ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ
ΣΧΟΛΗ ΑΓΡΟΝΟΜΩΝ ΤΟΠΟΓΡΑΦΩΝ ΜΗΧΑΝΙΚΩΝ-ΜΗΧΑΝΙΚΩΝ
ΓΕΩΠΛΗΡΟΦΟΡΙΚΗΣ
ΕΡΓΑΣΤΗΡΙΟ ΤΗΛΕΠΙΣΚΟΠΗΣΗΣ

Εφαρμογή Τεχνητών Νευρωνικών Δικτύων για την Αναγνώριση και Ταξινόμηση Δράσεων σε Δεδομένα Βίντεο

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

ΤΗΛΕΜΑΧΟΣ ΜΟΥΜΟΥΡΗΣ

Αθήνα, Ιούλιος 2022

NATIONAL TECHNICAL UNIVERSITY OF ATHENS
SCHOOL OF RURAL,SURVEYING AND GEOINFORMATICS ENGINEERING
REMOTE SENSING LABORATORY

Implementation of Artificial Neural Networks for Action Recognition and Classification in Video Data

THESIS PROJECT

Tilemachos Moumouris

Athens, July 2022

Εφαρμογή Τεχνητών Νευρωνικών Δικτύων για την Αναγνώριση και Ταξινόμηση Δράσεων σε Δεδομένα Βίντεο

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

Τηλέμαχος Μουμούρης

Επιβλέπων: Κωνσταντίνος Καράντζαλος
Αναπληρωτής Καθηγητής ΕΜΠ

Εγκρίθηκε από την τριμελή εξεταστική επιτροπή την 6η Ιουλίου 2022.

(Υπογραφή)	*(Υπογραφή)*	*(Υπογραφή)*
.....	.....	.....
Κωνσταντίνος Καράντζαλος	Αναστάσιος Δουλάμης	Μαρία Παπαδοπούλου
Αναπληρωτής Καθηγητής ΕΜΠ	Αναπληρωτής Καθηγητής ΕΜΠ	Καθηγήτρια ΕΜΠ

Αθήνα, Ιούλιος 2022

Υπεύθυνη Δήλωση

Βεβαιώνω ότι είμαι συγγραφέας αυτής της πτυχιακής εργασίας, και ότι κάθε βοήθεια την οποία είχα για την προετοιμασία της είναι πλήρως αναγνωρισμένη και αναφέρεται στην πτυχιακή εργασία. Επίσης έχω αναφέρει τις όποιες πηγές από τις οποίες έκανα χρήση δεδομένων, ιδεών ή λέξεων, είτε αυτές αναφέρονται ακριβώς είτε παραφρασμένες. Επίσης, βεβαιώνω ότι αυτή η πτυχιακή εργασία προετοιμάστηκε από εμένα προσωπικά ειδικά για τις απαιτήσεις του προγράμματος σπουδών του Τμήματος Αγρονόμων Τοπογράφων Μηχανικών-Μηχανικών Γεωπληροφορικής του Εθνικού Μετσόβιου Πολυτεχνείου.

Procedures on animal experiments were reviewed and approved by the relevant local ethics committee and studies were carried out in accordance with the European Union Directive 2010/63/EU on animal care and experimentation.

(Υπογραφή)

.....
Τηλέμαχος Μουμούρης

Περίληψη

Η Μηχανική Μάθηση, είναι ένας κλάδος ο οποίος έχει γνωρίσει σημαντική εξέλιξη τα τελευταία χρόνια. Παρατηρείται όλο και περισσότερο η χρήση Τεχνητών Νευρωνικών Δικτύων για την επίλυση καθημερινών προβλημάτων σε πληθώρα εφαρμογών, που υπό άλλες συνθήκες, θα ήταν δύσκολο και χρονοβόρο να επιλυθούν. Στην παρούσα ΔΕ, σκοπός ήταν η διερεύνηση των δυνατοτήτων των Τεχνητών Νευρωνικών Δικτύων να ταξινομήσουν τις δράσεις των πειραματόζωων κατά την διάρκεια του πειράματος της εξαναγκασμένης κολύμβησης. Η δοκιμασία αυτή έχει στόχο να αναγνωρίσει ποσοτικά και ποιοτικά τις κινήσεις που εκτελούνται από τους επιμύες για την μελέτη αντικαταθλιπτικών ουσιών. Τα πειράματα εκτελέστηκαν στο εργαστήριο Φαρμακολογίας της Ιατρικής Σχολής Αθηνών (ΕΚΠΑ) από τη συνεργαζόμενη ομάδα Νευροψυχοφαρμακολογίας. Για την εκτέλεση του πειράματος, τα πειραματόζωα, τοποθετούνται μέσα σε κύλινδρο και στην συνέχεια, καταγράφεται ο χρόνος της εκτέλεσης των κινήσεων Αναρρίχησης, Κολύμβησης, Ακινησίας καθώς και οι φορές που εκτελέστηκε Κατάδυση και Τίναγμα Κεφαλής. Αρχικά, δημιουργήθηκε Dataset που περιείχε στο σύνολό του βίντεο διάρκειας 16 ωρών και ταξινομημένα ως προς την κίνηση από ειδικούς του παραπάνω εργαστηρίου και από τον συγγραφέα της παρούσας ΔΕ σε συνεργασία με την ΥΔ Παυλίνα Παυλίδη. Στην συνέχεια, έγινε προ-επεξεργασία των δεδομένων και συγχρονισμός των ταξινομήσεων. Για την εφαρμογή Τεχνητών Νευρωνικών Δικτύων για την αναγνώριση των δράσεων του εκάστοτε επιμύ, αρχικά σχεδιάστηκαν αρχιτεκτονικές δικτύων που έκαναν χρήση απλών επιπέδων τρισδιάστατων συνελίξεων και εκτελέστηκε ενδελεχής πειραματισμός και μελέτη της απόδοσής τους, καθώς και βελτιστοποίηση. Στην συνέχεια έγινε πειραματισμός με την χρήση της αρχιτεκτονικής ResNet3D και στην συνέχεια εκτελέστηκε σειρά πειραμάτων με την αρχιτεκτονική R(2+1)D όπου και έγινε βελτιστοποίηση των υπερπαραμέτρων του δικτύου. Το αποτέλεσμα των παραπάνω ήταν η υλοποίηση ολοκληρωμένου αλγορίθμου για την ταξινόμηση δράσεων σε δεδομένα βίντεο πειραμάτων εξαναγκασμένης κολύμβησης.

Λέξεις Κλειδιά

Μηχανική Μάθηση, Τεχνητά Νευρωνικά Δίκτυα, Νευροψυχοφαρμακολογία, πείραμα εξαναγκασμένης κολύμβησης, συνελικτικά επίπεδα, ταξινόμηση δράσεων σε βίντεο

Abstract

Machine learning, has seen many advances over the recent years. Nowadays, the use of Artificial Neural Networks is increasing and many everyday problems that are time consuming, laborious and too hard to be solved by humans are tackled with the use of these Networks. The purpose of this thesis is the implementation of artificial neural networks for action recognition in video, more specifically in the Forced Swim Test (FST) in rats. This experiment is used in order to study the effects of antidepressant drugs by scoring the actions during these experiments, a task which is very time consuming. The experiments were carried out by the Neuropsychopharmacology Research Group in the Department of Pharmacology, Medical School (NKUA). The Dataset contained 16 hours of annotated videos. Before implementing artificial neural networks to classify the actions of animals, videos and ground truth classifications were synchronized. Firstly, simple architectures were implemented which used different numbers of 3D convolutional layers. After series of experiments, these networks were fine-tuned and compared. Next, the well-known architecture of ResNet3D was used to further study the use of residual neural networks for action recognition. Finally, the architecture R(2+1)D was studied and fine-tuned. The final result of this thesis is a complete algorithm for the classification of actions in videos of FST.

Keywords

Machine learning, Artificial Neural Networks, Neuropsychopharmacology, Forced Swim Test (FST), skip connections, convolutional layers, action classification in video data

Ευχαριστίες

Αρχικά θα ήθελα να ευχαριστήσω τον καθηγητή μου κ. Κωνσταντίνο Καράντζαλο για την επίβλεψη της παρούσας ΔΕ καθώς και για την ευκαιρία που μου έδωσε να εκπονήσω το παρόν θέμα.Επίσης, τους Χ. Δάλλα και Ν. Κόκρα για την παροχή των δεδομένων βίντεο καθώς και την ΥΔ Παυλίνα Παυλίδη για την καθοδήγησή της. Ακόμα, τους ΥΔ του Εργαστηρίου Τηλεπισκόπησης Ζ. Κανδυλάκη και Ι. Κακογεωργίου για την βοήθειά τους και τους πόρους που διέθεσαν για την περάτωση της ΔΕ. Επιπλέον, θα ήθελα να ευχαριστήσω τον Α. Βυθούλκα για την πολύτιμη καθοδήγηση και τον χρόνο, τον οποίο μου παρείχε με προθυμία. Κλείνοντας, θα ήθελα να ευχαριστήσω την οικογένεια και τους φίλους μου για την στήριξη που μου προσέφεραν όλο αυτό τον καιρό, χωρίς την οποία το δύσκολο αυτό έργο δεν θα μπορούσε να περατωθεί.

Περιεχόμενα

Κεφάλαιο 1

Εισαγωγή

1.1 Πρώτα βήματα προς την Έξυπνη Μηχανή

Ήδη από την κατασκευή του πρώτου υπολογιστή στο μυαλό των ανθρώπων υπήρχε η ιδέα της κατασκευής μιας μηχανής η οποία θα μπορούσε να αναπαράγει την ανθρώπινη συμπεριφορά και τις ανθρώπινες αντιδράσεις. Με την πάροδο των ετών δεν είναι λίγες οι περιπτώσεις όπου τα σύνορα μεταξύ των διάφορων επιστημονικών αντικειμένων έπαψαν να είναι διακριτά και επιστήμονες όπως ο Rosenblat ο οποίος είχε ασχοληθεί με την ψυχολογία, έδωσε το έναυσμα για την ραγδαία ανάπτυξη της τεχνητής νοημοσύνης όπως την ξέρουμε σήμερα με την δημοσίευσή του [1].

Η δημοσίευση αυτή θεωρείται ένα κομβικό σημείο στην προσπάθεια του ανθρώπου να δημιουργήσει μηχανές που μπορούν να αναπαράγουν κάποιου είδους ευφυή λειτουργία ή λειτουργία μάθησης. Επίσης, έγινε σημαντική προσπάθεια ώστε να διερευνηθούν οι μηχανισμοί οι οποίοι επιτρέπουν στους βιολογικούς οργανισμούς να επεξεργάζονται την πληροφορία από το περιβάλλον τους. Όλη η γνώση η οποία άρχισε να συσσωρεύεται βοήθησε στην ραγδαία ανάπτυξη τόσο νέων αλγορίθμων αλλά κατέδειξε και την ανάγκη για μεγαλύτερη υπολογιστική ισχύ ώστε να είναι δυνατή η επεξεργασία ολοένα μεγαλύτερου όγκου δεδομένων.

1.2 Μηχανική Μάθηση και Δεδομένα

Η Μηχανική Μάθηση είναι ένας όρος ο οποίος αναδείχθηκε μέσα από την ανάγκη να μπορέσουν οι ερευνητές από διάφορους κλάδους να δώσουν τις κατάλληλες εντολές στους υπολογιστές ώστε να εκτελέσουν διαδικασίες και να μπορέσουν να διδαχθούν από μια διαδικασία εκπαίδευσης πάνω σε μια ομάδα δεδομένων.

Τα δεδομένα τα οποία χρησιμοποιούνται για εκπαίδευση αυτή ανομάζονται δεδομένα εκπαίδευσης (Training Data). Ένα πολύ σημαντικό μέρος της εκπαίδευσης είναι η ύπαρξη δεδομένων, χωρίς τα οποία αυτή είναι αδύνατη. Για πρώτη φορά εμφανίστηκε η ανάγκη για την συλλογή και την προ-επεξεργασία των παραπάνω. Η ανάγκη αυτή ώθησε τους ερευνητές να αφιερώσουν χρόνο για την κατασκευή Dataset που θα χρησιμοποιηθούν για τις ανάγκες των δικτύων αυτών. Η όλο και μεγαλύτερη πολυπλοκότητα των δικτύων αυτών κατέδειξε την ανάγκη όχι μόνο για μεγαλύτερο όγκο δεδομένων αλλά και για μεγαλύτερη ταχύτητα επεξεργασίας τους από τους αλγορίθμους που δημιουργήθηκαν στην συνέχεια. Τα δεδομένα είναι ουσιαστικά η πρώτη ύλη πάνω στην οποία ο εκάστοτε αλγόριθμος θα μπορέσει να εκπαιδεύσει κάποιες παραμέτρους, ώστε να καταφέρει να ανταπεξέλθει καλύτερα στην εργασία που του έχει ανατεθεί. Στην παρούσα Διπλωματική Εργασία

θα αφιερωθεί ειδικό καφάλαιο για την περιγραφή της διαδικασίας συλλογής και προ-επεξεργασίας των δεδομένων, που θα χρησιμοποιηθούν για την εκπαίδευση των δικτύων.

1.3 Computer Vision

Με την ανάπτυξη της Μηχανικής Μάθησης δημιουργήθηκε και το πεδίο στην Επιστήμη των Υπολογιστών που αναφέρεται ως Όραση Υπολογιστών ή αλλιώς Computer Vision. Η Όραση Υπολογιστών θα μπορούσε να περιγραφεί ως η διαδικασία μέσα από την οποία ένα υπολογιστικό σύστημα αναγνωρίζει πρότυπα από κάποια σειρά εικόνων όπως τα frames κάποιου βίντεο ή από κάποια μεμονωμένη εικόνα. Η Όραση Υπολογιστών αντικατοπτρίζει την προσπάθεια των επιστημόνων να δώσουν την δυνατότητα στα υπολογιστικά συστήματα να αντιληφθούν το περιβάλλον με έναν τέτοιο τρόπο που να είναι ελεγχόμενος και να μπορεί να δώσει λύσεις σε προβλήματα τόσο της καθημερινότητας αλλά και περισσότερο εξειδικευμένα όπως πολύπλοκα προβλήματα κατάτμησης στον χώρο.

Η Όραση Υπολογιστών με το πέρασμα των ετών και την ανάπτυξη ισχυρών συστημάτων από την σκοπιά της υπολογιστικής ισχύος, συνδυάζεται σχεδόν κατά κανόνα με αλγόριθμους Βαθιάς Μάθησης. Ο τομέας αυτός όπως αναφέρεται και στο [5] από τον T.S. Huang κάνει προσπάθεια να μιμηθεί την ικανότητα των ζωντανών οργανισμών να αντιλαμβάνονται τον γύρω κόσμο μέσα από κάποια ερεθίσματα. Αυτά τα ερεθίσματα δεν περιορίζονται μόνο στα οπτικά αλλά και είναι δυνατό να είναι ακόμα και ήχοι. Τόσο οι οργανισμοί αλλά και τα υπολογιστικά συστήματα μετά από την ανάλυση αυτών των ερεθισμάτων καλούνται και λαμβάνουν κάποιες αποφάσεις. Οι μεν ζωντανοί οργανισμοί έχουν την δυνατότητα να ανταπεξέρχονται με μεγαλύτερη επιτυχία και οι δε υπολογιστές προσπαθούν να βελτιωθούν συνεχώς.

