

**Βασίλειος Στεργίου**

**Σύνθεση εικόνας σε βίντεο με χρήση  
Ανταγωνιστικών Νευρωνικών Δικτύων και  
της εξίσωσης Poisson**

Επιβλέπων: Νίκου Χριστόφορος

Ιωάννινα, Ιανουάριος, 2023



**ΤΜΗΜΑ ΜΗΧ. Η/Υ & ΠΛΗΡΟΦΟΡΙΚΗΣ  
ΠΑΝΕΠΙΣΤΗΜΙΟ ΙΩΑΝΝΙΝΩΝ**

---

**DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING**

# Ευχαριστίες

Στους γονείς μου, Πέτρο και Αναστασία, και στον αδερφό μου, Κωνσταντίνο

12/01/2023

Στεργίου Βασίλειος

# Πίνακας Περιεχομένων

Κεφάλαιο 1. Εισαγωγή.....	1
1.1 Σχετική έρευνα και στόχος της Διπλωματικής εργασίας.....	1
1.2 Δομή της Διπλωματικής εργασίας.....	4
Κεφάλαιο 2. Νευρωνικά Δίκτυα.....	6
2.1 Εισαγωγή.....	6
2.2 Εκπαίδευση ενός Νευρωνικού Δικτύου.....	8
2.3 Ο Αλγόριθμος Gradient Descend.....	11
2.4 Χρήση Νευρωνικών Δικτύων στο τομέα της Υπολογιστικής Όρασης .....	16
2.5 Η συμβολή των Νευρωνικών Δικτύων στο κλάδο της Ιατρικής.....	17
2.6 Η συμβολή των Νευρωνικών Δικτύων στο κλάδο της Βιομηχανίας.....	19
Κεφάλαιο 3. Το μοντέλο Monkey-Net.....	22
3.1 Δομή του μοντέλου Monkey-Net .....	22
3.2 Υπολογισμός των Χαρακτηριστικών Σημείων .....	25
3.3 Υπολογισμός του Πυκνού Πεδίου Κίνησης .....	28
3.4 Παραγωγή του νέου Καρέ .....	29
Κεφάλαιο 4. Το μοντέλο FOMM .....	31
4.1 Δομή του μοντέλου FOMM .....	31
4.2 Υπολογισμός του Πυκνού Πεδίου Κίνησης .....	34
4.3 Παραγωγή του τελικού Καρέ με χρήση του Χάρτη Απόκλισης .....	37
Κεφάλαιο 5. Μεθοδολογία.....	39
5.1 Η μέθοδος Seamless Cloning .....	39
5.2 Εφαρμογή της μεθόδου Seamless cloning στα παραγόμενα Καρέ .....	45

Κεφάλαιο 6. Αποτελέσματα.....	49
6.1 Παράμετροι του δικτύου και μετρικές απόδοσης .....	49
6.2 Εφαρμογή Γκαουσιανού φίλτρου στα παραγόμενα βίντεο.....	57
6.2.1 Αποτελέσματα μετρικών στο σύνολο <i>Fashion</i> .....	57
6.2.2 Αποτελέσματα μετρικών στο σύνολο <i>Taichi</i> .....	60
6.3 Εφαρμογή της μεθόδου seamless cloning στα παραγόμενα βίντεο.....	62
6.3.1 Αποτελέσματα μετρικών στο σύνολο <i>Fashion</i> .....	62
6.3.2 Αποτελέσματα μετρικών στο σύνολο <i>Taichi</i> .....	64
6.4 Εφαρμογή της μεθόδου seamless cloning και του Γκαουσιανού φίλτρου στα παραγόμενα βίντεο. ....	67
6.4.1 Αποτελέσματα μετρικών στο σύνολο <i>Fashion</i> .....	67
6.4.2 Αποτελέσματα μετρικών στο σύνολο <i>Taichi</i> .....	69
Βιβλιογραφία.....	72

# Περίληψη

Η χρήση Νευρωνικών Δικτύων, αναμφισβήτητα, αποτελεί μία από τις πιο διαδεδομένες μεθόδους επίλυσης προβλημάτων σε ευρέως διαδεδομένους και ανεπτυγμένους κλάδους, με πρωτοπόρους τους κλάδους της Πληροφορικής, της σύγχρονης Βιομηχανίας [8,13,14,16] και της Ιατρικής [19,20]. Η ευρεία αυτή χρήση τους έγκειται στη συμβολή της ραγδαίας τεχνολογικής ανάπτυξης με τη πάροδο των χρόνων, τόσο με τη παροχή πόρων από την άποψη του υλικού σχεδιασμένου για την βελτίωση της απόδοσης τους σε υπολογιστικά συστήματα ευρείας κλίμακας, όσο και από την ευρείας έκτασης έρευνα που έχει πραγματοποιηθεί σχετικά με την εφαρμογή τους για την επίλυση νέων προκλήσεων στους αναφερθέντες τομείς. Ειδικότερα, ένα μεγάλο μέρος της επίλυσης των προκλήσεων αυτών, ανάγονται στην επίλυση προβλημάτων που σχετίζονται με την Υπολογιστική Όραση (Computer Vision) και κυρίως σε αυτά που γίνεται χρήση Νευρωνικών Δικτύων Βαθιάς Μάθησης (Deep Learning). Το αντικείμενο της παρούσας Διπλωματικής Εργασίας, σχετιζόμενο με το κλάδο της Υπολογιστικής Όρασης, συνίσταται στη μελέτη της επίδρασης της εφαρμογής του μετασχηματισμού Seamless Cloning που προτάθηκε από τον Perez [1] στο κομμάτι της Επεξεργασίας Εικόνας (Image Processing), παρουσία Γκαουσιάνου φίλτρου και μη, με σκοπό τη βελτίωση των αποτελεσμάτων του μοντέλου First Order Motion Model (FOMM) [10] στο τομέα της κατασκευής βίντεο, δοθέντων μίας εικόνας-πηγής (source image) και ενός βίντεο-οδηγού (driving video).

**Λέξεις Κλειδιά:** Βαθιά Μάθηση, FOMM ,Seamless Cloning, Ανακατασκευή βίντεο, Φιλτράρισμα εικόνας

# Abstract

The use of Neural networks, undeniably, is included in the widely used problem solving methods in highly recognized and developed departments, with the most recognizable ones being those of Informatics, Economics and Industry [8,13,14,16] and Medicine [19,20],. The wide use of Neural Networks comes as the result of the rapid evolution of technology over the years, not only through the development of Neural-Network specified hardware to improve their functionality and their use in high-performance computational systems, but also from the extended research of their use in providing solution in new appearing challenges in the departments metioned above. Specifically, a big range of those new challenges, is inducted into Computer Vision related problems, and more importantly in those that Deep Learning Neural Networks are used. Having mentioned that, this Thesis revolves around applying the Seamless Cloning transformation that Perez suggests [1], which utilizes the Poisson Equation in Image Processing, with and without using a Gaussian Filter, in an attempt to improve the quality of the videos exported from the First Order Motion Model (FOMM) [10], given a source image and a driving video as inputs.

**Keywords:** Deep Learning, FOMM ,Seamless Cloning, Video reconstruction, Image Filtering

# Κεφάλαιο 1. Εισαγωγή

## 1.1 Σχετική έρευνα και στόχος της Διπλωματικής εργασίας

Η απόδοση κίνησης σε μία εικόνα-πηγή (source image) ανάγεται στη παραγωγή ενός νέου βίντεο, που πρόκειται για μία ακολουθία καρέ, με απόδοση των κινήσεων που ανιχνεύονται σε κάθε καρέ, ενός δοθέντος βίντεο-οδηγού (driving video), στην εικόνα-πηγή. Με το τρόπο αυτό, παράγεται ένα νέο καρέ, έχοντας ως βάση την εικόνα-πηγή, στο οποίο οι κινήσεις είναι οι αντίστοιχες του καρέ που χρησιμοποιήθηκε από το βίντεο-οδηγό [7,10]. Οι κινήσεις που πρέπει να αποδοθούν σε κάθε καρέ πρόκειται για Αφηνικούς Μετασχηματισμούς (Affine Transformation), σύμφωνα με την ορολογία του τομέα της Επεξεργασίας Εικόνας. Προκειμένου να γίνει όσο το δυνατόν πιο ποιοτική απόδοση των κινήσεων αυτών στην εικόνα-πηγή, δηλαδή να εφαρμοστεί ο αφηνικός μετασχηματισμός που ανιχνεύεται σε κάθε καρέ του βίντεο-οδηγού, εκπαιδεύονται Νευρωνικά Δίκτυα Βαθιάς Μάθησης σε σύνολα δεδομένων (datasets) που απαρτίζονται από βίντεο που περιέχουν ανθρώπινες κινήσεις, είτε αυτές προέρχονται από το ανθρώπινο σώμα είτε από το πρόσωπο, ώστε να καλύπτεται μία ευρεία ποικιλία από περιπτώσεις στις οποίες μπορεί να αποδοθεί κίνηση σε μία στατική εικόνα.

Με τη προσθήκη κατάλληλων συνόλων από βίντεο που περιέχουν κινήσεις του σώματος ή του προσώπου, είναι εφικτή η γενίκευση της απόδοσης κίνησης και σε περιπτώσεις που η στατική εικόνα δεν περιέχει απαραίτητα μία ανθρώπινη φυσιογνωμία, αλλά μπορεί να αφορά σχέδια ή μοντέλα στο χώρο των γραφικών, όπως για παράδειγμα το σύνολο BAIR που χρησιμοποιεί η ομάδα του Siarohin για την ανάπτυξη του First Order Motion Model [10].

Η διαδικασία της δημιουργίας ενός νέου βίντεο από μια στατική εικόνα-πηγή αποτελεί μία πρόκληση σε γενικότερα πλαίσια, καθώς απαιτεί τόσο τον ακριβή προσδιορισμό του αντικειμένου, δηλαδή εάν αυτό πρόκειται για πρόσωπο ή το ανθρώπινο σώμα [3], είτε στο δισδιάστατο (2D) είτε στο τρισδιάστατο (3D) χώρο, όσο και τον ακριβή προσδιορισμό των προτύπων που χαρακτηρίζουν τη κίνηση και την κατάλληλη απόδοση της στη στατική εικόνα. Ο ακριβής προσδιορισμός των προτύπων αυτών πρόκειται, στην ουσία, για την εκμάθηση Αφηνικών Μετασχηματισμών από τα βίντεο του συνόλου εκπαίδευσης, ώστε να είναι ακριβής η απόδοση της κίνησης στην εικόνα πηγή. Σύγχρονες έρευνες, σχετικές με την απόδοσης κίνησης από ένα βίντεο σε μία στατική εικόνα, αξιοποιούν τα Γενετικά Ανταγωνιστικά Δίκτυα (Generative Adversarial Model) [15,19], εν συντομία GAN, καθώς και Διακυμαίνοντες Αυτοκωδικοποιητές (Variational Autoencoders) [2], εν συντομία VAE.



Τα δίκτυα αυτά συγκαταλέγονται στις περιπτώσεις που η διαδικασία της μάθησης γίνεται υπό επίβλεψη (Supervised Learning), παρέχοντας μία ετικέτα (label) στην εικόνα-πηγής και στο καρέ του βίντεο-οδηγού που δέχονται, διευκολύνοντας την προσαρμογή των κινήσεων που ανιχνεύονται στο καρέ της εικόνας-πηγής. Παρά το γεγονός ότι δύνανται να αποδώσουν αποτελεσματικά την ανιχνευόμενη κίνηση από τα καρέ ενός βίντεο-οδηγού (driving video) που δίνεται ως είσοδος κάθε φορά, παρουσιάζουν σημαντική υποβάθμιση της ποιότητας του εξαγόμενου βίντεο και συνεπώς υπάρχει ευδιάκριτος θόρυβος στο αποτέλεσμα. Στη παρούσα Διπλωματική εργασία, γίνεται αναφορά στα δύο μοντέλα που αναπτύχθηκαν από την ομάδα του Siarohin, το Monkey-Net [7] και το μοντέλο FOMM [10], στη προσπάθεια σύνθεσης βίντεο από μία στατική εικόνα, δοθέντος ενός βίντεο πηγής, με τα μοντέλα αυτά να αξιοποιούν την εκμάθηση Αφηνικών Μετασχηματισμών με χρήση ενός δικτύου GAN και ενός Αυτοκωδικοποιητή.

Έχοντας ως βάση το μοντέλο FOMM του Siarohin, ο απώτερος στόχος είναι η ανακατασκευή (reconstruction) του παραγόμενου βίντεο στη φάση της μετα-επεξεργασίας αυτού (Post - processing), με χρήση της τεχνικής Seamless Clone, η οποία συνιστά έμπρακτη αξιοποίηση της εξίσωσης Poisson που προτάθηκε από τον Perez [1], προκειμένου να βελτιωθεί η ποιότητα των εξαγόμενων βίντεο του μοντέλου FOMM. Παράλληλα, στη παρούσα Διπλωματική εργασία γίνεται χρήση ενός Γκαουσιανού χαμηλοπερατού φίλτρου (Gaussian lowpass filter), το οποίο στο κλάδο της Υπολογιστικής Όρασης συντελεί στην εξομάλυνση (smoothing) των σημείων που παρουσιάζονται καμπές (edges), με ελάχιστο θόλωμα (blurring) στις περιοχές αυτές. Αξιοποιώντας τις αναφερθείσες τεχνικές επεξεργασίας εικόνας, που συντελούν στην μετα-επεξεργασία του παραγόμενου βίντεο, εξετάζεται και σε χαμηλότερο επίπεδο ο αντίκτυπος της χρήσης τους, μέσω των μεγεθών της Μέσης Ευκλείδειας Απόστασης (Average Euclidean Distance), της απώλειας L1 και της Δομικής Ομοιότητας (Structural Similarity) . Η μεταβολή των μεγεθών αυτών αποτελεί το γνώμονα με τον οποίο μπορούμε να αποφανθούμε για τη βέλτιστη αξιοποίηση των τεχνικών αυτών και κατά πόσο βελτιώνουν το εξαγόμενο αποτέλεσμα του μοντέλου FOMM.

## 1.2 Δομή της Διπλωματικής εργασίας

Η οργάνωση της παρούσας Διπλωματικής εργασίας έγκειται σε 6 κεφάλαια, καθένα από τα οποία συνιστά μία ομαλή μετάβαση στο αντικείμενο που αυτή πραγματεύεται, παρέχοντας το απαραίτητο υπόβαθρο, για τη κατανόηση των κρίσιμων σημείων αυτής.

Αρχικά, έχει προστεθεί μία σύντομη περίληψη, τόσο στην Ελληνική γλώσσα όσο και στην Αγγλική, σχετική με το πραγματευόμενο αντικείμενο, στην οποία γίνεται αναφορά στην αξιοποίηση των Νευρωνικών Δικτύων στο τομέα της Υπολογιστικής Όρασης (Computer Vision). Με την αναφορά στη χρήση των Νευρωνικών Δικτύων, γίνεται ομαλή μετάβαση στα Νευρωνικά Δίκτυα βαθιάς μάθησης που ειδικεύονται στην εκμάθηση αφηνικών μετασχηματισμών, με σκοπό τη δημιουργία ενός νέου βίντεο, δοθείσης μίας εικόνας πηγής (source image) και ενός βίντεο-οδηγού (driving video). Τα μοντέλα αυτά πρόκειται για τα Monkey-Net [7] και το μοντέλο First Order Motion Model [10].

Η περίληψη ολοκληρώνεται με επεξήγηση της συλλογιστικής πορείας που ακολουθήθηκε για την εκπόνηση της Διπλωματικής εργασίας, που συνίσταται στην εφαρμογή της εξίσωσης Poisson, σύμφωνα με τον τρόπο που προτείνεται από τον Perez [1], καθώς και της εφαρμογής χαμηλοπερατού Γκαουσιανού φίλτρου και τα αντίστοιχα συμπεράσματα που εξήχθησαν, ύστερα από τη χρήση τους.

Στο κεφάλαιο 2 παρουσιάζεται το απαραίτητο γνωσιακό υπόβαθρο για τα Νευρωνικά Δίκτυα, δηλαδή η γενική δομή τους, καθώς και ένας από τους βασικότερους αλγορίθμους εκπαίδευσης ενός Νευρωνικού Δικτύου, ο αλγόριθμος Gradient Descend. Συμπληρωματικά με τη δομή και το τρόπο λειτουργίας των Νευρωνικών Δικτύων, αναφέρονται οι κλάδοι που χρησιμοποιούν Νευρωνικά Δίκτυα Βαθιάς Μάθησης για την επίλυση των προκλήσεων που παρουσιάζονται σε αυτούς, με τους κλάδους αυτούς να περιλαμβάνουν της Ιατρικής και της Βιομηχανίας.

Αφότου έχει παρουσιαστεί το απαραίτητο γνωσιακό επίπεδο για να Νευρωνικά Δίκτυα και, κατ' επέκταση για τους κλάδους που χρησιμοποιούν Νευρωνικά Δίκτυα Βαθιάς Μάθησης για εφαρμογές που αφορούν το τομέα της Επεξεργασίας Εικόνας, ακολουθεί η παρουσίαση του κεφαλαίου 3, το οποίο σχετίζεται με τη παρουσίαση του μοντέλου Monkey – Net. Το μοντέλο Monkey-Net προτάθηκε από την ομάδα του Siarohin ως μία προσπάθεια εξαγωγής ενός νέου βίντεο από μία εικόνα-πηγή (source image), με βάση τις κινήσεις που ανιχνεύονται σε ένα βίντεο-οδηγό (driving video).

Μεταβαίνοντας στο κεφάλαιο 4, παρουσιάζεται το μοντέλο First Order Motion Model, εν συντομία FOMM, το οποίο προτάθηκε από τον Siarohin, βασισμένο στην αρχιτεκτονική του Monkey-Net, σχετιζόμενο με την έρευνα στο αντικείμενο της παρούσας διπλωματικής εργασίας.

Παράλληλα με την επεξήγηση του απαραίτητου γνωσιακού υπόβαθρου και τρόπου λειτουργίας του μοντέλου FOMM, το κεφάλαιο 5 συνιστά την αρχή της παρουσίασης των μεθόδων και τεχνικών που εφαρμόστηκαν στο μοντέλο FOMM και συγκεκριμένα της μεθόδου seamless cloning που προτάθηκε από τον Perez το 2003, η οποία συνιστά εφαρμογή της εξίσωσης Poisson, με πολλαπλές χρήσεις γενικότερα στο τομέα της Επεξεργασίας Εικόνας .

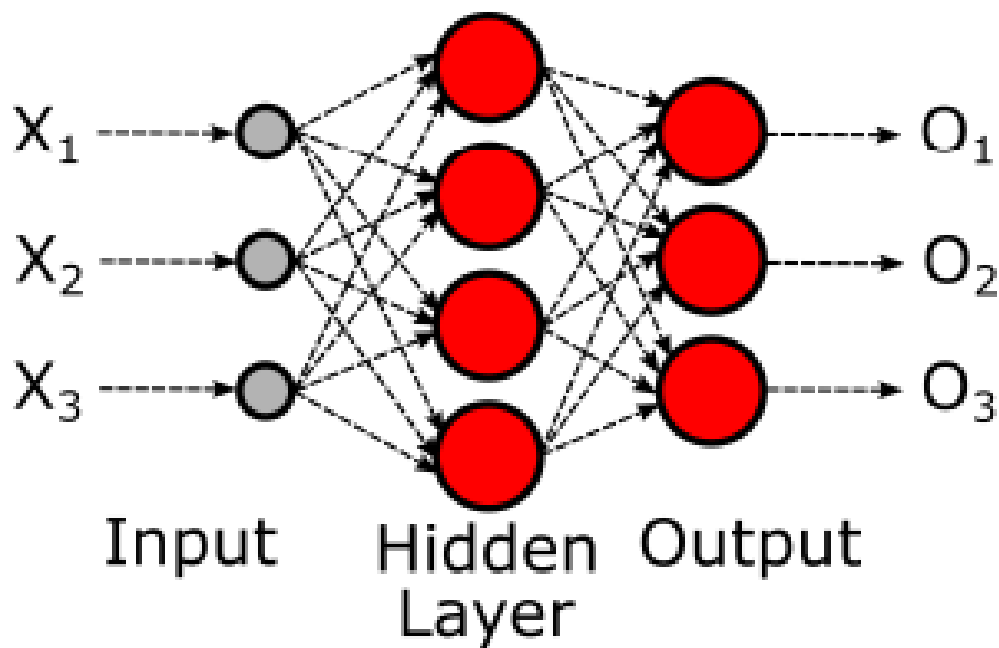
Το ερευνητικό μέρος της Διπλωματικής εργασίας ολοκληρώνεται με τα συμπεράσματα που εξήχθησαν κατά τη σύγκριση των μεθόδων που εφαρμόστηκαν στα εξαγόμενα βίντεο του μοντέλου FOMM, ύστερα από την αξιοποίηση της μεθόδου seamless cloning που επισημάνθηκε στο κεφάλαιο 5, με τα συμπεράσματα αυτά τα περιλαμβάνουν την επίδραση της εφαρμογής και μη του Γκαουσιανού φίλτρου.

Η βιβλιογραφία με τις πηγές από τις οποίες αξιοποιήθηκαν πληροφορίες σχετικές για την υλοποίηση της Διπλωματικής εργασίας συνιστά το τελευταίο μέρος της δομής που αυτή ακολουθεί.

## Κεφάλαιο 2.     Νευρωνικά Δίκτυα

### 2.1 Εισαγωγή

Τα Νευρωνικά Δίκτυα, αναφορικά με τη δομή που αυτά ακολουθούν, στηρίζεται σε αυτή του ανθρώπινου εγκεφάλου [11], καθώς αυτά αναπτύχθηκαν ως μία πρώιμη μέθοδος που αποσκοπούσε στην προσομοίωση των ικανοτήτων του. Οι ικανότητες αυτές πρόκειται για αυτές της μάθησης με χρήση παραδειγμάτων, είτε με παρουσία κάποιου επιβλέποντος (Supervised Learning) είτε όχι (Unsupervised Learning), για να παρέχει τη σωστή απάντηση και διορθώνεται κάποια ενδεχόμενη εσφαλμένη απάντηση, της απομνημόνευσης της πληροφορίας από ερεθίσματα που έχει δεχθεί μέχρι στιγμής και της ικανότητας γενίκευσης ώστε να μπορεί ο εγκέφαλος να δώσει την ανάλογη απόκριση, όταν δέχεται κάποιο ερέθισμα παρόμοιο με αυτά που έχει απομνημονεύσει. Η πληροφορία αυτή αποθηκεύεται στους νευρώνες του εγκεφάλου και συνιστά τη γνώση που έχει δεχθεί ο εγκέφαλος. Επιπλέον ικανότητες του εγκεφάλου είναι αυτές της ανάκλησης της πληροφορίας με συνδυασμό των ερεθισμάτων που δέχεται και η ανοχή στο θόρυβο, δηλαδή να βρίσκεται σε θέση αναγνωρίζει ένα ερέθισμα, ακόμα και αν το σήμα πληροφορίας που μεταφέρεται είναι αλλοιωμένο. Έχοντας τη δομή του ανθρώπινου εγκεφάλου και τις δυνατότητες του ως πρότυπα, η δομή των Νευρωνικών Δικτύων έγκειται σε αυτή που παρουσιάζεται στην Εικόνα 1 [20].



*Εικόνα 1. Δομή ενός Νευρωνικού Δικτύου, με παρουσίαση του επιπέδου εισόδου, των κρυμμένων επιπέδων, του επιπέδου εξόδου και των αντίστοιχων νευρώνων τους [20].*

Ένα Νευρωνικό Δίκτυο, σε αντιστοιχία με τον ανθρώπινο εγκέφαλο, διαθέτει νευρώνες (Neurons), οι οποίοι είναι οργανωμένοι σε επίπεδα (Layers). Οι νευρώνες δέχονται ερεθίσματα, τα οποία είναι γνωστά με τον όρο παραδείγματα στην ορολογία των Νευρωνικών Δικτύων. Τα επίπεδα διακρίνονται στο επίπεδο εισόδου, το οποίο πρόκειται για το επίπεδο που εισάγονται τα παραδείγματα στο δίκτυο, το επίπεδο εξόδου μέσω του οποίου αποφασίζεται η τιμή του δικτύου για τις εκάστοτε τιμές που δίνονται ως είσοδοι, και τα κρυμμένα επίπεδα που πρόκεινται για τα ενδιάμεσα ενός κρυμμένου επιπέδου. Τα κρυμμένα επίπεδα ενδέχεται να είναι περισσότερα του ενός, ανάλογα με τις δυνατότητες που θέλουμε να προσδώσουμε στον Νευρωνικό Δίκτυο και η αρμοδιότητα τους αφορά τη μεταβίβαση της πληροφορίας στους νευρώνες του επόμενου επιπέδου, ή ακόμα και νευρώνες κάποιου εκ των προηγούμενων επιπέδων, σε περίπτωση που υπάρχει ανάδραση της πληροφορίας. Τέλος, πρέπει να επισημανθεί ότι η πληροφορία που μεταβιβάζεται από τους νευρώνες ενός επιπέδου στους νευρώνες του επόμενου επιπέδου, εισέρχεται σε μία συνάρτηση ενεργοποίησης που έχει ανατεθεί στο επίπεδο αυτό.

Η μορφή της συνάρτησης ενεργοποίησης είναι γραμμική ή σιγμοειδής, με τις σιγμοειδείς συναρτήσεις να χρησιμοποιούνται στα κρυμμένα επίπεδα. Οι νευρώνες ενός επιπέδου συνδέονται με τους νευρώνες του προηγούμενου επιπέδου μέσω τυχαία ανατιθέμενων τιμών, τα λεγόμενα βάρη (weights), τα οποία στο πλήθος τους είναι όσες και οι νευρώνες του προηγούμενου επιπέδου. Στις τιμές αυτές περιλαμβάνεται και μία επιπλέον τιμή, η λεγόμενη πόλωση  $w_0$  που αντιστοιχεί σε μία είσοδο του κάθε νευρώνα με τιμή 1. Στο κεφάλαιο 2.2, όπου αναφέρεται η εκπαίδευση ενός Νευρωνικού Δικτύου, γίνεται αναφορά για το ρόλο της μορφής της συνάρτησης ενεργοποίησης που επιλέγουμε στα επίπεδα του Νευρωνικού Δικτύου.

