



ΤΜΗΜΑ ΜΗΧ. Η/Υ & ΠΛΗΡΟΦΟΡΙΚΗΣ  
ΠΑΝΕΠΙΣΤΗΜΙΟ ΙΩΑΝΝΙΝΩΝ

---

DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING  
UNIVERSITY OF IOANNINA

Διπλωματική Εργασία

---

## **Automatic video colorization using generative adversarial networks**

---

Σάμιος Γεώργιος  
ΑΜ: 3326

Επιβλέπων:  
Νίκου Χριστόφορος

Συνεπιβλέπων:  
Σφήκας Γεώργιος

Ιωάννινα, Φεβρουάριος 2022

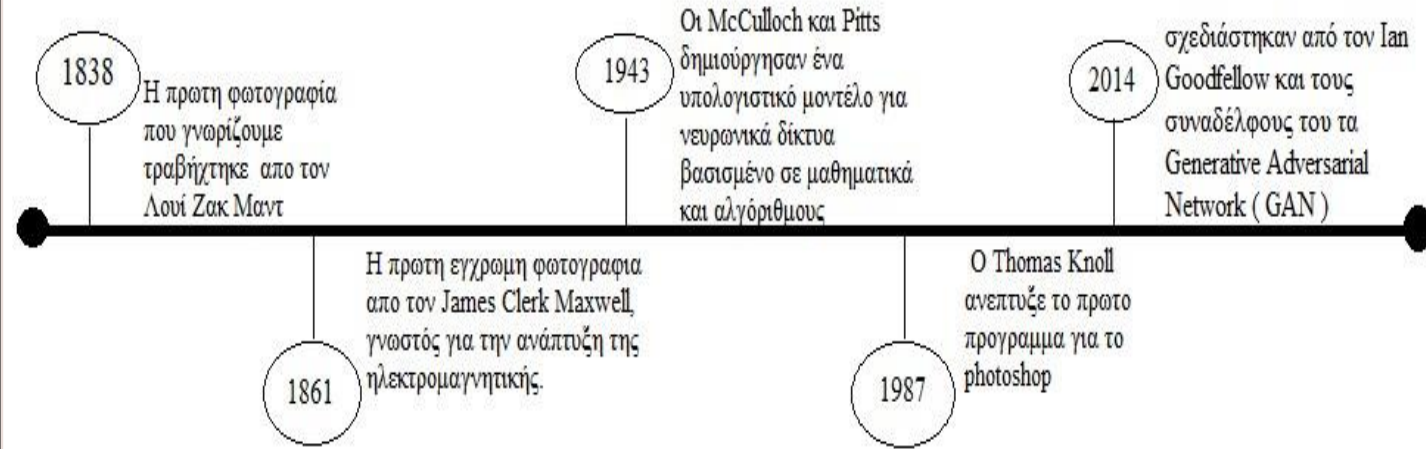


# Επιχρωματισμός ασπρόμαυρου Βίντεο & Στόχος Διπλωματικής

## Στόχος διπλωματικής εργασίας:

- Ανάλυση της τεχνικής GAN's για τον χρωματισμό ασπρόμαυρου βίντεο.
- Προσπάθεια υλοποίησης ενός εκτελέσιμου προγράμματος για χρωματισμό.

Εξέλιξη Νευρωνικών Δικτύων και  
Υπολογιστικής Όρασης  
+  
Προσπάθεια εξέλιξης χρωμάτων εικόνων  
↓  
Αυτόματος χρωματισμός βίντεο/εικόνων



# Στόχοι Χρωματισμού Ασπρόμαυρων Εικόνων

## Στους θεατές

- Περισσότερο ρεαλιστικές εικόνες
- Έμφαση στη λεπτομέρεια
- Ενεργοποίηση του ενδιαφέροντος του θεατή
- Καλύτερη ποιότητα εικόνας



## Στους επιστήμονες

- Διόρθωση λαθών και βελτίωση των ήδη υπάρχουσών τεχνολογιών επιχρωματισμού
- Ανάπτυξη νέων τεχνικών





# ΚΑΤΗΓΟΡΊΕΣ ΔΙΕΡΓΑΣΙΩΝ ΚΑΙ Ο ΕΠΙΧΡΩΜΑΤΙΣΜΟΣ ΩΣ ΔΙΕΡΓΑΣΊΑ

## Καλά ορισμένες διεργασίες:

- Επίλυση παζλ
- Απόδειξη μαθηματικών θεωρημάτων
- Επιτραπέζια παιχνίδια



## Διεργασίες ειδικών:

- Ιατρικές διαγνώσεις
- Σχεδίαση υπολογιστικού υλικού

## Τετριμμένες διεργασίες:

- Ομιλία
- Αντίληψη
- Γραφή
- Περπάτημα
- Αφή



# ΣΕ ΠΟΙΑ ΚΑΤΗΓΟΡΙΑ ΤΕΛΙΚΑ ΑΝΉΚΕΙ Ο ΕΠΙΧΡΩΜΑΤΙΣΜΌΣ ΑΣΠΡΌΜΑΥΡΟΥ ΒΊΝΤΕΟ;

Είναι πιο κοντά στις καλά ορισμένες διεργασίες; στις τετριμμένες διεργασίες; ή στις διεργασίες ειδικών.