Ήδη από το έτος 1996 ήταν εμφανής η αξία του τομέα της Όρασης Υπολογιστών, με τους ερευνητές να αναγνωρίζουν την πολυπλοκότητα των προβλημάτων που τα δίκτυα καλούνταν να λύσουν. Τέτοια προβλήμα αποτελούν αντικείμενο συνεχούς έρευνας και περισσότερο θα μπορούσαν να καλούνται προβληματισμοί που αποτελούν εφαλτήριο για νέες έρευνες.

1.4 Αναγνώριση Δράσεων-Action Recognition

Η αναγνώριση δράσεων(Action Recognition) είναι ένας εξαιρετικά σημαντικός τομέας της επιστήμης της Πληροφορικής που τα τελευταία χρόνια έχει εξελιχθεί σημαντικά. Ειδικά με την εξέλιξη των υπολογιστικών συστημάτων που γίνονται ολοένα και πιο ισχυρά, καθώς και με την εξέλιξη των Αρχιτεκτονικών των Νευρωνικών Δικτύων, είναι δυνατή η μελέτη των τεχνικών και η εξαγωγή αποτελεσμάτων που είναι συγκρίσιμα με αυτά που προκύπτουν από την ταξινόμηση μέσω ανθρώπου.

Ειδικότερα, η μελέτη και η ανάπτυξη τεχνικών που είναι δυνατό να αναγνωρίσουν αποτελεσματικά διαφορετικές κατηγορίες δράσεων είναι εξαιρετικής σημασίας για πολλούς τομείς, όχι μόνο της καθημερινότητας αλλά και της επιστήμης. Έτσι, εφαρμογές για Αναγνώριση Δράσεων χρησιμοποιούνται για παράδειγμα στην βελτίωση της ασφάλειας χώρων, στην μελέτη της συμπεριφοράς καταναλωτών, αλλά και στην γρηγορότερη ανάπτυξη φαρμάκων. Ειδικότερα, η αναγνώριση δράσεων τείνει να αποτελέσει ένα πολύ σημαντικό τομέα και η μελέτη των τεχνικών που μπορούν να δώσουν τα καλύτερα δυνατά αποτελέσματα είναι εξαιρετικά σημαντική. Ιδιαίτερη σημασία έχει η μελέτη της δυνατότητας των τεχνητών νευρωνικών δικτύων να αναγνωρίσουν δράσεις σε βίντεο χωρίς την εμπλοκή κάποιου χαρακτηριστικού προτύπου του παρασκηνίου. Τα πειράματα που θα παρουσιαστούν στην συνέχεια της ΔΕ έχουν την ιδιαιτερότητα ότι δεν υπάρχει κάποια πληροφορία

εκτός από τον επιμύ που κινείται μέσα στην δεξαμενή. Αυτό δίνει την δυνατότητα να μελετηθούν οι δυνατότητες των δικτύων στην αναγνώριση δράσεων ιδιαίτερα σε πειράματα που σχετίζονται με την Φαρμακολογία και συγκεκριμένα το πείραμα εξαναγκασμένης κολύμβησης.

Η μελέτη αντικαταθλιπτικών φαρμάκων βασίζεται στο πείραμα της εξαναγκασμένης κολύμβησης με χρήση επιμύων ως πειραματόζωα. Τα πειράματα που διεξάγονται με την διαδικασία αυτή έχουν στόχο την κατηγοριοποίηση της κίνησης του ζώου, ώστε να μελετηθεί η αποτελεσματικότητα του φαρμάκου στο νευρικό του σύστημα. Η συνεισφορά της Βαθιάς Μάθησης αναμένεται να είναι καταλυτικής σημασίας για την διεξαγωγή τέτοιων μελετών, αφού είναι δυνατή η αυτόματη αναγνώριση της κατηγορίας κίνησης και η εξαγωγή συμπερασμάτων σε πολύ μικρότερο χρόνο, με αποτέλεσμα την επίσπευση της όλης διαδικασίας και την μελέτη φαρμάκων γρηγορότερα και πιο αποτελεσματικά.

Κεφάλαιο 2

Θεωρητικό Υπόβαθρο

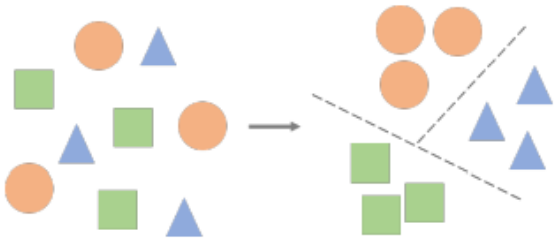
Το δεύτερο κεφάλαιο της παρούσας Διπλωματικής Εργασίας αναφέρεται σε όλο το θεωρητικό υπόβαθρο που αποτελεί την βάση για την μελέτη της συμπεριφοράς των Νευρωνικών Δικτύων. Αρχικά γίνεται αναφορά στους τύπους αλγόριθμων για μηχανική μάθηση. Επιπλέον, γίνεται αναφορά στον Τεχνητό Νευρώνα (Perceptron) που αποτελεί την βάση για την περαιτέρω κατανόηση των δικτύων. Στην συνέχεια αναλύονται οι συναρτήσεις κόστους και διαδικασία εκπαίδευσης μέσα από την περιγραφή των περισσότερο διαδεδομένων αλγορίθμων βελτιστοποίησης.

2.1 Τύποι Αλγόριθμων Μηχανικής Μάθησης

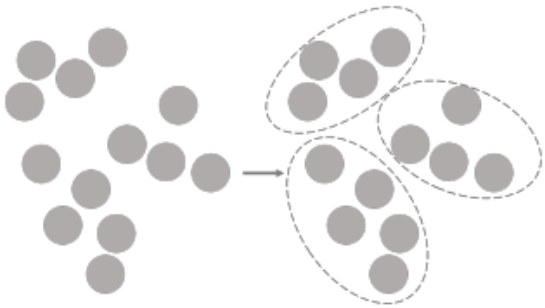
Σύμφωνα με το [45] η μηχανική μάθηση έχει ως σκοπό της, την αυτοματοποίηση διαδικασιών που βασίζονται σε κάποιο τύπο δεδομένων. Η Μηχανική Μάθηση όπως αναφέρθηκε παραπάνω αυτοματοποιεί διαδικασίες, με το να εντοπίζει πρότυπα μέσα σε πολλές φορές πολύπλοκα δεδομένα. Παρακάτω, αναφέρονται οι βασικές κατηγορίες αλγορίθμων Μηχανικής Μάθησης:

- Επιβλεπόμενη Μάθηση- Supervised Learning:Πρόκειται για την μέθοδο στην οποία τα δεδομένα έχουν την μορφή μεταβλητών εισόδου και επιθυμητές τιμές εξόδου. Όπως είναι φανερό για την εφαρμογή αυτής της μεθόδου, προϋπόθεση είναι η ύπαρξη μεγάλων σετ δεδομένων εκπαίδευσης.

- Μη Επιβλεπόμενη Μάθηση- Unsupervised Learning:Η τεχνική αυτή είναι η πλέον κατάλληλη όταν υπάρχουν τα δείγματα διαθέσιμα, αλλά όχι οι κλάσεις που αντιστοιχούν σε κάθε ομάδα. Με την Μη Επιβλεπόμενη Μάθηση μπορούν να εντοπιστούν ομάδες στα δεδομένα.



Σχήμα 2.1: Παράδειγμα Επιβλεπόμενης Μάθησης, τα δείγματα ταξινομούνται σε γνωστές κλάσεις.[45]



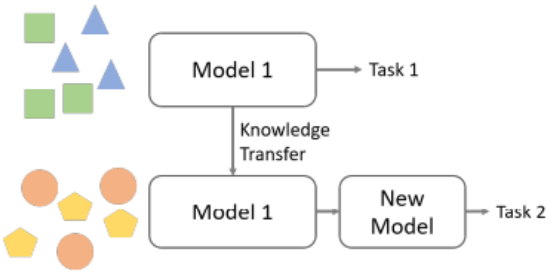
Σχήμα 2.2: Μη Επιβλεπόμενη Μάθηση, τα δείγματα ταξινομούνται σε συστάδες.[45]

- Ημιεπιβλεπόμενη Μάθηση- Semisupervised Learning:Εξαιρετικά σημαντικός τύπος μάθησης, είναι αυτός της Ημιεπιβλεπόμενης Μάθησης. Οι αλγόριθμοι που ανήκουν σε αυτήν την οικογένεια, χρησιμοποιούν έναν συνδυαστικό τρόπο για να κατηγοριοποιήσουν τα δείγματα. Αρχικά αυτά ομαδοποιούνται σε συστάδες χωρίς να λαμβάνονται υπόψη δεδομένα εκπαίδευσης, και στην συνέχεια με την χρήση μικρού αριθμού ταξινομημένων δειγμάτων κατατάσσονται όλες οι συστάδες που βρέθηκαν προηγουμένως.

Σχήμα 2.3: Ημιεπιβλεπόμενη Μάθηση, τα δείγματα ταξινομούνται σε συστάδες και στην συνέχεια ταξινομούνται με χρήση μικρού όγκου δεδομένων εκπαίδευσης.[45]

- Μεταφορά Μάθησης- Transfer Learning: Η μεταφορά μάθησης, είναι η τεχνική, κατά την οποία ένα δίκτυο εκπαιδεύεται αρχικά σε ένα σετ δεδομένων και τα βάρη που έχουν προκύψει από την διαδικασία αυτή χρησιμοποιούνται για να αρχικοποιηθούν κάποια άλλα και για κάποια

παραπλήσια διαδικασία ταξινόμησης. Η αρχική εκπαίδευση, συνήθως γίνεται σε κάποιο γνωστό Dataset που περιέχει μεγάλο αριθμό κατηγοριοποιημένων δεδομένων. Με τον τρόπο αυτό είναι δυνατή η μεταφορά της μάθησης και η επίτευξη καλύτερης σύγκλισης και σε μικρότερο χρόνο, αλλά με επιπλέον δεδομένα που πρέπει να είναι επίσης κατηγοριοποιημένα. Παρακάτω φαίνεται η σχηματική αναπαράσταση της τεχνικής αυτής.



Σχήμα 2.4: Η διαδικασία της Μεταφοράς Μάθησης (Transfer Learning).[45]

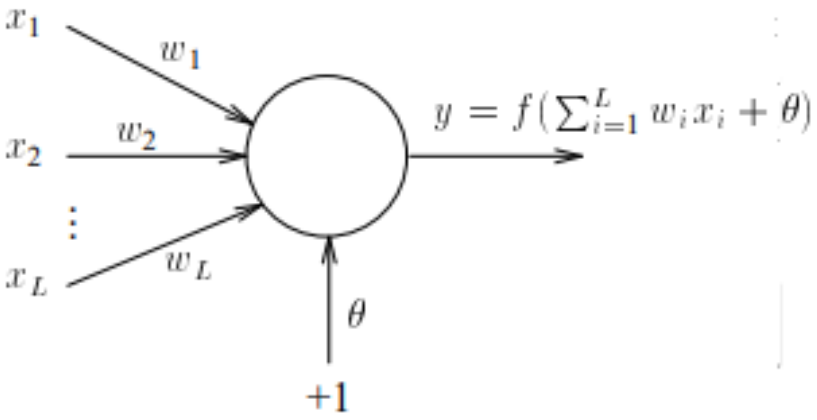
2.2 Ο Τεχνητός Νευρώνας-Perceptron

Το βασικό στοιχείο με το οποίο δομείται ένα απλό Νευρωνικό Δίκτυο,είναι ο Τεχνητός Νευρώνας (Perceptron). Ο Τεχνητός Νευρώνας εκφράζεται μαθηματικά ως:

$$y_k = f\left(\sum_{i=0}^{N} x_{ki} w_{ki}\right) \tag{2.1}$$

Στην σχέση 2.1 το y_k είναι η έξοδος του Τεχνητού Νευρώνα. Στην συνέχεια της σχέσης η συνάρτηση f αποτελεί την συνάρτηση ενεργοποίησης (Activasion Function), το x_{ki} είναι η είσοδος του νευρώνα και τέλος το w_{ki} αποτελεί το συναπτικό βάρος (Synaptic Weight).

Παρακάτω δίνεται η σχηματική αναπαράσταση του Τεχνητού Νευρώνα.



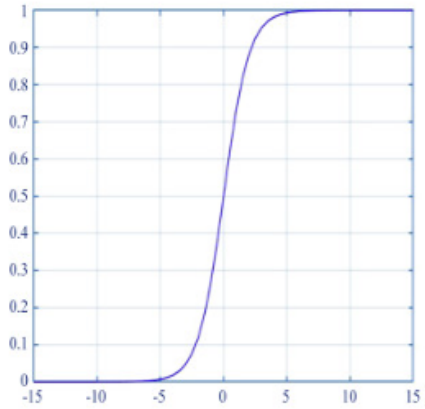
Σχήμα 2.5: Σχηματική αναπαράσταση Τεχνητού Νευρώνα[7]

2.2.1 Συνάρτηση Ενεργοποίησης

Παραπάνω έγινε αναφορά σε ένα εξαιρετικά σημαντικό μέρος του Τεχνητού Νευρώνα, την συνάρτηση ενεργοποίησης. Έχουν προταθεί αρκετές τέτοιες συναρτήσεις, με τις σημαντικότερες να είναι οι παρακάτω [8]:

1. Σιγμοειδής,

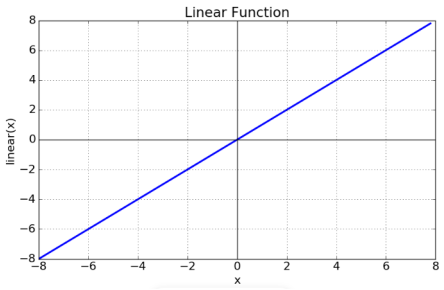
$$f(x) = \left(\frac{1}{(1 + exp^{-x})} \right) \tag{2.2}$$



Σχήμα 2.6: Σιγμοειδής συνάρτηση ενεργοποίησης [9]

2. Γραμμική,

$$f(x) = w^T x + b \tag{2.3}$$



Σχήμα 2.7: Γραφική παράσταση συνάρτησης ενεργοποίησης linear [47].

3. Rectified Linear Unit (ReLU),

$$f(x) = \begin{cases} 0 \text{ for } x < 0 \\ x \text{ for } x \geq 0 \end{cases} \tag{2.4}$$

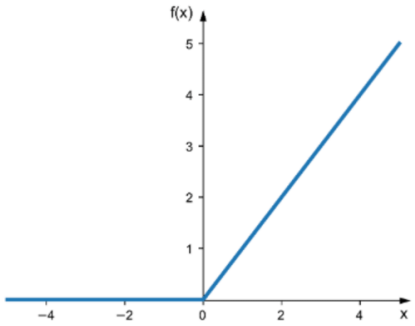


Figure 2.8: Γραφική παράσταση συνάρτησης ενεργοποίησης ReLU [9]

4. Softmax,

$$f\left(x_i\right) = \frac{\exp\left(x_i\right)}{\sum_j \exp\left(x_j\right)} \tag{2.5}$$

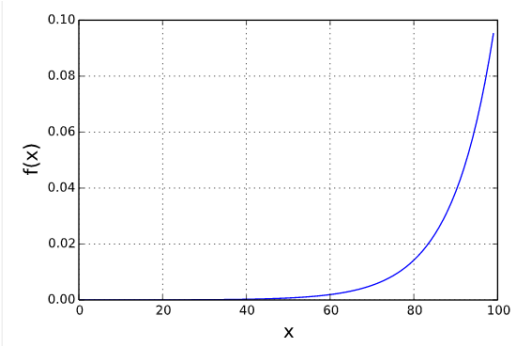
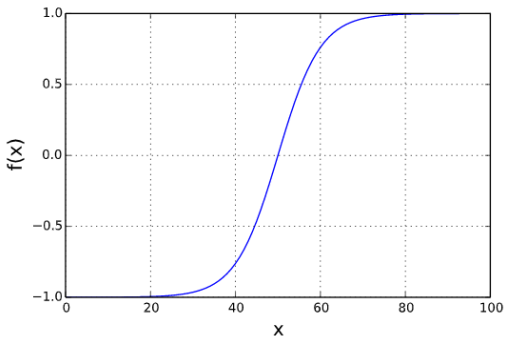


Figure 2.9: Γραφική παράσταση συνάρτησης ενεργοποίησης Softmax .