## 2.2 Εκπαίδευση ενός Νευρωνικού Δικτύου

Η εκπαίδευση ενός Νευρωνικού Δικτύου διακρίνεται από δύο εξίσου σημαντικές φάσεις, τη φάση του Ευθέως Περάσματος (Forward Pass) και της Οπισθοδιάδοσης (Backpropagation) του σφάλματος που προκύπτει στην έξοδο του δικτύου. Στη φάση του Ευθέως Περάσματος, επιλέγεται κάθε φορά, είτε τυχαία είτε ακολουθιακά, όπως συμβαίνει στη πλειοψηφία των περιπτώσεων εκπαίδευσης ενός Νευρωνικού Δικτύου, ένα παράδειγμα από ένα σύνολο παραδειγμάτων που χρησιμοποιούνται για την εκπαίδευση του Δικτύου, γνωστό με τον όρο Σύνολο Εκπαίδευσης (Training Set). Σε κάθε νευρώνα του πρώτου επιπέδου, υπολογίζεται η συνολική είσοδος  $u(x)$  (Σχέση (3)) στο νευρώνα αυτό ως το εσωτερικό γινόμενο μεταξύ του διανύσματος με τις τιμές που δίνονται ως είσοδοι στο Δίκτυο και του διανύσματος βαρών που χαρακτηρίζει τον κάθε νευρώνα του πρώτου επιπέδου, προσθέτοντας τη πόλωση  $w_0$  στο γινόμενο αυτό. Η συνολική είσοδος που υπολογίζεται στο νευρώνα, με τη σειρά της, αποτελεί είσοδος της Συνάρτησης Ενεργοποίησης που έχει οριστεί από το χρήστη του δικτύου σε κάθε στρώμα, η οποία είτε έχει σιγμοειδής μορφή είτε γραμμική, μέσω της οποίας υπολογίζεται η έξοδος  $o(x)$  του νευρώνα (Σχέση (4)).

Στις περισσότερες περιπτώσεις, χρησιμοποιείται μη γραμμική συνάρτηση ενεργοποίησης στα κρυμμένα επίπεδα, καθώς, σύμφωνα με το θεώρημα της Παγκόσμιας Προσέγγισης (Universal Approximation) [11], ένα Νευρωνικό Δίκτυο με τουλάχιστον ένα κρυμμένο επίπεδο με μη-γραμμική συνάρτηση ενεργοποίησης στα κρυμμένα επίπεδα μπορεί να προσεγγίσει σε πεπερασμένο πλήθος βημάτων την επιθυμητή συνάρτηση, ώστε να επιτευχθεί η επιθυμητή εκπαίδευση του δικτύου πάνω στο σύνολο εκπαίδευσης. Οι συχνά χρησιμοποιούμενες μορφές των σιγμοειδών συναρτήσεων ενεργοποίησης είναι η Λογιστική  $\sigma(x)$  (Σχέση (1)) και η Υπερβολική Εφαπτομένη ( $\tanh$ ) (Σχέση (2)), οι οποίες περιγράφονται από τις εξής παρακάτω σχέσεις, όπου  $a$  η κλίση της κάθε συνάρτησης με συνήθη τιμή  $a=1$  και το σύνολο τιμών τους να ανήκει στο διάστημα  $[0,1]$ , ενώ για τις γραμμικές συναρτήσεις ενεργοποίησης, οι συχνότερα χρησιμοποιούμενες είναι η γραμμική (linear) και η συνάρτηση Relu.

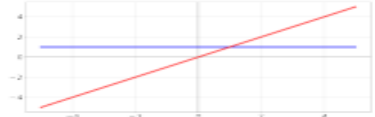


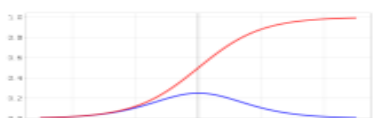
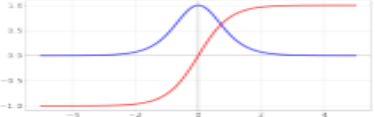
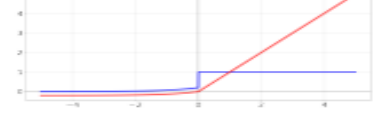
$$\sigma(x) = \frac{1}{1 + e^{-ax}} \quad (1)$$

$$\tanh(x) = \frac{e^{ax} - e^{-ax}}{e^{ax} + e^{-ax}} \quad (2)$$

$$u(x) = \sum_{i=1}^d w_i x_i + w_0 \quad (3)$$

$$o(x) = g(u(x)) \quad (4)$$

Στην Εικόνα 2 [19], παρουσιάζονται συγκεντρωμένες οι συναρτήσεις ενεργοποίησης που χρησιμοποιούνται με το αντίστοιχο πεδίο ορισμού και σύνολο τιμών τους, ανάλογα με τις ιδιαιτερότητες του εκάστοτε προβλήματος που καλείται να επιλύσει τον Νευρωνικό Δίκτυο.

Name	Function [Range]	Function (Red) and Derivative (Blue)
linear	$f(x) = a * x$ $\text{Range} : [-b, b]$	
Heaviside	$f(x) = \begin{cases} 0 & \text{if } x < 0 \\ 1 & \text{if } x \geq 0 \end{cases}$ $\text{Range} : [-\infty, \infty]$	
Rectified Linear Unit	$f(x) = \begin{cases} 0 & \text{if } x < 0 \\ x & \text{if } x \geq 0 \end{cases}$ $\text{Range} : [0, \infty]$	
Logistic/sigmoid	$f(x) = \frac{1}{1 + e^{-x}}$ $\text{Range} : [0, 1]$	
tanh	$f(x) = \tanh(x) = \frac{2}{1 + e^{-2x}} - 1$ $\text{Range} : [-1, 1]$	
elu	$f(x) = \begin{cases} a * (e^x - 1) & \text{if } x < 0 \\ x & \text{if } x \geq 0 \end{cases}$ $\text{Range} : [-\infty, \infty]$	

Εικόνα 2. Παραδείγματα γνωστών μορφών συναρτήσεων ενεργοποίησης, με τα αντίστοιχα πεδία ορισμού και το σύνολο τιμών τους [19]

Οι τιμές  $o(x)$  που υπολογίζονται σε κάθε νευρώνα, μέσω της συνάρτησης ενεργοποίησης του κρυμμένου επιπέδου, αποτελούν το διάνυσμα εισόδου του επόμενου, κατά σειρά, κρυμμένου επιπέδου, στο οποίο επαναλαμβάνεται η διαδικασία που περιγράφηκε νωρίτερα, μέχρις ότου οι τιμές που υπολογίζονται στα κρυμμένα επίπεδα να καταλήξουν στο επίπεδο εξόδου του Δικτύου. Η διαδικασία υπολογισμού τις συνολικής εισόδου  $u(x)$  και της αντίστοιχης εξόδου  $o(x)$  μπορεί να γίνει κατανοητή από την αναπαράσταση της Εικόνας 2 [19]. Στο επίπεδο εξόδου, γίνεται σύγκριση της εξόδου  $o(n)$  του δικτύου με την επιθυμητή έξοδο του  $t(n)$  και υπολογίζεται το σφάλμα  $\delta$  στην έξοδο ως η διαφορά  $o(n) - t(n)$ , όπου  $n$  ο αριθμός των νευρώνων στο επίπεδο εξόδου.



Αφότου υπολογιστεί το σφάλμα  $\delta$  στην έξοδο, ξεκινά η φάση της Οπισθοδιάδοσης (Backpropagation) στους νευρώνες του τελευταίου κρυμμένου επιπέδου. Το σφάλμα αυτό μεταδίδεται από το τελευταίο κρυμμένο επίπεδο στο πρώτο, υπολογίζοντας στην ουσία το σφάλμα σε κάθε νευρώνα του κάθε κρυμμένου επιπέδου. Η απαραίτητη μαθηματική ορολογία παρουσιάζεται στη παράγραφο 2.3, στο οποίο παρουσιάζεται ο αλγόριθμος *Gradient Descend*, που χρησιμοποιεί το υπολογιζόμενο ανά νευρώνα σφάλμα, έτσι ώστε να υπολογίσει τη παράγωγο του σφάλματος στον νευρώνα  $i$  ως προς τον βάρος  $j$  του νευρώνα, δηλαδή τη παράγωγο  $\frac{dE^n}{dw_{ij}^{(h)}}$ , ώστε να ελαχιστοποιηθεί το σφάλμα εκπαίδευσης, σε κάθε κρυμμένο επίπεδο  $h$  του Δικτύου

## 2.3 Ο Αλγόριθμος Gradient Descend

Ο αλγόριθμος *Gradient Descend* συνιστά έναν από τους πιο ευρέως χρησιμοποιούμενους αλγορίθμους βελτιστοποίησης στο τομέα της εκπαίδευσης των Νευρωνικών Δικτύων [25,26], καθώς δεν απαιτεί ιδιαίτερες συνθήκες οι οποίες αφορούν τις παραγώγους των Συναρτήσεων ενεργοποίησης, παρά μόνο τη πρώτη τους παράγωγο, η οποία υπάρχει καθώς τόσο για τις γραμμικές, όσο και για τις σιγμοειδείς συναρτήσεις, ορίζεται η πρώτη παράγωγος. Ακόμα ένας λόγος που χρησιμοποιείται είναι διότι έχει παρατηρηθεί ότι επιτυγχάνεται υψηλή ικανότητα γενίκευσης στα Νευρωνικά Δίκτυα, δηλαδή να παρέχεται σωστή η σωστή τιμή από το δίκτυο σε παραδείγματα που δεν ανήκουν σε αυτά του συνόλου Εκπαίδευσης, μέσω της ελαχιστοποίησης του σφάλματος εκπαίδευσης με τη χρήση του αλγορίθμου. Για ένα δοθέν Νευρωνικό Δίκτυο με  $H$  σε πλήθος κρυμμένα επίπεδα, ο αλγόριθμος Backpropagation υπολογίζει, αρχικά, το σφάλμα  $\delta_i^{H+1}$  σε κάθε νευρώνα  $i$  του επιπέδου  $H+1$ , το οποίο πρόκειται για το επίπεδο εξόδου, σύμφωνα με τους μαθηματικούς τύπους που παρουσιάζονται από τις Σχέσεις (5) και (6).

$$\delta_i^{(H+1)} = g' \left( u_i^{(H+1)} \right) (o_i - t_{ni}), i = 1, \dots, p \quad (5)$$

$$g' \left( u_i^{(H+1)} \right) = \begin{cases} (o_i - t_{ni}), & i = 1, \dots, p \text{ για γραμμική συν. ενεργοποίησης} \\ o_i(1 - o_i)(o_i - t_{ni}), & i = 1, \dots, p, \text{ για συγμοειδή συν. ενεργοποίησης} \end{cases} \quad (6)$$

Οι παραπάνω ορισμοί επεκτείνονται και για τα κρυμμένα επίπεδα  $h = H, \dots, 1$ , σε καθένα από τους νευρώνες  $i$  του καθενός, όπου  $i = 1, \dots, d^h$ , σύμφωνα με τους μαθηματικούς τύπους που παρουσιάζονται στις Σχέσεις (7) και (8).

$$\delta_i^{(h)} = g' \left( u_i^{(h)} \right) \sum_{j=1}^{d_{h+1}} w_{ji}^{(h+1)} \delta_j^{(h+1)}, i = 1, \dots, d_h \quad (7)$$

$$g' \left( u_i^{(h)} \right) = \begin{cases} 1, & i = 1, \dots, p \text{ για γραμμική συν. ενεργοποίησης} \\ g(u_i^{(h)}) (1 - g(u_i^{(h)})), & i = 1, \dots, p, \text{ για συγμοειδή συν. ενεργοποίησης} \end{cases} \quad (8)$$

Η μετάδοση του σφάλματος από το επίπεδο εξόδου προς το πρώτο κρυμμένο επίπεδο επαναλαμβάνεται είτε μέχρις ότου να χρησιμοποιηθούν όλα τα παραδείγματα του συνόλου εκπαίδευσης, όπου έχει ολοκληρωθεί μία εποχή του, εφαρμόζοντας τη λεγόμενη ομαδική ενημέρωση των βαρών του δικτύου, είτε κατά ένα προκαθορισμένο, από το χρήστη του δικτύου, αριθμό παραδειγμάτων που έχει δεχθεί μέχρι στιγμής το δίκτυο, δηλαδή τη λεγόμενη ενημέρωση κατά παρτίδες (batches). Και στις δύο περιπτώσεις που αναφέρθηκαν, υπολογίζεται η παράγωγος του σφάλματος  $\frac{dE^n}{dw_{ij}^{(h)}}$  του δικτύου, ως προς το κάθε βάρος  $j$  του νευρώνα  $i$  στο τρέχον επίπεδο  $h$ , με τη διαφορά ότι τα βάρη στην ομαδική ενημέρωση ενημερώνονται στο τέλος κάθε εποχής, ενώ στην ενημέρωση κατά συστάδες τα βάρη ενημερώνονται ανά  $N$  αριθμό παραδειγμάτων μέχρι να ολοκληρωθεί η εποχή.

Η παράγωγος του σφάλματος ως προς το κάθε βάρος  $j$  του νευρώνα  $i$  στο τρέχον επίπεδο  $h$ , καθώς και για τη πόλωση  $w_0$ , ορίζεται από τις από τις σχέσεις 9 και 10 αντίστοιχα.

$$\frac{dE^n}{dW_{ij}^{(h)}} = \delta_i^{(h)} g(u_j^{(h-1)}) \quad (9)$$

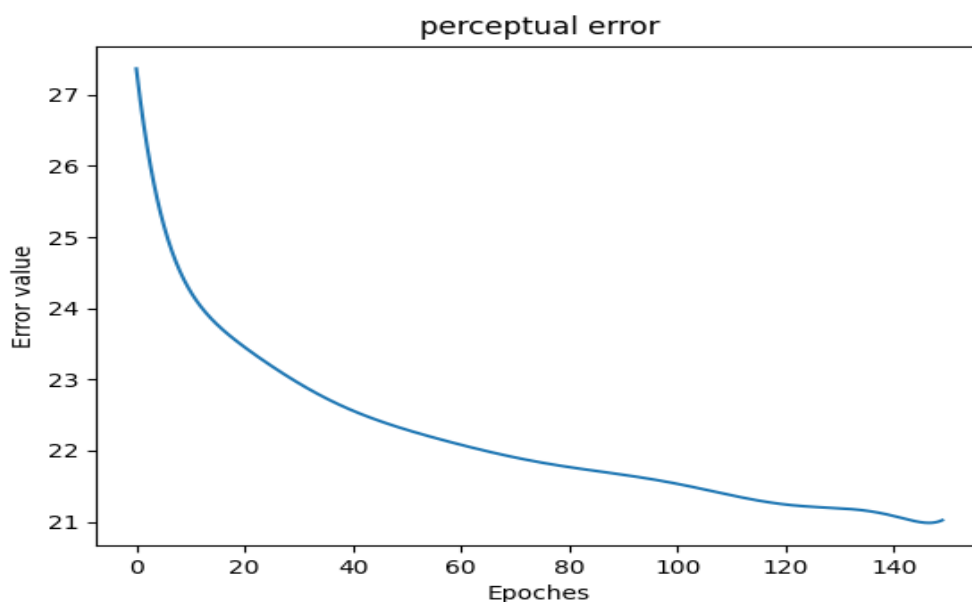
$$\frac{dE^n}{dW_{ij}^{(h)}} = \delta_i^{(h)} \quad (10)$$

Αξιοποιώντας τις μερικές παραγώγους που έχουν υπολογιστεί για τους κρυμμένους νευρώνες και τις πολώσεις σε κάθε κρυμμένο επίπεδο, το άθροισμα τους αποτελεί την ολική παράγωγο  $\frac{dE^n}{dW_i^{(h)}}$  του σφάλματος ως προς το κάθε νευρώνα  $i$  του κρυμμένου επιπέδου  $h$ . Η ενημέρωση των βαρών χρησιμοποιεί τη παράγωγο  $\frac{dE^n}{dW_i^{(h)}}$  και μία σταθερά  $\eta$ , ορισμένη από το χρήστη, η οποία ονομάζεται ρυθμός μάθησης στην ορολογία των Νευρωνικών Δικτύων. Τα βάρη και η πόλωση του κάθε κρυμμένου επιπέδου στην επόμενη επανάληψη  $t+1$  του αλγορίθμου Gradient Descend, ενημερώνονται σύμφωνα με το μαθηματικό ορισμό που παρουσιάζεται παρακάτω στη Σχέση (11).

$$w_i(t+1) := w_i(t) + \eta \frac{dE^n}{dW_i^{(h)}} \quad (11)$$

Το αναμενόμενο αποτέλεσμα της εφαρμογής του αλγορίθμου Gradient Descend για την εκπαίδευση ενός Νευρωνικού Δικτύου, σύμφωνα με τον ορισμό που περιγράφηκε νωρίτερα και την αντίστοιχη μαθηματική ορολογία που αφορά την ενημέρωση των βαρών, τείνει στη γραφική παράσταση που παρουσιάζεται στη Εικόνα 3, η οποία πρόκειται για τη γραφική πορεία του σφάλματος εκπαίδευσης του Νευρωνικού Δικτύου του μοντέλου FOMM για  $N = 150$  εποχές και ρυθμό μάθησης  $\eta = 2 \cdot 10^{-4}$  [10], όπου σε κάθε εποχή, η ενημέρωση των βαρών του δικτύου συνεπάγεται τη μείωση του σφάλματος εκπαίδευσης του δικτύου και, κατά συνέπεια, τη μεγιστοποίηση της ικανότητας γενίκευσης του.

Για τον υπολογισμό της ικανότητας γενίκευσης ενός Νευρωνικού Δικτύου, χρησιμοποιείται ένα σύνολο παραδειγμάτων, με μη-κοινά παραδείγματα μεταξύ αυτού και του συνόλου εκπαίδευσης, γνωστό με τον όρο Σύνολο Ελέγχου (Test Set). Παράλληλα, η ελαχιστοποίηση του σφάλματος εκπαίδευσης, παρά το γεγονός ότι μπορεί να επιτευχθεί με την εφαρμογή του αλγορίθμου Gradient Descend, εξαρτάται σε σημαντικό βαθμό, σε αρκετές περιπτώσεις, από τις παραμέτρους με τις οποίες εκτελείται ο αλγόριθμος.



Εικόνα 3 . Πορεία σφάλματος εκπαίδευσης στο μοντέλο FOMM με εφαρμογή του αλγορίθμου Gradient Descend

Το πλήθος των εποχών που διαρκεί η εκπαίδευση του δικτύου μπορεί να οδηγήσει στο φαινόμενο της υπο-εκπαίδευσής του, εάν το πλήθος των εποχών είναι σημαντικά μικρό, αλλά ενδέχεται και να οδηγήσει και στο φαινόμενο της υπερ-εκπαίδευσης, εάν το πλήθος των εποχών είναι σημαντικά μεγάλο. Και στις δύο περιπτώσεις, η επιλογή του ρυθμού μάθησης έχει εξίσου σημαντική επίδραση κατά την ενημέρωση των βαρών του δικτύου.

Η επιλογή πολύ μικρής τιμής του ρυθμού εκπαίδευσης  $\eta$  ενδέχεται να οδηγήσει σε σημαντικά μικρές μεταβολές των βαρών του δικτύου καθώς αυξάνονται οι εποχές, με αποτέλεσμα την ελαχιστοποίηση του σφάλματος σε μία συναρτησιακή τιμή αρκετά μεγαλύτερη της αναμενόμενης.

Από την άλλη πλευρά, η επιλογή σημαντικά μεγάλης συναρτησιακής τιμής επιφέρει ταχύτερη, ως προς το πλήθος των εποχών, σύγκλιση προς το ολικό ελάχιστο της συνάρτησης που ακολουθεί το σφάλμα, αλλά ενδέχεται να οδηγήσει στο φαινόμενο της υπερ-εκπαίδευσης, καθώς αυξάνονται η εποχές. Η επιλογή των επιθυμητών παραμέτρων παραμέτρων για την εκπαίδευση ενός Νευρωνικού Δικτύου, συνεπώς, έγκειται στις επιθυμητές ιδιότητες που επιθυμεί να επιτύχει ο χρήστης από τη χρήση του, γεγονός το οποίο συνεπάγεται την επιλογή της βέλτιστης κάθε φορά αρχιτεκτονικής του δικτύου. Έχοντας παρουσιάσει το βασικό γνωστικό υπόβαθρο που απαιτείται για τη κατανόηση της δομής και της μεθόδου εκπαίδευσης των Νευρωνικών Δικτύων, στη παράγραφο **2.4** παρουσιάζονται μερικά παραδείγματα από τη χρήση τους στο τομέα της Υπολογιστικής Όρασης (Computer Vision) και γενικότερα στον ευρύτερο κλάδο της Επεξεργασίας Εικόνας (Image Processing).

## 2.4 Χρήση Νευρωνικών Δικτύων στο τομέα της Υπολογιστικής Όρασης

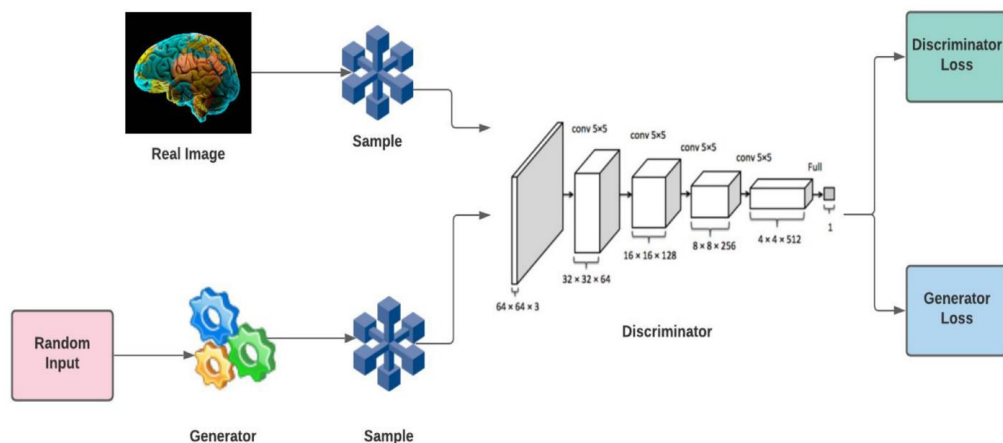
Τα Νευρωνικά Δίκτυα, λόγω της ευέλικτης δομής τους ως προς των αριθμό των κρυμμένων επιπέδων των κρυμμένων νευρώνων ανά επίπεδο, καθώς και των συναρτήσεων ενεργοποίησης που ανατίθεται στο κάθε επίπεδο, χρησιμοποιούνται ευρέως και από άλλους ανεπτυγμένους επιστημονικούς κλάδους για τη διεξαγωγή έρευνας σε περιπτώσεις, όπου απαιτείται με ανάλυση και κατηγοριοποίηση δεδομένων, σχετιζόμενα με σύνολα εικόνων. Ένας από τους μεγάλους αυτούς κλάδους συνιστά ο κλάδος της Ιατρικής [19,20,27], με το κομμάτι της έρευνας να επικεντρώνεται στην ανάλυση, κατηγοριοποίηση και σε πολλές φορές και τη ταξινόμηση εικόνων ως προς το περιεχόμενο τους, ώστε να επιτευχθεί σωστή διάγνωση ασθενειών και εξαγωγή στατιστικών δεδομένων, με βάση την εμφάνιση συγκεκριμένων χαρακτηριστικών στις εικόνες που επικεντρώνεται το ενδιαφέρον. Ο κλάδος της Βιομηχανίας συγκαταλέγεται εξίσου στους μεγάλους κλάδους στους μεγάλους κλάδους, όπου η χρήση Νευρωνικών Δικτύων και η παράλληλη αξιοποίηση των πορισμάτων στο κλάδο της Επεξεργασίας Εικόνας [13,14,16,24] μπορούν να οδηγήσουν στην αποφυγή πώλησης ελαττωματικών προϊόντων τους καταναλωτές, ελαχιστοποιώντας τις αρνητικές συνέπειες στα κέρδη της επιχείρησης που παράγει το προϊόν.

## 2.5 Η συμβολή των Νευρωνικών Δικτύων στο κλάδο της Ιατρικής

Όσον αφορά τη συσχέτιση του κλάδου της Ιατρικής με το κλάδο της Υπολογιστικής Όρασης, ο τομέας της Βιοιατρικής γεφυρώνει τους δύο κλάδους, μέσω του αντικειμένου που μελετάται από σχετική έρευνα πάνω σε αυτόν. Στο τομέα της Βιοιατρικής, η αντίστοιχη έρευνα παράγει εκατομμύρια δεδομένα, σχετιζόμενα με την ιατρική ορολογία, με τις κυριότερες μορφές να είναι σε αριθμητική μορφή (numerical data), όπως βιολογικούς δείκτες, χρονοσειρές, όπως καρδιογραφήματα και, κυρίως εικόνες με ιατρικά δεδομένα. Ο ρόλος του κλάδου της Υπολογιστικής Όρασης, με τη χρήση Νευρωνικών Δικτύων, πρόκειται για την εξαγωγή χρήσιμων χαρακτηριστικών από τις εικόνες αυτές, δοθέντων μεγάλης κλίμακας σύνολα εικόνων, σε κυτταρικό και μοριακό επίπεδο [19]. Τα χαρακτηριστικά αυτά ενδέχεται να σχετίζονται με την εύρεσης μορφολογικών και γεωμετρικών ευρημάτων στις εικόνες που πραγματεύεται η έρευνα, αλλά και πιθανώς την συσχέτιση μίας εικόνας ή περιοχών της εικόνας με τις υπόλοιπες, ώστε είτε αυτή να καταταγεί σε κάποια συγκεκριμένη κατηγορία, είτε για την εξαγωγή στατιστικών δεδομένων, σχετικά με την εμφάνιση των χαρακτηριστικών αυτών στις εικόνες, ώστε τα πορίσματα αυτά να συντελέσουν σε εφαρμογές της κυτταρικής βιολογίας, στην ανάπτυξη φαρμάκων, αλλά και σε εφαρμογές που γίνεται χρήση ακτινοβολίας στο ανθρώπινο σώμα.

Επιπλέον εφαρμογές που η χρήση Νευρωνικών Δικτύων διευκολύνει την διεξαγωγή της έρευνας είναι σε περιπτώσεις που απαιτείται τμηματοποίηση (Image Segmentation) [6,16,18] σε περιοχές γύρω από τα επιθυμητά χαρακτηριστικά που αυτά ανιχνεύονται στις εικόνες ενδιαφέροντος, μείωσης της διάστασης των δεδομένων (Principal Component Analysis), με σκοπό την εισαγωγή των δεδομένων σε ένα άλλο Νευρωνικό Δίκτυο, σε εφαρμογές σχετιζόμενες με μελέτη αλληλουχιών πρωτεϊνών και αναγνώριση προτύπων, μέσω Παραγωγικών Αντιπαραθετικών Δικτύων (GAN) [15,19] στο DNA, αλλά και για τη πρόβλεψη ύπαρξης νόσων όπως η Λευχαιμία [19,27].

Η δομή ενός δικτύου GAN μπορεί να γίνει κατανοητή μέσω της Εικόνας 4 που παρατίθεται παρακάτω, με το δίκτυο αυτό να αποτελείται από δύο βασικές οντότητες, την γεννήτρια (Generator) και του Διαχωριστή (Discriminator). Το δίκτυο δέχεται ως είσοδο μία εικόνα-πηγή (source image), ενώ παράλληλα η γεννήτρια παράγει μία τυχαία εικόνα, οι οποίες προωθούνται στο Διαχωριστή. Η αρμοδιότητα του Διαχωριστή πρόκειται για την αναγνώριση της εικόνας που παράγει η Γεννήτρια ως μη πραγματική, γεγονός το οποίο αποτελεί το στόχο της Γεννήτριας.