Το ερώτημα παραμένει  
ανοιχτό καθώς ανάλογα με  
την τεχνική που  
χρησιμοποιείται  
υπάρχουν ενέργειες που  
μπορούν να οριστούν  
αναλόγως την κατάσταση  
και άλλες που απαιτούν ένα  
βαθμό συλλογιστικής σκέψης  
και αντίληψης.

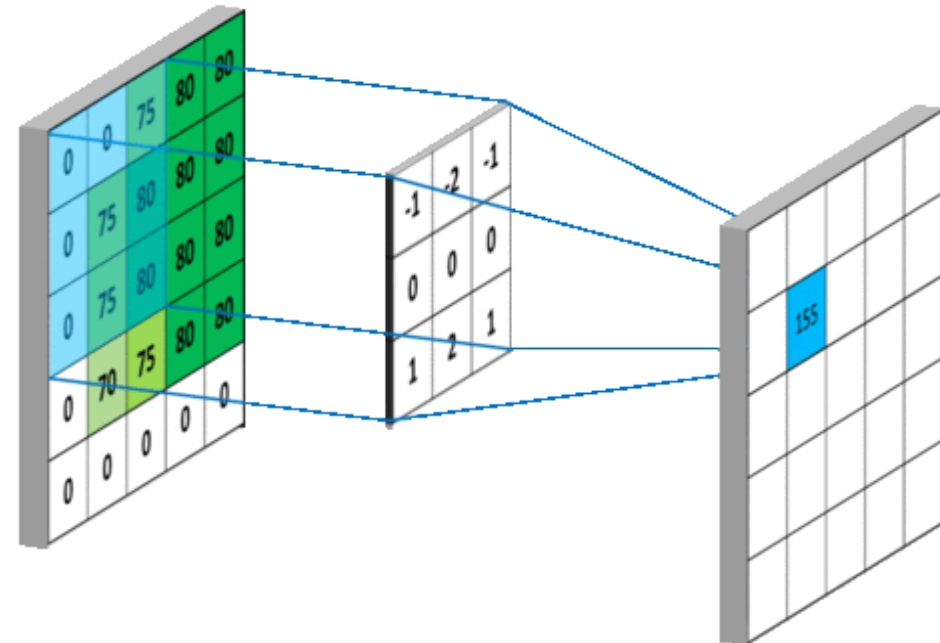
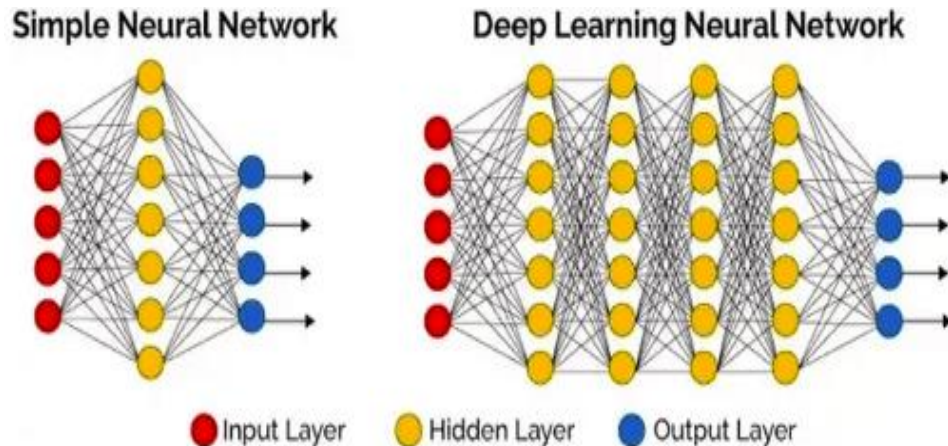


# ΤΙ ΕΙΝΑΙ ΤΑ ΝΕΥΡΩΝΙΚΑ ΔΙΚΤΥΑ ΚΑΙ ΓΙΑΤΙ ΕΙΝΑΙ “ΒΑΘΙΑ”;

Νευρωνικό δίκτυο ονομάζεται ένα κύκλωμα διασυνδεδεμένων νευρώνων το οποίο κατασκευάζεται για να εκπαιδεύεται και να επιλύει προβλήματα.

Τα βαθιά νευρωνικά δίκτυα είναι απλά νευρωνικά δίκτυα τα οποία περιέχουν πολλά επίπεδα υπολογισμού.

Τα Generative Adversarial Networks (GAN's) είναι παραγωγικά μοντέλα νευρωνικών δικτύων που βασίζονται στη βαθιά μάθηση.