5. Υπερβολική Εφαπτομένη (Hyperbolic Tangent)

$$f(x) = \left(\frac{e^x - e^{-x}}{e^x + e^{-x}}\right) \tag{2.6}$$



Σχήμα 2.10: Γραφική παράσταση συνάρτησης ενεργοποίησης Tanh .

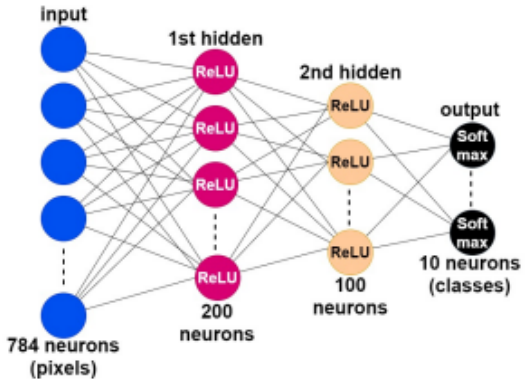
2.3 Πλήρως Συνδεδεμένα Νευρωνικά Δίκτυα

Η βασική αρχιτεκτονική η οποία αποτελείται από επιμέρους Τεχνητούς Νευρώνες και αποτελεί τον πρώτο τύπο δικτύου ο οποίος υλοποιήθηκε. Όπως αναφέρεται στο [11], τα δίκτυα αυτά έχουν ως κύριο στόχο την μίμηση των βιολογικών λειτουργιών που επιτελούνται από το νευρικό σύστημα απλών οργανισμών όπως για παράδειγμα το Caenorhabditis elegans worm [13]. Σε όλες τις περιπτώσεις που κάποιο υπολογιστικό σύστημα προσπαθεί να μιμηθεί κάποιο φυσικό αντίστοιχό του, είναι σημαντικό να τονιστεί ότι ακόμα και απλές λειτουργίες απαιτούν πολύ μεγαλύτερο αριθμό συνάψεων ώστε να μπορέσει το Τεχνητό Νευρωνικό Δίκτυο να αντιδρά με παρόμοιο τρόπο με τον οργανισμό. Τα Πλήρως Συνδεδεμένα Νευρωνικά Δίκτυα αποτελούνται από επίπεδα ή αλλιώς (Layer)Τεχνητών Νευρώνων. Τα επίπεδα αυτά κατηγοριοποιούνται σε:

1. Πρώτο Layer , το επίπεδο εισόδου στο δίκτυο

2. Τα διάφορα κρυμμένα επίπεδα (Hidden Layers)

3. Το επίπεδο εξόδου του δικτύου



Σχήμα 2.11: Αρχιτεκτονική Πλήρως Συνδεδεμένου Νευρωνικού Δικτύου[11]

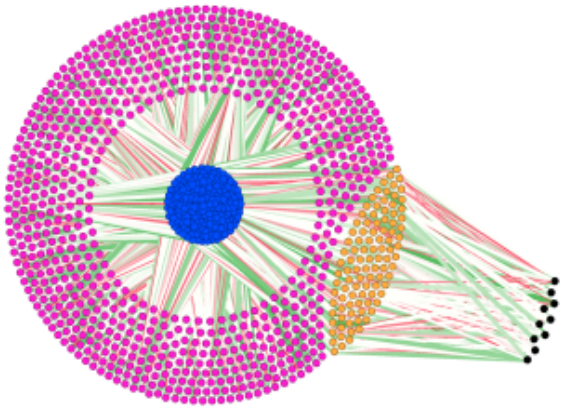
Όπως είναι φανερό από το Σχήμα 2.11, η εν λόγω αρχιτεκτονική χαρακτηρίζεται από το γεγονός ότι κάθε Τεχνητός Νευρώνας του τροφοδοτείται από τις εξόδους του προηγούμενου επιπέδου. Αυτή η διαδικασία συνεχίζεται για όλη την έκταση του δικτύου έως ότου ληφθεί η έξοδος, που μπορεί να είναι η ταξινόμηση της εισόδου σε κάποια κλάση. Η μαθηματική έκφραση της παραπάνω διαδικασίας, δίνεται ως:

$$p_k(t) = \sum_i o_j(t) w_{ik} + w_{0k} \tag{2.7}$$

Στην παραπάνω σχέση $p_j(t)$ είναι η έξοδος του Τεχνητού Νευρώνα k και o_j η έξοδος του ακριβώς προηγούμενου Νευρώνα, αφού όπως ειπώθηκε είναι άμεσα συνδεδεμένοι.

Τέλος, είναι σημαντικό να τονιστεί ότι ακόμα και δίκτυα τέτοιου είδους, που αποτελούν το πρώτο είδος αρχιτεκτονικής το οποίο κατασκευάστηκε, ανάλογα με τον αριθμό των κρυμμένων επιπέδων, είναι δυνατό να παρουσιάσουν σημαντική πολυπλοκότητα με στόχο την καλύτερη ταξινόμηση ή

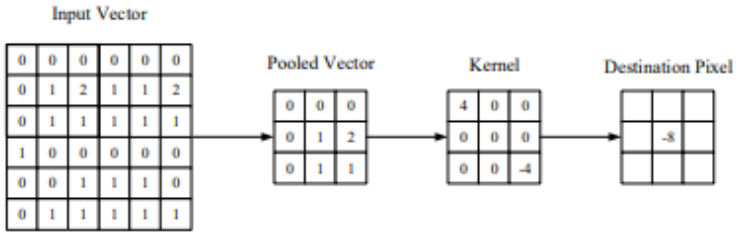
λήψη κάποιας απόφασης. Παρακάτω δίνεται η εικόνα που απεικονίζει το παραπάνω δίκτυο (Σχήμα 2.12) με όλα τα επίπεδα, καθώς και όλες τις μεταξύ τους συνδέσεις.



Σχήμα 2.12: Σχηματική αναπαράσταση FC Neural Network με τα διάφορα επίπεδα και τις συνδέσεις[11]

2.4 Συνελικτικά Νευρωνικά Δίκτυα(CNN)

Τα Συνελικτικά Νευρωνικά Δίκτυα(Convolutional Neural Networks)κατασκευάστηκαν ώστε να είναι δυνατή η αναγνώριση προτύπων σε εικόνες, αλλά και σε σειρές εικόνων όπως είναι τα frames των video.Όπως στην περίπτωση των Πλήρως Συνδεδεμένων Δικτύων, απαιτείται η ύπαρξη εκπαιδευόμενων παραμέτρων στο δίκτυο. Τον ρόλο αυτό παίζουν συνελικτικά φίλτρα, που έχουν την ιδιότητα να αναγνωρίζουν πρότυπα μετά από την διαδικασία της εκπαίδευσης. Τα φίλτρα αυτά με τις εκπαιδευόμενες παραμέτρους είναι σε θέση να αναγνωρίσουν διαφορετικά πρότυπα στην εικόνα ή στην σειρά εικόνων και να παράξουν τα feature maps τα οποία είναι ο συνδυασμός των συνελίξεων με το αντίστοιχο φίλτρο και της συνάρτησης ενεργοποίησης που χρησιμοποιείται.

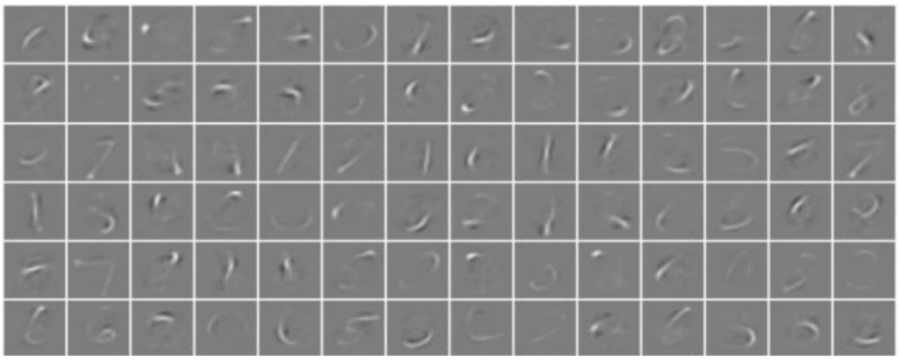


Σχήμα 2.13: Αναπαράσταση ενός συνελικτικού επιπέδου[14]

Όπως είναι φανερό από το Σχήμα 2.13, το τελικό εικονοστοιχείο εξαρτάται από τις τιμές του φίλτρου και από τις τιμές των εικονοστοιχείων της εισόδου. Αυτό οδηγεί στην τελική εξαγωγή χαρακτηριστικών όσο το δίκτυο προχωράει βαθύτερα.

2.4.1 Αριθμός φίλτρων ανά επίπεδο

Τα συνελικτικά νευρωνικά αποτελούνται από συνελικτικά επίπεδα, που με την χρήση των φίλτρων εξάγουν πληροφορία που περιέχεται στα δεδομένα εισόδου στο επίπεδο αυτό. Για να είναι δυνατή η εξαγωγή διαφόρων προτύπων από αυτά τα δεδομένα, κάθε επίπεδο, μεταξύ άλλων, κατά κανόνα περιέχει περισσότερα από ένα φίλτρα. Αυτός ο αριθμός, προκαθορίζεται και αποτελεί αναπόσπαστο μέρος της αρχιτεκτονικής των συνελικτικών νευρωνικών δικτύων. Κάθε φορά που μια εικόνα εισέρχεται ως είσοδος στο δίκτυο, το φίλτρο με τα αντίστοιχα βάρη ΄περνάει΄ από όλη την εικόνα εισόδου και στην συνέχεια ανάλογα με την συνάρτηση ενεργοποίησης που έχει επιλεχθεί θα είναι σε θέση να αναγνωρίζει την ύπαρξη ή όχι ενός συγκεκριμένου προτύπου στην εικόνα εισόδου. Σε αυτό το σημείο, αναφέρεται ότι η είσοδος εκτός από μεμονωμένες εικόνες (με την συνήθη μορφή πινάκων) μπορεί να είναι και ακολουθίες από frame και σε αυτήν την περίπτωση εκτός από την χωρική διάσταση στο φίλτρο υπάρχει ακόμα μια, αυτή του χρόνου και πλέον γίνεται λόγος για 3D συνελικτικά νευρωνικά δίκτυα.



Σχήμα 2.14: Τα Feature Maps μετά την ενεργοποίησή τους στο 1ο συνελικτικό επίπεδο ενός CNN για την αναγνώριση χειρόγραφων χαρακτήρων[14]

2.4.2 Βήμα του φίλτρου-Stride

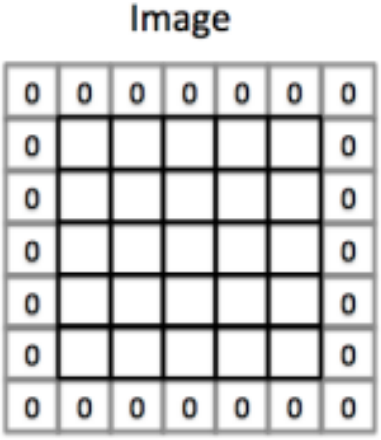
Όπως αναφέρθηκε τα φίλτρα έχουν τον κύριο ρόλο να εξάγουν πρότυπα από τα δεδομένα εισόδου. Τα φίλτρα (Kernels)συνήθως έχουν διαστάσεις 2x2, 3x3 ή και 5x5 , εάν πρόκειται για δίκτυα 2 διαστάσεων και αντίστοιχα 2x2x2, 3x3x3 ή και 5x5x5 όταν το συνελικτικό δίκτυο παίρνει σαν είσοδο δεδομένα τριών διαστάσεων. Για να είναι δυνατή η επεξεργασία όλης της χωρικής (και πιθανά χρονικής) διάστασης των δεδομένων εισόδου είναι απαραίτητο να γίνει η επιλογή του βήματος με το οποίο το φίλτρο κάθε φορά θα προχωράει ώστε να παραλάβει για επεξεργασία το επόμενο κομμάτι της εισόδου. Αυτή η υπερπαράμετρος ονομάζεται Βήμα (Stride).

2.4.3 Padding

Επόμενο σημαντικό μέρος για την κατανόηση των συνελικτικών νευρωνικών δικτύων είναι η έννοια της επένδυσης ή Padding.Όταν ένα φίλτρο κινείται κατά την χωρική διάσταση του πίνακα εισόδου, αυτή αλλάζει όσο προχωράει βαθύτερα στο δίκτυο. Η τεχνική της επένδυσης έχει ως αποτέλεσμα τον καλύτερο έλεγχο του μεγέθους των δεδομένων μετά από τη κάθε συνέλιξη και την καλύτερη σχεδίαση δικτύων με βαθύτερη αρχιτεκτονική.

$$\frac{(V-R)+2Z}{S+1} \tag{2.8}$$

Όπως αναφέρεται στο [14], η σχέση 2.8 πρέπει να έχει ως αποτέλεσμα έναν θετικό ακέραιο αριθμό. Όταν το αποτέλεσμα είναι τέτοιο όλοι οι νευρώνες θα είναι σε θέση να συνδεθούν σωστά και να μπορέσουν να επεξεργαστούν όλα τα δεδομένα εισόδου.



Σχήμα 2.15: Σχηματική αναπαράσταση επένδυσης με τιμές 0. [15]

2.4.4 Επίπεδο Συγκέντρωσης- Pooling Layers

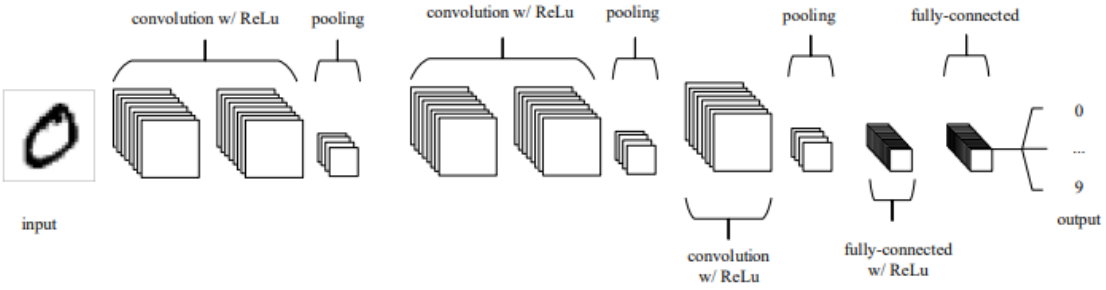
Τα επίπεδα αυτά δεν περιέχουν βάρη με την έννοια που αναφέρθηκε παραπάνω, αλλά επιτελούν λειτουργίες όπως το Average Pooling και το Max Pooling που αποτελούν και δύο από τα περισσότερο διαδεδομένα επίπεδα. Αυτά τα layer έχουν την ιδιότητα να μειώνουν τον αριθμό των παραμέτρων του δικτύου, καθιστώντας το λιγότερο περίπλοκο.