Εικόνα 4 . Δομή ενός δικτύου GAN με τις αντίστοιχες εισόδους και εξόδους της κάθε δομής του [15]



## **2.6 Η συμβολή των Νευρωνικών Δικτύων στο κλάδο της Βιομηχανίας**

Η χρήση Νευρωνικών Δικτύων Βαθιάς μάθησης, με συνδυασμό με την αξιοποίηση της έρευνας που πραγματοποιείται στο τομέα της Υπολογιστικής Όρασης, συνιστά παράγοντα ζωτικής σημασίας για το κλάδο της Βιομηχανίας.

Στο κλάδο της Βιομηχανίας χρησιμοποιούνται ευρέως συστήματα με ενσωματωμένες κάμερες τα οποία εποπτεύουν τη διαδικασία της παραγωγής, γνωστά και με τον όρο Συστήματα Όρασης (vision systems), με εφαρμογές των μεθόδων της Υπολογιστικής Όρασης στις εικόνες που αυτά εποπτεύουν για την ανίχνευση προτύπων [3,6,16].

Παράλληλα, η ανίχνευση ελαττωματικών προϊόντων με χρήση Νευρωνικών Δικτύων [13,14] οδηγεί στην ελαχιστοποίηση της παράδοσης τους στη καταναλωτική αγορά, ενώ παράλληλα κάνει εφικτή και λιγότερο χρονοβόρα την αυτοματοποίηση του ελέγχου για τα ελαττωματικά προϊόντα, σε σχέση με το χειρωνακτικό έλεγχο από το ανθρώπινο δυναμικό.

Ωστόσο, τα προβλήματα που συχνά παρουσιάζονται με αυτά τα συστήματα είναι η μειωμένη ανθεκτικότητά τους στην αλλαγή των περιβαλλοντικών συνθηκών στις εικόνες που δέχονται, όπως του φωτισμού του χώρου, του θορύβου που ενδέχεται να παρουσιάζεται σε αυτές, καθώς και τεχνικά ζητήματα, σχετιζόμενα με την εγκατάσταση και συντήρηση του υλικού (hardware) από το οποίο αποτελούνται τα συστήματα αυτά. Η χρήση Νευρωνικών Δικτύων, συνεπώς, ενδείκνυται ως ένα μέτρο για βελτίωση της ανίχνευσης λόγω της εκπαίδευσης του σε μεγάλα σύνολα από δεδομένα, με τα δεδομένα αυτά να πρόκειται για εικόνες προϊόντων στο κλάδο της βιομηχανίας.

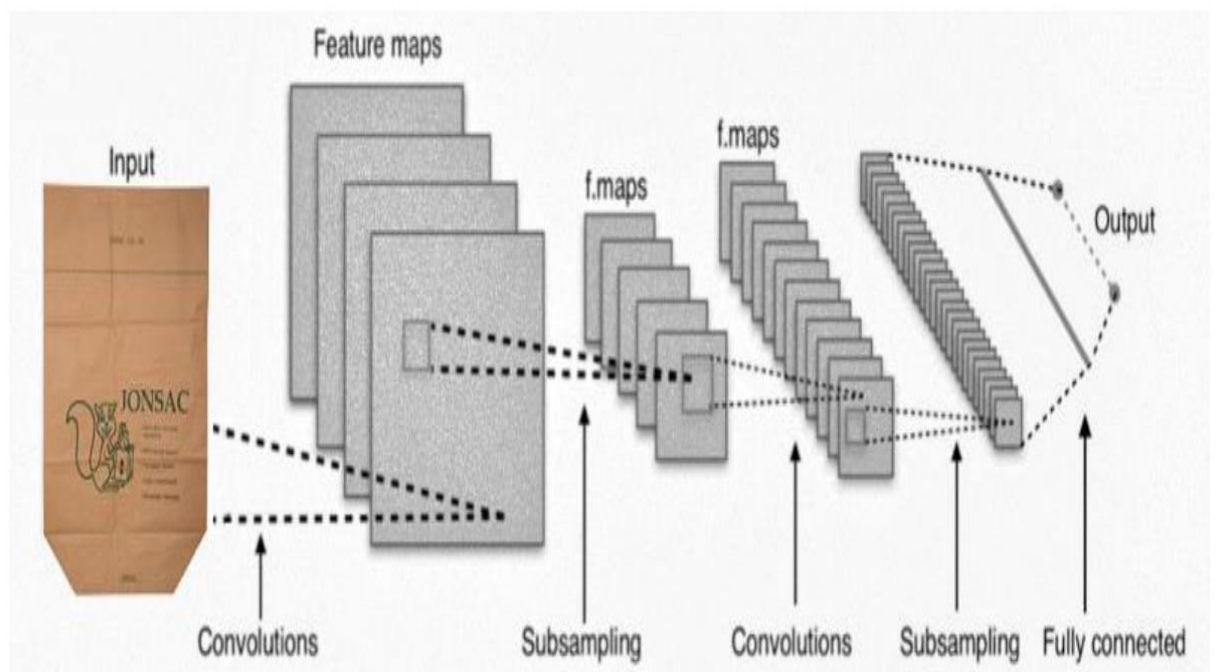
Με τη χρήση όσο το δυνατόν μεγαλύτερων συνόλων δεδομένων κατά την εκπαίδευση των δικτύων, όπως επισημάνθηκε και νωρίτερα στις παραγράφους **2.2** και **2.3**, συμβάλει στην ελαχιστοποίηση του σφάλματος γενίκευσης του δικτύου και συνεπώς τη μεγιστοποίηση της πιθανότητα να δώσει σωστή απόφαση στο εάν ένα προϊόν είναι ελαττωματικό ή όχι, γεγονός το οποίο επιδρά σημαντικά στην αποφυγή παράδοσης ελαττωματικών προϊόντων στους καταναλωτές με ενδεχόμενες επιπτώσεις και στα κέρδη του παραγωγού.



Εικόνα 5. Παραδείγματα εφαρμογής Νευρωνικών Δικτύων για την ανίχνευση ελαττωματικών συσκευασιών [13]

Στην Εικόνα 5 φαίνονται τα αποτελέσματα που εξήχθησαν από την έρευνα των Syberfeldt και Vuoluterä [13], όπου στην έρευνα τους χρησιμοποιούν μία εναλλακτική μορφή Νευρωνικών Δικτύων, γνωστή και ως Συνελικτικό Νευρωνικό Δίκτυο [6,13,16]. Τα δίκτυα αυτά αξιοποιούν μία σημαντική ιδιότητα στο χώρο της Επεξεργασίας Εικόνας, την ιδιότητα της Συνέλιξης, και η δομή τους διαφέρει από αυτή που περιγράφηκε μέχρι στιγμής γενικότερα για τα Νευρωνικά Δίκτυα. Τα δίκτυα CNN αποτελούνται από δύο είδους επίπεδα, τα Επίπεδα Συνέλιξης (Convolution layers) και τα Επίπεδα Ομαδοποίησης (Pooling Layers).

Στα Επίπεδα Συνέλιξης επιτυγχάνεται η μείωσης της διάστασης της εικόνας μέσω της συνέλιξης με φίλτρα (kernels) που αντιπροσωπεύουν χαρακτηριστικά που πρέπει να εντοπιστούν στην εικόνα. Η μειωμένη σε διάσταση εικόνα υπόκειται, εν συνεχεία, στη διαδικασία της ομαδοποίησης (Pooling), όπου παράγεται ένας Χάρτης Χαρακτηριστικών (Feature Map) με βάση τους πυρήνες, ώστε να υπολογιστούν οι γειτονιές που εντοπίζονται τα χαρακτηριστικά αυτά και να τροφοδοτηθούν στο επόμενο Επίπεδο Συνέλιξης. Η δομή ενός δικτύου CNN που μόλις περιγράφηκε, μπορεί να γίνει κατανοητή από την Εικόνα 6 [13] που παρουσιάζεται παρακάτω



Εικόνα 6 Δομή ενός Συνελικτικού Νευρωνικού Δικτύου [13]

## Κεφάλαιο 3. Το μοντέλο Monkey-Net

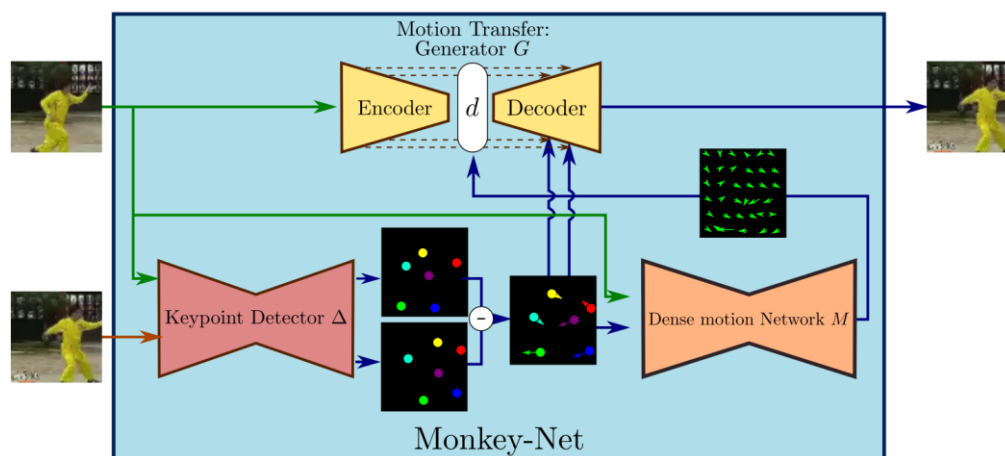
### 3.1 Δομή του μοντέλου Monkey-Net

Το μοντέλο Monkey – Net που επισημάνθηκε νωρίτερα, προτάθηκε από τον Siarohin [7] στα πλαίσια της έρευνας στο πρόβλημα της δημιουργίας ενός βίντεο από μία αρχική στατική εικόνα, στην οποία τα παραγόμενα καρέ προέρχονται από προσαρμογή των κινήσεων των καρέ του βίντεο πάνω στη στατική εικόνα. Η εξαγωγή της κίνησης που θα αποδοθεί στη στατική εικόνα διακρίνεται σε δύο φάσεις.

Η πρώτη φάση πρόκειται για τον εντοπισμό των Χαρακτηριστικών Σημείων του σώματος ή του προσώπου (Keypoints) στην εικόνα πηγή και στο εκάστοτε καρέ του βίντεο-οδηγού και η δεύτερη αφορά την εξαγωγή ενός Πυκνού Πεδίου Κίνησης (Dense Motion Field), το οποίο δείχνει τη πορεία της κίνησης των Χαρακτηριστικών Σημείων, συγκριτικά με τα Χαρακτηριστικά Σημεία που έχουν εντοπιστεί στην εικόνα-πηγή.

Στην Εικόνα 7 [7] παρουσιάζεται η αρχιτεκτονική πάνω στην οποία δομείται το μοντέλο Monkey-Net, με τις δομές που χρησιμοποιούνται να αποτελούν ανεξάρτητα Νευρωνικά Δίκτυα Βαθιάς Μάθησης, τα οποία, ωστόσο, τροφοδοτούν μεταξύ τους τις εξόδους τους, ώστε να γίνει η εξαγωγή του νέου καρέ.

Η εκπαίδευση του δικτύου Monkey-Net δεν απαιτεί κατηγοριοποίηση (labeling) στα Χαρακτηριστικά Σημεία που των βίντεο που χρησιμοποιούνται στο σύνολο εκπαίδευσης. Η μάθηση γίνεται με αυτό-επίβλεψη (self - supervised). Οι ετικέτες (labels) είναι απαραίτητες στα βίντεο του συνόλου ελέγχου, μέσω του οποίου αξιολογείται η απόδοση του με βάση τη τιμή της τελικής τιμής που έχει πάρει το σφάλμα ελέγχου.

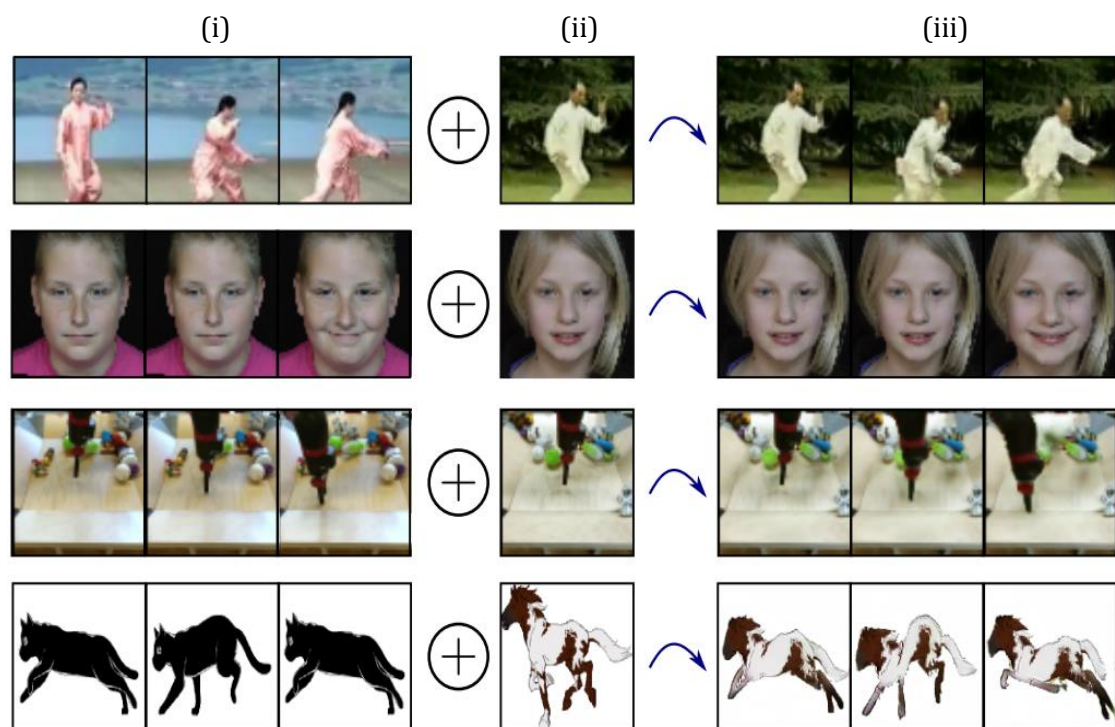


Εικόνα 7. Δομή του μοντέλου Monkey-Net με τις αντίστοιχες εισόδους και εξόδους του κάθε λειτουργικού μέρους του [7]

Το μοντέλο Monkey-Net, όπως γίνεται αντιληπτό από την Εικόνα 7 [7] αξιοποιεί τρία διαφορετικά Νευρωνικά Δίκτυα, με καθένα από αυτά να αναλαμβάνει διαφορετικές αρμοδιότητες για τη παραγωγή του εκάστοτε καρέ. Το δίκτυο Keypoint Detector είναι υπεύθυνο για την ανίχνευση των Χαρακτηριστικών Σημείων της στατικής εικόνας και του κάθε καρέ του βίντεο-οδηγού, με τη διαδικασία αυτή να πρόκειται για το πρώτο βήμα στη παραγωγή του νέου καρέ.

Μέσω της εύρεσης των Χαρακτηριστικών σημείων της εικόνας-πηγής και του τρέχοντος καρέ του βίντεο-οδηγού, προκύπτει ο χάρτης κίνησης των Χαρακτηριστικών Σημείων, αφαιρώντας τις θέσεις των χαρακτηριστικών σημείων του τρέχοντος καρέ του βίντεο-οδηγού από τα Χαρακτηριστικά Σημεία της εικόνας-πηγής.

Η διαφορά αυτή τροφοδοτείται στη συνέχεια στο δεύτερο Νευρωνικό Δίκτυο, το Dense Motion Network, μέσω του οποίου παράγεται το Πυκνό Πεδίο Κίνησης (Dense Motion Field), υποδεικνύοντας μία πρόβλεψη για τη πορεία των Χαρακτηριστικών Σημείων στο νέο καρέ. Εν τέλει, το Πυκνό Πεδίο Κίνησης τροφοδοτείται μαζί με τη στατική εικόνα σε μία γεννήτρια μεταφοράς κίνησης (Motion Transfer Generator). Η γεννήτρια χρησιμοποιεί ένα κωδικοποιητή (Encoder) που μετατρέπει σε διάνυσμα τη στατική εικόνα εισόδου, ώστε να αποδοθεί η κίνηση σε αυτή, παρουσία του Πυκνού Πεδίου Κίνησης, καθώς και έναν αποκωδικοποιητή (Decoder), μέσω του οποίου παράγεται το τελικό καρέ. Ενδεικτικά αποτελέσματα που εξάγονται από το μοντέλο Monkey - Net παρουσιάζονται στις εικόνες των στηλών (i) έως (iii), των οποίων το περιεχόμενο θα εξηγηθεί παρακάτω, από τα αριστερά προς τα δεξιά, με τη κάθε στήλη να αποτελεί την κάθε κατηγορία εικόνων.



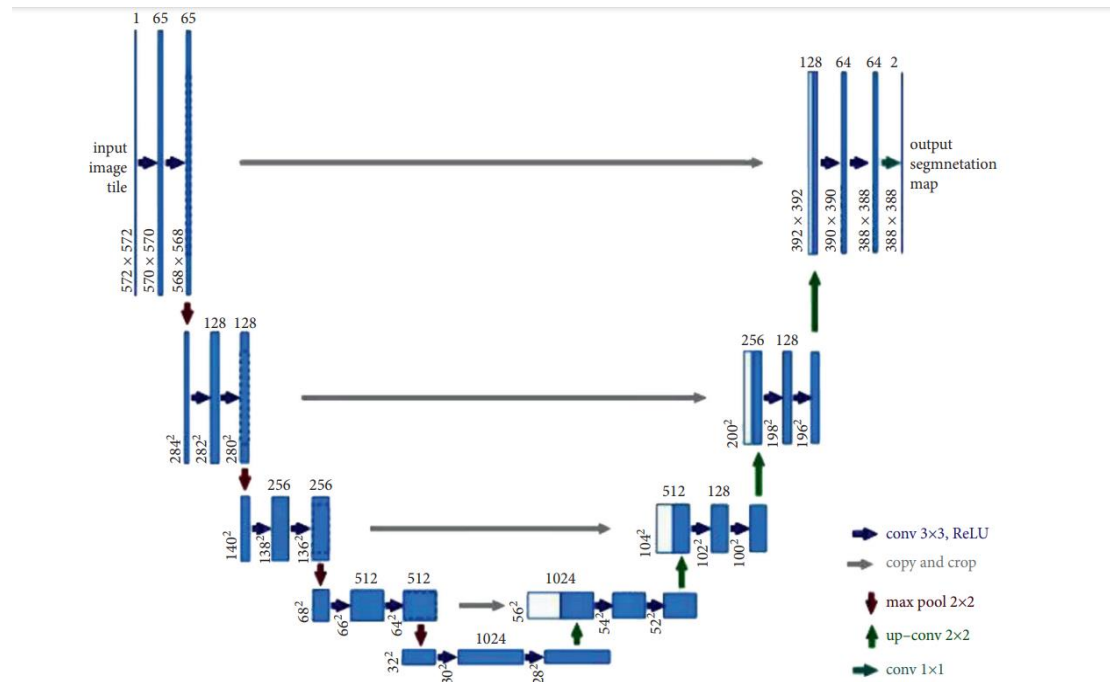
Εικόνα 8. Προσαρμογή κίνησης από τα καρέ του βίντεο οδηγού (i) στην εικόνα-πηγή (ii) και παραγωγή του νέου καρέ (iii) [7]

Όπως επισημάνθηκε και πρωτίτερα κατά την εξήγηση των Νευρωνικών Δικτύων που χρησιμοποιεί το μοντέλο Monkey-Net, μέσω του δικτύου Keypoint Detector υπολογίζονται τα Χαρακτηριστικά Σημεία (Keypoints) στο τρέχον καρέ του βίντεο-οδηγού και στη στατική εικόνα-πηγή, όπως παρουσιάζεται στις στήλες (i) και (ii) αντίστοιχα. Οι εικόνες της στήλης (iii) πρόκειται για αναπαραστάσεις της πρόβλεψης της κίνησης στο παραγόμενο καρέ από το δίκτυο Motion Transfer, το οποίο παράλληλα αξιοποιεί τα νέα χαρακτηριστικά σημεία που παράγονται από το δίκτυο Keypoint Detector και το Πυκνό Πεδίο Κίνησης (Dense Motion Field) που παράγεται από το δίκτυο Dense Motion Network [7].

Έχοντας περιγράψει τις δομές που απαρτίζουν το δίκτυο Monkey-Net και το ρόλο που αναλαμβάνει η κάθε μία στην παραγωγή, θα επισημανθεί η μέθοδος που ακολουθεί ο Siarohin και η ομάδα του για τον υπολογισμό των Χαρακτηριστικών Σημείων και του Πυκνού Πεδίου Κίνησης, με τελικό στόχο τη παραγωγή ενός νέου καρέ από τη χρήση αυτών.

## 3.2 Υπολογισμός των Χαρακτηριστικών Σημείων

Η αρχιτεκτονική πάνω στην οποία στηρίζεται το δίκτυο Keypoint Detector πρόκειται για αυτή του δικτύου U-Net [18], μίας εκ των μορφών που εφαρμόζονται τα Συνελικτικά Νευρωνικά Δίκτυα [6,13,16] σε εφαρμογές που απαιτείται αναγνώριση προτύπων, κατατάσσοντας σε μία από τις δοθείσες ετικέτες (labels) το κάθε εικονοστοιχείο (pixel) της εικόνας, ώστε να επιτευχθεί τμηματοποίηση (Segmentation) στις περιοχές που επικεντρώνεται το ενδιαφέρον. Στη περίπτωση του Keypoint Detector Network, η κατάταξη ενός pixel σε κάποια κατηγορία ανάγεται στο εάν αποτελεί μέρος ενός Χαρακτηριστικού Σημείου ή όχι και οι περιοχές των Χαρακτηριστικών Σημείων αποτελούν τα μόνα χαρακτηριστικά που περιέχονται στην εικόνα. Στην Εικόνα 8 [18], παρουσιάζεται η δομή ενός δικτύου U-Net, με παράλληλη επεξήγηση του τρόπου με τον οποίο οργανώνεται το κάθε επίπεδο (layer) και τις μεταξύ τους αλληλεπιδράσεις.



Εικόνα 9. Παράδειγμα δομή ενός δικτύου U-Net, από το επίπεδο εισόδου και τα συνελκτικά επίπεδα (αριστερά), προς το επίπεδο εξόδου [18].

Το δίκτυο U-Net συνίσταται από  $2 \cdot N$  συμμετρικά Συνελκτικά Επίπεδα, όπου τα  $N$  από αυτά συμβάλλουν στη μείωση των διαστάσεων της εικόνας εισόδου μέσω της πράξης της συνέλιξης με χρήση Συνελκτικού Φίλτρου (Convolutional Kernels), συνήθως μικρών διαστάσεων όπως 3x3 ή 2x2 όπως παρουσιάζεται και στο παράδειγμα. Η μείωση της διάστασης της εικόνας στοχεύει στο προσδιορισμό των περιοχών ενδιαφέροντος που υποδεικνύουν τα φίλτρα.

Για τη σύνθεση της τελικής εικόνας που οι περιοχές αυτές παρουσιάζονται ως ανεξάρτητα τμήματα, αξιοποιούνται  $N$  Συνελκτικά Επίπεδα, τα οποία σε αυτό το στάδιο δέχονται τη μειωμένη σε διάσταση εικόνα που έχει προκύψει. Σε κάθε επίπεδο που συμμετέχει στην ανασύνθεση της τελικής εικόνας, παράγεται μία εικόνα ίδιας διάστασης με αυτή του αντίστοιχου ομότιμου επιπέδου του δικτύου, την οποία αξιοποιεί με πρόσθεση στην εικόνα που έχει παραχθεί στο τωρινό επίπεδο. Η διαδικασία αυτή εφαρμόζεται μέχρι το επίπεδο εξόδου, όπου η εικόνα που προστίθεται στο επίπεδο αυτό πρόκειται για την εικόνα του επιπέδου εισόδου.



Στη περίπτωση του U-Net που χρησιμοποιεί ο Keypoint Detector του Siarohin [7], η τελική εικόνα που εξάγεται αφορά τοπικές αναπαραστάσεις  $H_k$  (Heatmaps) διάστασης  $H \times W$  για κάθε Χαρακτηριστικό Σημείο  $k$  που έχει ανιχνευθεί μέσω της μείωσης της διάστασης της αρχικής εικόνας που έχει προηγηθεί, με  $H$  να αποτελούν τις γραμμές και  $W$  τις στήλες της εικόνας και  $H_k \in [0,1]^{H \times W}$ . Για τη παραγωγή της εικόνας που περιέχει τα Χαρακτηριστικά Σημεία, ο Siarohin αξιοποιεί δύο βασικές μετρικές, την αναμενόμενη τιμή  $h_k$  των συντεταγμένων των pixels που απαρτίζουν τη περιοχή του κάθε Χαρακτηριστικού Σημείου και τη Συνδιακύμανση (Covariance)  $\Sigma_k$  των τιμών τους. Οι αντίστοιχοι μαθηματικοί ορισμοί που περιγράφουν τις παραπάνω σχέσεις παρουσιάζονται από τις Σχέσεις (12) και (13), όπου  $p$  κάθε pixel της εικόνας διάστασης  $U = H \times W$ .

$$h_k = \sum_{p \in U} H[p] \cdot p \quad (12)$$

$$\Sigma_k = \sum_{p \in U} H[p] \cdot (p - h_k) \cdot (p - h_k)^T \quad (13)$$

Με βάσεις τις σχέσεις 12 και 13 που αναφέρονται παραπάνω, η τιμή που αποδίδεται σε κάθε εικονοστοιχείο των τοπικών αναπαραστάσεων (Heatmaps) των Χαρακτηριστικών Σημείων δίνεται από της Σχέση (14) με  $\alpha$  μία σταθερά κανονικοποίησης (Normalization).

$$H_k(p) = \frac{1}{\alpha} e^{-(p - h_k) \Sigma_k^{-1} (p - h_k)} \quad (14)$$

### 3.3 Υπολογισμός του Πυκνού Πεδίου Κίνησης

Η εξαγωγή του Πυκνού Πεδίου Κίνησης (Dense Motion Field) συνδέεται άμεσα με τον υπολογισμό των Χαρακτηριστικών Σημείων, σύμφωνα με τη μαθηματική ορολογία που παρουσιάζεται στη παράγραφο 3.2. Στο δίκτυο Dense Motion Network που χρησιμοποιεί ο Siarohin [7], ο υπολογισμός του Πυκνού Πεδίου Κίνησης  $F$  υποθέτει ότι οι περιοχές των Χαρακτηριστικών Σημείων, που δίνονται ως έξοδοι του δικτύου U-Net [18] στο οποίο στηρίζεται το δίκτυο Keypoint Detector, είναι μη-επικαλυπτόμενες, δηλαδή κάθε περιοχή των χαρακτηριστικών σημείων είναι ανεξάρτητη από τις υπόλοιπες. Παράλληλα, σύμφωνα με τη μεθοδολογία που ακολουθεί ο Siarohin, το Πυκνό Πεδίο Κίνησης  $F$  συνιστά το άθροισμα δύο υπο-μεγεθών, του Τραχέος Πεδίου Κίνησης  $F_{coarse}$  και του Υπολειπόμενου Πεδίου Κίνησης  $F_{residual}$ . Ο υπολογισμός του  $F_{coarse}$  περιγράφεται στη συνέχεια από τη Σχέση (15).

$$F_{coarse} = \sum_{k=1}^{K+1} M_k \otimes \rho(h_k) \quad (15)$$

Αναλύοντας τη Σχέση (15), για καθένα από τα  $K$  Χαρακτηριστικά Σημεία που έχουν υπολογιστεί νωρίτερα, ο Siarohin ανάγει το πρόβλημα της εύρεσης του Πυκνού Πεδίου Κίνησης στον υπολογισμό  $k$  μασκών  $M_k \in R^{H \times W}$  και της ποσότητας  $\rho(h_k)$ , όπου  $\rho(h_k)$  πρόκειται για μετασχηματισμό που επιστέφει ένα διάνυσμα διαστάσεων  $H \times W$ , με κάθε στοιχείο  $i \in H \times W$  του διανύσματος να έχει ανατεθεί στην αναμενομένη τιμή  $h_k$  της περιοχής γύρω από το Χαρακτηριστικό Σημείο  $k$ .

