{lr}=0.0001$, $\beta_1=0.9$, $\beta_2=0.999$)
Τύπος βελτ/τών για Discriminator(s)	Adam, με παραμέτρους ($\text{lr}=0.0001$, $\beta_1=0.9$, $\beta_2=0.999$)
Σύνολο δεδομένων εκπαίδευσης	LookBook + ICRB (DeepFashion)
Μέγεθος συνόλου δεδομένων	81.2K εικόνες (ανθρώπων + ρούχων) 71.7K Ζεύγη εικόνων εξαγωγής ρούχου
Ανάλυση εικόνων	64×64px
Μέγεθος ομάδας	256 εικόνες/batch
Αριθμός epochs	348 epochs (88.200 επαναλήψεις)
Χρόνος Εκπαίδευσης	περίπου δώδεκα (12) ημέρες σε 12GB & 16GB GPUs

- 1 Σύνοψη της Εργασίας
- 2 Παραγωγική Μοντελοποίηση & GANs
- 3 Εφαρμογή σε Εικόνες Μόδας**
 - Αλλαγή Πόζας (PGPG - *PoseGAN*)
 - Εξαγωγή Ρούχου (PixelDTGAN)
 - Ταίριασμα Στιλ (DiscoGAN - CycleGAN)**
 - Παραγωγή Ρεαλιστικών Εικόνων (StyleGAN)
- 4 Αποτελέσματα & Αξιολόγηση

Ταίριασμα Στιλ (DiscoGAN - CycleGAN)

«Learning to Discover Cross-Domain Relations with Generative Adversarial Networks», Kim et al., 2017



Σχήμα 3: Αρχιτεκτονική Μοντέλου DiscoGAN

Ταίριασμα Στιλ (DiscoGAN - CycleGAN)

Σύνοψη Μοντέλου - Παράμετροι Εκπαίδευσης

Σύνοψη του μοντέλου CycleGAN	
Όνομα Μοντέλου	CycleGAN ^{DT}
Κωδικός Configuration	discogan
Εφαρμογή	Ταίριασμα στιλ μεταξύ παπουτσιών και τσαντών (μετατροπή από το ένα πεδίο στο άλλο)
Κατηγορία Εφαρμογής	Μη-Συζευγμένη Μετατροπή Εικόνας-σε-Εικόνα
Κατηγορία Παραγωγής	Υπό-αυθόκη (εικόνα)
Σχετικό Μοντέλο	DiscoGAN
Σχετικό Άρθρο	«Learning to Discover Cross-Domain Relations with Generative Adversarial Networks» [76]
Αριθμός GANs	δύο (2)
Αριθμός Generators	δύο (2) - G_{AB} (παπούτσια \rightarrow τσάντες) και G_{BA} (αντίστροφα)
Αριθμός Discriminators	δύο (2) - D_B (τσάντες) και D_A (παπούτσια)
Τύπος Generator(s)	2 × δίκτυο encoder-transformer-decoder (2 blocks στον encoder & decoder, και 9 blocks στον transformer)
Τύπος Discriminator(s)	2 × PatchGAN Discriminator (4 blocks, 16×16 receptive field, Κανονικοποίηση Φάσματος)
# Παραμέτρων Generator(s)	G_{BA} : 11.4M εκπαιδευσιμες παράμετροι
# Παραμέτρων Discriminator(s)	D_B : 1.6M εκπαιδευσιμες παράμετροι
# Παραμέτρων	D_A : 1.6M εκπαιδευσιμες παράμετροι
# Παραμέτρων	25.9M εκπαιδευσιμες παράμετροι συνολικά



Ενδεικτικές εικόνες που δίνονται στο μοντέλο από τον φορτωτή δεδομένων
Σύνολο Δεδομένων: **handbags2shoes**