Ειδικότερα, χρησιμοποιείται ένα ΄παράθυρο΄ 2x2 (σύνηθες μέγεθος) και από την περιοχή αυτή κρατείται η μεγαλύτερη τιμή (Max Pooling). Επιπλέον, αφού γίνεται λόγος για ΄παράθυρο΄ κάποιου μεγέθους, επιλέγεται και ένα βήμα ίσο με 2 τις περισσότερες φορές. Τα επίπεδα αυτά όπως έχει αναλυθεί δεν περιέχουν από μόνα τους εκπαιδευόμενα βάρη σε μορφή φίλτρου, αλλά με τις λειτουργίες που επιτελούν όπως η συγκέντρωση τελικά των μέγιστων τιμών της εισόδου τους, μικραίνουν τον όγκο των δεδομένων. Αυτό συνεπάγεται ότι ένα μέρος της πληροφορίας που αρχικά εισήλθε προς επεξεργασία θα χαθεί.

Όπως είναι φυσικό, οι παράμετροι του βήματος και του μεγέθους του παραθύρου είναι δυνατό να ρυθμιστούν τόσο με τρόπο τέτοιο ώστε να ελαχιστοποιηθεί η επικάλυψη στην χωρική ή και χρονική διάσταση, όσο και να επιλεχθούν τιμές που να δώσουν τιμές εξόδου από επικαλυπτόμενο τμήμα της εισόδου όπως για παράδειγμα βήμα ίσο με 2 και μέγεθος παραθύρου 3x3 . Με βάση την μελέτη που έχει δημοσιευτεί [14] ένα φίλτρο μεγαλύτερο από 3x3 θα οδηγούσε σε μικρότερη απόδοση του δικτύου, λόγω των χαρακτηριστικών που αναφέρθηκαν παραπάνω.

2.4.5 Πλήρως Συνδεδεμένα Επίπεδα

Στην περίπτωση του προβλήματος της ταξινόμησης κλάσεων από εικόνες ή σειρές τέτοιων όπως είναι η αλληλουχίες των frames των video στα τελευταία επίπεδα των Συνελικτικών Δικτύων υπάρχουν πλήρως συνδεδεμένα επίπεδα.



Σχήμα 2.16: Συνελικτικό Νευρωνικό δίκτυο με πλήρως συνδεδεμένα επίπεδα για την αναγνώριση χειρόγραφων αριθμών [14]

Στο Σχήμα 2.16, τα τελευταία επίπεδα είναι πλήρως συνδεδεμένα, σε πρώτο στάδιο με όλους τους προηγούμενους νευρώνες και στην συνέχεια με όλους τους νευρώνες της εξόδου,με αποτέλεσμα αναγνώριση των αριθμών. Το γεγονός ότι όλοι οι νευρώνες των τελευταίων επιπέδων συνδέονται με όλους του αμέσως προηγούμενους και τους επόμενους, προσθέτει εκπαιδεύσιμες παραμέτρους στο δίκτυο, αυξάνοντας την πολυπλοκότητά του [16].

2.4.6 Συνάρτηση κόστους- Loss Function

Όλα τα παραπάνω αποτελούν τον τρόπο με τον οποίο το δίκτυο επεξεργάζεται όλα τα δεδομένα εισόδου και στην έξοδό του δίνει τελικά ένα αποτέλεσμα που πρέπει να ανταποκρίνεται στα δεδομένα βάσης, προερχόμενα από κάποιον έμπειρο μελετητή στην περίπτωση της επιβλεπόμενης ταξινόμησης (Supervised Learning). Για να είναι δυνατή η βελτίωση της απόδοσης του τεχνητού νευρωνικού δικτύου, πρέπει να είναι δυνατή η αξιολόγηση του αποτελέσματος μέσα από μια συνάρτηση που να αποδίδει την επίδοσή του. Αυτή, ονομάζεται συνάρτηση κόστους και αποτελεί έναν ενεργό τομέα έρευνας που εξελίσσεται συνεχώς με πολλές βελτιώσεις. Επιπλέον, η συνάρτηση κόστους που επιλέγεται για ένα δίκτυο πρέπει να είναι η κατάλληλη ώστε να επιτευχθεί η καλύτερη δυνατή σύγκλιση. Στο πρόβλημα της ταξινόμησης σε πολλές κλάσεις από εικόνες, οι κλάσεις μπορεί να εκφραστούν ως:

$$y_i = [y_{i,1}, \ldots, y_{i,C}] \tag{2.9}$$

Με το διάνυσμα y_i να περιέχει τις ετικέτες από την κάθε κλάση [17]. Ακόμα, η διαδικασία της ταξινόμησης θα μπορούσε να μοντελοποιηθεί με την παρακάτω σχέση:

$$F(x_i) = g(f(x_i)) \tag{2.10}$$

Όπου γίνεται η ταξινόμηση με βάση την συνάρτηση $f(x_i) = p_i$ και την συνάρτηση $g()$, η οποία

δίνει την πιθανότητα η εικόνα να ανήκει σε μια κλάση. Η πιθανότητα μια εικόνα είσοδος στο δίκτυο να ανήκει σε μια από τις κλάσεις δίνεται από:

$$p_i = \begin{cases} y & \text{if } y = 1 \\ 1 - y & \text{otherwise} \end{cases} \tag{2.11}$$

Παρακάτω δίνονται οι σημαντικότερες συναρτήσεις κόστους που χρησιμοποιούνται για την ταξινόμηση κλάσεων επό εικόνες.

1. Cross Entropy Loss, η συνάρτηση αυτή προέρχεται από την απόκλιση Kullback-Leibler και είναι η βασικότερη από όλες τις συναρτήσεις κόστους [18]. Εδώ γίνεται προσπάθεια να ελαχιστοποιηθεί η απόκλιση ανάμεσα από την προβλεπόμενη κατανομή και την πραγματική.

$$\text{CEL} = -\sum \log\left(p_i\right) \tag{2.12}$$

2. Focal Loss,εδώ υπάρχει μια επιπλέον παράμετρος γ που στόχο έχει την καλύτερη ταξινόμηση στις κλάσεις που το δίκτυο δεν είναι σε θέση να ανταπεξέλθει με μεγάλη επιτυχία.

$$\text{FL} = -\sum \left(1 - p_i\right)^{\gamma} \log\left(p_i\right) \tag{2.13}$$

3. Hamming loss (HAL), Αυτή η συνάρτηση κόστους, έχει ως κύριο στόχο την μείωση του ποσοστού των λανθασμένα ταξινομημένων κλάσεων σε σχέση με το σύνολό τους.

$$\text{HAL} = \frac{1}{C} \sum_{c=1}^{C} y_{i,c} \oplus g\left(p_{i,c}\right) \tag{2.14}$$

4. SparseMax loss (SML), η συνάρτηση αυτή έχει ενσωματωμένη την συνάρτηση ενεργοποίησης Sparse Max.Στόχος της είναι να δίνει ως μοναδική κλάση αυτήν με την μεγαλύτερη πιθανότητα και να μηδενίζει όλες τις άλλες.

$$\text{SML} = -y_i^T z_i + \frac{1}{2} \sum_{j \in S} \left(z_{i,j}^2 - \tau^2\left(z_i\right)\right) + \frac{1}{2} \left\|y_i\right\|^2 \tag{2.15}$$

Από τα παραπάνω είναι εμφανές ότι η συνάρτηση κόστους είναι ένα πολύ σημαντικό κομμάτι του τεχνητού νευρωνικού δικτύου και η σωστή επιλογή της έχει άμεσο αντίκτυπο στην σωστή ταξινόμηση των κλάσεων που ζητούνται ως αποτέλεσμα από αυτό.

2.4.7 Αλγόριθμος βελτιστοποίησης- Optimizer

Η εκπαίδευση του νευρωνικού δικτύου είναι άρρηκτα συνδεδεμένη με τον αλγόριθμο βελτιστοποίησης, μέσα από τον οποίο κατά την διαδικασία της εκπαίδευσης αλλάζουν οι παράμετροι του με τρόπο τέτοιο ώστε να είναι σε θέση να προβλέψει την σωστή κλάση στο τελικό επίπεδο. Όπως συμβαίνει και με την περίπτωση των συναρτήσεων κόστους οι αλγόριθμοι αυτοί αποτελούν αντικείμενο έρευνας και υπόκεινται σε συνεχείς βελτιώσεις.

1. Vanilla Gradient Descent,στην πιο απλή περίπτωση με δεδομένο ότι υπάρχει ένα σετ δεδομένων εκπαίδευσης T, ο αλγόριθμος βελτιστοποίησης έχει την μορφή:

$$\boldsymbol{\theta}^{(\tau)} = \boldsymbol{\theta}^{(\tau-1)} - \eta \cdot \nabla_{\boldsymbol{\theta}} \mathcal{L}\left(\boldsymbol{\theta}^{(\tau-1)}; \mathcal{T}\right) \tag{2.16}$$

Το $\theta^{(\tau)}$ είναι η παράμετρος που αλλάζει, στην επανάληψη τ, η $\theta^{(\tau-1)}$ η παράμετρος στην αμέσως προηγούμενη επανάληψη, το η είναι ο ρυθμός μάθησης και το $\nabla_{\boldsymbol{\theta}} \mathcal{L}\left(\boldsymbol{\theta}^{(\tau-1)}; \mathcal{T}\right)$, είναι η παράγωγος της συνάρτησης κόστους [19]. Εδώ όλο το σετ δεδομένων εκπαίδευσης περνάει μέσα από το δίκτυο πριν γίνει η αλλαγή στα βάρη του, γεγονός που τον κάνει μη αποτελεσματικό για μεγάλα dataset. Παρακάτω δίνεται ο ψευδοκώδικας για τον αλγόριθμο αυτό:

Algorithm 1 Vanilla Gradient Descent

Require: Training Set: $\mathcal{T}$; Learning Rate η; Normal Distribution Std: σ.
Ensure: Model Parameter $\boldsymbol{\theta}$

```
Initialize parameter with Normal distribution θ ~ N(0, σ²)
Initialize convergence tag = False
while tag == False do
    Compute gradient ∇_θ L(θ; T) on the training set T
    Update variable θ = θ − η · ∇_θ L(θ; T)
    if convergence condition holds then
        tag = True
    end if
end while
Return model variable θ
```

Σχήμα 2.17: Ψευδοκώδικας Vanilla Gradient Descent [19]

2. Stochastic Gradient Descent, αυτός αποτελεί βελτιωμένη εκδοχή του προηγούμενου, αφού εδώ η αλλαγή των παραμέτρων γίνεται σε πολλά σημεία του σετ δεδομένων και όχι μετά το σύνολό του.

$$\boldsymbol{\theta}^{(\tau)} = \boldsymbol{\theta}^{(\tau-1)} - \eta \cdot \nabla_{\boldsymbol{\theta}} \mathcal{L}\left(\boldsymbol{\theta}^{(\tau-1)}; (\mathbf{x}_i, \mathbf{y}_i)\right) \tag{2.17}$$

Το $\theta^{(\tau)}$ είναι η παράμετρος η οποία ενημερώνεται, στην επανάληψη τ, η $\theta^{(\tau-1)}$ η παράμετρος στην αμέσως προηγούμενη επανάληψη, το η είναι ο ρυθμός μάθησης και το$\nabla_{\boldsymbol{\theta}} \mathcal{L}\left(\boldsymbol{\theta}^{(\tau-1)}; \mathcal{T}\right)$, είναι η παράγωγος της συνάρτησης κόστους [19]. Εδώ δεν περνάει όλο το σετ δεδομένων εκπαίδευσης μέσα από το δίκτυο,αφού γίνει ανακάτεμα των δειγμάτων, πριν ενημερωθούν τα βάρη του. Το βασικό μειονέκτημα του αλγορίθμου αυτού είναι ότι παρόλο που είναι σε θέση να εκπαιδεύσει δίκτυα με μεγάλο όγκο δεδομένων, η συχνή αλλαγή των παραμέτρων μπορεί να προκαλέσει αρνητικές συνέπειες και τελικά να μην είναι σε θέση να συγκλίνει στο ολικό ελάχιστο της συνάρτησης κόστους. Παρακάτω δίνεται ο ψευδοκώδικας για τον αλγόριθμο αυτό:

Algorithm 2 Stochastic Gradient Descent

Require: Training Set $\mathcal{T}$; Learning Rate η; Normal Distribution Std: σ.
Ensure: Model Parameter $\boldsymbol{\theta}$

```
Initialize parameter with Normal distribution θ ~ N(0, σ²)
Initialize convergence tag = False
while tag == False do
    Shuffle the training set T
    for each data instance (x_i, y_i) ∈ T do
        Compute gradient ∇_θ L(θ; (x_i, y_i)) on the training instance (x_i, y_i)
        Update variable θ = θ − η · ∇_θ L(θ; (x_i, y_i))
    end for
    if convergence condition holds then
        tag = Truc
    end if
end while
Return model variable θ
```

Σχήμα 2.18: Ψευδοκώδικας Stochastic Gradient Descent [19]

3. Mini-batch Gradient Descent, στην περίπτωση αυτή το σετ δεδομένων χωρίζεται σε μικρές ομάδες, τις λεγόμενες παρτίδες (batch). Αυτός ο αλγόριθμος μπορεί να βελτιστοποιήσει δίκτυα με μεγάλο όγκο δεδομένων. Επίσης, δεν παρουσιάζει το μειονέκτημα της προηγούμενης μεθόδου όπου μπορεί να υπάρξουν δυσκολίες στην σύγκλιση.

$$\boldsymbol{\theta}^{(\tau)} = \boldsymbol{\theta}^{(\tau-1)} - \eta \cdot \nabla_{\boldsymbol{\theta}} \mathcal{L}\left(\boldsymbol{\theta}^{(\tau-1)}; \mathcal{B}\right) \tag{2.18}$$

Στην σχέση 2.18, υπάρχουν τα ίδια στοιχεία με παραπάνω (2.17), με την διαφορά την παρτίδα ως υποσύνολο του dataset $\mathcal{B} \subset \mathcal{T}$. Όπως και παραπάνω ο αλγόριθμος είναι ο εξής:

Algorithm 3 Mini-batch Gradient Descent

Require: Training Set $\mathcal{T}$; Learning Rate η; Normal Distribution Std σ; Mini-batch Size b.
Ensure: Model Parameter $\boldsymbol{\theta}$

Initialize parameter with Normal distribution $\boldsymbol{\theta} \sim N(0, \sigma^2)$
Initialize convergence $tag = False$
while $tag == False$ **do**
  Shuffle the training set $\mathcal{T}$
  for each mini-batch $\mathcal{B} \subset \mathcal{T}$ **do**
    Compute gradient $\nabla_{\boldsymbol{\theta}} \mathcal{L}(\boldsymbol{\theta}; \mathcal{B})$ on the mini-batch $\mathcal{B}$
    Update variable $\boldsymbol{\theta} = \boldsymbol{\theta} - \eta \cdot \nabla_{\boldsymbol{\theta}} \mathcal{L}(\boldsymbol{\theta}; \mathcal{B})$
  end for
  if convergence condition holds **then**
    $tag = True$
  end if
end while
Return model variable $\boldsymbol{\theta}$

Σχήμα 2.19: Ψευδοκώδικας Mini-batch Gradient Descent [19]

2.4.8 Αλγόριθμοι βελτιστοποίησης βασιζόμενοι στην Ορμή

Μια άλλη κατηγορία αλγορίθμων είναι αυτοί που κατά την διάρκεια της εκπαίδευσης, αλλάζουν τις εκπαιδευόμενες παραμέτρους του δικτύου λαμβάνοντας υπόψη όχι μόνο τις τελευταίες παραγώγους από την εκπαίδευση αλλά και τις προηγούμενες. Αυτό δίνει την δυνατότητα στον αλγόριθμο να αποφύγει διακυμάνσεις στην διαδικασία της σύγκλισης και να είναι αποτελεσματικότερος.