Παρά το γεγονός ότι έχουν υπολογιστεί  $k$  Χαρακτηριστικά Σημεία νωρίτερα, η μεθοδολογία του Siarohin χρησιμοποιεί επιπλέον μία μάσκα  $M_{k+1}$  που αντιστοιχεί στη τιμή  $\rho(0)$ , ώστε να συμπεριληφθεί και το στατικό υπόβαθρο (background).

Εν τέλει, ο υπολογισμός του Τραχέος Πεδίου Κίνησης  $F_{coarse}$  προκύπτει από το άθροισμα των εσωτερικών γινομένων των μεγεθών  $M_k$  και  $\rho(h_k)$ , για καθένα από τα Χαρακτηριστικά Σημεία  $k$ . Παράλληλα το Δίκτυο Dense Motion Network προβλέπει το Υπολειπόμενο Πεδίο Κίνησης  $F_{residual}$ , το οποίο συμβάλει στην απόδοση πιο ομαλής κίνησης, μέσω του αθροίσματος  $F_{coarse} + F_{residual}$  που αναφέρθηκε νωρίτερα.

### 3.4 Παραγωγή του νέου Καρέ

Η παραγωγή του νέου καρέ με την απόδοση της κίνησης του  $i$ -οστού καρέ του βίντεο οδηγού (driving video) στην εικόνα πηγή (source image) συνδυάζει τις τοπικές αναπαραστάσεις (Heatmaps) των Χαρακτηριστικών Σημείων που έχουν παραχθεί από το δίκτυο, την αναμενόμενη τιμή  $h_k$  των συντεταγμένων τους και τη Συνδιακύμανση (Covariance)  $\Sigma_k$  των τιμών τους, με αξιοποίηση του Πυκνού Πεδίου Κίνησης του δικτύου Dense Motion Field. Σε πρώτο στάδιο, το δίκτυο Keypoint Detector, εκτός από τον υπολογισμό των τοπικών αναπαραστάσεων των Χαρακτηριστικών Σημείων του  $i$ -οστού καρέ, υπολογίζει τις τοπικές αναπαραστάσεις των  $k$  Χαρακτηριστικών σημείων της εικόνας-πηγής και του πρώτου καρέ του βίντεο-οδηγού, αξιοποιώντας τα για την εξαγωγή των αντίστοιχων τιμών  $h_k^{source}$  και  $h_k^{frame}$ .

Οι συντεταγμένες για τα Χαρακτηριστικά Σημεία καθενός από τα παραγόμενα καρέ προκύπτουν από τη μεταφορά των συντεταγμένων των Χαρακτηριστικών Σημείων της εικόνας πηγής κατά τη σχετική απόσταση των συντεταγμένων των Χαρακτηριστικών Σημείων του  $i$ -στου καρέ του βίντεο-οδηγού από το πρώτο καρέ του. Συνεπώς, τα νέα Χαρακτηριστικά Σημεία δίνονται από την ακόλουθη σχέση.

$$h_k^{s'} = h_k^s + (h_k^t - h_k^1) \quad (16)$$

Οι τιμές που ανατίθενται σε καθένα από τα pixels αποτελούν τις τιμές της Αναμενόμενης Τιμής από το καρέ του βίντεο-οδηγού, κατά τη μέθοδο που ακολουθεί ο Siarohin και η οποία περιεγράφηκε στη παράγραφο **3.2**. Οι τοπικές αναπαραστάσεις που έχουν υπολογιστεί με βάση τη παραπάνω σχέση, καθώς και η εικόνα-πηγή, τροφοδοτούνται στο Dense Motion Network για την εξαγωγή του Πυκνού Πεδίου Κίνησης και στη συνέχεια στο δίκτυο Motion Transfer Generator για τη παραγωγή του νέου καρέ.

Κλείνοντας το κεφάλαιο **3**, που εξηγήθηκε η μέθοδος που ακολουθεί ο Siarohin μέσω του μοντέλου Monkey-Net για την δημιουργία ενός βίντεο, δοθέντων μίας στατικής εικόνας και ενός βίντεο-οδηγού, γίνεται η μετάβαση στο μοντέλο First Order Motion Model [10], το οποίο αναπτύχθηκε επίσης από τον Siarohin, με γνώμονα την αρχιτεκτονική του Monkey-Net, στα πλαίσια της έρευνας της σύνθεσης βίντεο με χρήση Νευρωνικών Δικτύων από μία στατική εικόνα [7,9,10,12,21].

## Κεφάλαιο 4. Το μοντέλο FOMM

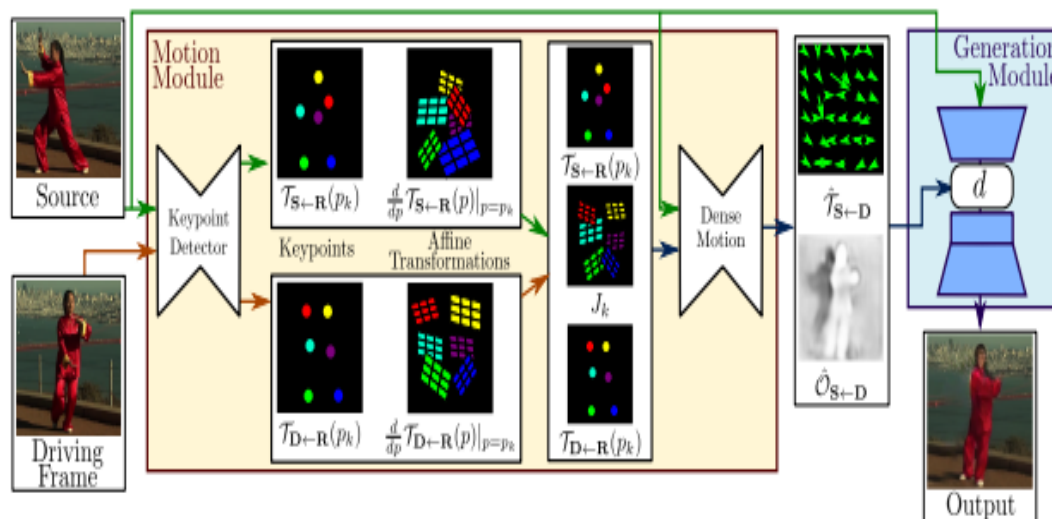
### 4.1 Δομή του μοντέλου FOMM

Στα πλαίσια της έρευνας σχετικά με τη εκπαίδευση Νευρωνικών Δικτύων πάνω σε σύνολα από στάσεις του ανθρώπινου σώματος και του προσώπου, με απώτερο στόχο την απόδοση κίνησης σε μία στατική εικόνα με βάση της κινήσεις που ανιχνεύονται σε ένα βίντεο-οδηγό, ο Siarohin και η ομάδα του ανέπτυξαν το μοντέλο First Order Motion Model, εν συντομία FOMM [10], στηριζόμενοι κατά βάση στην αρχιτεκτονική του Monkey-Net [7]. Το μοντέλο FOMM, όπως και στη περίπτωση του μοντέλου Monkey-Net, εκπαιδεύεται σε σύνολα από βίντεο που περιέχουν κινήσεις, ώστε να μπορεί να αναγνωρίσει τους Αφηνικούς Μετασχηματισμούς που χαρακτηρίζουν αυτές τις κινήσεις. Η διαφορά μεταξύ των δύο μοντέλων έγκειται στο γεγονός ότι το μοντέλο FOMM δέχεται και ένα σύνολο από χαρακτηριστικά σημεία παράλληλα με τους αφηνικούς μετασχηματισμούς των βίντεο.

Η βασική αδυναμία του Monkey-Net που οδήγησε τον Siarohin και την ομάδα του στην απόφαση αυτή είναι το γεγονός ότι το Monkey-Net παρουσιάζει μειωμένη απόδοση και χαμηλής ποιότητας παραγωγής νέου καρέ σε περιπτώσεις που υπάρχει μεγάλη απόσταση μεταξύ ενός ή περισσότερων Χαρακτηριστικών Σημείων από το προηγούμενο καρέ του βίντεο-οδηγού και περιπτώσεις που οι ανιχνευόμενες κινήσεις είναι σύνθετοι αφηνικοί μετασχηματισμοί.

Η μάθηση του νέου αυτού μοντέλου γίνεται, επίσης, με αυτό-επίβλεψη (Self-Supervised), με γνώμονα τα Χαρακτηριστικά Σημεία του Συνόλου Εκπαίδευσης ως ετικέτες (labels) για τη διόρθωση των σφαλμάτων του μοντέλου FOMM. Παράλληλα, μία σημαντική διαφορά μεταξύ των δύο μοντέλων είναι ότι στο μοντέλο FOMM, η χρήση του Δικτύου Πυκνής Κίνησης (Dense Motion Network) [4,10,15] παράγει ένα επιπλέον αποτέλεσμα, που είναι ο Χάρτης Απόκλισης (Occlusion Map), μία μάσκα που υποδεικνύει περιοχές στο καρέ που παράγεται και που οι οποίες δεν είναι ορατές στην εικόνα-πηγή.

Ο χάρτης αυτός έχει ιδιαίτερη σημασία για την πρόβλεψη των τιμών των pixel στις περιοχές που δεν ανήκουν στα χαρακτηριστικά σημεία, κυρίως στις περιπτώσεις που η μετακίνηση των Χαρακτηριστικών Σημείων έχει μεγάλη απόκλιση σε σχέση με το προηγούμενο καρέ. Ο στόχος είναι η βελτίωση της απόδοσης σωστών προβλεπόμενων τιμών (impainting) στις περιοχές αυτές, όπως θα περιγραφεί στην παράγραφο 4.3. Η αρχιτεκτονική του μοντέλου FOMM που περιεγράφηκε νωρίτερα μπορεί να συνοψιστεί στο σχήμα που παρουσιάζεται στην Εικόνα 10 [10], με τις δομές που χρησιμοποιούνται στο Monkey-Net να παραμένουν κοινές και για τα δύο μοντέλα, αλλά να υπάρχουν διαφορές μεταξύ των αποτελεσμάτων υπολογίζονται σε κάθε μοντέλο.



Εικόνα 10. Δομή του μοντέλου First Order Motion Model, με τα αντίστοιχα εξαγόμενα αποτελέσματα σε κάθε δομή [10]

Στη παράγραφο 4.2 που ακολουθεί, γίνεται η επεξήγηση των σταδιακών υπολογισμών των δομών που απαρτίζουν το μοντέλο FOMM, που είναι απαραίτητοι για την εξαγωγή του τελικού αποτελέσματος.

Στην Εικόνα 11 παρουσιάζονται ενδεικτικά τα εξαγόμενα αποτελέσματα των δομών στις στήλες (i) έως (vi) που χρησιμοποιεί το μοντέλο FOMM με αντίστοιχη ανάλυση των τους. Οι στήλες (i) και (ii) παρουσιάζουν τα Χαρακτηριστικά Σημεία που έχουν προβλεφθεί για την εικόνα-πηγή  $S$  και το τρέχον καρέ  $D$  του βίντεο-οδηγού αντίστοιχα, μέσω του δικτύου Keypoint Detector. Η στήλη (iii) πρόκειται για ένα τυχαία επιλεγόμενο καρέ αναφοράς  $R$ , το οποίο χρησιμοποιείται για τη προσέγγιση της μεταφοράς των Χαρακτηριστικών Σημείων από το τρέχον καρέ  $D$  του βίντεο-οδηγού προς την εικόνα-πηγή  $S$  με τη χρήση του αναπτύγματος Taylor πρώτου βαθμού, με τη μεθοδολογία που περιγράφεται διεξοδικά στις παραγράφους 5.2 και 5.3. Η πρόβλεψη των νέων Χαρακτηριστικών Σημείων παρουσιάζεται στις εικόνες της στήλης (iv), με τη νέα στάση της πρόβλεψης να εξάγεται από το δίκτυο Dense Motion Network, λαμβάνοντας, παράλληλα, ως είσοδο τα Χαρακτηριστικά Σημεία που έχουν προβλεφθεί από το δίκτυο Keypoint Detector. Το τελικό αποτέλεσμα του μοντέλου FOMM συνιστούν οι εικόνες της στήλης (v), στις οποίες έχει χρησιμοποιηθεί και ο Χάρτης Απόκλισης που υπολογίζεται από το δίκτυο Dense Motion Network (στήλη (vi)) για την απόδοση τιμών στα pixels της τελικής εικόνας που δεν έχει γίνει σωστή πρόβλεψη των τιμών τους, μέσα από τη χρήση της τεχνικής inpainting.



Εικόνα 11. Είσοδοι και εξαγόμενα αποτελέσματα του μοντέλου FOMM, σύμφωνα με την εξήγηση του κεφαλαίου 4.1 [10]

## 4.2 Υπολογισμός του Πυκνού Πεδίου Κίνησης

Στη παράγραφο 4.1 αναφέρθηκε ότι μία από τις διαφορές μεταξύ των μοντέλων FOMM και Monkey-Net έγκειται στον υπολογισμό του Χάρτη Απόκλισης (Occlusion Map)  $T_{S \leftarrow D}: R^2 \leftarrow R^2$  από το Δίκτυο Πυκνής Κίνησης, παράλληλα με το Πεδίο Πυκνής Κίνησης (Dense Motion Field) που υπολόγιζε και το μοντέλο Monkey-Net. Στη περίπτωση του μοντέλου FOMM, ο Χάρτης Πυκνής Κίνησης  $T_{S \leftarrow D}: R^2 \leftarrow R^2$  αντιστοιχίζει τη θέση κάθε pixel των Χαρακτηριστικών σημείων του εκάστοτε καρέ οδηγού προς τις αντίστοιχες συντεταγμένες της εικόνας-πηγής, προκειμένου να γίνει αντιληπτός ο τρόπος κίνησης που πρέπει να αποδοθεί στην εικόνα-πηγή, με βάση το τρέχον καρέ του βίντεο-οδηγού.

Ο υπολογισμός αυτός αξιοποιεί το ανάπτυγμα Taylor πρώτης τάξης, υποθέτοντας ότι υπάρχει ένα καρέ αναφοράς  $R$ , το οποίο έχει επιλεγεί τυχαία μεταξύ των καρέ του βίντεο οδηγού.

Συνεπώς, ο υπολογισμός του Χάρτη Πυκνής Κίνησης  $T_{S \leftarrow D}$  διακρίνεται σε δύο επιμέρους υπολογισμούς με χρήση του αναπτύγματος Taylor, τον Χάρτη Πυκνής Κίνησης  $T_{S \leftarrow R}$  και τον χάρτη  $T_{R \leftarrow D}$ .

Θεωρώντας την εικόνα-πηγή ή το τρέχον καρέ του βίντεο-οδηγού ως  $X$ , το ανάπτυγμα Taylor που προσεγγίζει το πυκνό πεδίο κίνησης δίνεται από τη Σχέση (17) [10], με σημείο αναφοράς  $p$  καθένα από τα χαρακτηριστικά σημεία  $p_1, \dots, p_k$ .

$$T_{X \leftarrow R}(p) = T_{X \leftarrow R}(p_k) + \left( \frac{d}{dp} T_{X \leftarrow R}(p) \Big|_{p = p_k} \right) (p - p_k) + o(\|p - p_k\|) \quad (17)$$



Έχοντας υπολογίσει την προσέγγιση του Πυκνού Πεδίου Κίνησης μέσω του αναπτύγματος Taylor πρώτου βαθμού σε σημείο αναφοράς  $p \in [p_1, p_k]$ , ο αντίστοιχος Ιακωβιανός Πίνακας (Jacobian Matrix), που συνιστά μία εκ των εισόδων της παραλλαγής του δικτύου Dense Motion Field του μοντέλου FOMM, δίνεται από τη Σχέση (18) [10] ως εξής.

$$T_{X \leftarrow R}(p) \approx \left\{ \left\{ T_{X \leftarrow R}(p_1), \left( \frac{d}{dp} T_{X \leftarrow R}(p) \right) \Big|_{p=p_1} \right\}, \dots, \left\{ T_{X \leftarrow R}(p_k), \left( \frac{d}{dp} T_{X \leftarrow R}(p) \right) \Big|_{p=p_k} \right\} \right\} \quad (18)$$

Δοθείσας της Σχέσης (17) που προσεγγίζει το ανάπτυγμα Taylor πρώτου βαθμού και της Σχέσης (18) για τον υπολογισμό του αντίστοιχου Ιακωβιανού Πίνακα για ένα καρέ X, μπορούμε επιπλέον να υπολογίσουμε και τα μεγέθη  $T_{S \leftarrow R}$  και  $T_{D \leftarrow R}$ , με εφαρμογή των μαθηματικών ορισμών. Για τον υπολογισμό του του  $T_{S \leftarrow D}$ , απαιτείται επίσης και ο υπολογισμός του πίνακα  $T_{R \leftarrow D}$ , όπως μπορεί να γίνει αντιληπτό και από την Εικόνα 10 [10]. Υποθέτοντας ότι ο πίνακας  $T_{D \leftarrow R}$  είναι τοπικά αντιστρέψιμος γύρω από μία γειτονιά του κάθε Χαρακτηριστικού Σημείου, προκύπτει ότι  $T_{R \leftarrow D} = T_{D \leftarrow R}^{-1}$ . Αξιοποιώντας, παράλληλα, το κάθε Χαρακτηριστικό Σημείο  $z_k$  του τρέχοντος καρέ D του βίντεο-οδηγού, υπό την προϋπόθεση ότι οι συντεταγμένες του  $z_k$  αντιστοιχούν σε αυτές του καρέ αναφοράς R. Έτσι, με βάση τις παραπάνω σχέσεις, η προσέγγιση του αναπτύγματος  $T_{S \leftarrow D}$  στο σημείο  $z \in [z_1, z_k]$  δίνεται από τη Σχέση (19) και  $J_k$  τον αντίστοιχο Ιακωβιανό Πίνακα να δίνεται από τη Σχέση (20).

$$T_{S \leftarrow D}(z) = T_{S \leftarrow R}(p_k) + J_k (z - T_{D \leftarrow R}(p_k)) \quad (19)$$

$$J_k = \left( \frac{d}{dp} T_{S \leftarrow R}(p) \Big|_{p=p_k} \right) \left( \frac{d}{dp} T_{D \leftarrow R}(p) \Big|_{p=p_k} \right)^{-1} \quad (20)$$

Ο υπολογισμός των προσεγγίσεων  $T_{S \leftarrow R}(p)$  και  $T_{D \leftarrow R}(p)$ , με  $p \in [p_1, p_k]$ , πραγματοποιείται μέσω του δικτύου U-Net [18], το οποίο αποτελεί τη βάση υλοποίησης του δικτύου Keypoint Detector των μοντέλων Monkey-Net και FOMM [7,10], και κατά συνέπεια πρόκειται για προσεγγίσεις των Τοπικών Αναπαραστάσεων  $H_k$  (Heatmaps) γύρω από το κάθε χαρακτηριστικό σημείο  $p_i$ . Η κάθε Τοπική Αναπαράσταση, στη συνέχεια, αποτελεί υπόδειξη στο δίκτυο Dense Motion Network για τις περιοχές που θα εφαρμοστούν οι προβλεπόμενοι μετασχηματισμοί στην δοθείσα εικόνα-πηγή S. Χρησιμοποιώντας τις  $K+1$  μάσκες  $M_i$ ,  $i \in [1, k]$ , με τις  $k$  μάσκες να αφορούν το Κάθε χαρακτηριστικό Σημείο που έχει προβλεφθεί από το δίκτυο Keypoint Detector και τη μάσκα  $M_0$  το υπόβαθρο (background) της εικόνας, η τελική προσέγγιση από τη μετάβαση των Χαρακτηριστικών Σημείων του καρέ D του βίντεο οδηγού σε αυτά της εικόνας-πηγής S πραγματοποιείται από το δίκτυο Dense Motion Network, αξιοποιώντας τις προσεγγίσεις  $T_{S \leftarrow R}$ ,  $T_{D \leftarrow R}$  και τον Ιακωβιανό Πίνακα  $J_k$  που εξήχθησαν από το δίκτυο Keypoint Detector. Η μαθηματική ορολογία που περιγράφει το αποτέλεσμα παρατίθεται μέσω της Σχέσης (21) [10].

$$\hat{T}_{S \leftarrow D} \approx M_0 + \sum_{k=1}^k M_k(T_{S \leftarrow R}(p_k) + J_k(z - T_{D \leftarrow R}(p_k))) \quad (21)$$

## 4.3 Παραγωγή του τελικού Καρέ με χρήση του Χάρτη Απόκλισης

Όπως αναλύθηκε διεξοδικά και στη παράγραφο 5.2, με αξιοποίηση της εικόνας-πηγής  $S$ , του τρέχοντος καρέ  $D$  του βίντεο-οδηγού και ενός τυχαία επιλεγμένου καρέ αναφοράς  $R$ , το δίκτυο Keypoint Detector παράγει τις προσεγγίσεις  $T_{S \leftarrow R}$  και  $T_{D \leftarrow R}$ , οι οποίες λαμβάνονται ως είσοδοι του δικτύου Dense Motion Network μαζί με τον Ιακωβιανό Πίνακα  $J_k$ , για την εξαγωγή του τελικού καρέ  $\hat{D} = \hat{T}_{S \leftarrow D}$ , υποδεικνύοντας ταυτόχρονα την προσαρμογή των Χαρακτηριστικών Σημείων του τρέχοντος καρέ στην εικόνα-πηγή. Ωστόσο, σε αρκετές περιπτώσεις είναι επόμενο τα pixels της εικόνας εξόδου  $\hat{D}$  να μην έχουν κοντινές τιμές με αυτές της εικόνας-εισόδου, γεγονός που οφείλεται στη μη-ορθή ευθυγράμμιση (misalignment) του κάθε εικονοστοιχείου της εικόνας-πηγής με το τελικό αποτέλεσμα  $\hat{D}$ .

Για την αντιμετώπιση του προβλήματος αυτό που παρουσιάζεται, ο Siarohin και η ομάδα του χρησιμοποιούν τη τεχνική της περιτύλιξης χαρακτηριστικών (feature warping)[12]. Η τεχνική αυτή χρησιμοποιεί δύο συνελκτικά blocks δειγματοληψίας (down-sampling convolutional blocks) διαστάσεων  $H' \times W'$  για την εξαγωγή ενός Χάρτη Χαρακτηριστικών (feature map)  $\xi \in R^{H' \times W'}$  για την διαδικασία της περιτύλιξης (warping).

Εν συνεχεία, ο χάρτης  $\xi$  εφαρμόζεται στο καρέ  $\hat{T}_{S \leftarrow D}$ . Παράλληλα, στην εικόνα που έχει εφαρμοστεί η τεχνική feature warping εξακολουθούν να υπάρχουν τμήματα που δεν έχουν αποδοθεί τιμές στα pixels τους και συνεπώς οι η απόδοση τιμών τους πρέπει να επιτευχθεί με βάση το περιεχόμενο των γειτονικών pixels (impainting).

Για το λόγο αυτό, η εκδοχή του δικτύου Dense Motion Network στο μοντέλο FOMM παράγει και το Χάρτη Απόκλισης  $\hat{O}_{S \leftarrow D} \in [0,1]^{H' \times W'}$ , όπως επισημάνθηκε και στη παράγραφο 4.1 κατά την επεξήγηση της αρχιτεκτονικής που αυτό ακολουθεί, όπου η εφαρμογή του στο προηγούμενο αποτέλεσμα υποδεικνύει πλέον τις περιοχές της τελικής εικόνας που πρέπει να γίνει η απόδοση κατάλληλων τιμών με τη τεχνική του impainting.

Η διαδικασία εξαγωγής της τελικής εικόνας  $\xi'$  μετά από την εφαρμογή του Χάρτη Απόκλισης (Occlusion Map), μπορεί να συνοψιστεί από τη Σχέση (22) [10] που τη περιγράφει με την αντίστοιχη μαθηματική ορολογία.

$$\xi' = \hat{O}_{S \leftarrow D} \odot f_w(\xi, \hat{T}_{S \leftarrow D}) \quad (22)$$

Με  $f_w(x, y)$  να αποτελεί τον τελεστή εφαρμογής της τεχνικής περυτιλίσματος μεταξύ του αποτελέσματος  $\hat{T}_{S \leftarrow D}$  του Dense Motion Network και του Χάρτη Χαρακτηριστικών  $\xi$ , και τον τελεστή  $\odot$  το γινόμενο Hadamard [5,22] ανάμεσα στο Χάρτη Απόκλισης  $\hat{O}_{S \leftarrow D}$  και του αποτελέσματος του τελεστή  $f_w(\xi, \hat{T}_{S \leftarrow D})$ .

## Κεφάλαιο 5. Μεθοδολογία

### 5.1 Η μέθοδος Seamless Cloning

Τα προβλήματα που σχετίζονται με το μετασχηματισμό μιας εικόνας στο πεδίο του χώρου, δηλαδή τη μεταβολή των τιμών των εικονοστοιχείων της με χρήση μετασχηματισμών ή φίλτρων (Kernels), μπορούν να κατανεμηθούν σε δύο κατηγορίες χωρικών μετασχηματισμών, τους τοπικούς μετασχηματισμούς (local transformations) ή τους ολικούς (global transformations), ανάλογα με το εάν ο μετασχηματισμός αφορά μέρος της εικόνας ή ολόκληρη την εικόνα αντίστοιχα.

Η μέθοδος seamless cloning, που προτάθηκε από τον Perez το 2003 [1] και συνιστά τη βάση για την επικόλληση τμημάτων μεταξύ εικόνων από γνωστές εφαρμογές επεξεργασίας εικόνας, κατατάσσεται στους τοπικούς χωρικούς μετασχηματισμούς, καθώς δοθείσας μίας εικόνας-πηγής  $f$  (source image), ενός τμήματος μίας εικόνας-προορισμού  $g$  (destination image) και μίας προσδιορισμένης περιοχής στη εικόνα-πηγή με τη μορφή δυαδικής μάσκας  $\Omega$  (binary mask), στοχεύει στη προσαρμογή του τμήματος αυτού της εικόνας προορισμού στη περιοχή της εικόνας-πηγής. Η προσαρμογή πραγματοποιείται με χρήση παρεμβολής στη προσδιορισμένη περιοχή, γεγονός με το οποίο εξασφαλίζεται η διαφάνεια (transparency) στη περιοχή και την ρεαλιστική ανάμιξη του τμήματος της εικόνας-προορισμού στην εικόνα-πηγή.