Παράμετροι εκπαίδευσης του μοντέλου CycleGAN	
Όνομα Μοντέλου	CycleGAN ^{DT}
Κωδικός Configuration	discogan
Συναρτ. Κόστους Generator	2 × Κόστος Κυκλικής Συνοχής (L_1 Loss, με βάρος $\lambda_{cycle} = 10$) + 2 × Ταυτοτικό Κόστος (L_1 Loss, με βάρος $\lambda_{identity} = 5$) + 2 × Αντιπαραθετικό Κόστος Ελαχίστων Τετραγώνων (MSE)
Συναρτ. Κόστους Discriminator(s)	2 × Ελαχίστων Τετραγώνων (MSE)
Αριθμός βελτ/τών για Generator(s)	ένας (1)
Αριθμός βελτ/τών για Discriminator(s)	ένας (1) (από κοινού εκπαίδευση των D_A και D_B)
Τύπος βελτ/τών για Generator(s)	Adam, με παραμέτρους ($\text{lr}=0.0002, \beta_1=0.9, \beta_2=0.999$) LR Scheduler: OnPlateau(factor: 0.99, cooldown: 200) ανά βήμα (όχι epoch)
Τύπος βελτ/τών για Discriminator(s)	Adam, με παραμέτρους ($\text{lr}=0.0002, \beta_1=0.9, \beta_2=0.999$) LR Scheduler: OnPlateau(factor: 0.99, cooldown: 200) ανά βήμα (όχι epoch)
Σύνολο δεδομένων εκπαίδευσης	handbags2shoes
Μέγεθος συνόλου δεδομένων	138.8K εικόνες τσαντών + 50K εικόνες παπουτσιών
Ανάλυση εικόνων	71.7K ζεύγη εικόνων εξαγωγής ρούχου
Μέγεθος ομάδας	64×64px
Αριθμός epochs	32 εικόνες /batch
Χρόνος Εκπαίδευσης	189 epochs (142.200 επαναλήψεις)
Χρόνος Εκπαίδευσης	περίπου τρεις (3) ημέρες σε 24GB GPU

1 Σύνοψη της Εργασίας

2 Παραγωγική Μοντελοποίηση & GANs

3 Εφαρμογή σε Εικόνες Μόδας

Αλλαγή Πόζας (PGPG - *PoseGAN*)

Εξαγωγή Ρούχου (PixelDTGAN)

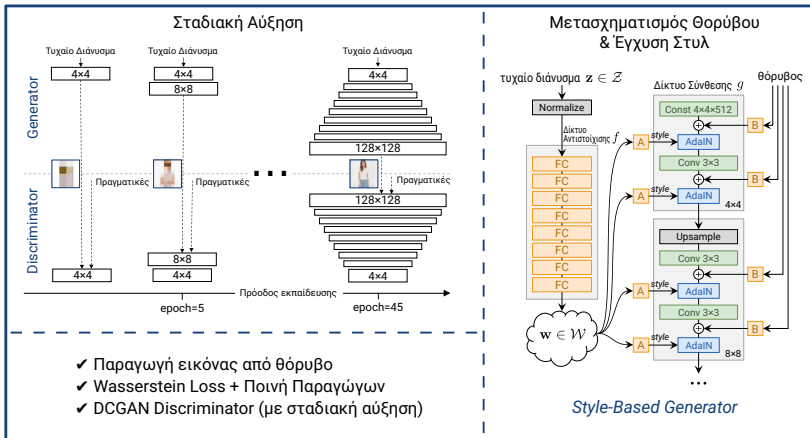
Ταίριασμα Στιλ (DiscoGAN - *CycleGAN*)

Παραγωγή Ρεαλιστικών Εικόνων (*StyleGAN*)

4 Αποτελέσματα & Αξιολόγηση

Παραγωγή Ρεαλιστικών Εικόνων (StyleGAN)

Αρχιτεκτονική Μοντέλου

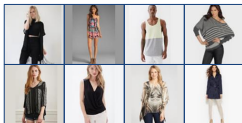


Σχήμα 4: Στοιχεία της αρχιτεκτονικής του μοντέλου StyleGAN^{GT}

Παραγωγή Ρεαλιστικών Εικόνων (StyleGAN)