1. Η σχέση που περιγράφει τα παραπάνω με δεδομένο ότι βασίζονται στον αλγόριθμο που συζητήθηκε και παραπάνω, τον Mini-batch Gradient Descent,δίνοντας τον αλγόριθμο Momentum based Mini-batch Gradient Descent.

$$\boldsymbol{\theta}^{(\tau)} = \boldsymbol{\theta}^{(\tau-1)} - \eta \cdot \Delta \mathbf{v}^{(\tau)} \tag{2.19}$$

Όπου $\theta^{(\tau)}$ και $\theta^{(\tau-1)}$ τα διανύσματα των τρεχουσών παραμέτρων και των προηγούμενων και $\Delta \mathbf{v}^{(\tau)} = \rho \cdot \Delta \mathbf{v}^{(\tau-1)} + (1-\rho) \cdot \nabla_{\boldsymbol{\theta}} \mathcal{L}\left(\boldsymbol{\theta}^{(\tau-1)}\right)$ είναι ο όρος με τον οποίο η έννοια της ορμής εισάγεται στον αλγόριθμο.

Algorithm 4 Momentum based Mini-batch Gradient Descent

Require: Training Set $\mathcal{T}$; Learning Rate η; Normal Distribution Std σ; Mini-batch Size b; Momentum Term Weight: ρ.
Ensure: Model Parameter $\boldsymbol{\theta}$

Initialize parameter with Normal distribution $\boldsymbol{\theta} \sim N(0, \sigma^2)$
Initialize Momentum term $\Delta \mathbf{v} = \mathbf{0}$
Initialize convergence $tag = False$
while $tag == False$ **do**
  Shuffle the training set $\mathcal{T}$
  for each mini-batch $\mathcal{B} \subset \mathcal{T}$ **do**
    Compute gradient $\nabla_{\boldsymbol{\theta}} \mathcal{L}(\boldsymbol{\theta}; \mathcal{B})$ on the mini-batch $\mathcal{B}$
    Update term $\Delta \mathbf{v} = \rho \cdot \Delta \mathbf{v} + (1-\rho) \cdot \nabla_{\boldsymbol{\theta}} \mathcal{L}(\boldsymbol{\theta}; \mathcal{B})$
    Update variable $\boldsymbol{\theta} = \boldsymbol{\theta} - \eta \cdot \Delta \mathbf{v}$
  end for
  if convergence condition holds **then**
    $tag = True$
  end if
end while
Return model variable $\boldsymbol{\theta}$

Σχήμα 2.20: Ψευδοκώδικας Momentum based Mini-batch Gradient Descent [19]

2. Adam,ο αλγόριθμος αυτός αποτελεί μια ειδική περίπτωση και συγκεντρώνει πολλά επιθυμητά χαρακτηριστικά ώστε η εκπαίδευση να είναι αποτελεσματική αλλά και αποδοτική. Εδώ γίνεται

αποθήκευση των πρώτων παραγώγων και των τετραγώνων τους, δύο ποσότητες που με το πέρασμα των εποχών εκπαίδευσης μειώνονται. Έτσι, σε σχέσεις έχουμε τις ποσότητες αυτές σε διανύσματα αντίστοιχα ως:

$$\mathbf{m}^{(\tau)} = \beta_1 \cdot \mathbf{m}^{(\tau-1)} + (1-\beta_1) \cdot \mathbf{g}^{(\tau-1)} \tag{2.20}$$

$$\mathbf{v}^{(\tau)} = \beta_2 \cdot \mathbf{v}^{(\tau-1)} + (1-\beta_2) \cdot \mathbf{g}^{(\tau-1)} \odot \mathbf{g}^{(\tau-1)} \tag{2.21}$$

Όπου $\mathbf{g}^{(\tau-1)} = \nabla_{\boldsymbol{\theta}} \mathcal{L}\left(\boldsymbol{\theta}^{(\tau-1)}\right)$. Στην συνέχεια γίνεται επανυπολογισμός των παραπάνω διανυσμάτων ως:

$$\hat{\mathbf{m}}^{(\tau)} = \frac{\mathbf{m}^{(\tau)}}{1-\beta_1^{\tau}} \tag{2.22}$$

$$\hat{\mathbf{v}}^{(\tau)} = \frac{\mathbf{V}^{(\tau)}}{1-\beta_2^{\tau}} \tag{2.23}$$

Τέλος, η ενημέρωση των παραμέτρων του δικτύου γίνεται με βάση την σχέση:

$$\boldsymbol{\theta}^{(\tau)} = \boldsymbol{\theta}^{(\tau-1)} - \frac{\eta}{\sqrt{\hat{\mathbf{v}}^{(\tau)}} + \epsilon} \odot \hat{\mathbf{m}}^{(\tau)} \tag{2.24}$$

Σε αυτό το σημείο πρέπει να αναφερθεί ότι ο αλγόριθμος αυτός συνδυάζει πολλές τεχνικές και δίνει αρκετά καλά αποτελέσματα σε πληθώρα τύπων δικτύων. Κλείνοντας, παρουσιάζεται και σε αυτήν την περίπτωση ο ψευδοκώδικας του αλγορίθμου Adam .

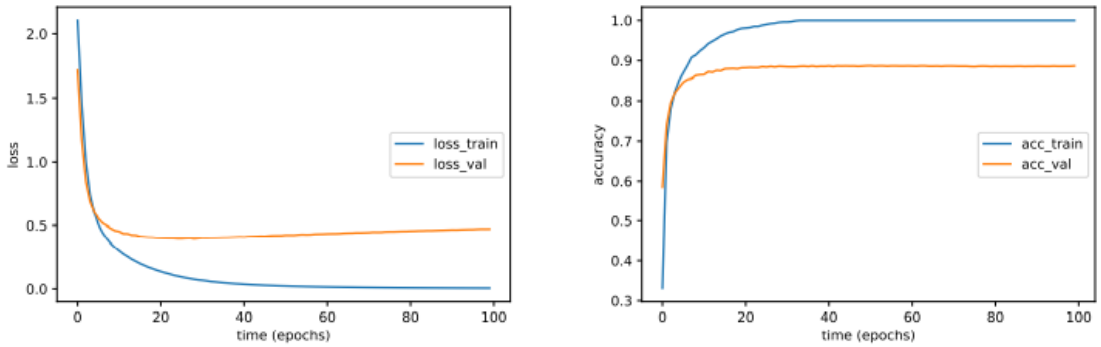
Algorithm 9 Adam

Require: Training Set $\mathcal{T}$; Learning Rate η; Normal Distribution Std σ; Mini-batch Size b; Decay Parameters β_1, β_2.
Ensure: Model Parameter $\boldsymbol{\theta}$
Initialize parameter with Normal distribution $\boldsymbol{\theta} \sim N(0, \sigma^2)$
Initialize vector $\mathbf{m} = \mathbf{0}$
Initialize vector $\mathbf{v} = \mathbf{0}$
Initialize step $\tau = 0$
Initialize convergence $tag = False$
while $tag == False$ **do**
  Shuffle the training set $\mathcal{T}$
  for each mini-batch $\mathcal{B} \subset \mathcal{T}$ **do**
    Update step $\tau = \tau + 1$
    Compute gradient vector $\mathbf{g} = \nabla_{\boldsymbol{\theta}} \mathcal{L}(\boldsymbol{\theta}; \mathcal{B})$ on the mini-batch $\mathcal{B}$
    Update vector $\mathbf{m} = \beta_1 \cdot \mathbf{m} + (1-\beta_1) \cdot \mathbf{g}$
    Update vector $\mathbf{v} = \beta_2 \cdot \mathbf{v} + (1-\beta_2) \cdot \mathbf{g} \odot \mathbf{g}$
    Rescale vector $\hat{\mathbf{m}} - \mathbf{m}/(1-\beta_1^{\tau})$
    Rescale vector $\hat{\mathbf{v}} = \mathbf{v}/(1-\beta_2^{\tau})$
    Update variable $\boldsymbol{\theta} = \boldsymbol{\theta} - \frac{\eta}{\sqrt{\hat{\mathbf{v}}}+\epsilon} \odot \hat{\mathbf{m}}$
  end for
  if convergence condition holds **then**
    $tag = True$
  end if
end while
Return model variable $\boldsymbol{\theta}$

Σχήμα 2.21: Ψευδοκώδικας αλγορίθμου Adam [19]

2.4.9 Υπερπροσαρμογή - Overfitting

Ένα τεχνητό νευρωνικό δίκτυο κατά την διάρκεια της εκπαίδευσής του, μαθαίνει να αναγνωρίζει πρότυπα σε εικόνες και να τις κατηγοριοποιεί σε κλάσεις ανάλογα με το πρόβλημα που προσεγγίζεται κάθε φορά. Το σύνολο των δεδομένων εκπαίδευσης χωρίζεται σε υποσύνολα που χρησιμοποιούνται για την εκπαίδευση αλλά και για την δοκιμή. Το δίκτυο σε αυτό το σημείο της διαδικασίας, δεν έχει κάποια επαφή με τα δεδομένα ελέγχου. Στην διαδικασία του test ένα δίκτυο με την ικανότητα να γενικεύει θα είναι σε θέση να αναγνωρίσει πρότυπα και να ταξινομήσει σωστά τα δεδομένα δοκιμής σε κλάσεις. Αντίθετα όταν παρατηρείται Overfitting του δικτύου, αυτό έχει εκπαιδευτεί μεν αλλά δεν είναι σε θέση να πετύχει αρκετά καλή ακρίβεια στα δεδομένα ελέγχου.



Σχήμα 2.22: Διαγραμματική απεικόνιση κόστους και ακρίβειας σε μοντέλο που υπερπροσαρμόζεται [20]

Παραπάνω είναι φανερό το πρόβλημα που περιγράφεται με την βοήθεια των 2 διαγραμμάτων. Και στις 2 περιπτώσεις το δίκτυο φαίνεται να αναγνωρίζει σχεδόν τέλεια τα δεδομένα εισόδου, αλλά δεν είναι σε θέση να ανταπεξέλθει με τον ίδιο αποτελεσματικό τρόπο στα δεδομένα Validation.Για την αποφυγή αυτού του φαινομένου έχουν προταθεί διάφορες τεχνικές, που αποσκοπούν είτε στην βελτίωση του μοντέλου ή στην χρήση τεχνικών ώστε να μπορεί να γενικεύει με τον καλύτερο δυνατό τρόπο. Παρακάτω αναφέρονται οι κυριότερες από αυτές τις τεχνικές [21].

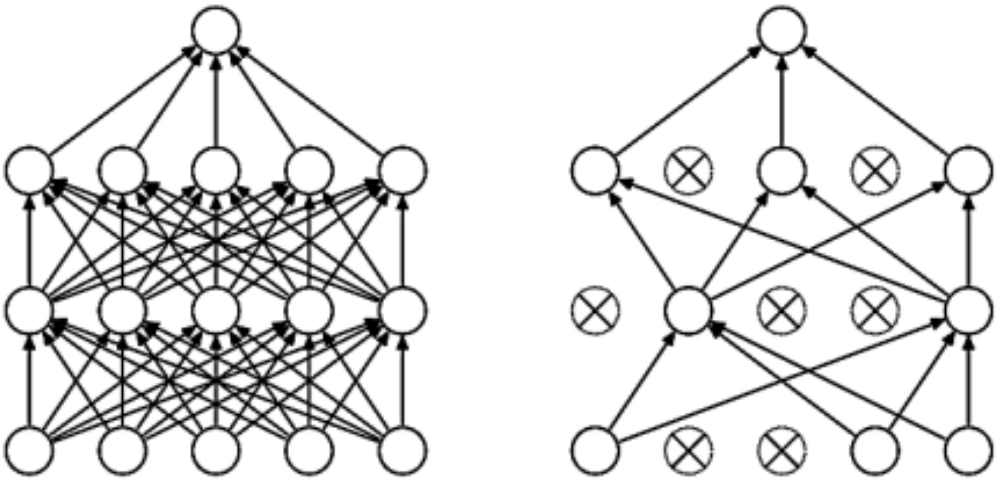
1. Επαύξηση δεδομένων εκπαίδευσης- Data Augmentation . Η τεχνική αυτή επεξεργάζεται τα δεδομένα εκπαίδευσης με διάφορα όπως:

   - Περιστροφή της εικόνας σε κάποιο διάστημα μοιρών
   - Τυχαίο κόψιμο
   - Μετακίνηση ως προς τους 2 άξονες
   - Κύλιση

   Τα παραπάνω δίνουν μεγαλύτερο αριθμό δεδομένων για εκπαίδευση και είναι δυνατό να βοηθήσουν ιδιαίτερα σε περιπτώσεις όπου αυτά είναι περιορισμένα [22]. Η επαύξηση των δεδομένων όμως δεν επιλύει το πρόβλημα που δημιουργείται όταν το δίκτυο καλείται να ταξινομήσει δεδομένα που δεν έχει συναντήσει προηγουμένως.

2. Απόσυρση- Dropout , η τεχνική αυτή κατά την περίοδο της εκπαίδευσης βγάζει εκτός του δικτύου κάποιους από τους νευρώνες. Αυτή η ποσότητα καθορίζεται ως ένα ποσοστό και

μπορεί να κάνει το δίκτυο επί της ουσίας να εκπαιδευτεί πάνω σε διαφορετικές αρχιτεκτονικές παράλληλα, βγάζονταις κάθε φορά κάποιες παραμέτρους του εκτός της διαδικασίας.



Σχήμα 2.23: Δίκτυο πριν και μετά την εφαρμογή της τεχνικής Dropout [23]

3. Batch Normalization,στην τεχνική αυτή γίνεται κανονικοποίηση των δεδομένων που εισέρχονται στο τεχνητό νευρωνικό δίκτυο ανά παρτίδες. Αυτή η κανονικοποίηση διαμορφώνει έτσι τις εισόδους ώστε να έχουν μηδενικό μέσο όρο και μοναδιαία τυπική απόκλιση [23].

4. Transfer Learning- Μεταφορά Μάθησης: Με την πρόοδο της τεχνολογίας σε ό,τι αφορά την υπολογιστική ισχύ, έγινε δυνατή η εκπαίδευση Deep Neural Networks σε μεγάλο όγκο δεδομένων. Αυτά τα δίκτυα που έχουν ήδη εκπαιδευτεί έχουν κάποια βάρη που χρησιμοποιούνται ώστε να αρχικοποιηθούν αυτά σε άλλα δίκτυα και να χρησιμοποιηθούν σε άλλες εφαρμογές. Ουσιαστικά, το νέο δίκτυο με παρόμοια δομή δανείζει τις αρχικές του παραμέτρους, μεταφέροντας αυτά τα οποία έχει μάθει και έτσι το νέο δίκτυο προσαρμόζεται ευκολότερα και πιο γρήγορα.