Η μαθηματική ορολογία που χαρακτηρίζει τη μέθοδο προσδιορίζεται στη συνέχεια, με τη τις σχέσεις που ακολουθούν να είναι άμεσα συνυφασμένες με την εξίσωση Poisson. Θεωρούμε  $S$  ένα κλειστό υποσύνολο του χώρου  $\mathbb{R}^2$ , στον οποίο είναι ορισμένη μία οποιαδήποτε εικόνα διαστάσεων  $H \times W$ , με  $H, W \in \mathbb{Z}$ , καθώς και μία κλειστή περιοχή  $\Omega \subseteq S$  με όριο  $\partial\Omega$ . Θεωρούμε, επίσης,  $f^*$  μία γνωστή βαθμωτή συνάρτηση ορισμένη στο  $S$ ,  $f$  και  $g$  δύο άγνωστες βαθμωτές συναρτήσεις στο  $\Omega$  και  $\mathbf{v}$  ένα διάνυσμα πεδίου, επίσης ορισμένο στο  $\Omega$ . Ο όρος παρεμβολής που προκύπτει από τη παρεμβολή της συνάρτησης  $f$  και της συνάρτησης  $f^*$  είναι μια λύση που ικανοποιεί το πρόβλημα ελαχιστοποίησης που περιγράφεται από τη Σχέση (23) [1] που ακολουθεί.

$$\min_f \iint_{\Omega} |\nabla f|^2 \text{ με } f|_{\partial\Omega} = f^*|_{\partial\Omega} \quad (23)$$

Με  $\nabla = \begin{bmatrix} \frac{d}{dx} & \frac{d}{dy} \end{bmatrix}$  τον τελεστή βάθμωσης (gradient operator). Ο ελαχιστοποιητής που προέρχεται από τη Σχέση (24) πρέπει επίσης να ικανοποιεί και την εξίσωση Euler – Lagrange που παρατίθεται παρακάτω.

$$\Delta f = 0 \text{ στο } \Omega, \text{ με } f|_{\partial\Omega} = f^*|_{\partial\Omega} \quad (24)$$

Με  $\Delta = \frac{d^2}{dx^2} + \frac{d^2}{dy^2}$  να αποτελεί τον τελεστή Laplace. Σε εφαρμογές που σχετίζονται με την επεξεργασία εικόνας, η αξιοποίηση των παραπάνω σχέσεων χωρίς κάποια άλλη επεξεργασία να ακολουθεί, έχει ως αποτέλεσμα τη παραγωγή μίας θολής και μη-ικανοποιητικής παρεμβολής μεταξύ των δύο εικόνων.

Προκειμένου να αντιμετωπιστεί το παραπάνω πρόβλημα, χρησιμοποιείται το διάνυσμα  $\mathbf{v}$  ως διάνυσμα-οδηγός (guidance vector), με παράλληλη αξιοποίηση της τεχνικής inpainting για την επίτευξη της διαφάνειας στη περιοχή που πραγματοποιείται η παρεμβολή. Το πρόβλημα εύρεσης του κατάλληλου ελαχιστοποιητή που αποτελεί λύση της Σχέσης (23) επεκτείνεται στην επίλυση της Σχέσης (25).

$$\min_f \iint_{\Omega} |\nabla f - \mathbf{v}|^2 \mu \varepsilon f | \partial\Omega = f^* | \partial\Omega \quad (25)$$

Με τις παραπάνω ρίζες, με τη σειρά τους, να ικανοποιούν τις οριακές συνθήκες Dirichlet [23]

$$\Delta f = \operatorname{div} \mathbf{v} \text{ στο } \Omega, \mu \varepsilon \tilde{f} | \partial\Omega = f^* | \partial\Omega \quad (26)$$

Όπου  $\operatorname{div} \mathbf{v} = \frac{du}{dx} + \frac{dv}{dy}$  η απόκλιση (divergence) του διανύσματος  $\mathbf{v} = (u, v)$ . Η τελευταία σχέση, η οποία αξιοποιεί τις Σχέσεις (23) και (24) που παρουσιάστηκαν νωρίτερα, συνιστά τη βάση για την υλοποίηση της εξίσωσης Poisson για την επεξεργασία διακριτών εικόνων. Για καθέναν από τα τρία κανάλια (color channels) που απαρτίζεται μία RGB εικόνα και με το υποσύνολο  $S$  να συνιστά τις τιμές των εικονοστοιχείων που απαρτίζουν την εικόνα διαστάσεων  $H \times W$ , επιλύονται τρεις εξισώσεις με αξιοποίηση των συνθηκών (23) έως (25).

Παράλληλα, γίνεται χρήση μίας συνάρτησης διόρθωσης  $\tilde{f}$ , με αποτέλεσμα η συνάρτηση  $f$  να είναι  $f = g + \tilde{f}$ . Συνεπώς, η εξίσωση Poisson της Σχέσης (26), ακολουθώντας την εξίσωση Laplace, πρέπει να ικανοποιεί τις οριακές συνθήκες Dirichlet [23] για τη παρακάτω περίπτωση.

$$\Delta \tilde{f} = 0 \text{ στο } \Omega, \text{ με } \tilde{f}|_{\partial\Omega} = (f^* - g)|_{\partial\Omega} \quad (27)$$

Το αποτέλεσμα που προκύπτει με τη χρήση των παραπάνω σχέσεων και της συνάρτησης διόρθωσης  $\tilde{f}$  συνιστά την αντιμετώπιση της μη-ικανοποιητικής παρεμβολής  $f^* - g$  μεταξύ των δύο δοθέντων εικόνων, στο ζητούμενο όριο  $\partial\Omega$  που πρέπει να προσαρμοστεί το τμήμα από την εικόνας-στόχου στην εικόνας-πηγή.

Η επίλυση του προβλήματος ελαχίστων τιμών της Σχέσης (27) και της εξίσωσης Poisson με τις οριακές συνθήκες Dirichlet, μπορεί να προσαρμοστεί στην επίλυση ενός προβλήματος εύρεσης ελαχίστων τιμών στο διακριτό χώρο. Χωρίς βλάβη της γενικότητας, αναθέτουμε τον συμβολισμό  $S$  στο σύνολο από τα pixels μίας εικόνας διαστάσεων  $H \times W$ , με το σύνολο αυτό να περιλαμβάνει είτε όλα τα pixels της εικόνας είτε ακόμα και ένα υποσύνολο αυτών, τον συμβολισμό  $p$  για το κάθε pixel που ανήκει στο  $S$ , καθώς επίσης και τον συμβολισμό  $\Omega$  για τη περιοχή στην εικόνα-πηγή που πρόκειται να προσαρτηθεί ένα τμήμα της εικόνας-προορισμού. Για καθένα από τα pixels  $p \in S$ , θεωρούμε ως  $N_p$  μία διακριτή γειτονιά από pixels, τέτοια ώστε οι γείτονες του τρέχοντος pixel να είναι το πρώτο pixel προς κάθε μία από τις 4 κατευθύνσεις. Για κάθε γειτονικό pixel  $q \in N_p$ , συνεπώς, έχουμε το πολύ 4 ζεύγη από pixels της μορφής  $\langle p, q \rangle$ . Για το όριο  $d\Omega$  στη περιοχή  $\Omega$  ισχύει  $d\Omega = \{p \in S \setminus \Omega \mid N_p \cap \Omega \neq \emptyset\}$  και  $f_p$  η συναρτησιακή τιμή του τρέχοντος pixel. Οι συναρτησιακές τιμές που πρέπει να υπολογιστούν, με το συνδυασμό των παραπάνω δεδομένων, είναι οι τιμές των pixels στη περιοχή  $\Omega$ , δηλαδή οι τιμές  $f|_{\Omega} = \{f_p, p \in \Omega\}$ .



Με τις συνθήκες Dirichlet να είναι ορισμένες στο όριο της αυθαίρετης περιοχής  $\Omega$ , προτιμάται η επίλυση του προβλήματος εύρεσης ελαχίστων τιμών της Σχέσης (25) στον διακριτό χώρο, έναντι της Σχέσης (26). Το πεπερασμένο πρόβλημα της Σχέσης (25) μετατρέπεται σε πρόβλημα βελτιστοποίησης τετραγωνικής μορφής, στη μορφή που παρουσιάζεται μέσω της Σχέσης (28).

$$\min_{f|_{\Omega}} \sum_{\langle p,q \rangle \cap \Omega \neq \emptyset} (f_p - f_q - v_{pq})^2, \text{ με } f_p = f_q^*, \forall p \in d\Omega \quad (28)$$

Με  $v_{pq} = \rightarrow_{pq}$  να αποτελεί τη προβολή  $\mathbf{v} \left( \frac{p+q}{2} \right)$  στη κατευθυνόμενη ακμή  $[p,q]$ .

Η ρίζα του προβλήματος βελτιστοποίησης τετραγωνικής μορφής ταυτόχρονα ικανοποιεί και τη γραμμική εξίσωση της Σχέσης (29) [1].

$$\forall p \in \Omega, \quad |N_p|f_p - \sum_{q \in N_p \cap \Omega} f_q = \sum_{q \in N_p \cap d\Omega} f_q^* + \sum_{q \in N_p} v_{pq} \quad [29]$$

Σε περίπτωση που η περιοχή  $\Omega$  περιέχει pixels που βρίσκονται στις άκρες του συνόλου  $S$ , για το μέγεθος της γειτονιάς  $N_p$  ισχύει ότι  $|N_p| < 4$ , καθώς στη περίπτωση αυτή η περιοχή  $\Omega$  δεν μπορεί να περιλαμβάνει pixels σε θέσεις εκτός των διαστάσεων της εικόνας. Συνεπώς, η παραπάνω γραμμική εξίσωση που πρέπει να ικανοποιείται αποκτά την εξής μορφή

$$\forall p \in \Omega, \quad |N_p|f_p = \sum_{q \in N_p} v_{pq} \quad (30)$$

Έχοντας παρουσιάσει την απαραίτητη μαθηματική ορολογία και το γνωστικό υπόβαθρο που απαιτείται για την κατανόηση της μεθόδου seamless cloning και τον τρόπο που αξιοποιεί την εξίσωση Poisson, ακολουθούν παραδείγματα εικόνων ύστερα από την εφαρμογή της μεθόδου.

Η Εικόνα 12 [1] που ακολουθεί περιλαμβάνει εικόνες-προορισμού, εκ των οποίων τμήματα αυτών έχουν επιλεγεί για να προσαρτηθούν σε μία δοθείσα εικόνα-πηγής



*Εικόνα 12. Παραδείγματα εικόνων-προορισμού, με τις αντίστοιχες περιοχές προς ενσωμάτωση στην εικόνα-πηγή [1]*

Χωρίς να χρησιμοποιείται η μέθοδος seamless cloning, το αποτέλεσμα της προσάρτησης τους θα ήταν μια απλή επικόλληση (paste) των τμημάτων αυτών στην εικόνα-πηγή, καθώς δεν χρησιμοποιείται παρεμβολή μεταξύ των τμημάτων και της εικόνας-πηγής, γεγονός το οποίο είναι εμφανές μέσω της Εικόνας 12 [1] που παρατίθεται στο αριστερό τμήμα. Το αποτέλεσμα της χρήσης της seamless cloning συνιστά η Εικόνα 13 [1] που ακολουθεί, όπου η χρήση παρεμβολής μεταξύ των τμημάτων από τις εικόνες-προορισμού και της εικόνας-πηγής μπορεί να παρατηρηθεί πλέον και οπτικά.



Εικόνα 14. Το αποτέλεσμα της επικόλλησης των εικόνων-προορισμού στην εικόνα-πηγή, χωρίς την εφαρμογή παρεμβολής (αριστερά) και ύστερα από την εφαρμογή παρεμβολής (δεξιά) [1]

Στη παράγραφο 5.2 που ακολουθεί, περιγράφεται η μέθοδος με την οποία ανακατασκευάζονται τα καρέ του βίντεο που εξάγει το μοντέλο FOMM με χρήση της μεθόδου seamless cloning, με την εικόνα-πηγή στη περίπτωση αυτή να αποτελεί το πρώτο καρέ του εξαγόμενου βίντεο, την εικόνα-προορισμού να αποτελεί το τρέχον καρέ του βίντεο, εκτός του πρώτου, και τη περιοχή που θα προσαρτηθεί στην εικόνα-πηγή να υποδεικνύεται από τον χάρτη απόκλισης, ο οποίος παράλληλα υποδεικνύει και τις περιοχές του καρέ που είναι απαραίτητη η εφαρμογή και της μεθόδου inpainting.

## 5.2 Εφαρμογή της μεθόδου Seamless cloning στα παραγόμενα Καρέ

Η μεθοδολογία με την οποία δομείται η μέθοδος seamless cloning, βασισμένη στην ικανοποίηση των προβλημάτων βελτιστοποίησης και της εξίσωσης Poisson που αναφέρθηκαν στο προηγούμενο κεφάλαιο, επεκτείνεται στη παρούσα Διπλωματική εργασία με τον συνδυασμό των αποτελεσμάτων που εξάγονται από το μοντέλο FOMM, με στόχο την ανακατασκευή του χαμηλής ποιότητας παραγόμενου βίντεο και, κατ' επέκταση, την βελτίωση των ποιοτικών μεγεθών AED, L1 και SSIM μέσω της διαδικασία αυτής.

Για καθένα από τα καρέ του παραγόμενου βίντεο διαστάσεων  $H \times W$  και  $H, W \in \mathbb{Z}$ , με εξαίρεση το πρώτο που αποτελεί την εικόνα-πηγή (source image), το τρέχον καρέ του βίντεο συνιστά την εικόνα-προορισμού (destination-image) και η περιοχή της εικόνας-πηγής που θα εφαρμοστεί η παρεμβολή που αξιοποιεί η μέθοδος προσδιορίζεται από τον Χάρτη Απόκλισης  $\hat{O}_{S \leftarrow D}$  που εξάγεται από το μοντέλο FOMM παράλληλα με το παραγόμενο βίντεο, όπως αναφέρθηκε και νωρίτερα στο κεφάλαιο 4.

Ο ρόλος του Χάρτη Απόκλισης  $\hat{O}_{S \leftarrow D}$ , κατά τη χρήση της μεθόδου, αφορά τόσο για τη δυαδική μάσκα που προσδιορίζει τη περιοχή που θα εφαρμοστεί η παρεμβολή μεταξύ τμήματος του τρέχοντος καρέ και του πρώτου, όσο και ως προς το μέγεθος που υποδεικνύει τις περιοχές στην εικόνα που πρέπει να εφαρμοστεί η τεχνική inpainting.

Εφαρμόζοντας τη μέθοδο με τα παραπάνω δεδομένα ως εισόδους της μεθόδου, παρατηρείται αισθητή μείωση του θορύβου και βελτίωση της ποιότητας του καρέ σε σχέση με το αρχικό, αλλά το πρόβλημα που παρουσιάζεται είναι ο διατήρηση του πρώτου καρέ στο υπόβαθρο (background) του νέου καρέ, όπως παρουσιάζεται στην Εικόνα 14.



Εικόνα 14. Αποτέλεσμα της εφαρμογής της μεθόδου *seamless cloning* (δεξιά), χωρίς κάποια επιπρόσθετη τροποποίηση.

Προκειμένου να αφαιρεθεί το υπόβαθρο από το παραγόμενο καρέ, το οποίο προέρχεται από τη πρώτο καρέ του παραγόμενου βίντεο, χρησιμοποιήθηκε η εξής μεθοδολογία. Προκειμένου να εξακριβωθούν ποια από τα pixels του πρώτου καρέ, που παράλληλα συνιστά και την εικόνα-πηγή της μεθόδου, αποτελούν το μπροστινό τμήμα της (foreground) και ποια το υπόβαθρό της (background), έγινε χρήση της τεχνικής κατοφλίωσης (thresholding).

Σύμφωνα με τη τεχνική αυτή, για κάθε pixel  $p \in F$ , όπου  $F$  μία δοθείσα εικόνα διαστάσεων  $H \times W$  και  $t$  η τιμή του κατωφλίου (threshold), τα pixels για τα οποία ισχύει  $p \leq t$  ανατίθενται στη τιμή 255, ενώ σε αντίθετη περίπτωση στη τιμή 0.

Η τεχνική αυτή παράγει ως αποτέλεσμα μία δυαδική μάσκα  $M$ , στόχος της οποίας είναι η κατηγοριοποίηση των pixels του μπροστινού τμήματος της εικόνας από αυτά του υποβάθρου. Η διαδικασία εξαγωγής του υποβάθρου από την εικόνα-πηγή ολοκληρώνεται με τη παραγωγή της νέας εικόνας  $F^*$  που αποτελείται μόνο από τα pixels του μπροστινού μέρους της, μέσω της εφαρμογής του δυαδικού τελεστή XOR μεταξύ της δυαδικής μάσκας και της εικόνας-πηγής. Το παραπάνω αποτέλεσμα δίνεται από τη Σχέση (31) που ακολουθεί.

$$F^* = F \oplus M \quad (31)$$

Η αντιπαράθεση του βίντεο που εξάγει το μοντέλο FOMM με το προϊόν της μετα-επεξεργασίας (post processing) του κάθε καρέ μπορεί να γίνει αντιληπτή στην Εικόνα 15 που παρατίθενται παρακάτω, όπου στο αριστερό τμήμα παρουσιάζεται το αρχικό καρέ του βίντεο του μοντέλου FOMM και στο δεξιό τμήμα το επεξεργασμένο καρέ. Συμπληρωματικά με την Εικόνα 15, από την οποία παρατηρείται βελτίωση της ποιότητας του παραγόμενου βίντεο και πιο ευκρινής αναπαράσταση των κινήσεων που καλείται να προσαρμόσει το μοντέλο FOMM, παρατίθεται η Εικόνα 16 στην οποία παράλληλα γίνεται αντιληπτό το impainting που πραγματοποιείται στις περιοχές που υποδεικνύει ο Χάρτης Απόκλισης  $\hat{O}_{S \leftarrow D}$  με τη χρήση της μεθόδου seamless cloning [1].



Εικόνα 15. Αποτέλεσμα της εφαρμογής της μεθόδου *seamless cloning*, με αφαίρεση του *background* της εικόνας-πηγής. Στο αριστερό τμήμα παρουσιάζεται το παραγόμενο καρέ ενός βίντεο στο οποίο δεν έχουν εφαρμοστεί οι παρπάνω μέθοδοι, ενώ στο δεξιό τμήμα παρατίθεται το ίδιο καρέ, ύστερα από τη χρήση των μεθόδων.



Εικόνα 16. Επιπλέον παράδειγμα ανακατασκευής ενός καρέ, με πιο αισθητή τη χρήση της μεθόδου *seamless cloning*

Με τη παρουσίαση της μεθόδου *seamless cloning* και του τρόπου που αυτή αξιοποιήθηκε στη παρούσα διπλωματική εργασία για την ανακατασκευή και βελτίωση της ποιότητας των βίντεο που παράγονται από το μοντέλο FOMM, γίνεται μετάβαση στο κεφάλαιο 6, το οποίο πρόκειται και για το τελευταίο κεφάλαιο της διπλωματικής εργασίας, στο οποίο παρουσιάζονται οι μετρήσεις των μεγεθών AED, L1 και SSIM μεταξύ των αρχικών βίντεο που παράγονται και των ανακατασκευασμένων.



## Κεφάλαιο 6. Αποτελέσματα

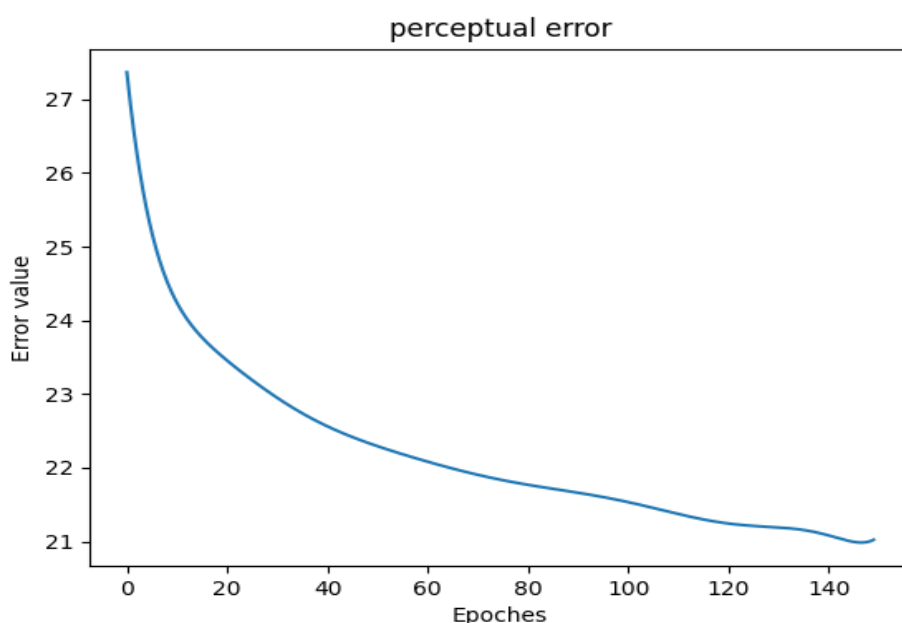
### 6.1 Παράμετροι του δικτύου και μετρικές απόδοσης

Για τη διεξαγωγή της εκπαίδευσης των Νευρωνικών Δικτύων που απαρτίζουν το μοντέλο FOMM, έχουν χρησιμοποιηθεί προσεγγιστικά 3000 βίντεο από δύο διαφορετικά σύνολα δεδομένων (training datasets). Τα σύνολα αυτά πρόκειται για τα σύνολα Taichi και Fashion [7,10], αποτελούμενα από βίντεο που χαρακτηρίζονται κινήσεις του ανθρωπίνου σώματος και εκφράσεων του προσώπου, με τις διαστάσεις των βίντεο αυτών να διακυμαίνονται μεταξύ  $256 \times 256$  και  $512 \times 512$  στη πλειοψηφία τους. Ως προς τις παραμέτρους εκπαίδευσης του μοντέλου, τόσο το δίκτυο Keypoint Detector όσο και το δίκτυο Dense Motion Field έχουν εκπαιδευτεί για αριθμό εποχών  $N = 150$  εποχές, ρυθμό μάθησης (learning rate)  $\eta = 2 \cdot 10^{-4}$  και αριθμό επαναλήψεων  $nr = 150$ . Ο αναφερθέντας ρυθμός μάθησης είναι κοινός για τη γεννήτρια και τον διαχωριστή του δικτύου Dense Motion Field, το οποίο πρόκειται για ένα δίκτυο GAN [15], όπως αναφέρθηκε και στο κεφάλαιο 4 κατά την επεξήγηση των δομών του μοντέλου FOMM [7,10].

Η παράμετρος  $nr$  συμβάλει στη βελτίωση της εκπαίδευσης του δικτύου μέσω της μεγέθυνσης της τρέχουσας εποχής κατά  $nr$  φορές και συνιστά μία από τις ειδικές παραμέτρους του μοντέλου.

Ο χρόνος που απαιτείται για την εκπαίδευση του δικτύου με χρήση των παραπάνω παραμέτρων τείνει, κατά προσέγγιση, σε 5 ολόκληρες ημέρες όταν αρχίσει η εκπαίδευση του μοντέλου. Για την επαλήθευση της ικανοποιητικής εκπαίδευσης του μοντέλου, έχει γίνει γραφική αναπαράσταση της πορείας του σφάλματος εκπαίδευσης για κάθε μία από τις συνιστώσες του μοντέλου, με αντίστοιχη επεξήγηση των γραφικών παραστάσεων των κάτωθι εικόνων.

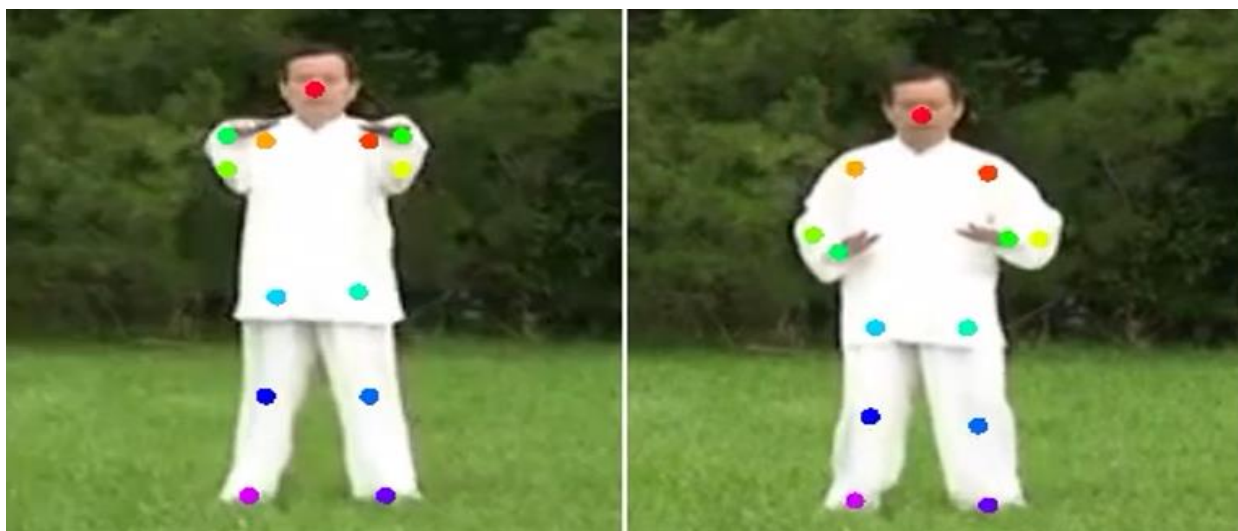
Σχετικά με το Σφάλμα Ομοιότητας (Perceptual Error), το οποίο κατά την εκπαίδευση Νευρωνικών Δικτύων για την αναγνώριση προτύπων ταυτίζεται με την έννοια του σφάλματος εκπαίδευσης στη γενικότερη ορολογία των Νευρωνικών Δικτύων, παρουσιάζει καθοδική πορεία, όπως γίνεται αντιληπτό από τη γραφική παράσταση της Εικόνας 17.



Εικόνα 17. Γραφική αναπαράσταση της πορείας του σφάλματος εκπαίδευσης του μοντέλου FOMM

Τα πρότυπα που καλείται να αναγνωρίσει το μοντέλο FOMM πρόκεινται για τα μέρη του ανθρωπίνου σώματος και του προσώπου στην εικόνα-πηγή και στα καρέ του βίντεο-οδηγού, με σκοπό τη σωστή απόδοση των Χαρακτηριστικών Σημείων (Keypoints) στις περιοχές των εικόνων που αυτά ανιχνεύονται. Η σταδιακή μείωση του σφάλματος με τη πάροδο των εποχών υποδεικνύει την αποτελεσματική απόδοση των Χαρακτηριστικών Σημείων, με την Εικόνα 18, η οποία συνιστά απόδειξη στον παραπάνω ισχυρισμό.