# ΠΕΡΙΓΡΑΦΉ ΔΟΜΉΣ GAN ΜΟΝΤΕΛΩΝ

Η αρχιτεκτονική του μοντέλου GAN περιλαμβάνει δύο υπομοντέλα:

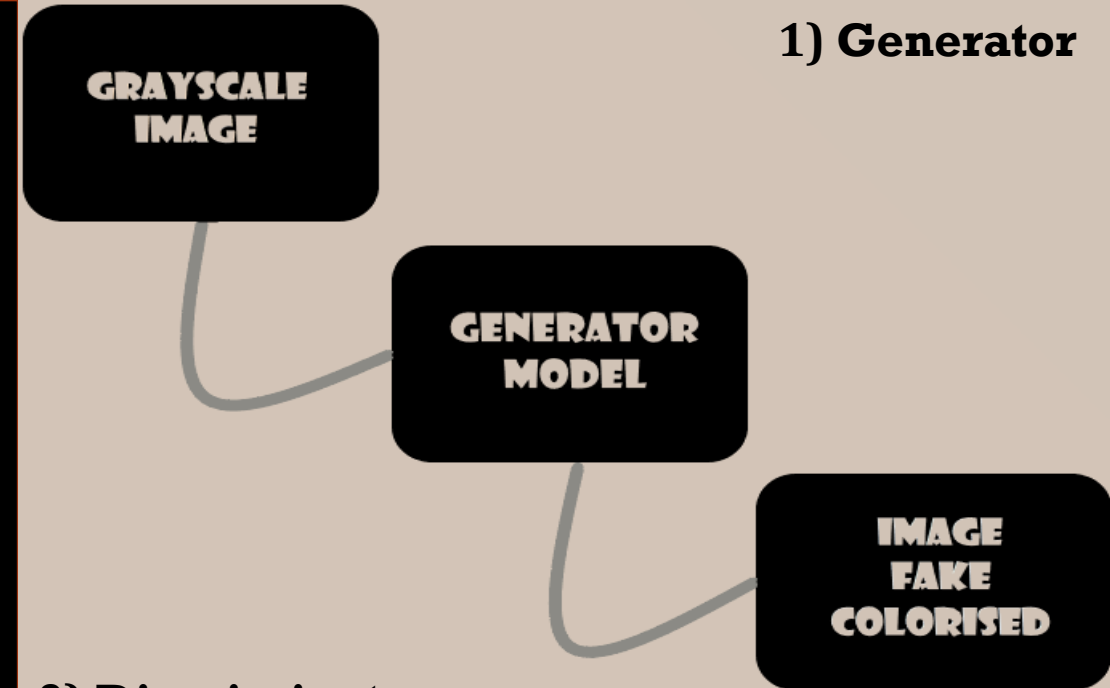
- **Γεννήτρια:** Μοντέλο που χρησιμοποιείται για τη δημιουργία νέων εύλογων παραδειγμάτων από τον τομέα του προβλήματος.
- **Διακριτής:** Μοντέλο που χρησιμοποιείται για την ταξινόμηση παραδειγμάτων ως πραγματικών ( από τον τομέα ) ή πλαστών ( δημιουργημένα ).

Τα δύο μοντέλα, το Generator και το Discriminator, εκπαιδεύονται μαζί.

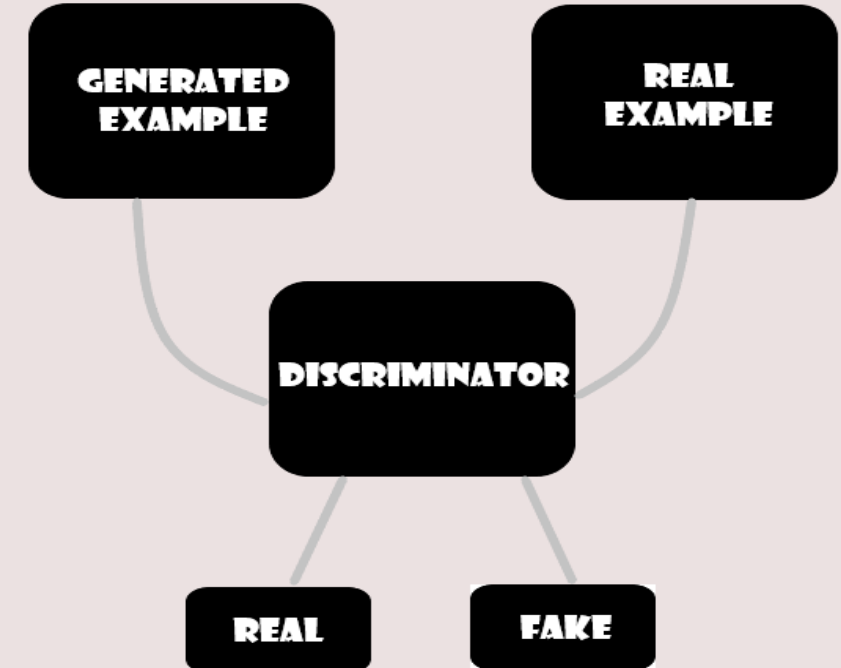
Ο Generator δημιουργεί μια παρτίδα πλαστών δειγμάτων και αυτά, μαζί με τα πραγματικά παραδείγματα από τον τομέα, παρέχονται στον υπεύθυνο διάκρισης-Discriminator και ταξινομούνται ως αληθινά ή πλαστά.