Σύνοψη Μοντέλου - Παράμετροι Εκπαίδευσης

Σύνοψη του μοντέλου StyleGAN	
Όνομα Μοντέλου	StyleGAN ^{GT}
Κωδικός Configuration	default_z512
Εφαρμογή	Παραγωγή ρεαλιστικών εικόνων ανθρώπων σε φωτογραφικές μόδας
Κατηγορία Εφαρμογής	Παραγωγή Εικόνας από Θόρυβο
Κατηγορία Παραγωγής	χωρίς συνθήκη
Σχετικό Μοντέλο	StyleGAN
Σχετικό Άρθρο	«A Style-Based Generator Architecture for Generative Adversarial Networks» [91]
Αριθμός GANs	ένα (1)
Αριθμός Generators	έναν (1) - αποτελούμενος από το δίκτυο αντιστοίχισης και το δίκτυο σύνθεσης
Αριθμός Discriminators	έναν (1)
Τύπος Generator(s)	Style-based με σταδιακή αύξηση progressive growing (5 blocks + έγχυση θορύβου + επιβολή στιλ στις στρώσεις AdaIN)
Τύπος Discriminator(s)	StyleGAN Discriminator με σταδιακή αύξηση (5 blocks, 128x128 receptive field)



Ενδεικτικές εικόνες που δίνονται στο μοντέλο από τον φορτωτή δεδομένων
Σύνολο Δεδομένων: **Fashion-Image Synthesis Benchmark (DeepFashion)**

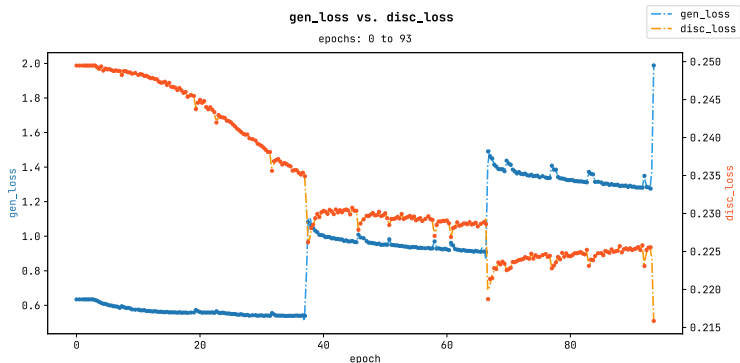
Παράμετροι εκπαίδευσης του μοντέλου StyleGAN	
Όνομα Μοντέλου	StyleGAN ^{GT}
Κωδικός Configuration	default_z512
Συναρτ. Κόστους Generator	Wasserstein (WGAN) (αντιπαραθετική)
Συναρτ. Κόστους Discriminator(s)	Wasserstein (WGAN) + Ποινή Παραγώγων με βάρος $\lambda_{DP}=10$
Αριθμός βελτ/τών για Generator(s)	έναν (1)
Αριθμός βελτ/τών για Discriminator(s)	έναν (1)
Τύπος βελτ/τών για Generator(s)	Adam, με παραμέτρους ($\text{lr}=0.0002, \beta_1=0.9, \beta_2=0.999$) LR Scheduler: OnPlateau(factor: 0.99, cooldown: 100) ανά βήμα (όχι epoch)
Τύπος βελτ/τών για Discriminator(s)	Adam, με παραμέτρους ($\text{lr}=0.0001, \beta_1=0.9, \beta_2=0.999$) LR Scheduler: OnPlateau(factor: 0.99, cooldown: 100) ανά βήμα (όχι epoch)
Σύνολο δεδομένων εκπαίδευσης	FISB (DeepFashion)
Μέγεθος συνόλου δεδομένων	79K εικόνες (χρησιμοποιήθηκαν 53.2K εικόνες με χρώμα παρασκήνιο πιο λευκό από #000000 (hex))
Ανάλυση εικόνων	128x128px
Μέγεθος ομάδας	28 εικόνες/batch (στην ανάλυση 128x128)
Αριθμός epochs	149 epochs (186.750 επαναλήψεις)
Χρόνος Εκπαίδευσης	περίπου επτά (7) μέρες σε 24GB GPUs
# Παραμέτρων Generator(s)	58.5M εκπαιδούμενες παράμετροι (G1: 276M, G2: 117.4M)
# Παραμέτρων Discriminator(s)	11.3M εκπαιδούμενες παράμετροι
# Παραμέτρων	69.8M εκπαιδούμενες παράμετροι συνολικά