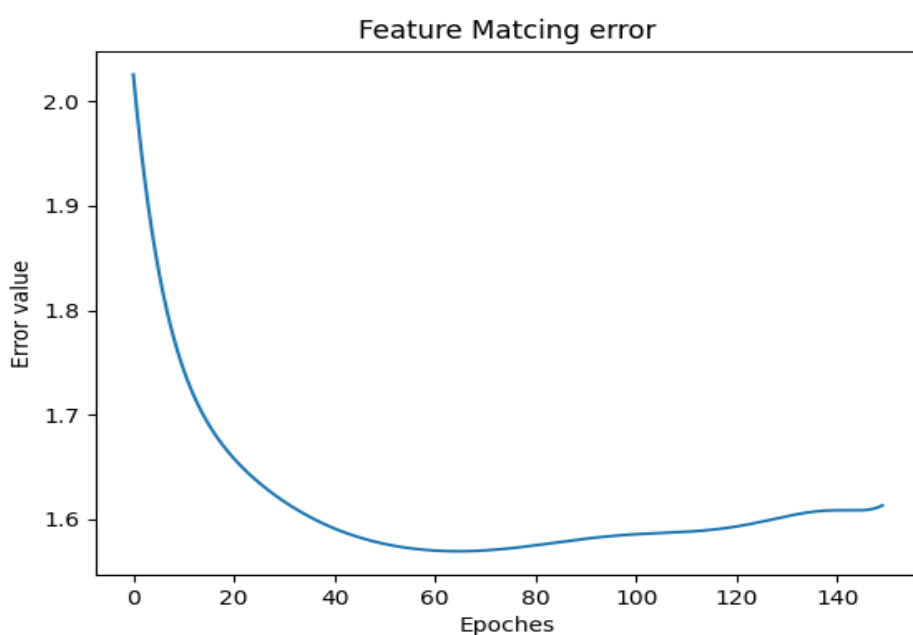




Εικόνα 18. Παράδειγμα απόδοσης Χαρακτηριστικών Σημείων στην εικόνα-πηγή (αριστερά) και στο τρέχον καρέ του βίντεο-οδηγού (δεξιά)

Το μέγεθος που αναλύεται στη συνέχεια πρόκειται για το Σφάλμα Εύρεσης Χαρακτηριστικών (Feature Matching error) [24], μέσω του οποίου συγκρίνεται το καρέ που παράγει σε κάθε επανάληψη το μοντέλο με το αναμενόμενο αποτέλεσμα, προκειμένου να ενημερωθούν κατάλληλα τα βάρη του δικτύου στο τέλος κάθε εποχής, με απότερο σκοπό την ελαχιστοποίηση του σφάλματος.

Το σφάλμα Feature Matching παρουσιάζει σταδιακή μείωση μέχρι, προσεγγιστικά, την εποχή 80, από την οποία παρατηρείται μία ανοδική πορεία του, διατηρούμενη ωστόσο σε χαμηλά επίπεδα σε σχέση με το αρχικό σφάλμα.

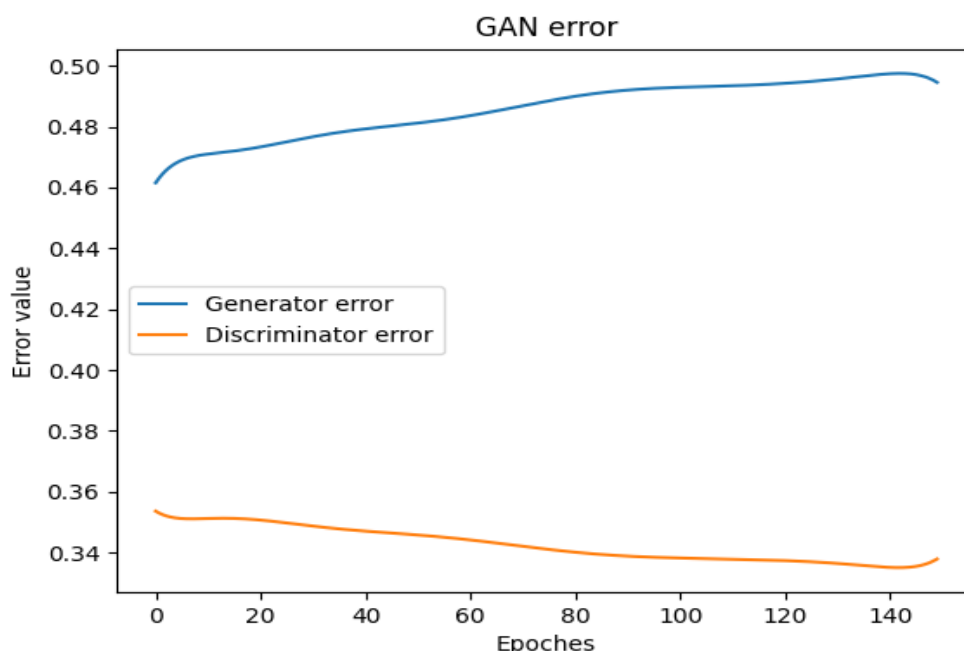


Εικόνα 19. Γραφική παράσταση της πορείας του Σφάλματος Εύρεσης Χαρακτηριστικών

Η αύξηση του συγκεκριμένου σφάλματος, με βάση τη γραφική παράσταση της εικόνας 19, συνιστά και τον λόγο που γίνεται χρήση της μεθόδου seamless cloning, ως μετα-επεξεργασία της του κάθε καρέ του παραγόμενου βίντεο για την αντιμετώπιση της αύξησης του. Τα αρχικά καρέ που παράγονται από το μοντέλο FOMM, βάσει των οποίων γίνεται η σύγκριση των επεξεργασμένων καρέ με αυτά στη παράγραφο 5.2, αποτελούν παραδείγματα μέσω των οποίων είναι εμφανής η αύξηση του.

Το τελευταίο μέγεθος που αφορά το μοντέλο FOMM είναι το σφάλμα του δικτύου Dense Motion Field, το οποίο ακολουθεί τη δομή ενός δικτύου GAN [15]. Τα δίκτυα GAN, στα οποία έγινε αναφορά στη παράγραφο 4.1 κατά τη περιγραφή των δομών του μοντέλου FOMM, αποτελούνται από μία γεννήτρια (generator) και ένα διαχωριστή (discriminator). Ο ρόλος της γεννήτριας είναι κάθε φορά η παραγωγή μίας τυχαίας εικόνας, την οποία ο διαχωριστής καλείται να αναγνωρίσει η εικόνα πρόκειται για ψεύτικη ή αληθινή, ως προς τα παραδείγματα που έχει εκπαιδευτεί. Καθώς αυξάνονται οι εποχές, στόχος στο δίκτυο GAN είναι η γεννήτρια να βρεθεί σε θέση να παράξει ένα καρέ, κοντινό ως προς το περιεχόμενο με αυτό του συνόλου εκπαίδευσης, προκειμένου ο διαχωριστής να μη δύναται να την αναγνωρίσει ως ψεύτικο δεδομένο και να τη κατατάξει ως αληθινό. Ερμηνεύοντας τη γραφική παράσταση της Εικόνας 20, που πρόκειται για την πορεία της εκπαίδευσης του δικτύου Dense Motion Field, η γεννήτρια τείνει στην παραγωγή καρέ, τα οποία τείνουν ολοένα και περισσότερο προς τα παραδείγματα που έχει εκπαιδευτεί το μοντέλο αντί για ένα καρέ τυχαίου περιεχομένου, με άμεση συνέπεια την αύξηση του σφάλματος της γεννήτριας.

Παράλληλα, ο διαχωριστής τείνει στη κατάταξη του νέου καρέ ως πραγματικό δεδομένο ως πραγματικό, καθώς το νέο καρέ έχει μεγάλη ομοιότητα με αυτά του συνόλου εκπαίδευσης, μειώνοντας έτσι το σφάλμα του διαχωριστή.



Εικόνα 20. Γραφική παράσταση της πορείας της γεννήτριας και του διαχωριστή του δικτύου Dense Motion Network

Σε άμεση σύνδεση με την ικανοποιητική απόδοση των Χαρακτηριστικών Σημείων στην εικόνα-πηγή και στο τρέχον καρέ του βίντεο-οδηγού, το οποίο επιτυγχάνεται με την ελαχιστοποίηση του Εύρεσης Χαρακτηριστικών (Feature Matching error) του δικτύου Keypoint Detector, η συμπεριφορά του σφάλματος του δικτύου Dense Motion Field είναι αναμενόμενη και εξαρτώμενη από την ποιότητα της εκπαίδευσης του δικτύου Keypoint Detector. Για κάθε καρέ του βίντεο-οδηγού, παράγεται ένα καρέ όπου η κίνηση πραγματοποιείται από την τα Χαρακτηριστικά Σημεία της εικόνας-πηγής προς τα χαρακτηριστικά σημεία του καρέ του βίντεο-οδηγού, με το καρέ αυτό να τείνει να κατατάσσεται ως πραγματικό δεδομένο, καθώς έχει μεγάλη ομοιότητα με την εικόνα-πηγή.

Με την ολοκλήρωση της παρουσίασης και ανάλυσης των γραφικών παραστάσεων της πορείας του σφάλματος για τα δίκτυα που απαρτίζουν το μοντέλο FOMM, ακολουθεί η ανάλυση των μεγεθών που υποδεικνύουν την ύπαρξη βελτίωσης του επεξεργασμένου βίντεο με τη χρήση της μεθόδου seamless cloning, συγκριτικά με το αρχικό βίντεο που εξάγει το μοντέλο.

Τα μεγέθη αυτά, τα οποία παρουσιάζονται και στη παράγραφο **1.1** κατά τη περιγραφή των στόχων που τέθηκαν στη παρούσα Διπλωματική Εργασία, πρόκεινται για τη την απώλεια L1, τη Μέση Ευκλείδεια Απόσταση (Average Euclidean Distance) και το δείκτη Δομικής Ομοιότητας (Structural Similarity). Για τα μεγέθη αυτά, παρατίθεται μία σύντομη περιγραφή της σκοπιμότητας τους, σχετικά με τις διεξαχθείσες μετρήσεις.

- **Απώλεια L1:**

Η απώλεια L1 συνιστά ένα ποσοτικό μέγεθος για την εξαγωγή πληροφοριών χαμηλού επιπέδου μεταξύ του καρέ  $\hat{y}$  που παράγεται από το μοντέλο FOMM και της πρόβλεψης  $y$  της δομής του καρέ. Με άλλα λόγια, το μέγεθος αυτό μετρά το πόσο διαφέρουν το εξαγόμενο καρέ με αυτό που προβλέπει το μοντέλο και συνεπώς πρόκειται για την κανονικοποιημένη Ευκλείδεια Απόσταση μεταξύ των δύο καρέ. Η συναρτησιακή τιμή της απώλειας L1 δίνεται από τη Σχέση (32) που ακολουθεί.

$$L_1 = \frac{\|\hat{y} - y\|_2^2}{C \cdot H \cdot W} \quad (32)$$

Όπου  $C$  το πλήθος χρωματικών καναλιών (color channels) των δύο καρέ,  $H$  το πλήθος γραμμών των δύο καρέ και  $W$  το πλήθος στηλών τους.

- **Μέση Ευκλείδεια Απόσταση (Average Euclidean Distance):**

Η Μέση Ευκλείδεια Απόσταση, εν συντομία AED με βάση την Αγγλική ορολογία, πρόκειται για τη μέση τιμή της απώλειας  $L_1$  μεταξύ των συγκρινόμενων καρέ του παραχθέντος βίντεο και του επεξεργασμένου, για καθένα από τα  $N$  καρέ των συγκρινόμενων βίντεο. Η τιμή του μεγέθους αυτού δίνεται από τη παρακάτω Σχέση (33).

$$AED = \frac{\sum_i^N \sum_1^H \sum_1^W L_1(\hat{y}_i, y_i)}{N} \quad (33)$$

- **Δείκτης Δομικής Ομοιότητας (Structural Similarity):**

Ο Δείκτης Δομικής Ομοιότητας πρόκειται για το μέγεθος μέσω του οποίου εκφράζεται το ποσοστό ομοιότητας μεταξύ δύο συγκρινόμενων εικόνων. Ο υπολογισμός του στη δείκτη αυτού στη παρούσα διπλωματική εργασία αφορά την ομοιότητα μεταξύ των ανακατασκευασμένων καρέ του παραγόμενου βίντεο με τη πρόβλεψη που εξάγει το μοντέλο FOMM για το εκάστοτε καρέ, προκειμένου να γίνει αντιληπτή η βελτίωση ή υποβάθμιση της ποιότητας του ανακατασκευασμένου βίντεο σε σχέση με το αρχικό. Ο υπολογισμός του Δείκτη Δομικής Ομοιότητας αποτελεί συνάρτηση της φωτεινότητας (luminance), της αντίθεσης (contrast) και της δομής (structure) μεταξύ των συγκρινόμενων εικόνων. Ο υπολογισμός των τριών αναφερθέντων συνιστωσών παρουσιάζεται από τις ακόλουθες σχέσεις.

- **Φωτεινότητα (luminance):** Η συνιστώσα της σύγκρισης της φωτεινότητας μεταξύ δύο εικόνων  $x, y$ , διαστάσεων  $H \times W$  και οι δύο, δίνεται από τη σχέση

$$l(x, y) = \frac{2\mu_x\mu_y + c_1}{\mu_x^2 + \mu_y^2 + c_1} \quad (34)$$

Με  $\mu_x, \mu_y$  να αποτελούν τη μέση τιμή των εικονοστοιχείων των εικόνων  $x, y$  αντίστοιχα και  $c_1 = 0.01 \cdot L$ , όπου  $L = 255$  το εύρος τιμών που δύνανται να ανατεθούν στα εικονοστοιχεία της εκάστοτε εικόνας.

- **Αντίθεση (contrast):** Η συνιστώσα της σύγκρισης της αντίθεσης μεταξύ δύο εικόνων  $x, y$ , διαστάσεων  $H \times W$  και οι δύο, δίνεται από τη σχέση

$$c(x, y) = \frac{2\sigma_x\sigma_y + c_2}{\sigma_x^2 + \sigma_y^2 + c_2} \quad (35)$$

Με  $\sigma_x, \sigma_y$  να αποτελούν τη διακύμανση των τιμών των εικονοστοιχείων των εικόνων  $x, y$  αντίστοιχα και  $c_2 = 0.03 \cdot L$ , όπου  $L = 255$  το εύρος τιμών που δύνανται να ανατεθούν στα εικονοστοιχεία της εκάστοτε εικόνας.

- **Δομή (structure):** Η συνιστώσα της σύγκρισης της δομής μεταξύ δύο εικόνων  $x, y$ , διαστάσεων  $H \times W$  και οι δύο, δίνεται από τη σχέση

$$s(x, y) = \frac{\sigma_{xy} + c_3}{\sigma_x \cdot \sigma_y + 3} \quad (36)$$

Με  $\sigma_x, \sigma_y$  να αποτελούν τη διακύμανση των τιμών των εικονοστοιχείων των εικόνων  $x$  και  $y$  αντίστοιχα,  $c_3 = \frac{c_2}{2}$  και  $\sigma_{xy}$  τη τιμή της συνδιακύμανσης των εικόνων  $x$  και  $y$ .

Εν τέλει, η τιμή του Δείκτης Δομικής Ομοιότητας μεταξύ των δύο συγκρινόμενων μεγεθών είναι η απόλυτη τιμή του γινομένου των τριών συνιστωσών, που δίνεται από τη παρακάτω σχέση

$$SSIM(x, y) = \|l(x, y) \cdot c(x, y) \cdot s(x, y)\| \quad (37)$$

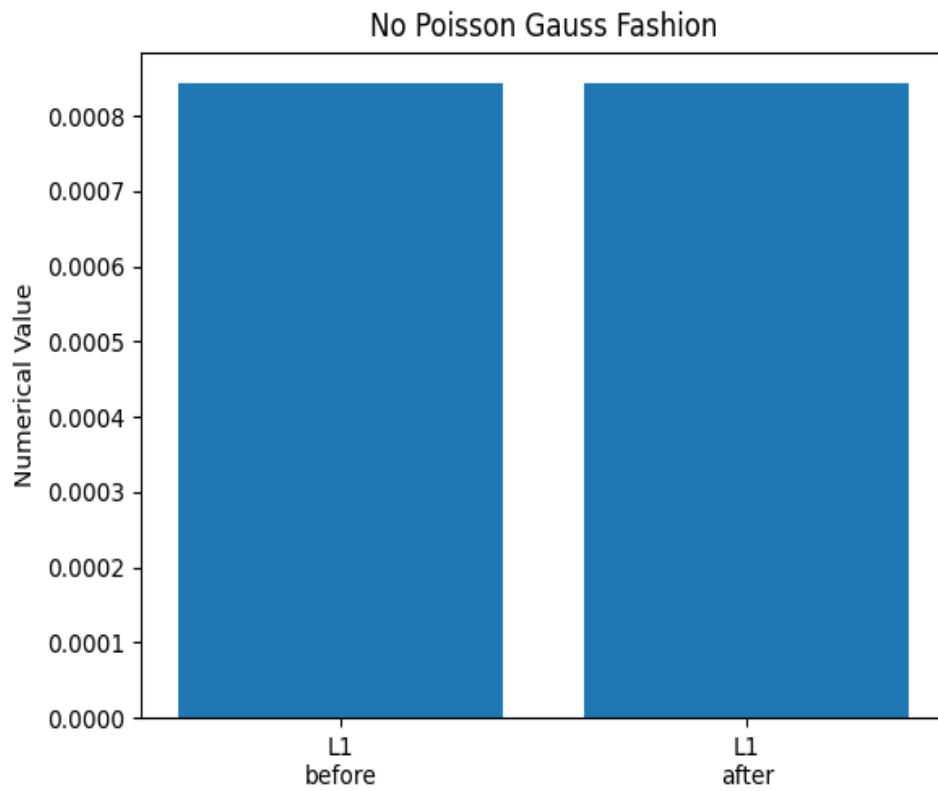
Με την ολοκλήρωση της παρουσίασης των εξεταστέων μεγεθών για την εξαγωγή των πορισμάτων της παρούσας διπλωματικής εργασίας, ακολουθούν οι μετρήσεις των μεγεθών αυτών που πραγματοποιήθηκαν στις περιπτώσεις συνδυασμού των μεθόδων *seamless cloning* και του Γκαουσιανού φίλτραρίσματος και της εφαρμογής της κάθε τεχνικής ξεχωριστά, συγκριτικά με τις τιμές των μεγεθών χωρίς την εφαρμογή των παραπάνω μεθόδων. Παράλληλα, παρατίθενται και οι γραφικές παραστάσεις σε μορφή ραβδογραμμάτων (*barplots*) με σκοπό την οπτική παρουσίαση της μεταβολής των μεγεθών ενδιαφέροντος.

## **6.2 Εφαρμογή Γκαουσιανού φίλτρου στα παραγόμενα βίντεο.**

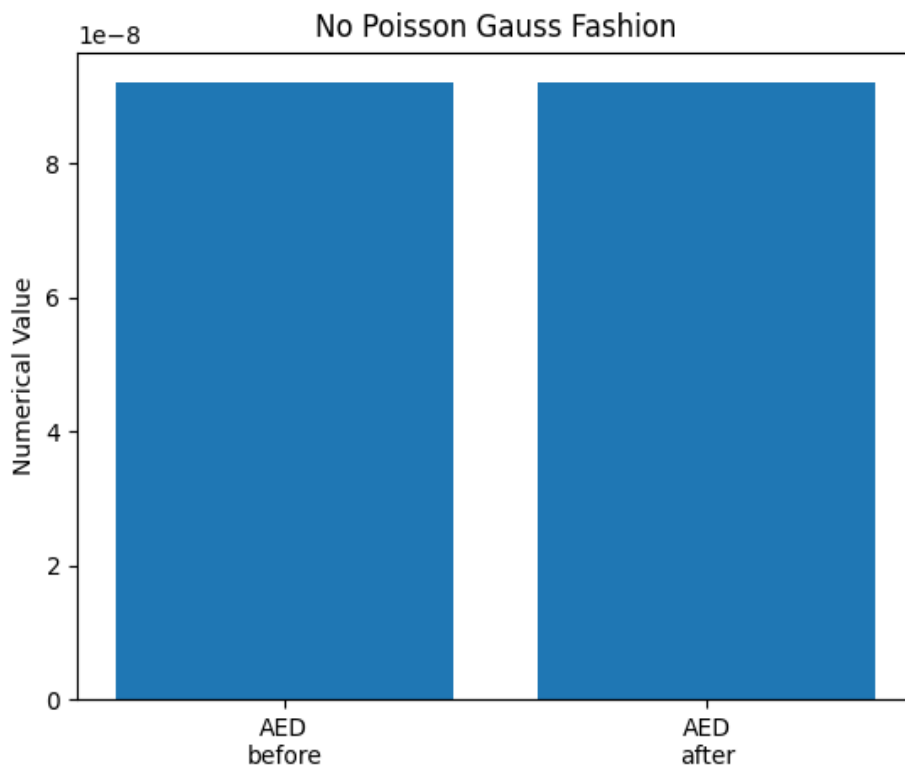
### **6.2.1 Αποτελέσματα μετρικών στο σύνολο Fashion**

Τα ποσοτικά αποτελέσματα που αναλύονται στη παρούσα παράγραφο αφορούν τη μεταβολή των μεγεθών *L1*, *AED* και *SSIM*, τα οποία έχουν υπολογιστεί με βάση τους μαθηματικούς ορισμούς που περιγράφηκαν στη παράγραφο **6.1**. Οι γραφικές παραστάσεις που ακολουθούν υποδεικνύουν τις μεταβολές που παρατηρήθηκαν στα μεγέθη κατά την ανακατασκευή των καρτέ των βίντεο που παράγονται από το μοντέλο *FOMM* στο σύνολο *Fashion* με χρήση Γκαουσιανού φίλτρου, στο οποίο η τιμή της τυπικής απόκλισης είναι  $\sigma = 2.5$ .

Αρχικά, παρατίθεται το γράφημα σύγκρισης της μεταβολής του μεγέθους *L1* για τα βίντεο που παράγονται ως έξοδοι από το μοντέλο *FOMM* και τα ανακατασκευασμένα βίντεο, ύστερα από την εφαρμογή του Γκαουσιανού φίλτρου και ακολουθεί το γράφημα σύγκρισης που αφορά τη Μέση Ευκλείδεια Απόσταση (*AED*).

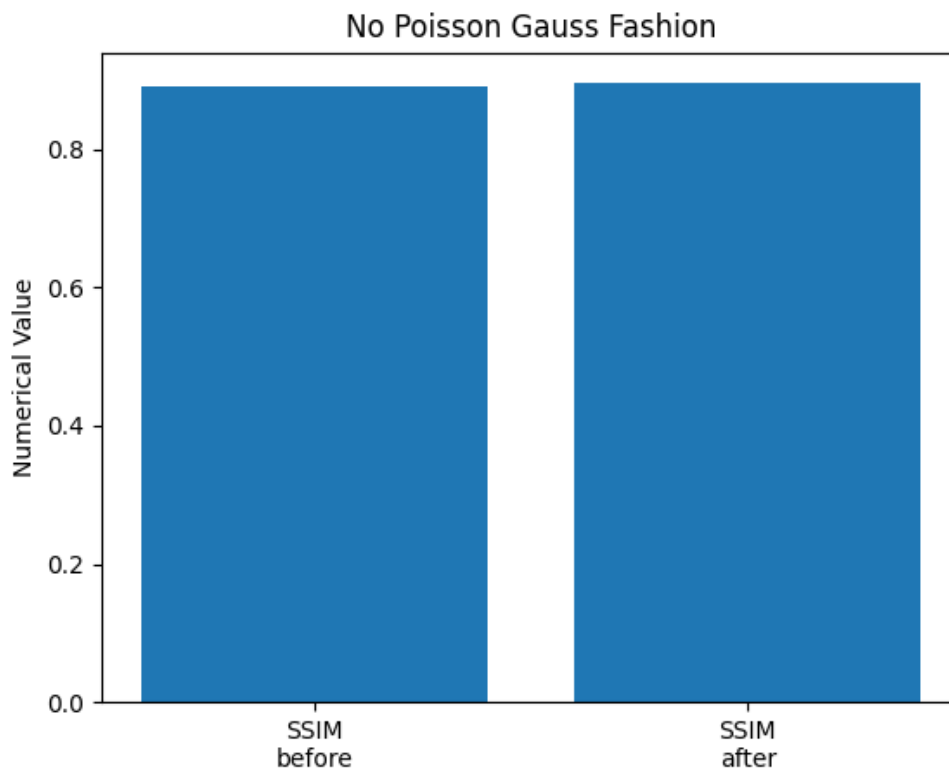


Εικόνα 21. Σύγκριση της τιμής της μετρικής L1 των αρχικών βίντεο του συνόλου Fashion (αριστερά) με τα ανακατασκευασμένα (δεξιά), παρουσία Γκαουσιανού φίλτρου



Εικόνα 22. Σύγκριση της τιμής της μετρικής AED των αρχικών βίντεο του συνόλου Fashion (αριστερά) με τα ανακατασκευασμένα (δεξιά), παρουσία Γκαουσιανού φίλτρου



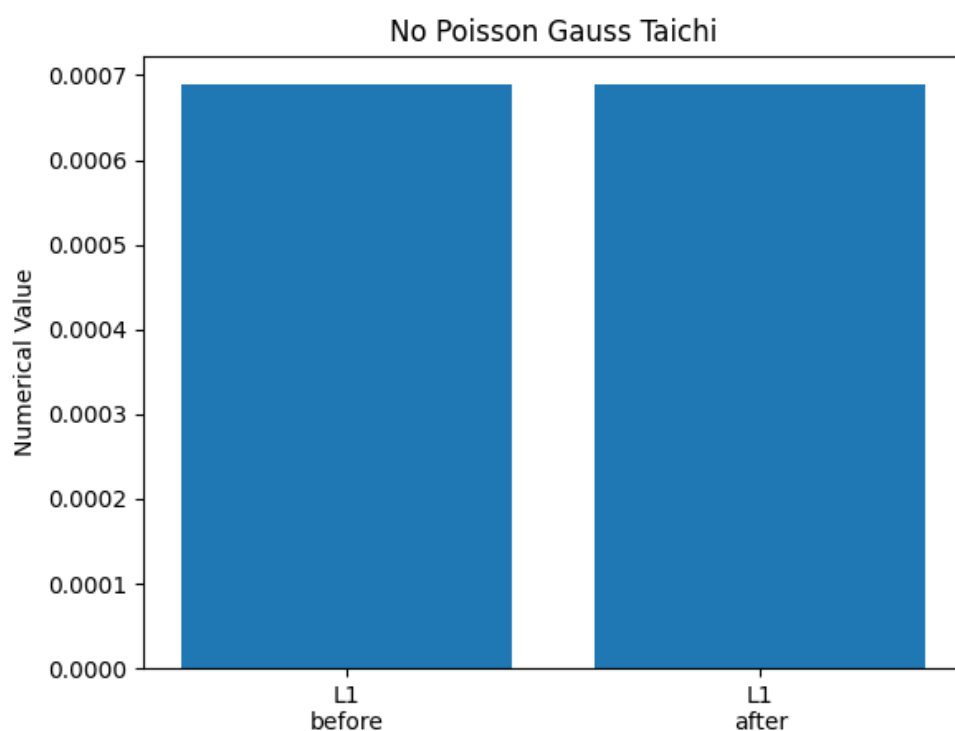


Εικόνα 23. Σύγκριση της τιμής της μετρικής SSIM των αρχικών βίντεο του συνόλου Fashion (αριστερά) με τα ανακατασκευασμένα (δεξιά), παρουσία Γκαουσιανού φίλτρου