Αφού γίνει η διάκριση και βρεθεί ένα δείγμα ως πλαστό τότε ο Discriminator “τιμωρεί” τον Generator με ένα σφάλμα και επιτυγχάνεται η εκπαίδευση του

## 1) Generator



## 2) Discriminator





# ΥΛΟΠΟΙΗΣΗ ΠΡΟΓΡΑΜΜΑΤΟΣ

Η διαδικασία υλοποίησης του παραπάνω προγράμματος περιλαμβάνει τα παρακάτω βασικά στάδια:

1. Λήψη του συνόλου δεδομένων.
2. Χωρισμός των δεδομένων σε σύνολα εκπαίδευσης-επικύρωσης-τεσταρίσματος.
3. Επιλογή της κατάλληλης αρχιτεκτονικής για την Γεννήτρια(**Generator**).
4. Επιλογή της κατάλληλης αρχιτεκτονικής για τον Διακριτή(**Discriminator**).
5. Εκπαίδευση του δικτύου.
6. Αξιολόγηση των αποτελεσμάτων του δικτύου και επανάληψη της παραπάνω διαδικασίας σε περίπτωση που τα αποτελέσματα δεν είναι τα επιθυμητά.

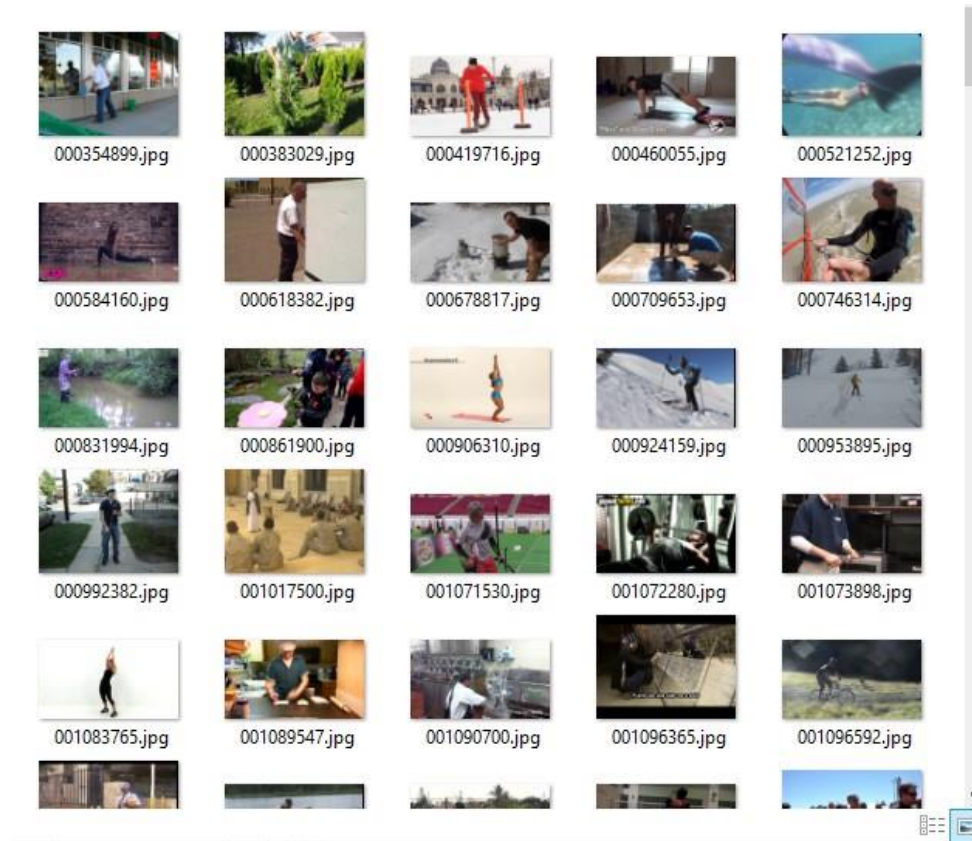




# ΛΗΨΗ ΣΥΝΟΛΟΥ ΕΚΠΑΙΔΕΥΣΗΣ

Αρχικά το σύστημα μας εκπαιδεύτηκε και δοκιμάστηκε σε χίλια δεδομένα από μια βάση δεδομένων, με ανθρώπους ,που λήφθηκε από το διαδίκτυο.

1. Για την εκπαίδευση αρχικά κατεβάσαμε από το διαδίκτυο ένα Dataset με εικόνες από ανθρώπους και τοπία.
2. Αποθηκεύουμε τις εικόνες, αλλά επειδή το σύνολο περιέχει πολλές χιλιάδες εικόνες προσαρμόζουμε τον αριθμό τους στις δικές μας απαιτήσεις και κρατάμε τις εικόνες με διαστάσεις 256\*256.



1500 φωτογραφίες



# ΧΩΡΙΣΜΟΣ ΣΥΝΟΛΩΝ ΕΚΠΑΙΔΕΥΣΗΣ-ΕΠΙΚΥΡΩΣΗΣ-ΤΕΣΤΑΡΙΣΜΑΤΟΣ

Χωρίζουμε τις παραπάνω εικόνες σε σύνολο εκπαίδευσης σύνολο επικύρωσης και σύνολο τεσταρίσματος.