- 1 Σύνοψη της Εργασίας
- 2 Παραγωγική Μοντελοποίηση & GANs
- 3 Εφαρμογή σε Εικόνες Μόδας
- 4 **Αποτελέσματα & Αξιολόγηση**
 - Αλλαγή Πόζας (PGPG - *PoseGAN*)
 - Εξαγωγή Ρούχου (PixelDTGAN)
 - Ταίριασμα Στιλ (CycleGAN)
 - Παραγωγή Ρεαλιστικών Εικόνων (StyleGAN)

- 1 Σύνοψη της Εργασίας
- 2 Παραγωγική Μοντελοποίηση & GANs
- 3 Εφαρμογή σε Εικόνες Μόδας
- 4 **Αποτελέσματα & Αξιολόγηση**
 - Αλλαγή Πόζας (PGPG - *PoseGAN*)
 - Εξαγωγή Ρούχου (PixelDTGAN)
 - Ταίριασμα Στιλ (CycleGAN)
 - Παραγωγή Ρεαλιστικών Εικόνων (StyleGAN)

Αλλαγή Πόζας (PGPG - PoseGAN)

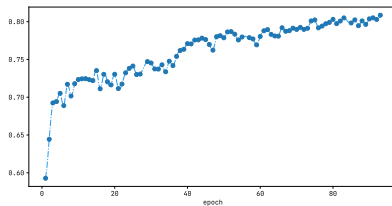
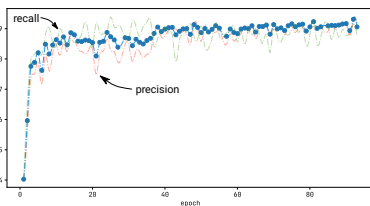
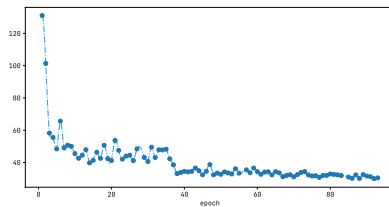
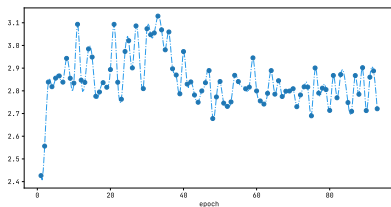
Καμπύλες Εκπαίδευσης



Σχήμα 5: Καμπύλες εκπαίδευσης του PoseGAN. Φαίνεται η εξέλιξη των συναρτήσεων κόστους του Generator και του Discriminator ως προς τα epochs της εκπαίδευσης και οι απότομες μεταβολές αυτών κατά την αλλαγή του λ_{recon} .

Αλλαγή Πόζας (PGPG - PoseGAN)

Καμπύλες Εξέλιξης Μετρικών Αξιολόγησης



Σχήμα 6: Καμπύλες εξέλιξης των μετρικών αξιολόγησης του *PoseGAN* κατά τη διάρκεια της εκπαίδευσής του.

Αλλαγή Πόζας (PGPG - PoseGAN)

Παραγωγές Εικόνων



Σχήμα 7: Παραγωγές του μοντέλου PoseGAN, ανάλυσης 128x128px.

Αλλαγή Πόζας (PGPG - PoseGAN)

Τελικές Μετρικές Αξιολόγησης - Σύγκριση με το σχετικό άρθρο

Πίνακας 1: Συγκρίσεις μετρικών του *PoseGAN* με του PGPG από το σχετικό άρθρο. Όπως φαίνεται σε όλες τις μετρικές το μοντέλο μας παρουσιάζει υπεροχή με εξαίρεση την οριακά χειρότερη μετρική Inception Score στο test set.