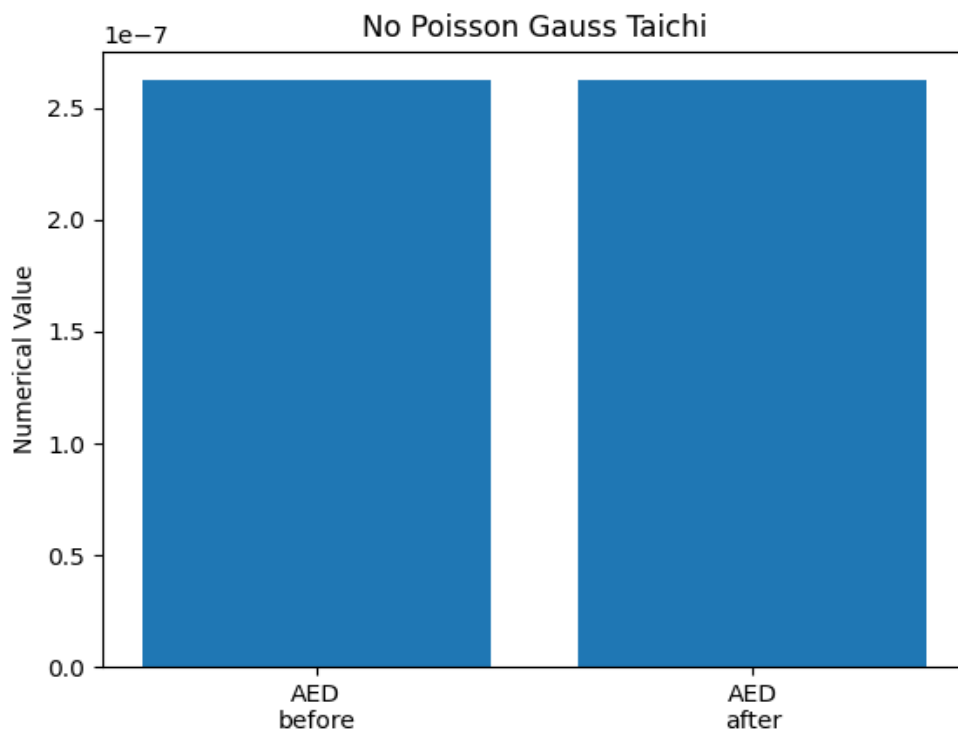
Παρατηρώντας τα γραφήματα, μέσω των οποίων πραγματοποιείται η σύγκριση των μεγεθών, ο δείκτης ομοιότητας SSIM, κατά την εξαγωγή των βίντεο και ύστερα από την επεξεργασία τους με χρήση Γκαουσιανού φίλτρου, παραμένει σε σταθερά επίπεδα με τη τιμή του τείνει στο 80%, γεγονός που υποδεικνύει επαρκή ομοιότητα τόσο των επεξεργασμένων, όσο και των αρχικών βίντεο, με τη πρόβλεψη του μοντέλου. Σε επίπεδο καρέ, η σύγκριση των τιμών της απώλειας L1 διατηρείται εξίσου σε σταθερή τιμή που τείνει στο μηδέν. Κατά συνέπεια, τα καρέ του παραγόμενου βίντεο και του επεξεργασμένου τείνουν ως προς την ομοιότητα του στο εκτιμώμενο καρέ του μοντέλου FOMM. Όσον αφορά τη σύγκριση των τιμών της Μέσης Ευκλείδιας Απόστασης και στις δύο περιπτώσεις, δηλαδή τη μέση τιμή της απώλειας L1 από τα καρέ του παραγόμενων και του ανακατασκευασμένων βίντεο, παρατηρείται βελτίωση της μετρικής αυτής στα ανακατασκευασμένα βίντεο, συγκριτικά με τα αρχικά.

## 6.2.2 Αποτελέσματα μετρικών στο σύνολο Taichi

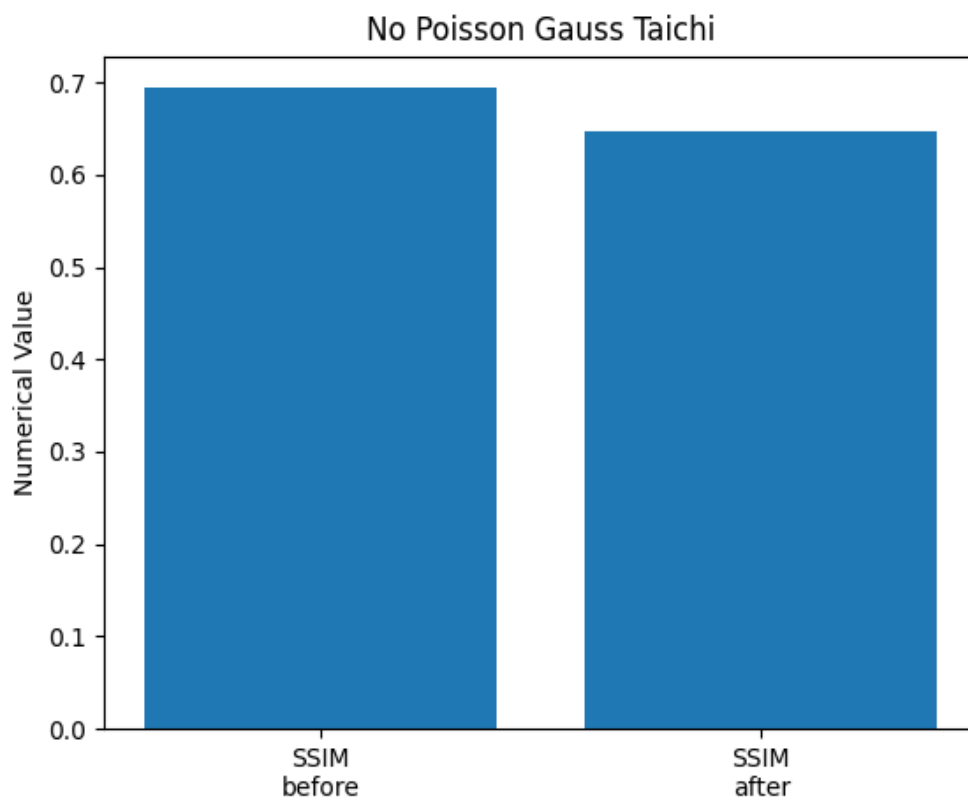
Οι γραφικές παραστάσεις που ακολουθούν υποδεικνύουν τις μεταβολές που παρατηρήθηκαν στα μεγέθη κατά την ανακατασκευή των καρτέ των βίντεο που παράγονται από το μοντέλο FOMM στο σύνολο Taichi με χρήση Γκαουσιανού φίλτρου, στο οποίο η τιμή της τυπικής απόκλισης είναι  $\sigma = 2.5$ , κατά τον ίδιο τρόπο που πραγματοποιήθηκαν και οι μετρήσεις της προηγούμενης παραγράφου στα βίντεο του συνόλου Fashion. Οι γραφικές παραστάσεις που παρουσιάστηκαν για το σύνολο Fashion επεκτείνονται και στο σύνολο Taichi.



Εικόνα 24. Σύγκριση της τιμής της μετρικής  $L1$  των αρχικών βίντεο του συνόλου Taichi (αριστερά) με τα ανακατασκευασμένα (δεξιά), παρουσία Γκαουσιανού φίλτρου



Εικόνα 26. Σύγκριση της τιμής της μετρικής AED των αρχικών βίντεο του συνόλου Taichi (αριστερά) με τα ανακατασκευασμένα (δεξιά), παρουσία Γκαουσιανού φίλτρου



Εικόνα 25. Σύγκριση της τιμής της μετρικής SSIM των αρχικών βίντεο του συνόλου Taichi (αριστερά) με τα ανακατασκευασμένα (δεξιά), παρουσία Γκαουσιανού φίλτρου.

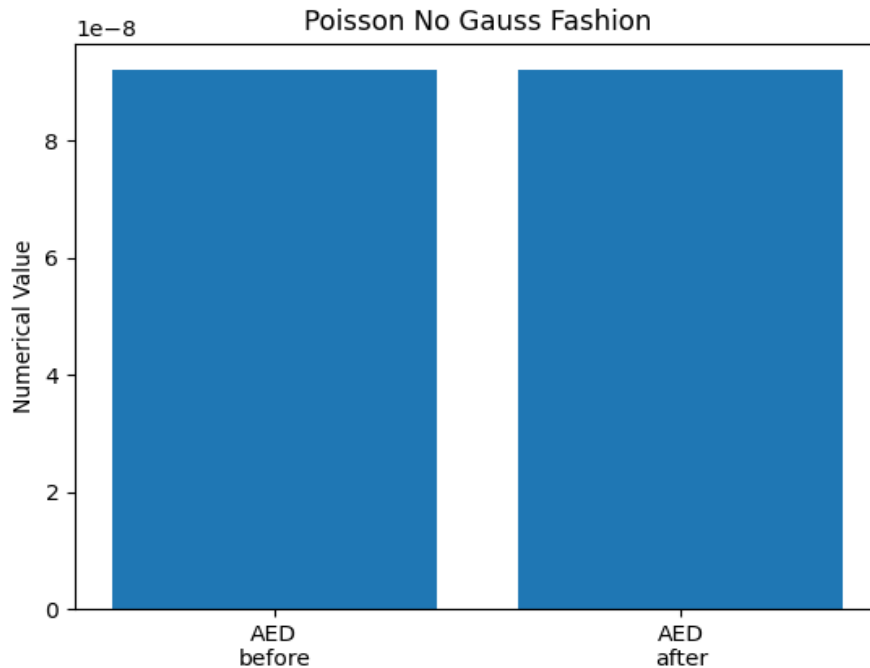
Παρατηρώντας τα γραφήματα, μέσω των οποίων πραγματοποιείται η σύγκριση των μεγεθών, ο δείκτης ομοιότητας SSIM, κατά την εξαγωγή των βίντεο και ύστερα από την επεξεργασία τους με χρήση Γκαουσιανού φίλτρου, παρουσιάζει πτωτική τάση και συγκεκριμένα η τιμή του για τα ανακατασκευασμένα βίντεο τείνει στο 60%, ενώ για τα αρχικά βίντεο η τιμή του προσεγγίζει το 80%. Η μείωση της τιμής του δείκτη ομοιότητας συνεπάγεται τη μη-αποτελεσματική ανακατασκευή των βίντεο του συνόλου Taichi με χρήση του Γκαουσιανού φίλτρου. Ωστόσο, η τιμή της απώλειας L1 διατηρείται σε σταθερή τιμή που τείνει στο μηδέν και η τιμή της Μέσης Ευκλείδιας Απόστασης παρουσιάζεται βελτιωμένη στα ανακατασκευασμένα βίντεο, συγκριτικά με τα αρχικά. Λαμβάνοντας υπόψιν τις ανωτέρω παρατηρήσεις στη μεταβολή των συγκρινόμενων μεγεθών, το συμπέρασμα που εξάγεται είναι ότι για τα ανακατασκευασμένα βίντεο, παρά την ομοιότητα των καρέ τους με τη πρόβλεψη που εξάγει το μοντέλο FOMM, η ποιότητα των ανακατασκευασμένων καρέ με χρήση Γκαουσιανού φίλτρου υποβαθμίζεται σημαντικά σε σχέση με τα αρχικά βίντεο.

## **6.3 Εφαρμογή της μεθόδου seamless cloning στα παραγόμενα βίντεο.**

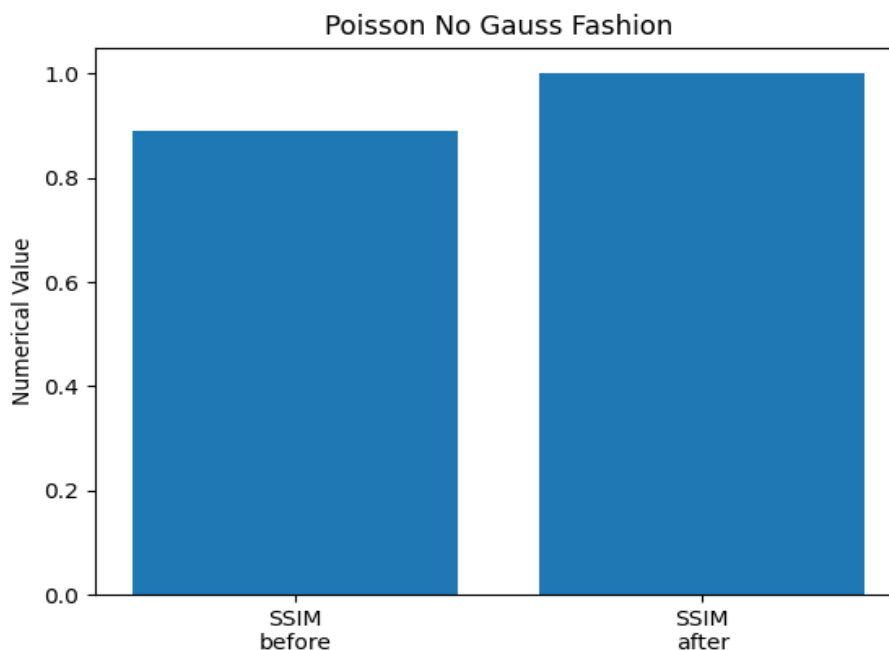
### **6.3.1 Αποτελέσματα μετρικών στο σύνολο Fashion**

Οι γραφικές παραστάσεις που ακολουθούν υποδεικνύουν τις μεταβολές που παρατηρήθηκαν στα μεγέθη κατά την ανακατασκευή των καρέ των βίντεο που παράγονται από το μοντέλο FOMM στο σύνολο Fashion με χρήση της μεθόδου seamless cloning, κατά το τρόπο που αυτή περιγράφηκε στη παράγραφο 5.2 για τα βίντεο που παράγονται από το μοντέλο FOMM.

Στις συγκεκριμένες μετρήσεις, η απώλεια L1, τόσο στα παραγόμενα βίντεο όσο και στα επεξεργασμένα, τείνει στο 0 και στις δύο περιπτώσεις, με αποτέλεσμα να μην είναι εφικτή η γραφική αναπαράσταση και σύγκριση των δύο μεγεθών. Συνεπώς, το ενδιαφέρον επικεντρώνεται στη μελέτη της πορείας των μεγεθών AED και SSIM.



Εικόνα 27. Σύγκριση της τιμής της μετρικής AED των αρχικών βίντεο του συνόλου Fashion (αριστερά) με τα ανακατασκευασμένα (δεξιά), ύστερα από την εφαρμογή της μεθόδου seamless cloning

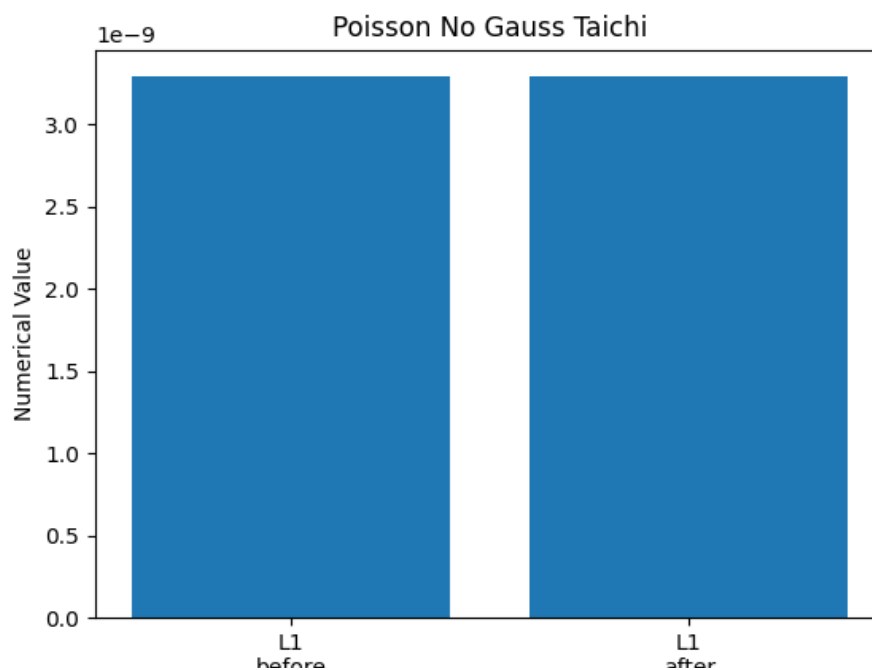


Εικόνα 28. Σύγκριση της τιμής της μετρικής SSIM των αρχικών βίντεο του συνόλου Fashion (αριστερά) με τα ανακατασκευασμένα (δεξιά), ύστερα από την εφαρμογή της μεθόδου seamless cloning.

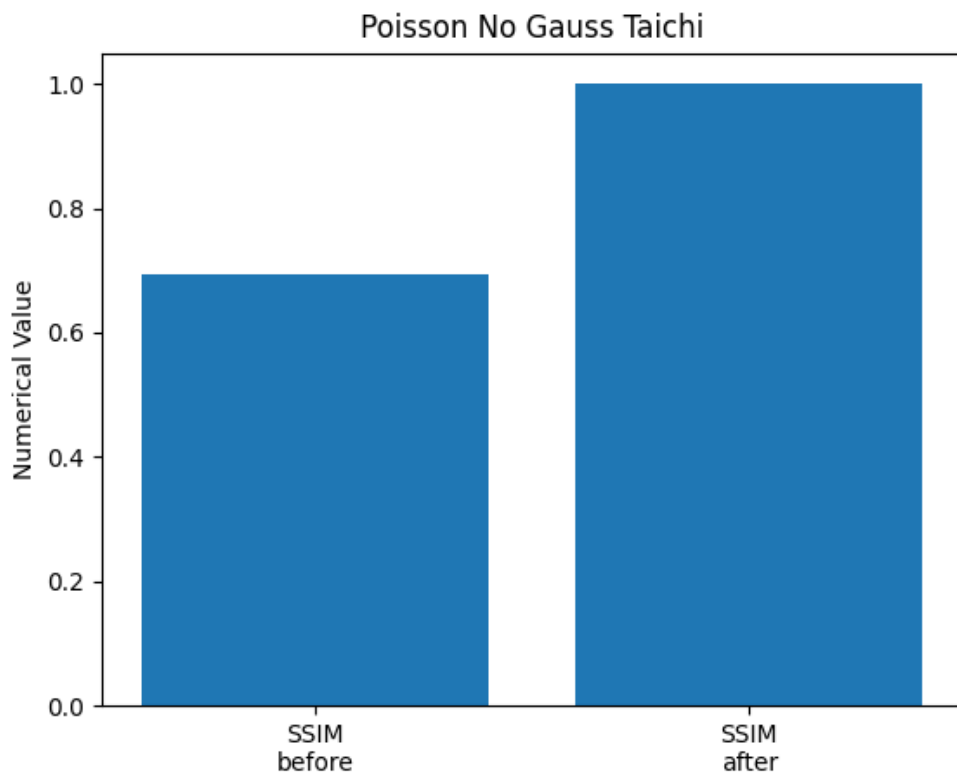
Εφαρμόζοντας τη μέθοδο seamless cloning στα παραγόμενα βίντεο του συνόλου Fashion, έχοντας ως βάση σύγκρισης τις γραφικές παραστάσεις που παρατέθηκαν, παρατηρείται αύξηση του Δείκτη Δομικής Ομοιότητας στα επεξεργασμένα βίντεο, με την τιμή του να προσεγγίζει το 100%, γεγονός το οποίο συνεπάγεται ότι η πρόβλεψη του καρέ από το μοντέλο και το ανακατασκευασμένο καρέ είναι σχεδόν πανομοιότυπα. Παράλληλα, η Μέση Ευκλείδεια Απόσταση παρουσιάζει μείωση για τα επεξεργασμένα βίντεο, εν συγκρίσει με τη τιμή της για τα αρχικά βίντεο.

### 6.3.2 Αποτελέσματα μετρικών στο σύνολο Taichi

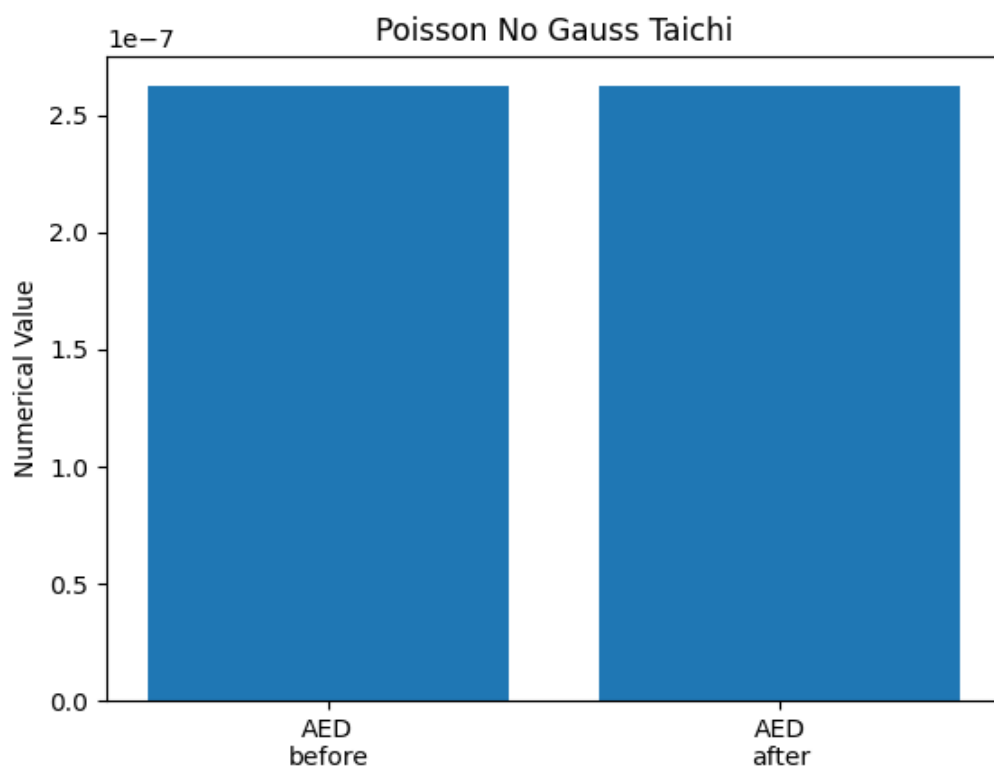
Οι γραφικές παραστάσεις που ακολουθούν υποδεικνύουν τις μεταβολές που παρατηρήθηκαν στα μεγέθη ενδιαφέροντος κατά την ανακατασκευή των καρέ των βίντεο που παράγονται από το μοντέλο FOMM στο σύνολο Taichi με χρήση της μεθόδου seamless cloning.



Εικόνα 29. Σύγκριση της τιμής της μετρικής L1 των αρχικών βίντεο του συνόλου Taichi (αριστερά) με τα ανακατασκευασμένα (δεξιά), ύστερα από την εφαρμογή της μεθόδου seamless cloning



Εικόνα 30. Σύγκριση της τιμής της μετρικής SSIM των αρχικών βίντεο του συνόλου Taichi (αριστερά) με τα ανακατασκευασμένα (δεξιά), ύστερα από την εφαρμογή της μεθόδου *seamless cloning*



Εικόνα 31. Σύγκριση της τιμής της μετρικής AED των αρχικών βίντεο του συνόλου Taichi (αριστερά) με τα ανακατασκευασμένα (δεξιά), ύστερα από την εφαρμογή της μεθόδου *seamless cloning*

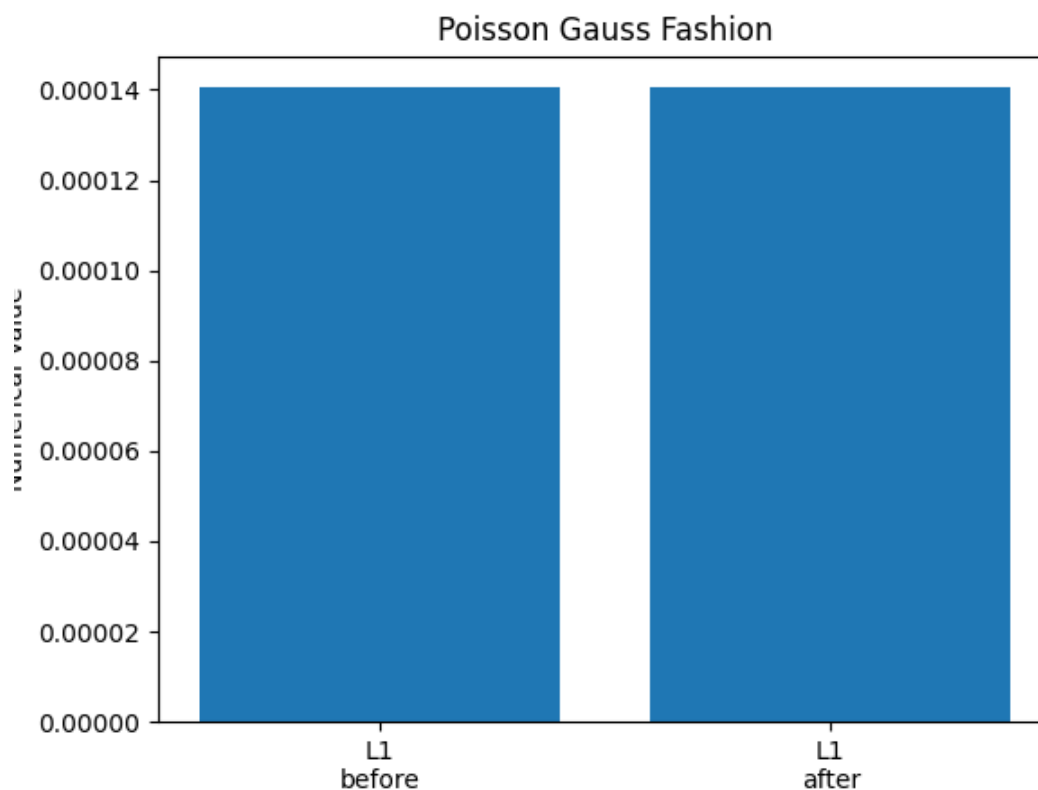
Η εφαρμογή της μεθόδου seamless cloning στα παραγόμενα βίντεο του συνόλου Taichi, έχοντας ως βάση σύγκρισης τις γραφικές παραστάσεις που παρατέθηκαν, επιφέρει σημαντική αύξηση στη τιμή του Δείκτη Δομικής Ομοιότητας στα επεξεργασμένα βίντεο, με τη τιμή του να προσεγγίζει το 100% όπως και στη περίπτωση της τιμής του συνόλου Fashion. Στο σύνολο Fashion η αρχική τιμή του Δείκτη Δομικής Ομοιότητας τείνει στο 80% με τη πρόβλεψη του μοντέλου FOMM για το εκάστοτε καρέ, ενώ στο σύνολο Taichi η παρατηρούμενη τιμή του δείκτη προσγγίζει το 70% αρχικά. Με βάση το πόρισμα αυτό, η βελτίωση της ποιότητας των παραγόμενων βίντεο είναι περισσότερο παρατήριση στα βίντεο του συνόλου Taichi, ύστερα από την εφαρμογή της μεθόδου seamless cloning. Όσον αφορά τις μετρικές L1 και AED στη προκειμένη περίπτωση, οι αντίστοιχες τιμές των μεγεθών παρουσιάζουν σταθερότητα συγκριτικά με τις αρχικές τιμές τους, σε χαμηλές συναρτησιακές τιμές. Λαμβάνοντας υπόψιν τις μεταβολές των μεγεθών, προκύπτει το συμπέρασμα ότι τα καρέ του αρχικού και του ανακατασκευασμένου βίντεο προσεγγίζουν τη πρόβλεψη του μοντέλου, με τη χρήση της μεθόδου seamless cloning να συνιστά παράγοντα σημαντικής βελτίωσης της ποιότητας του επεξεργασμένου βίντεο.