250 αρχεία, 0 φακέλους



1.000 αρχεία, 0 φακέλους

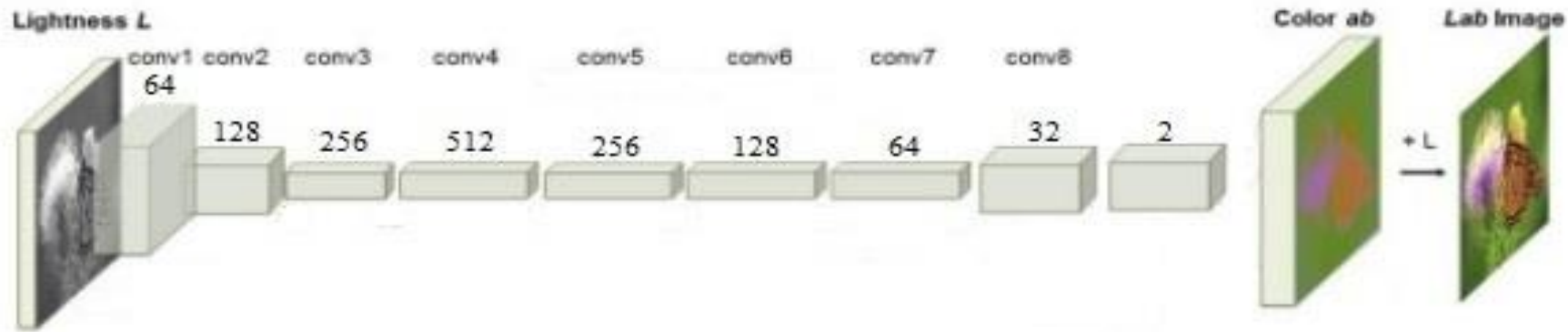


250 αρχεία, 0 φακέλους

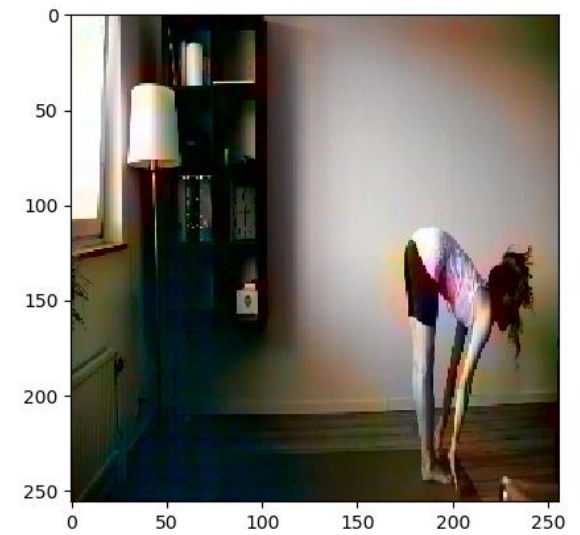
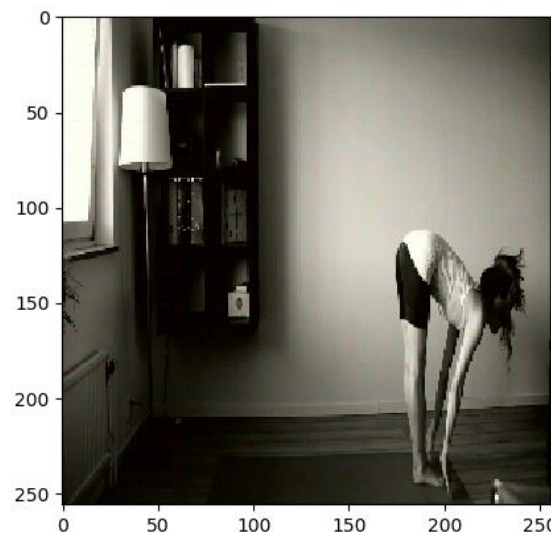


# ΕΠΙΛΟΓΗ ΑΡΧΙΤΕΚΤΟΝΙΚΗΣ GENERATOR

Η επιλογή για την κατάλληλη αρχιτεκτονική είναι μια δύσκολη διαδικασία, καθώς για την λήψη της τελικής απόφασης για το ποια αρχιτεκτονική θα χρησιμοποιηθεί για να φέρουμε εις πέραν τον στόχο του έργου πρέπει να εκπαιδευτούν όλα τα δίκτυα και να ελεγχθούν ως προς την ικανότητα γενίκευσης.



Σκοπός της γεννήτριας (Generator) είναι ο ψευδής χρωματισμός μιας ασπρόμαυρης εικόνας ώστε να προσπαθήσει να ξεγελάσει τον διακριτή.





# LAB VS RGB

Για την γρηγορότερη εκπαίδευση του δικτύου λόγο χαμηλών δυνατοτήτων υπολογιστή που διαθέταμε, χρησιμοποιήσαμε Lab εικόνες σαν έξοδο της γεννήτριας αντί για RGB.