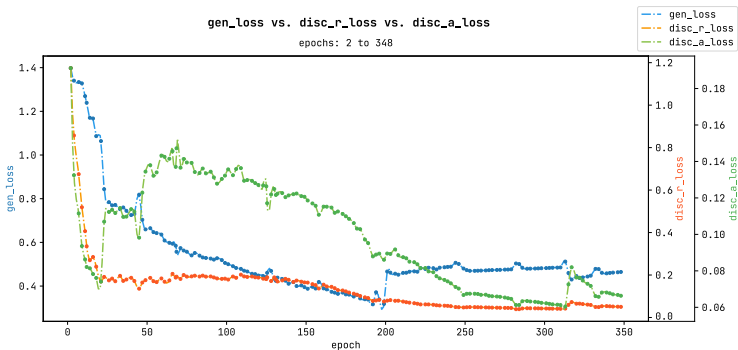
	PGPG από [2]	<i>PoseGAN</i> ^{GT}
FID	-	16.19 (train) - 26.50 (test)
IS	3.09	3.83 (train) - 2.96 (test)
SSIM	0.762	0.803 (train) - 0.769 (test)
Precision	-	0.886(train) - 0.835 (test)
Recall	-	0.880(train) - 0.864 (test)
F1	-	0.882(train) - 0.849 (test)

[2] L. Ma, X. Jia, Q. Sun, *et al.*, «Pose Guided Person Image Generation», *arXiv e-prints*, arXiv:1705.09368, arXiv:1705.09368, May 2017. arXiv: 1705 . 09368 [cs.CV]

- 1 Σύνοψη της Εργασίας
- 2 Παραγωγική Μοντελοποίηση & GANs
- 3 Εφαρμογή σε Εικόνες Μόδας
- 4 Αποτελέσματα & Αξιολόγηση**
 - Αλλαγή Πόζας (PGPG - *PoseGAN*)
 - Εξαγωγή Ρούχου (PixelDTGAN)
 - Ταίριασμα Στιλ (CycleGAN)
 - Παραγωγή Ρεαλιστικών Εικόνων (StyleGAN)

Εξαγωγή Ρούχου (PixelDTGAN)

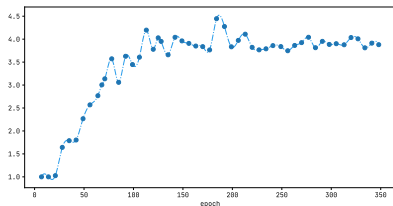
Καμπύλες Εκπαίδευσης



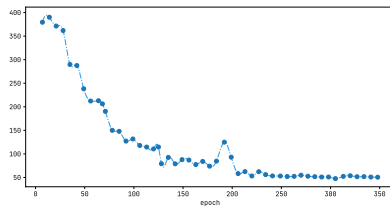
Σχήμα 8: Καμπύλες εκπαίδευσης του PixelDTGAN^{GT}. Φαίνεται η εξέλιξη των συναρτήσεων κόστους του Generator και των δύο Discriminators ως προς τα epochs της εκπαίδευσης και οι απότομες μεταβολές αυτών κατά την αλλαγή του λ_{recon} .

Εξαγωγή Ρούχου (PixelDTGAN)

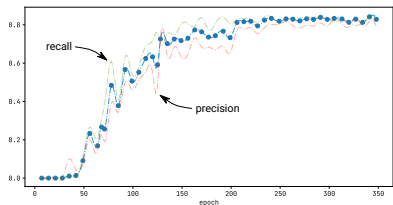
Καμπύλες Εξέλιξης Μετρικών Αξιολόγησης



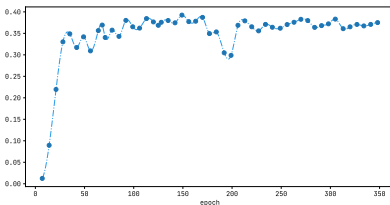
(α) Inception Score (IS)



(β) Fréchet Inception Distance (FID)



(γ) F₁-Score



(δ) Structural Similarity Index (SSIM)

Σχήμα 9: Καμπύλες εξέλιξης των μετρικών αξιολόγησης του PixelDTGAN^{GT} κατά τη διάρκεια εκπαίδευσης.