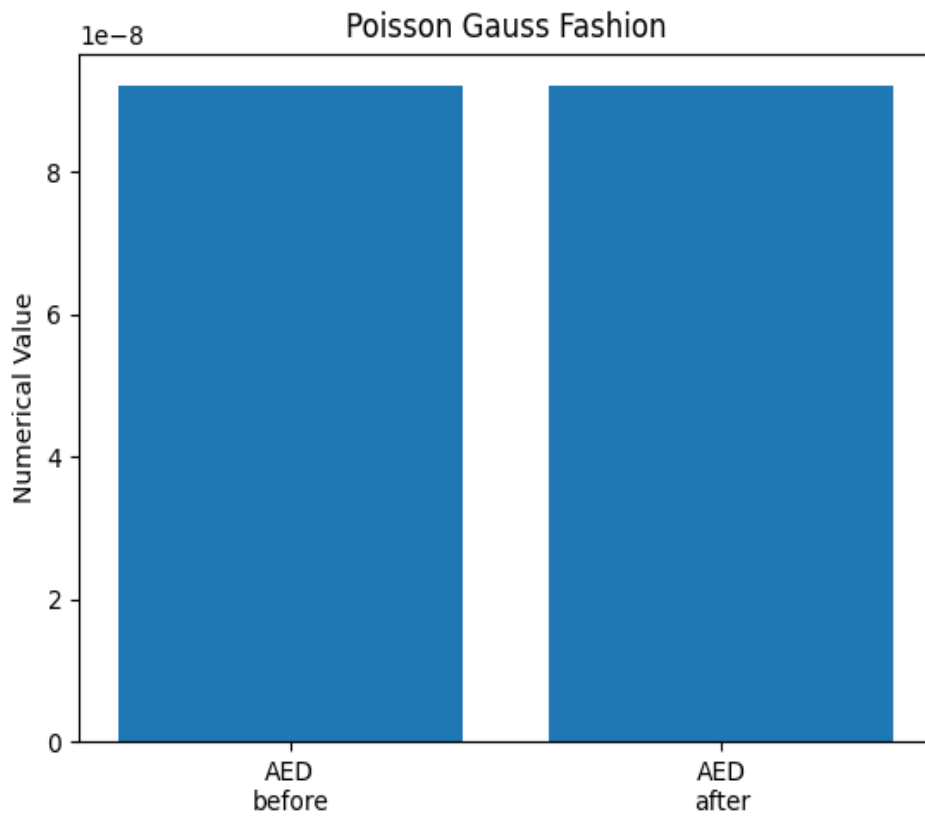
## 6.4 Εφαρμογή της μεθόδου seamless cloning και του Γκαουσιανού φίλτρου στα παραγόμενα βίντεο.

### 6.4.1 Αποτελέσματα μετρικών στο σύνολο Fashion

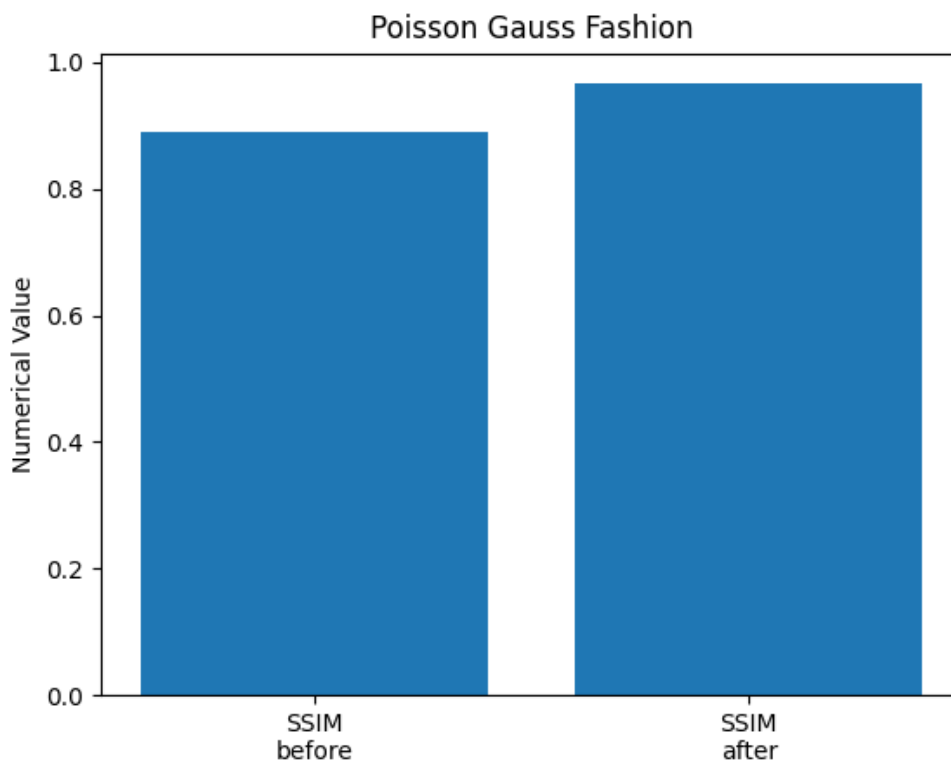
Η τελευταία εκ των εξεταστέων περιπτώσεων σχετικά με τις μετρήσεις των μεγεθών L1, AED και SSIM αφορά τον συνδυασμό της μεθόδου seamless cloning με τη χρήση Γκαουσιανού φίλτρου, με τη τιμή της τυπικής απόκλισης του φίλτρου να διατηρείται στη τιμή  $\sigma = 2.5$ , όπως και στις προηγούμενες περιπτώσεις. Στις γραφικές παραστάσεις που ακολουθούν παρουσιάζεται η πορεία των μεγεθών και, εν συνεχεία, τα πορίσματα της σύγκρισης των μεγεθών για τα βίντεο του συνόλου Fashion και ύστερα στα βίντεο του συνόλου Taichi.



Εικόνα 32. Σύγκριση της τιμής της μετρικής L1 των αρχικών βίντεο του συνόλου Fashion (αριστερά) με τα ανακατασκευασμένα (δεξιά), ύστερα από την εφαρμογή της μεθόδου seamless cloning και του Γκαουσιανού φίλτρου



Εικόνα 33 . Σύγκριση της τιμής της μετρικής AED των αρχικών βίντεο του συνόλου Fashion (αριστερά) με τα ανακατασκευασμένα (δεξιά), ύστερα από την εφαρμογή της μεθόδου seamless cloning και του Γκαουσιανού φίλτρου



Εικόνα 34. Σύγκριση της τιμής της μετρικής SSIM των αρχικών βίντεο του συνόλου Fashion (αριστερά) με τα ανακατασκευασμένα (δεξιά), ύστερα από την εφαρμογή της μεθόδου seamless cloning και του Γκαουσιανού φίλτρου

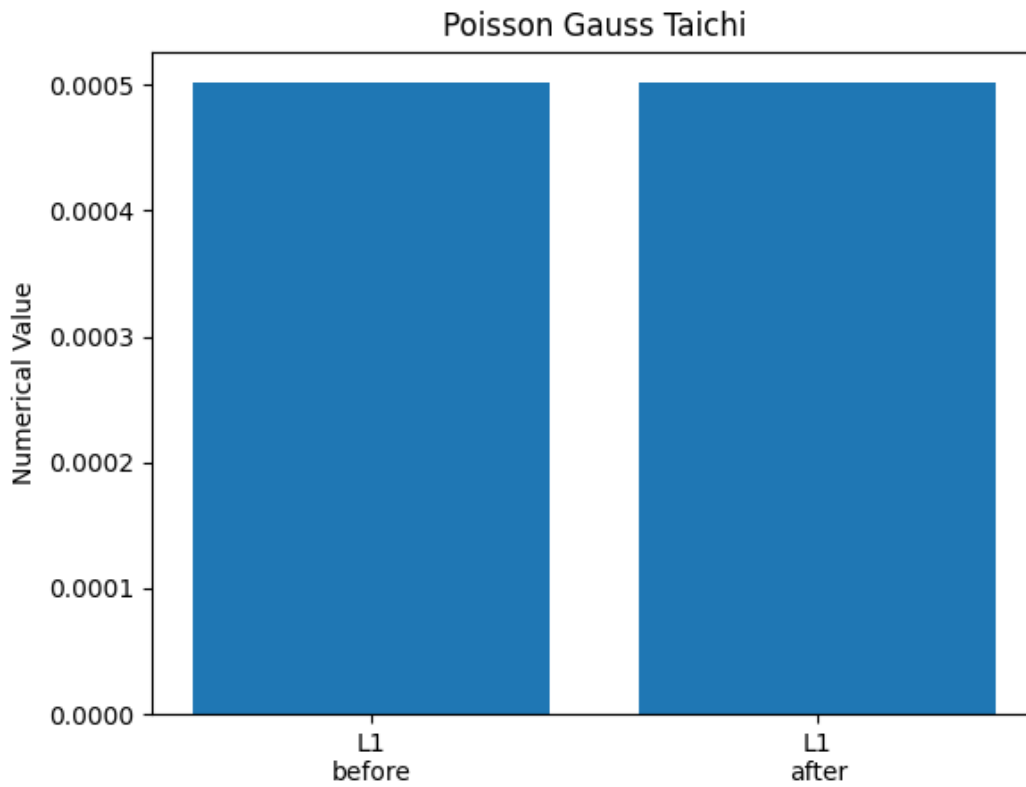
Παρατηρώντας τις γραφικές παραστάσεις της πορείας των μεγεθών L1, AED και SSIM στη περίπτωση που χρησιμοποιούνται παράλληλα η μέθοδος seamless cloning και το Γκαουσιανό φίλτρο, η τιμή του Δείκτη Δομικής Ομοιότητας τείνει στο 100%, όπως γίνεται αντιληπτό σε και στις προηγούμενες περιπτώσεις εφαρμογής της μεθόδου seamless cloning, με την αρχική τιμή του δείκτη για τα παραγόμενα βίντεο να προσεγγίζει το 80%. Η τιμή του δείκτη L1 παρουσιάζει σταθερότητα σε τιμή κοντά στο μηδέν, όμοια με τις περιπτώσεις που χρησιμοποιούνται ξεχωριστά η μέθοδος seamless cloning και το Γκαουσιανό φίλτρο για την ανακατασκευή των καρέ του παραγόμενου βίντεο. Ωστόσο, συγκριτικά με τις προηγούμενες μετρήσεις που αφορούν τη μεταβολή του μεγέθους AED για το σύνολο Fashion, στη περίπτωση αυτή η Μέση Ευκλείδεια Απόσταση παρουσιάζει σταθερότητα, ενώ στις υπόλοιπες περιπτώσεις παρατηρείται μείωση του μεγέθους αυτού. Συνοψίζοντας τα αποτελέσματα της μεταβολής των μετρικών στη παρούσα περίπτωση, η ανακατασκευή των καρέ των αρχικών βίντεο παρουσιάζει βελτίωση στη ποιότητα τους.

Κρίνοντας από την αύξηση του δείκτη ομοιότητας, ενώ παράλληλα το περιεχόμενο των καρέ είναι κοντινό, συγκεκριμένα σχεδόν όμοιο, με αυτό των καρέ που προβλέπονται από το μοντέλο FOMM με βάση τη διατήρηση των δεικτών L1 και AED για τα ανακατασκευασμένα και τα αρχικά βίντεο.

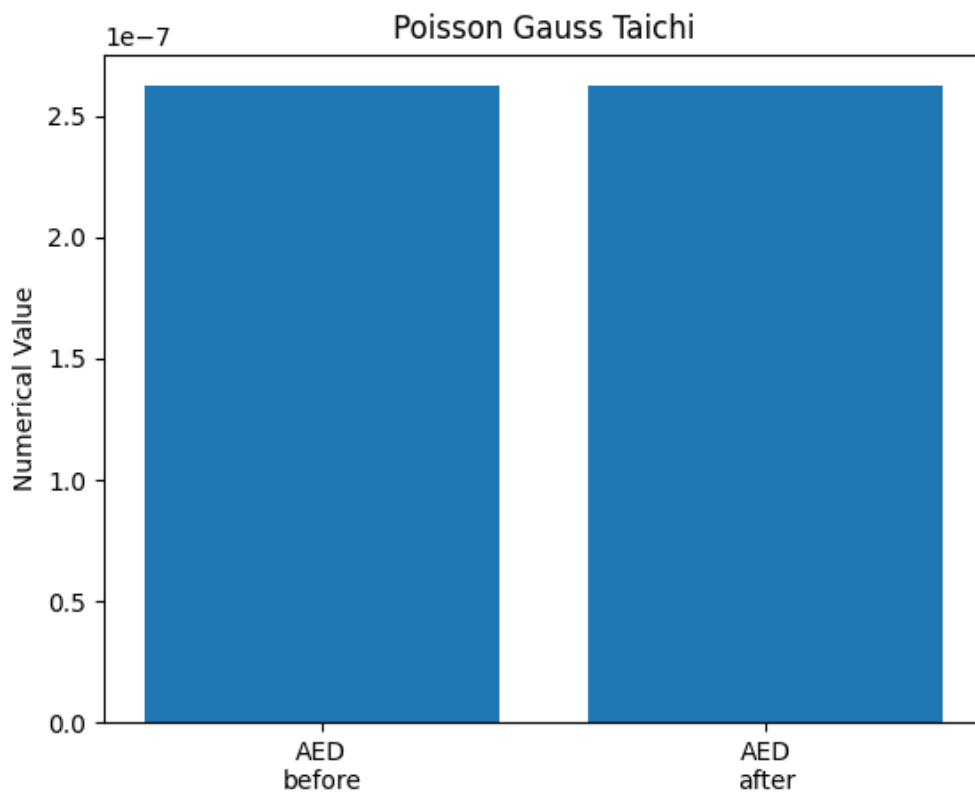
Η παρουσίαση και ανάλυση των μετρικών που διεξήχθησαν ολοκληρώνεται στη παράγραφο **6.4.2**, με τις εν λόγω συγκρίσεις των μεγεθών να αφορούν την επίδραση της εφαρμογής των μετασχηματισμών στο σύνολο Taichi.

#### **6.4.2 Αποτελέσματα μετρικών στο σύνολο Taichi**

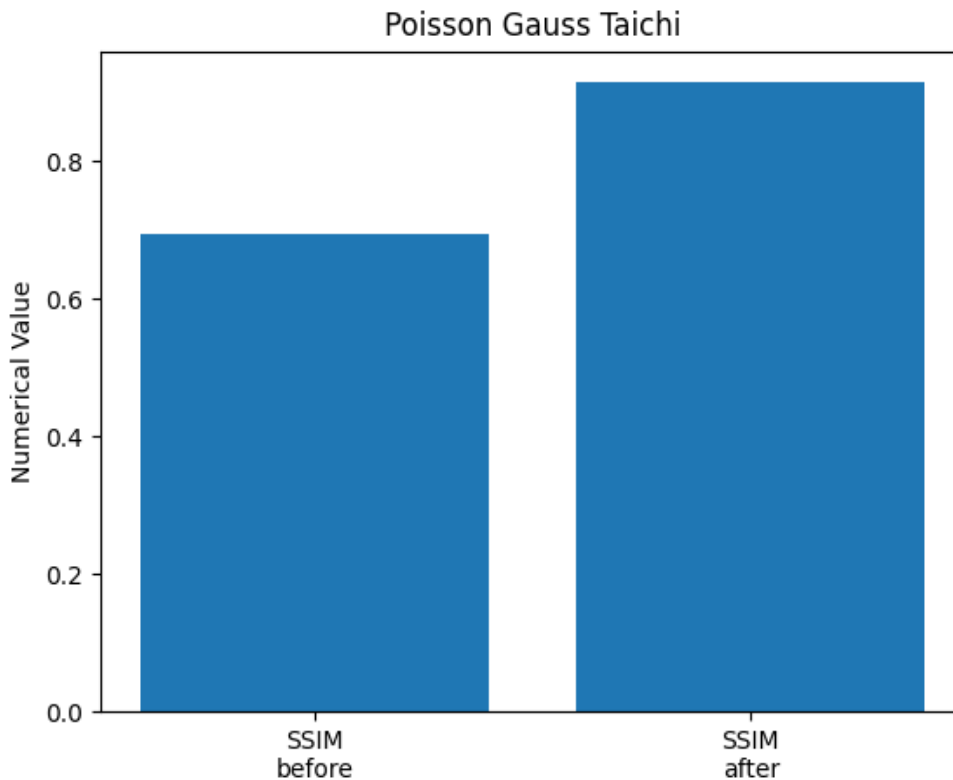
Οι μεταβολές στις μετρήσεις μεταξύ των αρχικά παραγόμενων και των ανακατασκευασμένων βίντεο του συνόλου Taichi, στην περίπτωση εφαρμογής της μεθόδου seamless cloning και του Γκαουσιανού φίλτρου, αποτελούν τις τελευταίες συγκρίσεις, σχετικές με την επίδειξη της επίδρασης των μεθόδων στα επεξεργασμένα βίντεο.



Εικόνα 35. Σύγκριση της τιμής της μετρικής L1 των αρχικών βίντεο του συνόλου Taichi (αριστερά) με τα ανακατασκευασμένα (δεξιά), ύστερα από την εφαρμογή της μεθόδου seamless cloning και του Γκαουσιανού φίλτρου



Εικόνα 36. Σύγκριση της τιμής της μετρικής AED των αρχικών βίντεο του συνόλου Taichi (αριστερά) με τα ανακατασκευασμένα (δεξιά), ύστερα από την εφαρμογή της μεθόδου seamless cloning και του Γκαουσιανού φίλτρου



Εικόνα 37. Σύγκριση της τιμής της μετρικής SSIM των αρχικών βίντεο του συνόλου Taichi (αριστερά) με τα ανακατασκευασμένα (δεξιά), ύστερα από την εφαρμογή της μεθόδου seamless cloning και του Γκαουσιανού φίλτρου

Μελετώντας τις γραφικές παραστάσεις που παρατίθενται, σχετικά με τη μεταβολής των μεγεθών L1, AED και SSIM, η τιμή του Δείκτη Δομικής Ομοιότητας τείνει στο 100%, όπως και στις προηγούμενες περιπτώσεις χρήσης της μεθόδου seamless cloning, με την αρχική τιμή του δείκτη να τείνει στο 70%, για τα παραγόμενα βίντεο και χωρίς κάποια προγενέστερη επεξεργασία. Παράλληλα, η τιμή του μεγέθους L1 παραμένει σταθερή αλλά σε χαμηλή συναρτησιακή τιμή, το οποίο υποδεικνύει την αποτελεσματικότητα της ανακατασκευής των καρέ του εκάστοτε βίντεο του συνόλου Taichi. Το πόρισμα αυτό, όπως και στις προηγούμενες συγκρίσεις των μεγεθών για τις αντίστοιχες χρησιμοποιούμενες μεθόδους, μπορεί να υποστηριχθεί και από την εξίσου σταθερή αλλά σχεδόν μηδενική τιμή της Μέσης Ευκλείδιας Απόστασης (AED). Οι παραπάνω τιμές των μεθόδων, ύστερα από την παράλληλη χρήση της μεθόδου seamless cloning και του Γκαουσιανού φίλτρου, υποδεικνύουν ότι η ποιότητα των ανακατασκευασμένων καρέ στα βίντεο του συνόλου Taichi παρουσιάζει βελτίωση, με γνώμονα τις προηγούμενες μετρήσεις των μεγεθών, καθώς και ομοιότητα με την πρόβλεψη των καρέ που εξάγει το μοντέλο FOMM, δεδομένων των σχεδόν μηδενικών τιμών που φέρουν τα μεγέθη L1 και AED.

# Βιβλιογραφία

- [1] Pérez, Patrick & Gangnet, Michel & Blake, Poisson image editing, ACM Trans. Graph.. 22. 313-318. 10.1145/1201775.882269., 2003.
- [2] Kingma, Diederik & Welling, Max, Auto-Encoding Variational Bayes. ICLR., ICLR. , 2013.
- [3] Bulat, Adrian & Tzimiropoulos, Georgios, Bulat, AdrianHow Far are We from Solving the 2D & 3D Face Alignment Problem? (and a Dataset of 230,000 3D Facial Landmarks), 10.1109/ICCV.2017.116., 2017.
- [4] Villegas, Ruben & Yang, Jimei & Zou, Yuliang & Sohn, Sungryull & Lin, Xunyu & Lee, Honglak, Learning to Generate Long-term Future via Hierarchical Prediction., 2017.
- [5] Elrawy, Amr & Kishka, prof & Saleem, Mohammed & Abul-Dahab, Mohamed, On Hadamard and Kronecker Products Over Matrix of Matrices, General Letters in Mathematics. 4. 13-22. 10.31559/GLM2016.4.1.3. , 2018.
- [6] Kim, H. & Kim, J. & Jung, H., Convolutional neural network based image processing system, Journal of Information and Communication Convergence Engineering. 16. 160-165. 10.6109/jicce.2018.16.3.160. , 2018.
- [7] Siarohin, Aliaksandr & Lathuilière, Stéphane & Tulyakov, Segey & Ricci, Elisa & Sebe, Nicu, Animating Arbitrary Objects via Deep Motion Transfer, 2018.

- [8] Wei, Chun-Chun & Yeh, Chung-Hsing & Wang, Ian & Walsh, Bernie & Lin, Yang-Cheng, Deep Neural Networks for New Product Form Design, 653-657. 10.5220/0007933506530657. , 2019.
- Grigorev, Artur & Sevastopolsky, Artem & Vakhitov, Alexander & Lempitsky, Victor,
- [9] Coordinate-Based Texture Inpainting for Pose-Guided Human Image Generation, 12127-12136. 10.1109/CVPR.2019.01241. , 2019.
- [10] Siarohin, Aliaksandr & Lathuilière, Stéphane & Tulyakov, Sergey & Ricci, Elisa & Sebe, Nicu, First Order Motion Model for Image Animation., 2020.
- [11] Lu, Yulong & Lu, Jianfeng, A Universal Approximation Theorem of Deep Neural Networks for Expressing Distributions, 2020.
- [12] Shalev, Yoav & Wolf, Lior, Image Animation with Perturbed Masks., 2020.
- [13] Syberfeldt, Anna & Vuoluterä, Fredrik, Image Processing based on Deep Neural Networks for Detecting Quality Problems in Paper Bag Production, Procedia CIRP. 93. 1224-1229. 10.1016/j.procir.2020.04.158. , 2020.
- [14] Papacharalampopoulos, Alexios & Petridis, Demitris & Stavropoulos, Panagiotis, Experimental Investigation of rubber extrusion process through vibrational testing, Procedia CIRP. 93. 1236-1240. 10.1016/j.procir.2020.04.160. , 2020.
- [15] Aggarwal, Alankrita & Mittal, Mamta & Battineni, Gopi, Generative adversarial network: An overview of theory and applications, International Journal of Information Management. 100004. 10.1016/j.jjime.2020.100004. , 2021.
- [16] Wu, Hao & Zhou, Zhi , Using Convolution Neural Network for Defective Image Classification of Industrial Components, Mobile Information Systems. 2021. 1-8. 10.1155/2021/9092589. , 2021.
- [17] van Dyck, Leonard & Kwitt, Roland & Denzler, Sebastian & Gruber, Walter , Comparing Object Recognition in Humans and Deep Convolutional Neural Networks—An Eye Tracking Study, Frontiers in Neuroscience. 15. 750639. 10.3389/fnins.2021.750639. , 2021.

- [18] Yin, Xiao-Xia & Sun, Le & Fu, Yuhang & Lu, Ruiliang & Zhang, Yanchun , U-Net-Based Medical Image Segmentation, *Journal of Healthcare Engineering*. 2022. 1-16. 10.1155/2022/4189781. , 2022.
- [19] Nagy, Brigitta & Galata, Dorián & Farkas, Attila & Nagy, Zsombor, Application of Artificial Neural Networks in the Process Analytical Technology of Pharmaceutical Manufacturing—a Review, *The AAPS Journal*. 24. 10.1208/s12248-022-00706-0. , 2022.
- [20] Weiss, Romano & Karimijafarbigloo, Sanaz & Roggenbuck, Dirk & Rödiger, Stefan , Applications of Neural Networks in Biomedical Data Analysis, *Biomedicines*. 10. 10.3390/biomedicines10071469. , 2022.
- [21] Drobyshev, Nikita & Chelishev, Jenya & Khakhulin, Taras & Ivakhnenko, Aleksei & Lempitsky, Victor & Zakharov, Egor, MegaPortraits: One-shot Megapixel Neural Head Avatars, 10.48550/arXiv.2207.07621. , 2022.
- [22] Weitao, Jiang & Hu, Haifeng , Hadamard Product Perceptron Attention for Image Captioning, *Neural Processing Letters*. 1-18. 10.1007/s11063-022-10980-w. , 2022.
- [23] Droniou, Jerome & Eymard, Robert & Thierry, Gallouet & Guichard, Cindy & Herbin, Raphaële , Dirichlet Boundary Conditions, 10.1007/978-3-319-79042-8\_2. , 2018.
- [24] Yang, Chun & Vadlamani, Ananth & Soloviev, Andrey & Veth, Michael & Taylor, Clark, Feature matching error analysis and modeling for consistent estimation in vision-aided navigation, *Navigation*. 65. 10.1002/navi.265. , 2018.
- [25] Tang, Geng & Xue, Xiaoming & Chen, Xinbiao & Wang, Ruoheng & Zhang, Chu., The short-term interval prediction of wind power using the deep learning model with gradient descend optimization, *Renewable Energy*. 155. 10.1016/j.renene.2020.03.098, 2020.
- [26] Li, Chaoshun & Tang, Geng & Xue, Xiaoming & Chen, Xinbiao & Wang, Ruoheng & Zhang, Chu, Deep interval prediction model with gradient descend optimization method for short-term wind power prediction, 2019.
- [27] Hanczar, Blaise & Zehraoui, Farida & Issa, Tina & Arles, Mathieu, Biological interpretation of deep neural network for phenotype prediction based on gene expression, *BMC Bioinformatics*. 21. 10.1186/s12859-020-03836-4., 2020.