## Γιατί Lab αντί RGB?

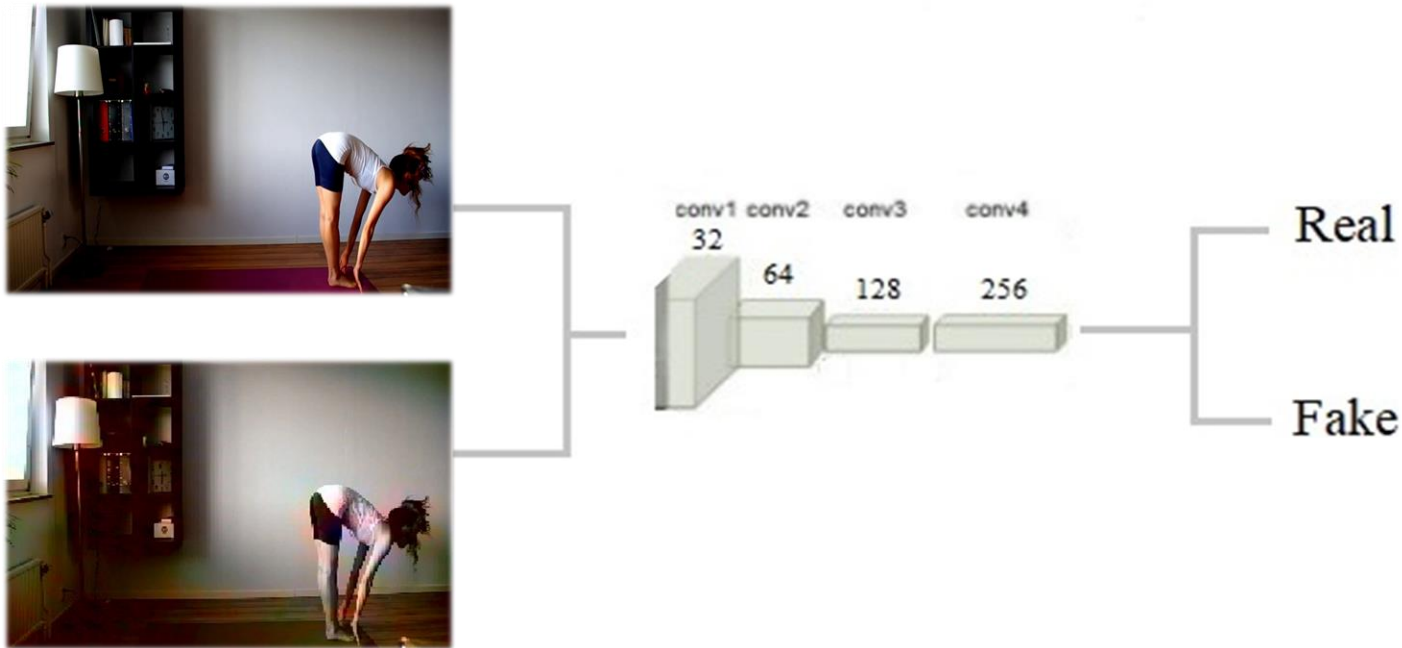
Εάν χρησιμοποιούσαμε rgb εικόνες το δίκτυο μας θα έπαιρνε ως είσοδο την ασπρόμαυρη εικόνα και θα έπρεπε να κάνει πρόβλεψη τριών (3) καναλιών του r, του g και του b.

Σε αντίθεση με την χρήση Lab εικόνων που ως είσοδο δίνουμε το L κανάλι (που είναι η εικόνα σε κλίμακα του γκρι) και θέλουμε να γίνει πρόβλεψη των άλλων δυο (2) καναλιών  $a^*$  και  $b^*$ .



# ΕΠΙΛΟΓΗ ΑΡΧΙΤΕΚΤΟΝΙΚΗΣ DISCRIMINATOR

Η επιλογή για την κατάλληλη αρχιτεκτονική, και η λήψη της τελικής απόφασης για το ποια αρχιτεκτονική θα χρησιμοποιηθεί τελικά είναι παρόμοια διαδικασία με αυτήν της γεννήτριας .



Σκοπός του διακριτή (Discriminator) είναι η αναγνώριση των αληθινών φωτογραφιών από τα πλαστά αντίγραφα που παράγει η γεννήτρια. Μέσω του διακριτή γίνεται και η εκπαίδευση της γεννήτριας καθώς κάθε φορά που η έξοδος της δεν “ξεγελάει” τον διακριτή, τότε στέλνει το σφάλμα στην στην γεννήτρια.



# ΕΚΠΑΪΔΕΥΣΗ ΤΟΥ ΓΑΝ ΔΙΚΤΥΟΥ

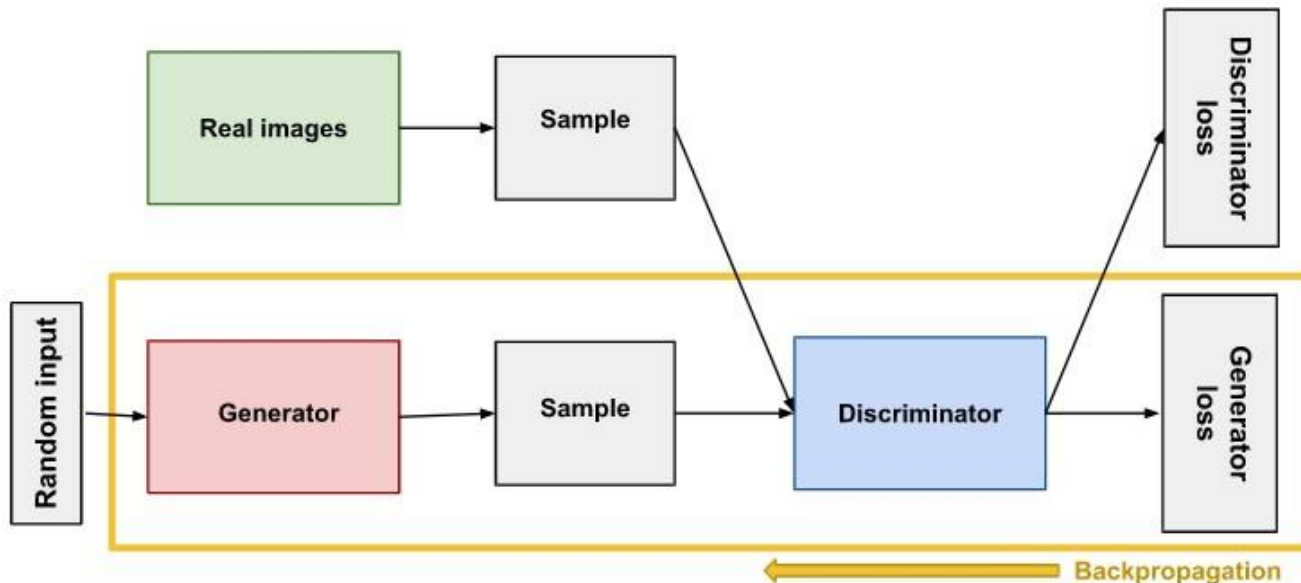
Για την κατασκευή και την εκπαίδευση ολόκληρου του δικτύου ακολουθούνται τα παρακάτω βήματα:

**Για την κατασκευή:**

- Δημιουργία συνόλου εκπαίδευσης, επικύρωσης και τεσταρίσματος (Dataset)
- Κατασκευή γεννήτριας (Generator)
- Κατασκευή Διακριτή (Discriminator)
- Συνένωση δικτύων

**Για την εκπαίδευση:**

- Εντοπισμός έγχρωμης εικόνας
- Μετατροπή σε ασπρόμαυρη
- Εισαγωγή φωτογραφίας στην γεννήτρια και παραγωγή της πλαστής φωτογραφίας
- Εισαγωγή πλαστής και αληθινής φωτογραφίας στον διακριτή.
- Απόφαση διακριτή για αληθινή και πλαστή φωτογραφία
- Υπολογισμός σφάλματος για την γεννήτρια και τον διακριτή.
- Διάδοση του σφάλματος της γεννήτριας ώστε να γίνει ενημέρωση βαρών.





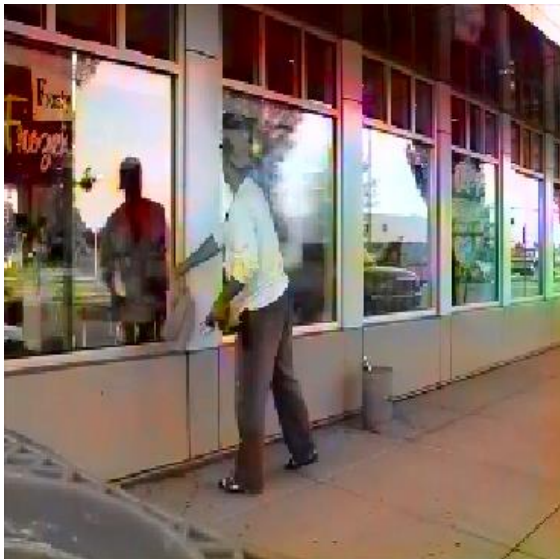
# ΑΞΙΟΛΟΓΗΣΗ ΤΩΝ ΑΠΟΤΕΛΕΣΜΑΤΩΝ ΤΟΥ ΣΥΣΤΗΜΑΤΟΣ



Για την αξιολόγηση των αποτελεσμάτων του συστήματος, θα χρησιμοποιήσουμε εικόνες από το σύνολο επικύρωσης.

Κριτήρια αξιολόγησης:

- Ακρίβεια
- Ευρωστία
- Αξιοπιστία
- Ευαισθησία
- Προσαρμοστικότητα
- Αποτελεσματικότητα

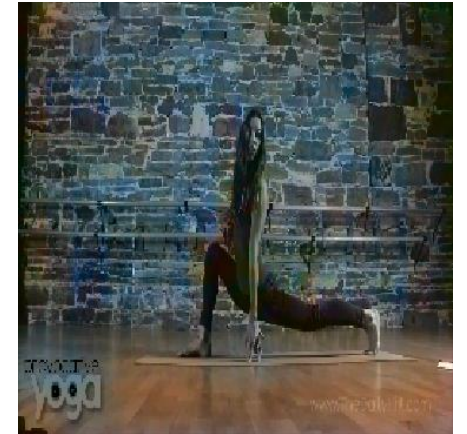


Μετρικές που χρησιμοποιούνται για την αξιολόγηση χρωματισμού εικόνας:

- **PSNR:** Το PSNR μετριέται σε ντεσιμπέλ. Μια υψηλότερη τιμή PSNR υποδηλώνει υψηλότερη ποιότητα ανακατασκευής.
- **SSIM:** Η SSIM κυμαίνεται από 0 έως 1. Η τιμή 1 υποδηλώνει ταυτότητα της αρχικής και της παραγόμενης εικόνας.



# ΧΡΩΜΑΤΙΣΜΟΣ ΑΣΠΡΟΜΑΥΡΟΥ ΒΙΝΤΕΟ



Παρακάτω παρουσιάζεται ένα παράδειγμα χρωματισμού ενός ασπρόμαυρου βίντεο με βάση την τελική μορφή του εκπαιδευμένου δικτύου.