Εξαγωγή Ρούχου (PixelDTGAN)

Παραγωγές Εικόνων



Σχήμα 10: Παραγωγές της υλοποίησής μας του μοντέλου PixelDTGAN^{GT} μετά το epoch 345. Όλες οι εικόνες είναι ανάλυσης 64×64px.

Εξαγωγή Ρούχου (PixelDTGAN)

Τελικές Μετρικές Αξιολόγησης - Σύγκριση με το σχετικό άρθρο

Πίνακας 2: Συγκρίσεις μετρικών του PixelDTGAN^{GT} με του PixelDTGAN από το σχετικό άρθρο. Όπως φαίνεται σε όλες τις μετρικές το μοντέλο μας παρουσιάζει υπεροχή αν και οι συγγραφείς επέλεξαν να μην αξιολογήσουν εξονυχιστικά το μοντέλο τους.

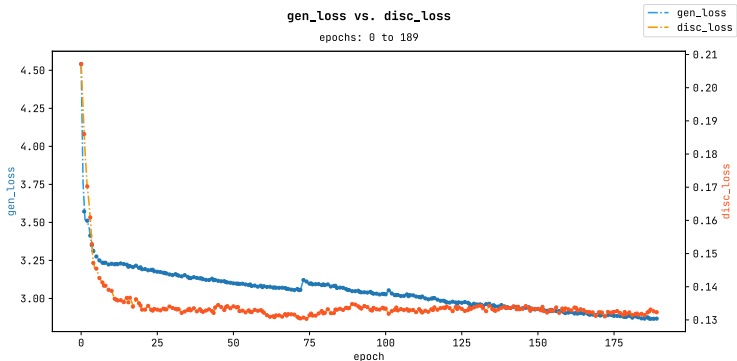
	PixelDTGAN από [1]	PixelDTGAN ^{GT}
FID	-	36.04 (test)
IS	-	4.04 (test)
SSIM	0.21	0.381 (test)
Precision	-	0.781 (test)
Recall	-	0.814 (test)
F1	-	0.798 (test)

[1] D. Yoo, N. Kim, S. Park, *et al.*, «Pixel-Level Domain Transfer», *arXiv e-prints*, arXiv:1603.07442, arXiv:1603.07442, Mar. 2016. arXiv: 1603.07442 [cs.CV]

- 1 Σύνοψη της Εργασίας
- 2 Παραγωγική Μοντελοποίηση & GANs
- 3 Εφαρμογή σε Εικόνες Μόδας
- 4 Αποτελέσματα & Αξιολόγηση**
 - Αλλαγή Πόζας (PGPG - *PoseGAN*)
 - Εξαγωγή Ρούχου (PixelDTGAN)
 - Ταίριασμα Στιλ (CycleGAN)**
 - Παραγωγή Ρεαλιστικών Εικόνων (StyleGAN)

Ταίριασμα Στιλ (CycleGAN)

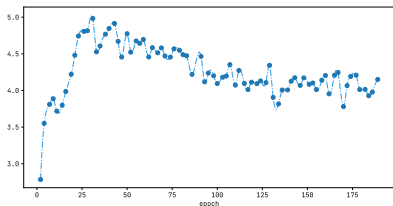
Καμπύλες Εκπαίδευσης



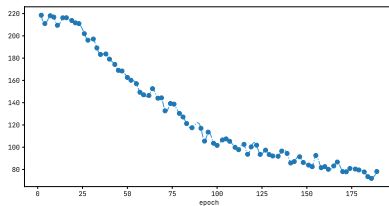
Σχήμα 11: Καμπύλες εκπαίδευσης του μοντέλου CycleGAN^{GT}. Φαίνεται η εξέλιξη των συναρτήσεων κόστους των Generators και των Discriminators ως προς τα epochs της εκπαίδευσης του μοντέλου.