2.4.10 Αξιολόγηση Αποτελεσμάτων

Στην παρούσα ΔΕ είναι σημαντικό να εξεταστεί η ακρίβεια με την οποία ταξινομούνται αυτόματα τα δείγματα και να εξαχθούν σημαντικά στατιστικά στοιχεία για το αποτέλεσμα. Για τον λόγο αυτό θα χρησιμοποιηθούν τα παρακάτω:

1. Πίνακας σύγχυσης-Confusion Matrix: Ο πίνακας αυτός έχει μέγεθος $N * N$, με N να είναι ο αριθμός των κλάσεων που πρόκειται να ταξινομηθούν από το μοντέλο. Κάθε κελί του πίνακα περιέχει τον αριθμό των δειγμάτων τα οποία εκτιμήθηκαν ότι ανήκουν σε μια κατηγορία έστω k και ανήκουν στην αληθή κατηγορία m. Όταν ισχύει $k = m$ τότε η πρόβλεψη του μοντέλου είναι σωστή και όπως είναι εμφανές αποτυπώνεται πάνω στην διαγώνιο του πίνακα. Εκτός από τα σωστά ταξινομημένα δείγματα, διακρίνονται και άλλες δύο κατηγορίες πρόβλεψης και αυτές είναι τα Ψευδώς Αληθή(False Positive) και τα Ψευδώς Λανθασμένα(False Negative). Τα μεν πρώτα τοποθετούνται κάτω από την διαγώνιο και τα δεύτερα επάνω από αυτή.

2. F-Score:

$$\frac{2 * TP}{2 * TP + FP + FN} \tag{2.25}$$

Είναι ένας συνδυαστικός τρόπος να απεικονιστεί η επίδοση του μοντέλου, αφού λαμβάνει υπόψη τόσο την ανάκληση, όσο και την ακρίβεια.

3. Accuracy:

$$\frac{TP + TN}{TP + TN + FP + FN} \tag{2.26}$$

Δίνει την συνολική εικόνα του μοντέλου, δηλαδή πόσο πετυχημένο ήταν στην ταξινόμηση τόσο των αληθών όσο και των ψευδών κλάσεων.

4. Recall:

$$\frac{TP}{TP + FN} \tag{2.27}$$

Με τον υπολογισμό της ανάκλησης, δίνεται εικόνα των σωστών προβλέψεων σε σχέση με το σύνολο των σωστών που θα ήταν δυνατό να γίνουν.

5. Precision:

$$\frac{TP}{TP + FP} \tag{2.28}$$

Δίνει την ακρίβεια του μοντέλου με βάση τα στοιχεία που ταξινομήθηκαν σωστά σε σχέση με αυτά που ταξινομήθηκαν λάθος.

2.5 Πείραμα Εξαναγκασμένης Κολύμβησης-FST

Η δοκιμασία της εξαναγκασμένης κολύμβησης, είναι ένα πείραμα κατά το οποίο αξιολογείται η επίδραση αντικαταθλιπτικών φαρμάκων σε μύες και επιμύες. Συγκεκριμένα, για το πείραμα αυτό το πρωτόκολλο που προτάθηκε αρχικά όριζε τα παρακάτω σύμφωνα με την [50]:

1. Τοποθετείται νερό μέσα σε διαφανείς κυλίνδρους με ύψος 15 εκατοστά και σε θερμοκρασία 25 βαθμών Κελσίου.

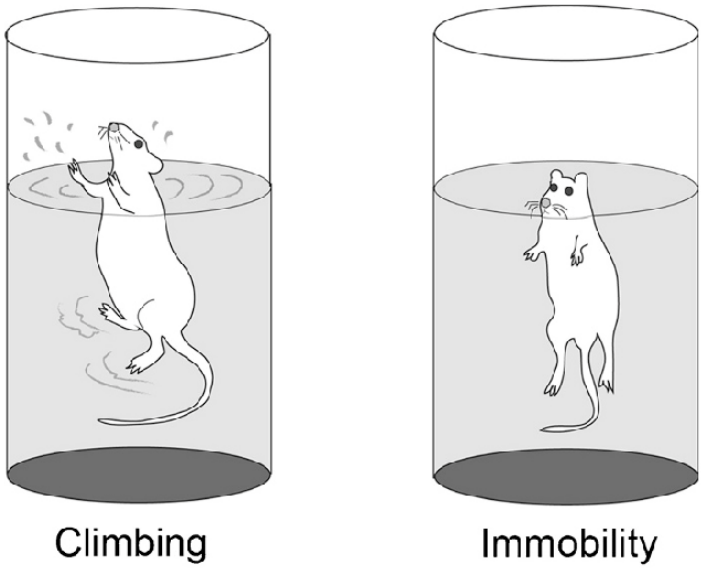
2. Το πειραματόζωο τοποθετείται μέσα στον κύλινδρο για 2 με 3 λεπτά, στην αρχική φάση της δοκιμασίας. Η αρχική περίοδος έχει ως στόχο την εξοικείωση του επιμύ με τον χώρο του πειράματος. Στην συνέχεια, τα πειραματόζωα απομακρύνονται από τον κύλινδρο.

3. Την επόμενη ημέρα της δοκιμασίας, οι επιμύες, τοποθετούνται ξανά στον κύλινδρο και το τελικό πείραμα λαμβάνει χώρα, με σκοπό την ποσοτική μέτρηση της κινητικότητας του εκάστοτε πειραματόζωου.

Το πρωτόκολλο που αναφέρθηκε, σχεδιάστηκε ώστε να μετρηθεί η κίνηση και η ακινησία του επιμύ. Σύμφωνα με τον Porsolt στις [49, 50] με το παραπάνω πείραμα μπορούν να εξαχθούν σημαντικά συμπεράσματα για την συμπεριφορά των πειραματόζωων πριν και μετά την λήψη αντικαταθλιπτικών φαρμάκων. Όταν ο επιμύς δεν έχει λάβει κάποια φαρμακευτική ουσία, όπως αναστολείς της μονοαμινοξειδάσης MAO η διάρκεια της ακινησίας είναι μεγαλύτερη από ότι σε σχέση με την περίπτωση που έχει χορηγηθεί μια τέτοια ουσία.

Το παραπάνω πρωτόκολλο, βελτιώθηκε με την [51] όπου συμπεριλήφθηκαν περισσότερες κατηγορίες ειδικά σε αυτές που χαρακτηρίζονται ως ενεργητικές με στόχο την μελέτη επιπλέον ουσιών όπως νοραδρενεργικών φάρμακών και σεροτονινεργικών αντικαταθλιπτικών. Σε αυτές διαχωρίστηκαν οι κινήσεις της αναρρίχησης και της κολύμβησης, με τις κατηγορίες που καταγράφονται τελικά να είναι:

1. Αναρρίχηση-Climbing: Κατά την κίνηση αυτή, ο επιμύς κινείται με το σώμα κάθετο κοντά στα τοιχώματα του κυλίνδρου και με κινήσεις των άκρων προσπαθεί να διαφύγει από την δεξαμενή.
2. Κολύμβηση-Swimming: Το πειραματόζωο κινείται σε οριζόντια θέση με κίνηση των άκρων του.
3. Ακινησία-Immobility: Κατά την ακινησία, ο επιμύς παραμένει τελείως ακίνητος.
4. Απότομο τίναγμα του κεφαλιού-Head Shake: Πρόκειται για μια ποιοτική κίνηση, κατά την οποία το πειραματόζωο τινάζει το κεφάλι του.
5. Κατάδυση-Diving: Είναι μια χαρακτηριστική κίνηση κατά την οποία ο επιμύς εκτελεί κατάδυση κάτω από την επιφάνεια του νερού στην δεξαμενή.



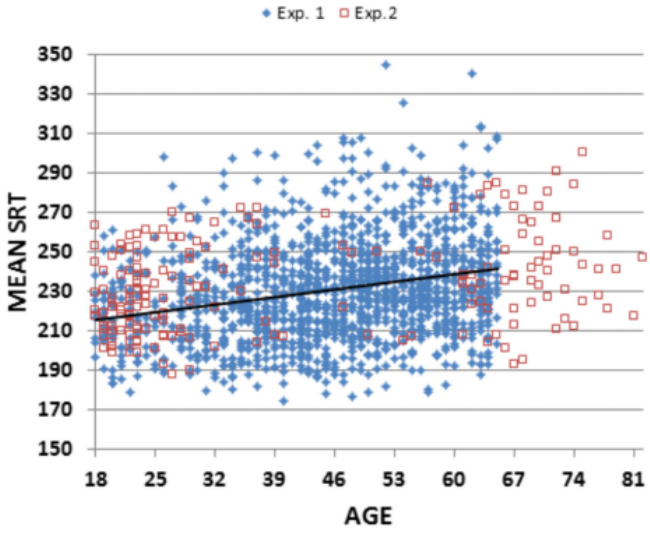
Σχήμα 2.24: Απεικόνιση του πειράματος εξαναγκασμένης κολύμβησης με τις κατηγορίες Αναρρίχησης και Ακινησίας. [48]

2.6 Χρόνος Αντίδρασης Παρατηρητή

Στην επιβλεπόμενη ταξινόμηση, η ύπαρξη ενός σετ δεδομένων που είναι πλήρως και αληθώς ταξινομημένα είναι ζωτικής σημασίας για την σωστή λειτουργία του μοντέλου. Στην περίπτωση των

εικόνων που έχουν μόνο χωρικές διαστάσεις, η ταξινόμηση για την δημιουργία ενός σετ δεδομένων που ανταποκρίνεται στην πραγματικότητα στο σύνολό του, δεν είναι τόσο απαιτητικό. Όλα όμως αλλάζουν, όταν το σετ δεδομένων που ταξινομείται για την δημιουργία του αρχικού Training Dataset αφορά δεδομένα που εξελίσσονται μέσα στον χρόνο όπως τα βίντεο. Εδώ δεν είναι δυνατό να παραληφθεί ο παράγοντας του μεγέθους του Dataset που θα ταξινομηθεί από τους εκάστοτε ειδικούς, καθώς η κόπωση του παρατηρητή αναμένεται να έχει επίδραση στην επίδοση της χειροκίνητης ταξινόμησης. Επίσης, εάν γίνει η υπόθεση ότι η οι παρατηρητές είναι έμπειροι και δεν πρόκειται να επηρεαστούν από κάποιον εξωτερικό παράγοντα, είναι σημαντικό να ληφθεί υπόψη ο παράγοντας του χρόνου αντίδρασης κατά την ταξινόμηση των δειγμάτων.

Η μελέτη [57] δίνει σημαντικά στοιχεία για τον χρόνο αντίδρασης, ξεκινώντας ορίζοντάς τον ως εξής: Simple Reaction Time είναι ο ελάχιστος χρόνος που χρειάζεται ώστε να υπάρξει αντίδραση σε ένα ερέθισμα και πρόκειται για έναν τρόπο μέτρησης της ταχύτητας με την οποία επεξεργάζεται το ερέθισμα που δόθηκε. Ο χρόνος αντίδρασης, σύμφωνα με την παραπάνω δεν επηρεάζεται από το φύλο του παρατηρητή του ερεθίσματος ή το επίπεδο μόρφωσης, αλλά σημαντικό ρόλο φαίνεται να παίζει η ηλικία του παρατηρητή, με την αύξηση να υπολογίζεται σε 0.50 second/year.



Σχήμα 2.25: Τα αποτελέσματα των πειραμάτων που δείχνουν την αύξηση του χρόνου αντίδρασης ανάλογα με την ηλικία του παρατηρητή σε 2 διαφορετικά πειράματα [57].

Η αντίδραση του ταξινομητή, είναι σημαντική κατά την διαδικασία ταξινόμησης βίντεο, ιδιαίτερα όταν πρόκειται για βίντεο με framerate μεγαλύτερο των 100 fps . Στην περίπτωση της μελέτης, είναι δυνατό να ληφθεί ως αποδεκτός χρόνος αντίδρασης τα 230ms για τους παρατηρητές που ταξινόμησαν τα βίντεο. Επιπλέον, η ύπαρξη του χρόνου αυτού, είναι δυνατό να παράξει περιοχές στην διάρκεια του βίντεο που η ταξινόμηση δεν είναι σωστή λόγω καθυστέρησης στην αντίδραση του παρατηρητή. Λαμβάνοντας υπόψη τα παραπάνω είναι επιβεβλημένη η μελέτη μέσω πειραμάτων της επίδρασης του Simple Reaction Time στην ταξινόμηση των βίντεο.

Κεφάλαιο 3

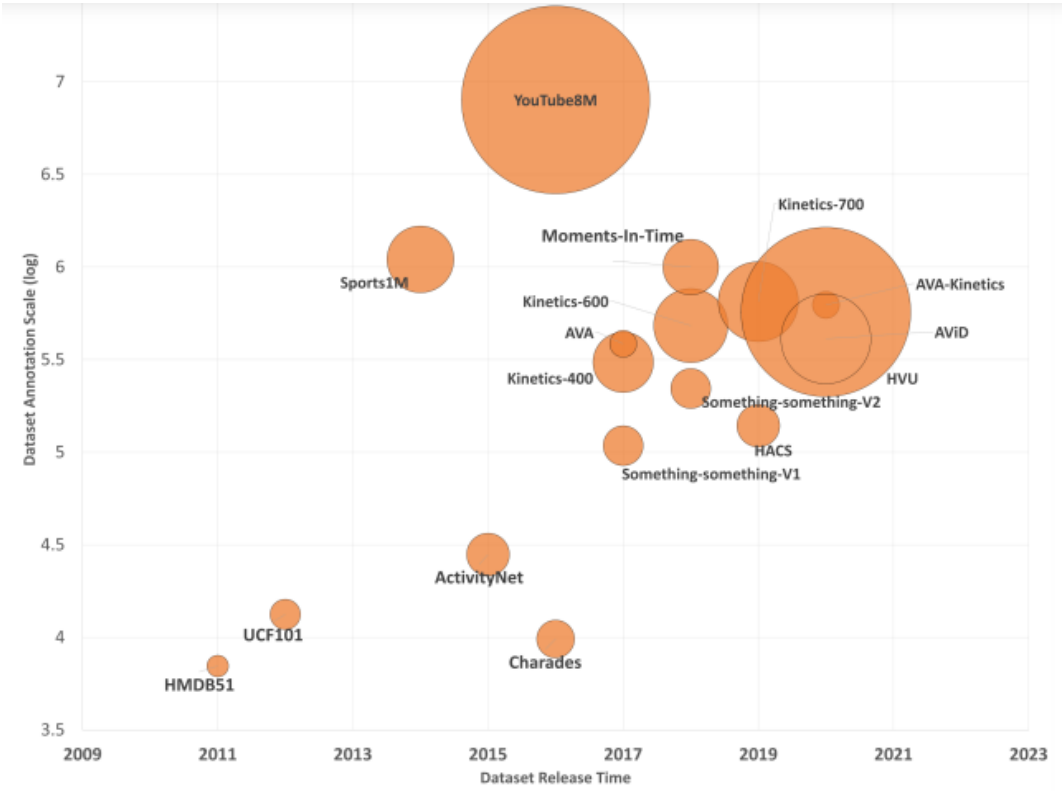
Αρχιτεκτονικές Αναγνώρισης Δράσης και Benchmark Datasets

Σε αυτό το κεφάλαιο γίνεται αναλυτική παρουσίαση των αρχιτεκτονικών που χρησιμοποιούνται για την αναγνώριση δράσεων σε βίντεο. Ο τομέας αυτός έχει γνωρίσει μεγάλη ανάπτυξη ιδιαίτερα τα τελευταία χρόνια με την ανάπτυξη ολο και περισσότερο ισχυρών υπολογιστικών συστημάτων. Αυτά έχουν την δυνατότητα να επεξεργάζονται δεδομένα με πολύ μεγαλύτερη ταχύτητα και συνεπώς να επιτρέπουν την εκπαίδευση δικτύων με αρκετά εκατομμύρια παραμέτρους. Επίσης, σημαντικό ρόλο έχουν διαδραματίσει οι κάρτες γραφικών (GPU) , που έχουν την δυνατότητα να επιταχύνουν δραματικά τον χρόνο εκπαίδευσης, εκτελώντας παράλληλα πολλούς υπολογισμούς. Επίσης, παρατίθενται τα κυριότερα Datasets που χρησιμοποιούνται για την αξιολόγηση της επίδοσης των εκάστοτε αρχιτεκτονικών.