Prediction



# ΠΡΟΒΛΗΜΑΤΑ ΚΑΙ ΙΔΕΕΣ ΒΕΛΤΙΩΣΗΣ ΤΟΥ ΣΥΣΤΗΜΑΤΟΣ

1. Ένα από τα σημαντικότερα προβλήματα ήταν η έλλειψη μιας καλής κάρτας γραφικών με μεγάλη μνήμη ram



Για να καταφέρουμε να αντιμετωπίσουμε αυτό το πρόβλημα πρέπει να αντικαταστήσουμε τα γραφικά του επεξεργαστή με μια κάρτα γραφικών με 6GB και πάνω μνήμη ram

2. Ποιότητα η χρόνος;  
Εκπαίδευση του δικτύου με εικόνες με διαστάσεις 128\*128 τότε ο συνολικός χρόνος είναι μικρότερος αλλά η ποιότητα εικόνας χειρότερη, ενώ με 256\*256 διαστάσεις έχουμε περισσότερο χρόνο εκπαίδευσης αλλά καλύτερη ποιότητα εικόνας.



OR

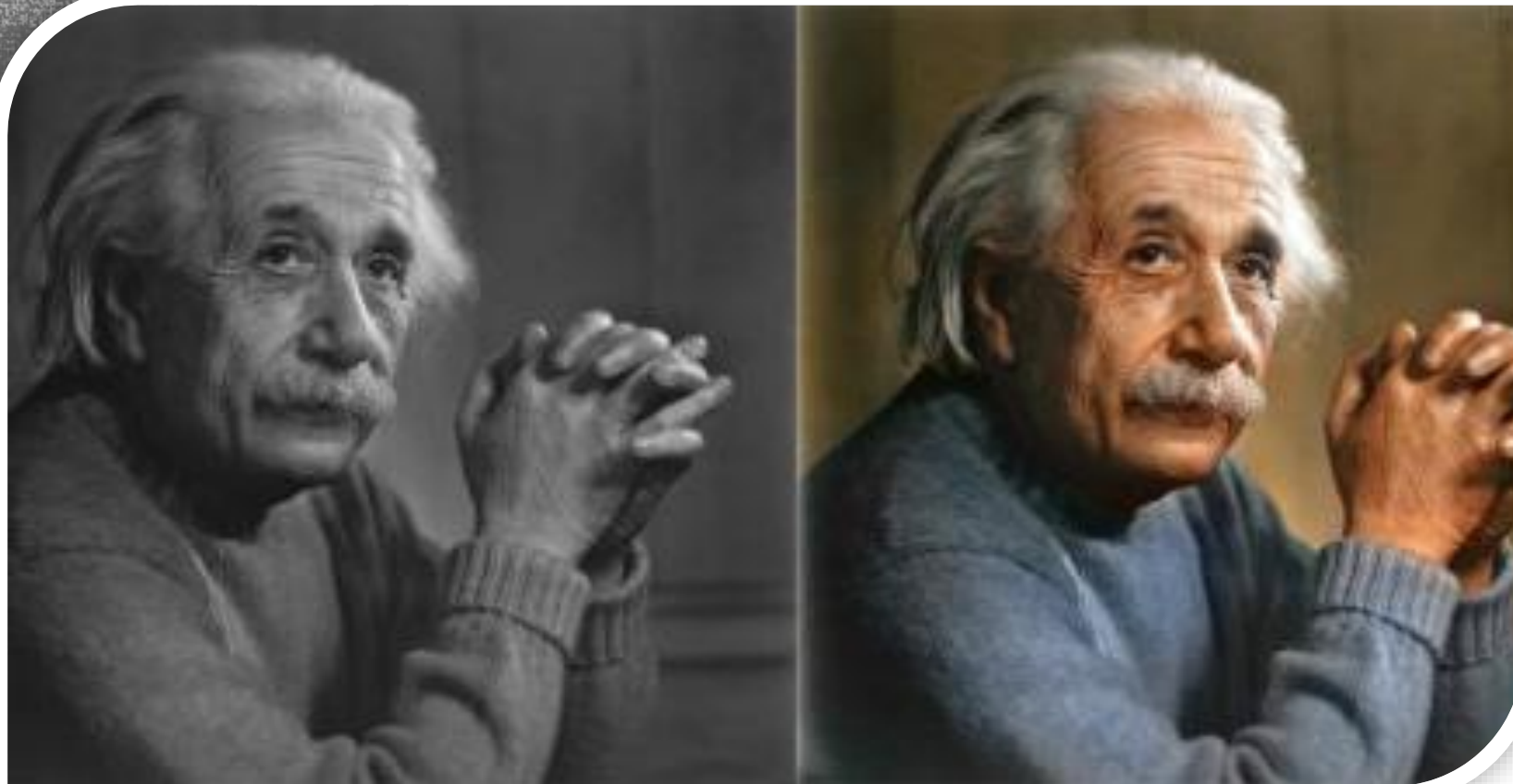


## Προτάσεις βελτίωσης

1. Χρήση καλής κάρτας γραφικών.
2. Χρήση μεθόδου pix2pix.
3. Χρήση μικρότερων δικτύων γεννήτριας και διακριτή.
4. Εκπαίδευση με μεγαλύτερο σύνολο δεδομένων.







Σας ευχαριστώ για τον χρόνο σας!

Σάμιος Γεώργιος  
Φεβρουάριος 2022