Ταίριασμα Στιλ (CycleGAN)

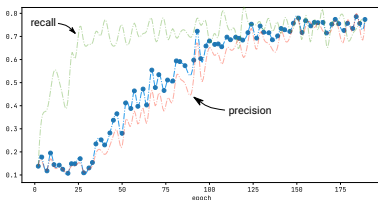
Καμπύλες Εξέλιξης Μετρικών Αξιολόγησης



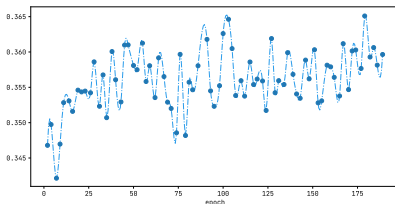
(α) Inception Score (IS)



(β) Fréchet Inception Distance (FID)



(γ) F₁-Score



(δ) Structural Similarity Index (SSIM)

Σχήμα 12: Καμπύλες εξέλιξης των μετρικών αξιολόγησης του μοντέλου CycleGAN^{GT} κατά τη διάρκεια της εκπαίδευσής του.

Ταίριασμα Στιλ (CycleGAN)

Παραγωγές Εικόνων



Σχήμα 13: Παραγωγές του μοντέλου CycleGAN^{GT} μετά το epoch 188. Όλες οι εικόνες είναι ανάλυσης 64×64px.

Ταίριασμα Στιλ (CycleGAN)

Τελικές Μετρικές Αξιολόγησης

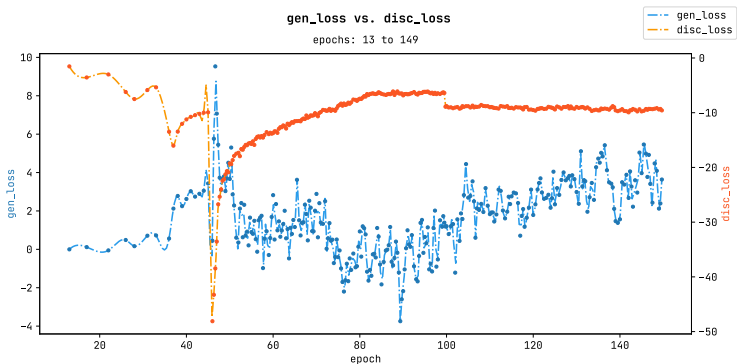
Πίνακας 3: Τελικές μετρικές αξιολόγησης του CycleGAN^{GT}.

	CycleGAN ^{GT}
FID	65.78 (train) - 66.06 (test)
IS	4.48 (train) - 4.42 (test)
SSIM	0.355 (train) - 0.353 (test)
Precision	0.607 (train) - 0.607 (test)
Recall	0.920 (train) - 0.918 (test)
F1	0.731 (train) - 0.730 (test)

- 1 Σύνοψη της Εργασίας
- 2 Παραγωγική Μοντελοποίηση & GANs
- 3 Εφαρμογή σε Εικόνες Μόδας
- 4 Αποτελέσματα & Αξιολόγηση
 - Αλλαγή Πόζας (PGPG - *PoseGAN*)
 - Εξαγωγή Ρούχου (PixelDTGAN)
 - Ταίριασμα Στιλ (CycleGAN)
 - Παραγωγή Ρεαλιστικών Εικόνων (StyleGAN)

Παραγωγή Ρεαλιστικών Εικόνων (StyleGAN)

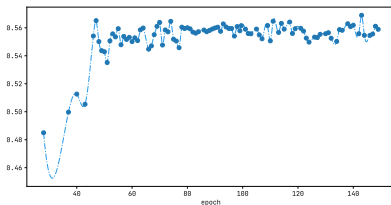
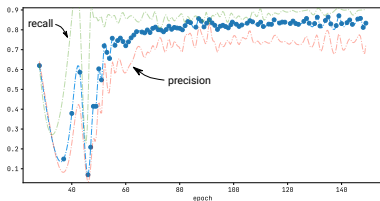
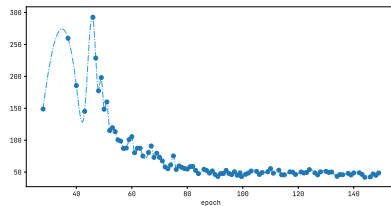
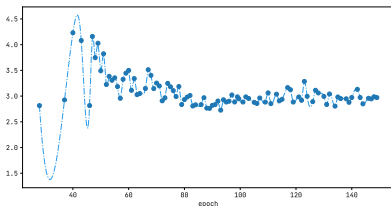
Καμπύλες Εκπαίδευσης