3.1 Datasets για Προβλήματα Αναγνώρισης Δράσεων

Τα τελευταία χρόνια υπάρχει σημαντική τάση για ανάπτυξη ποιοτικών σετ δεδομένων που να είναι ταυτόχρονα μεγάλα σε έκταση. Η ενασχόληση των ερευνητών με τον τομέα της αναγνώρισης δράσεων είχε ως αποτέλεσμα την δημοσίευση σετ δεδομένων που περιέχουν βίντεο με πολλές χιλιάδες κλάσεις. Παρακάτω δίνεται σε εικόνα η εξέλιξη των δεδομένων τα τελευταία χρόνια. Παρακάτω δίνεται διαγραμματικά το μέγεθος και το έτος δημοσίευσης του κάθε σετ δεδομένων, καθώς και τα βασικά τους χαρακτηριστικά [25]. Παρακάτω δίνονται σύντομα κάποια από τα σετ αυτά και τα χαρακτηριστικά τους.

- HMDB51 : Αυτό το σετ δεδομένων προέρχεται από ταινίες και δημιουργήθηκε το 2011 . Περιλαμβάνει 6.849 clip και αφορούν 51 κατηγορίες δράσης.
- UCF 101: Είναι η βελτιωμένη έκδοση του σετ δεδομένων UCF50 και περιλαμβάνει video με 101 κατηγορίες δράσεων.
- Sports1M : Δημοσιεύτηκε το 2014 και αποτελεί το πρώτο σετ δεδομένων που έχει 1 εκατομμύριο βίντεο από το Youtube και αντιπροσωπεύουν 487 κατηγορίες δράσης.



Σχήμα 3.1: Αναπαράσταση των σετ δεδομένων που έχουν δημοσιευτεί ανά χρονιά [25]

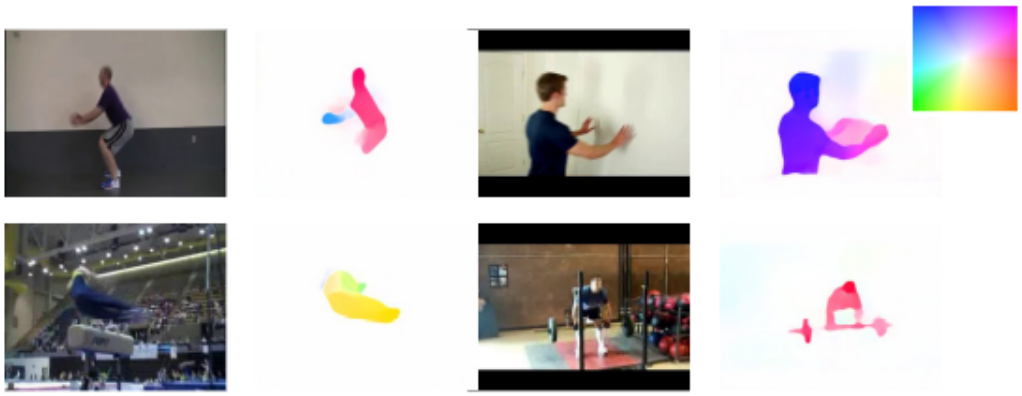
- ActivityNet : Δημοσιεύτηκε το 2015 και είναι το πρώτο από μια σειρά Datasets με συνεχείς βελτιώσεις, που τα νεότερα περιέχουν βίντεο χωρισμένα σε (split) 10.024 training videos, 4.926 validation videos και τέλος, 5.044 test videos.
- YouTube8M : Όπως δηλώνει και το όνομα του σετ αυτού, αποτελείται από 8 εκατομμύρια βίντεο και πηγή τους είναι το Youtube.Χαρακτηρίζεται από την μεγάλη συνολική διάρκεια των βίντεο καθώς και από τον μεγάλο αριθμό διαφορετικών κατηγοριών δράσης, που είναι συνολικά 3.862 .

3.2 Σύγχρονες Αρχιτεκτονικές Αναγνώρισης Δράσεων με Τεχνητά Νευρωνικά Δίκτυα

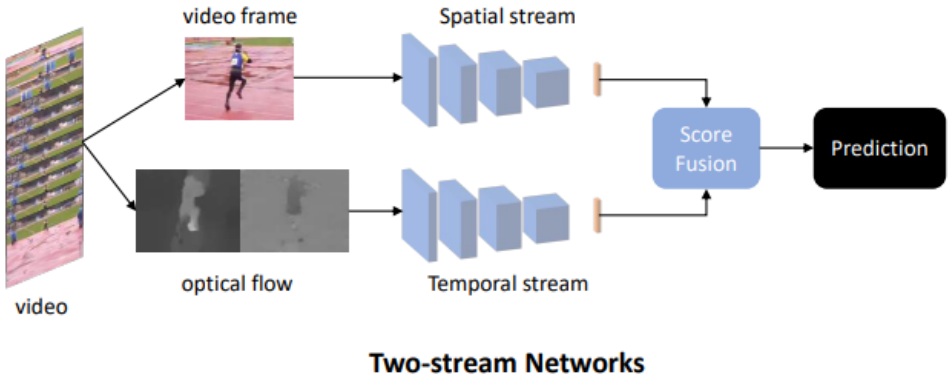
3.2.1 Αρχιτεκτονική Διπλής Ροής

Η χρήση τεχνητών νευρωνικών δικτύων άλλαξε τα δεδομένα στην έρευνα για την αναγνώριση δράσεων από το 2014. Τότε, με την [26] προτάθηκε η αρχιτεκτονική με διπλή ροή δεδομένων. Βασική αρχή για την κατασκευή αυτής της αρχιτεκτονικής, είναι η υπόθεση ότι θα είναι ένα ανάλογο της λειτουργίας της περιοχής του εγκεφάλου για την όραση, που χωρίζεται σε δύο τμήματα: αυτό που αναγνωρίζει το αντικείμενο και αυτό που αναγνωρίζει την κίνηση.

Ιδιαίτερη σημασία έχει η χρήση εικόνων οπτικής ροής στο δίκτυο αυτό αλλά και σε άλλα που εκμεταλλεύονται τις ιδιότητες της. Η οπτική ροή, μπορεί να περιγράψει με αρκετά καλό αποτέλεσμα το πρότυπο της κίνησης διάφορων αντικειμένων, επιφανειών αλλά και άλλων χωρικών χαρακτηριστικών όπως γωνίες [25]. Οι εικόνες οπτικής ροής είναι σε θέση να απομονώνουν το παρασκήνιο της εικόνας, αφού δεν παρουσιάζει κάποια κίνηση και αυτό συμβάλλει καθοριστικά στην άυξηση του ποσοστού επιτυχίας του δικτύου.



Σχήμα 3.2: Εικόνες οπτικής ροής και αντιστοιχία με την κατεύθυνση [25]



Σχήμα 3.3: Αρχιτεκτονική διπλής ροής[25]

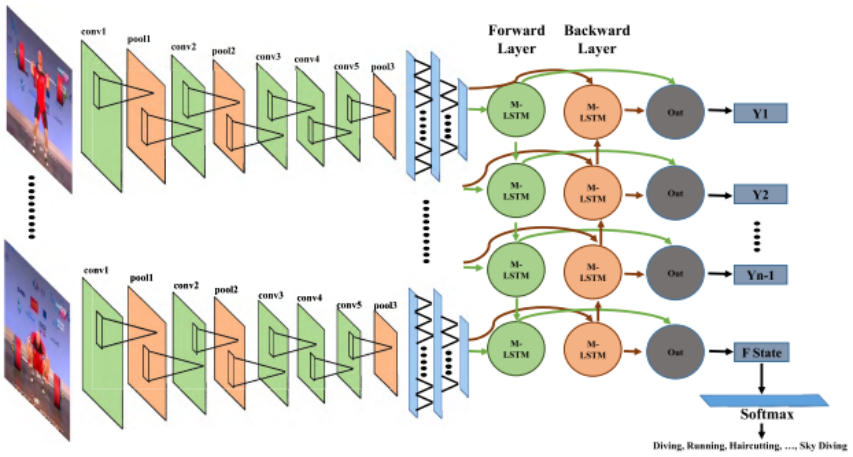
Όπως δίνεται στο σχήμα 3.3, το δίκτυο αυτό αποτελείται από δύο χωριστά μέρη στα οποία ρέει η πληροφορία και επεξεργάζεται. Ειδικότερα, το δίκτυο δέχεται ως είσοδό του frames και εικόνες οπτικής ροής. Το δίκτυο λαμβάνει ως είσοδο εικόνες αλλά και ίδιες που αφορούν την προϋπολογισμένη οπτική ροή. Στην συνέχεια, οι πρώτες αφορούν την αναγνώριση χωρικών χαρακτηριστικών και οι εικόνες οπτικής ροής την χρονική διάσταση του προβλήματος. Όταν επεξεργαστούν, χωρίζουν την δράση που αναπαριστάται στα δεδομένα εισόδου σε κλάσεις λαμβάνοντας υπόψη την μέση τιμή της πρόβλεψης από τα 2 κλαδιά του δικτύου. Η εμφάνιση του δικτύου αυτού ήταν πάρα πολύ σημαντική, διότι αποτελεί τον συνδετικό κρίκο ανάμεσα στις παλαιότερες τεχνικές αναγνώρισης δράσης και στην χρήση νευρωνικών δικτύων για την ίδια εργασία. Επιπλέον δείχθηκε ότι η χρονική διάσταση είναι πολύ σημαντική στην αναγνώριση δράσεων και δίνει την δυνατότητα στο τεχνητό νευρωνικό δίκτυο να ταξινομεί δράσεις με αρκετά καλό ποσοστό επιτυχίας. Η πρώτη προσπάθεια για την κατασκευή ενός δικτύου που να περιέχει 2 ροές δεδομένων ήταν αρκετά σημαντική, αλλά γρήγορα φάνηκε η ανάγκη για βελτίωση. Οι βελτιώσεις περιλαμβάνουν τα παρακάτω:

1. Η πρώτη απόπειρα για κατασκευή δικτύου 2 ροών χρησιμοποιούσε μια σχετικά ΄ρηχή αρχιτεκτονική΄ με μικρό αριθμό συνελικτικών επιπέδων. Προτάθηκε μια πιθανή βελτίωση αυτής της μορφής δικτύου με την χρήση περισσότερων επιπέδων, που όμως δεν ήταν σίγουρο ότι θα έφερνε το επιθυμητό αποτέλεσμα. Αυτό συμβαίνει αφού όπως αναφέρθηκε και παραπάνω υπάρχει το πρόβλημα της Υπερπροσαρμογής. Σημαντική ήταν η προσφορά της δημοσίευσης [28], με την οποία ο Wang et al , καταδεικνύει τρόπους για την καλύτερη εκπαίδευση και την αποφυγή της υπερπροσαρμογής, όπως:
   - cross-modality initialization
   - synchronized batch normalization
   - corner cropping
   - multi-scale cropping
   - Τέλος, η χρήση μεγαλύτερου ποσοστού απόσυρσης (Dropout)

2. Βελτίωση του σταδίου συνένωσης των 2 ροών: Στο σχήμα 3.3, δίνεται η βασική αρχιτεκτονκή των Τεχνητών Νευρωνικών Δικτύων 2 Ροών. Όπως φαίνεται ένα από τα βασικά στάδια της διαδικασίας της επεξεργασίας των δεδομένων είναι η συνένωση των 2 ροών ώστε να παρθεί η τελική ταξινόμηση σε κλάσεις. Για την βελτίωση αυτού του σταδίου έχουν γίνει πολλές σημαντικές έρευνες [31] ώστε να είναι όσο το δυνατόν αποτελεσματικότερη. Η πρώτη προσέγγιση στο πρόβλημα της βέλτιστης επιλογής του σταδίου συνένωσης ήταν η συνένωση μετά από την επεξεργασία κάθε ροής ξεχωριστά. Στα [32] και [33] αναφέρεται ότι η μέθοδος που χρησιμοποιείται στα προηγούμενα είναι μεν ικανοποιητική, αλλά το καλύτερο αποτέλεσμα θα μπορούσε να επιτευχθεί με περισσότερες συνδέσεις ανάμεσα στα 2 κλαδιά του δικτύου. Με τον τρόπο αυτό ξεπερνιούνται προβλήματα όπως αυτό του Vanishing Gradient Descent , που αναφέρεται στην δυσκολία στην εκπαίδευση των δικτύων με πολύ βαθιά αρχιτεκτονική. Οι βελτιώσεις αυτές οδήγησαν αναπόφευκτα στην αρχιτεκτονική του ResNet [34].

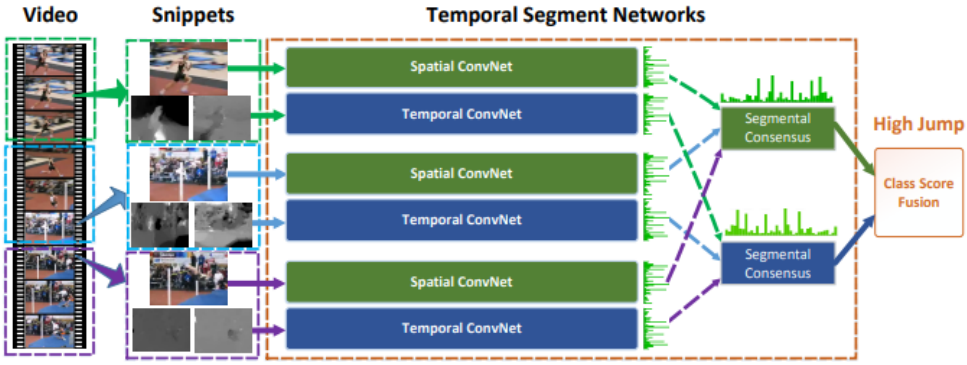
3. Αναδρομικά Νευρωνικά Δίκτυα-Recurrent neural networks : Τα δίκτυα 2 ροών δέχθηκαν αρκετές βελτιώσεις, που στόχευαν στην ενίσχυση της ικανότητάς τους να αναγνωρίζουν χαρακτηριστικά που διαθέτουν και χρονική διάσταση. Για τον λόγο αυτό χρησιμοποιήθηκαν δίκτυα Long Short Term Memory στους κλάδους του, με τελικό στάδιο την ένωση αυτών όπως αναφέρθηκε παραπάνω. Η έξοδος των συνελίξεων από το κάθε ένα κλάδο του αρχικού δικτύου μπαίνει ως είσοδος σε ένα δίκτυο Long Short Term Memory . Οι αρχικές δοκιμές με απλά δίκτυα της παραπάνω συνδυαστικής μορφής έδειξαν μικρή βελτίωση, αλλά με την δημοσίευση [35] έγινε σημαντική πρόοδος στον τομέα της αναγνώρισης δράσεων με την χρήση της συγκεκριμένης αρχιτεκτονικής.