Σχήμα 14: Καμπύλες εκπαίδευσης του μοντέλου StyleGAN^{GT}. Φαίνεται η εξέλιξη των συναρτήσεων κόστους του Generator και του Discriminator ως προς τα epochs εκπαίδευσης. Αρχικές απότομες μεταβολές: σταδιακή αύξηση. Η πτώση στο epoch=98: αλλαγή βήματος εκπαίδευσης.

Παραγωγή Ρεαλιστικών Εικόνων (StyleGAN)

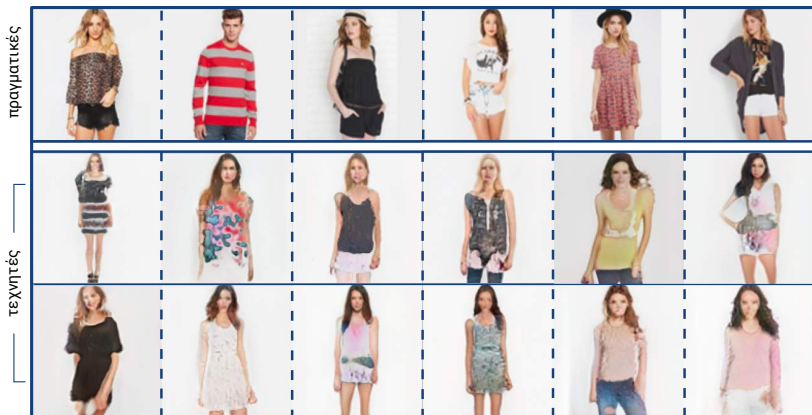
Καμπύλες Εξέλιξης Μετρικών Αξιολόγησης



Σχήμα 15: Καμπύλες εξέλιξης των μετρικών αξιολόγησης του μοντέλου StyleGAN^{GT} κατά τη διάρκεια της εκπαίδευσής του.

Παραγωγή Ρεαλιστικών Εικόνων (StyleGAN)

Παραγωγές Εικόνων



Σχήμα 16: Παραγωγές του μοντέλου StyleGAN^{GT} μετά το epoch 140. Όλες οι εικόνες είναι ανάλυσης 128×128px.

Παραγωγή Ρεαλιστικών Εικόνων (StyleGAN)

Τελικές Μετρικές Αξιολόγησης

Πίνακας 4: Τελικές μετρικές αξιολόγησης του StyleGAN^{GT}.

	StyleGAN ^{GT}
FID	34.21 (train) - 33.01 (test)
IS	2.979 (train) - 2.987 (test)
SSIM	0.562 (train) - 0.561 (test)
Precision	0.783 (train) - 0.779 (test)
Recall	0.864 (train) - 0.824 (test)
F1	0.831 (train) - 0.801 (test)

Σας ευχαριστώ για την προσοχή σας!

Ευχαριστώ θερμά τον καθ. κ. Μήτκα και τον κ. Αντώνη Χρυσόπουλο για την εμπιστοσύνη και τη διαρκή καθοδήγηση τους. Επίσης ευχαριστώ τον κ. Αλέξανδρο Κυπριανίδη και τον κ. Σωτήρη Τσαρούχη για την υποστήριξη και τις συμβουλές τους.

Ο κώδικας της παρούσας διπλωματικής εργασίας, ο οποίος ξεπερνάει τις 15.5K LoC, δίνεται ανοιχτά αλλά χωρίς ευθύνη στο ακόλουθο αποθετήριο κώδικα του GitHub:

`github.com/achariso/gans-thesis`

Πρότυπο παρουσίασης: <https://github.com/Trinkle23897/THU-Beamer-Theme>