4. Κατάτμηση στην χρονική διάσταση: Ένα μειονέκτημα που παρατηρήθηκε στις παραπάνω μορφές αρχιτεκτονικών ήταν η χαμηλή επίδοση σε ό,τι αφορά την αναγνώριση χαρακτηριστικών που είναι σημαντικά για την αναγνώριση δράσης και εκτυλίσσονται σε μεγάλη χρονική διάρκεια. Για να ξεπεραστεί αυτό το πρόβλημα, προτάθηκε από τον Wang et al [29] η εξής τεχνική: το βίντεο χωρίζεται σε μικρότερα με την ιδιότητα του ότι όλα τα νέα μέρη του είναι ομοιόμορφα κατανεμημένα στην χρονική διάρκειά του και στην συνέχεια με την τεχνική αυτή



Σχήμα 3.4: Συνδυαστική αρχιτεκτονική 2 ροών και LSTM [35]

λαμβανεται ένα frame από κάθε τμήμα και εξάγονται χαρακτηριστικά που κατανέμονται σε όλη την χρονική διάρκεια του βίντεο. Στην συνέχεια το τελικό αποτέλεσμα λαμβάνεται μέσα από μια διαδικασία όπως Max pooling ή Average pooling.

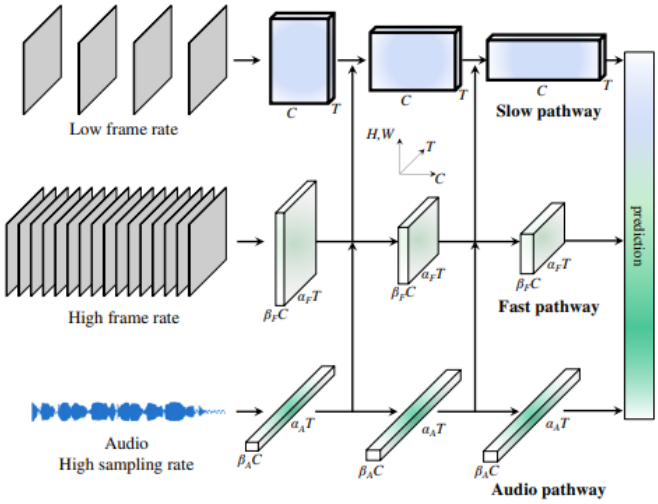


Σχήμα 3.5: Αναπαράσταση αρχιτεκτονικής TSN [29]

Η τεχνική αυτή παρουσιάζει ένα πάρα πολύ σημαντικό πλεονέκτημα, διότι επιτρέπει την επεξεργασία βίντεο και την ταξινόμηση δράσεων σε πολύ μεγαλύτερα χρονικά βίντεο και πλέον σχεδόν όλες οι προσεγγίσεις αρχιτεκτονικών 2 ροών ακολουθούν αυτήν την τεχνική.

5. Δίκτυα πολλαπλών ροών-Multi-stream networks: Μια ακόμα βελτίωση στα δίκτυα αυτά ήταν ήρθε από την διαπίστωση ότι η αναγνώριση της δράσης σε ένα βίντεο έρχεται και μέσα από την στάση του σώματος αλλά και από τον ήχο που συνοδεύει την δράση και το παρασκήνιο. Αρχικά προτάθηκε η αρχιτεκτονική Inflated 3D [37], με την παραπάνω να πετυχαίνει αξιοσημείωτα αποτελέσματα. Επιπλέον, σημαντική ήταν η βελτίωση που ήρθε με την εισαγωγή στην διαδικασία της αναγνώρισης δράσης και της πληροφορίας του παρασκηνίου [30]. Ο ήχος είναι ακόμα ένας παράγοντας που είναι σε θέση να προσφέρει αρκετή πληροφορία για την επιτυχή αναγνώριση της δράσης σε βίντεο. Τα δίκτυα αυτά ονομάζονται Audio Slow Fast [36].

   Η συγκεκριμένη τεχνική κάνει χρήση αρχικά 2 ροών που αφορούν το οπτικό μέρος της



Σχήμα 3.6: Αρχιτεκτονική Audiovisual Slow fast Network [36]

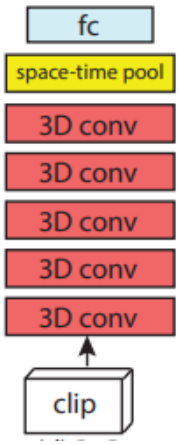
πληροφορίας για την αναγνώριση της δράσης όπως και τα αρχικά, με την διαφορά ότι η δειγματοληψία των frames γίνεται με διαφορετικό ρυθμό στον κάθε ένα κλάδο και έναν ακόμα κλάδο που γίνεται αξιοποίηση της ηχητικής πληροφορίας με την μορφή log-mel-spectrogram , σε δισδιάσταση μορφή, με τον ένα άξονα να αναπαριστά τον χρόνο και τον άλλο την συχνότητα του σήματος. Στο δίκτυο αυτό είναι εμφανείς και οι ενδιάμεσες συνδέσεις ανάμεσα στις ροές, ώστε να ενσωματωθεί η ηχητική πληροφορία στις άλλες 2 ή στο τελικό για ταξινόμηση της δράσης.

3.2.2 Συνελικτικά Νευρωνικά Δίκτυα 3D

Στην προηγούμενη παράγραφο έγινε αναφορά στις σημαντικότερες αρχιτεκτονικές που αφορούν την αναγνώριση δράσεων με την χρήση Τεχνητών Νευρωνικών Δικτύων. Στην πλειοψηφία τους, όλα αφορούν αρχιτεκτονικές που για να αναγνωρίσουν δράσεις με χρονική διάσταση χρησιμοποιούν διάφορες τεχνικές ώστε να είναι δυνατή η αναγνώριση με την χρήση συνελίξεων σε 2 διαστάσεις. Ο συνδετικός κρίκος ανάμεσα στα δύο στάδια εξέλιξης ήταν η αρχιτεκτονική Inflated 3D . Για την περαιτέρω βελτίωση της επίδοσης των συνελικτικών δικτύων τριών διαστάσεων, οι έρευνες συνεχίστηκαν με σημαντικότερες, τις παρακάτω αρχιτεκτονικές και τεχνικές:

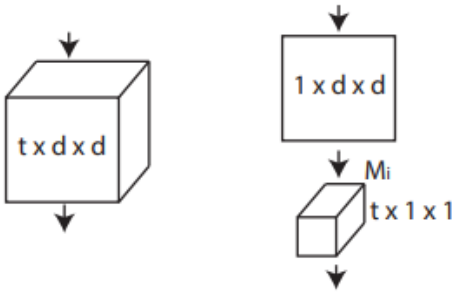
1. Η πρώτη προσπάθεια για την αναγνώριση δράσεων με αυτόν τον τύπο δικτύου έγινε αρχικά με την προσαρμογή του δικτύου ResNet (2D) σε ResNet3D.Η θεωρία των ερευνητών ήταν ότι τα χαρακτηριστικά της δράσης θα ήταν δυνατόν να αναγνωριστούν μέσα από την αντικατάσταση όλων των συνελίξεων που μέχρι πρότινος ήταν 2 διαστάσεων (Σχήμα 3.7) με αντίστοιχες τριών διαστάσεων, προσθέτοντας και την χρονική [38].

2. Ομαδοποίηση των 2D CNN και 3D CNN: Τα δίκτυα με συνελίξεις 2 διαστάσεων, έχουν ήδη δοκιμαστεί και έχουν πετύχει εξαιρετικές επιδόσεις σε πολλά σετ δεδομένων. Είναι σημαντικό να τονιστεί η σημασία των σετ δεδομένων ειδικά για τα προηγούμενα δίκτυα, που είναι ευρέως μελετημένα. Η έρευνα, επικεντρώθηκε στην ενοποίηση της έρευνας για τα δισδιάστατα δίκτυα και τα νεότερα των τριών διαστάσεων. Εξαιρετικής σημασίας ήταν η αρχιτεκτονική R(2+1)D.Η τεχνική αυτή συνδέει τις δύο αρχιτεκτονικές χωρίζοντας την μια συνέλιξη σε τρεις διαστάσεις, σε 2 μικρότερες μειώνοντας το υπολογιστικό κόστος και



Σχήμα 3.7: Σχηματική απεικόνιση αρχιτεκτονικής ResNet3D(έχουν παραληφθεί οι ενδιάμεσες ενώσεις των επιπέδων)[40]

διευκολύνοντας την εκπαίδευση. Ειδικότερα, τα φίλτρα έχουν την μορφή 1×3×3 για την χωρική διάσταση και 3×1×1 για την χρονική.

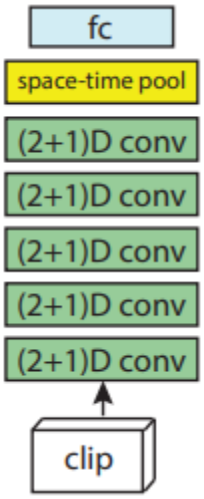


Σχήμα 3.8: Παραδείγματα συνελίξεων τριών διαστάσεων αριστερά και την ισοδύναμη 1×d×d + t×1×1 στα δεξιά. [40]

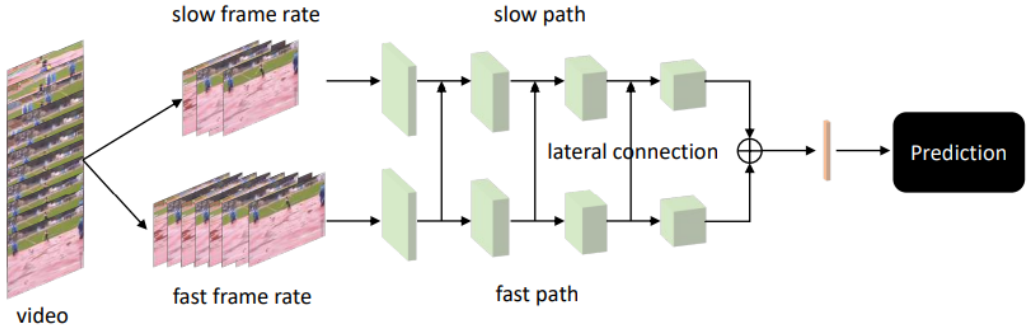
Η αρχιτεκτονική R(2+1)D (Σχήμα 3.8, 3.9) έχει δώσει εξαιρετικά αποτελέσματα που σε συνδυασμό με την εύκολη προσαρμογή της σε διάφορες εργασίες για ταξινόμηση δράσεων, την κάνει ιδανική για πειραματισμό και μελέτη. Ακόμα, σύμφωνα με την [40] δίνει όχι μόνο μικρότερο testing error, αλλά και training error.

Ένα άλλο σημαντικό πλεονέκτημα της αρχιτεκτονικής αυτής είναι ότι παρουσιάζει μικρότερο Training Error σε σχέση με αυτήν του ResNet3D ειδικά όταν το βάθος του δικτύου μεγαλώνει. Το γεγονός αυτό δηλώνει την δυνατότητα αυτής της αρχιτεκτονικής να μπορεί να βελτιστοποιηθεί ειδικά με την αύξηση του βάθους του δικτύου [40].

3. Slow Fast Networks : Η αρχιτεκτονική αυτή αναφέρθηκε και παραπάνω (Αρχιτεκτονικές 2 Ροών) για τον λόγο ότι παρόλο που υπάρχουν 2 ροές για την οπτική πληροφορία (1η ροή), υπάρχει και ακόμα μία για την ηχητική (2η ροή). Η καταταγή της σε αυτήν την ενότητα, έγκειται στο γεγονός ότι χρησιμοποιεί τα δίκτυα 2 διαστάσεων για να απλοποιήσει σε ό,τι αφορά την υπολογιστική ισχύ την αναγνώριση χωρο-χρονικών προτύπων. Με την τεχνική αυτή, λαμβάνονται δείγματα από ένα βίντεο με 2 διαφορετικούς ρυθμούς. Έτσι επιτυγχάνεται η αναγνώριση χωρο-χρονικών προτύπων με μικρότερο υπολογιστικό κόστος.

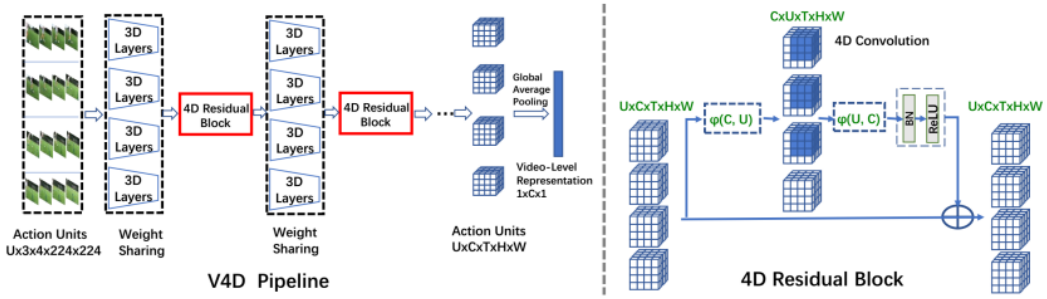


Σχήμα 3.9: Σχηματική απεικόνιση αρχιτεκτονικής R(2+1)D(έχουν παραληφθεί οι ενδιάμεσες ενώσεις των επιπέδων) [40]



Σχήμα 3.10: Απεικόνιση αρχιτεκτονικής Slowfast Network[25]

4. V4D : Η αρχιτεκτονική αυτή εισάγει ακόμα μια διάσταση στην διαδικασία της αναγνώρισης δράσεων. Ο βασικός σκοπός της διάστασης αυτής είναι να μπορέσει να μοντελοποιήσει με τον καλύτερο δυνατό τρόπο πρότυπα μέσα στην διάρκεια βίντεο. Η διάσταση που εισάγεται στην αρχιτεκτονική αυτή ονομάζεται Action Unit [56] και έχει την παρακάτω μορφή:

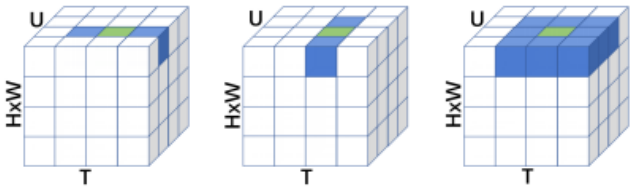


Σχήμα 3.11: Αρχιτεκτονική V4D[56]

Η παραπάνω αρχιτεκτονική μπορεί να δώσει καλύτερη αναπαράσταση των προτύπων σε ό,τι αφορά τα χρονικά πρότυπα με την χρήση της αναφερθείσας τέταρτης διάστασης. Η txtxtxt συνέλιξη εκφράζεται ως:

$$o_j^{uthw} = b_j + \sum_{c}^{C_{in}} \sum_{s=0}^{S-1} \sum_{p=0}^{P-1} \sum_{q=0}^{Q-1} \sum_{r=0}^{R-1} \mathcal{W}_{jc}^{spqr} v_c^{(u+s)(t+p)(h+q)(w+r)} \tag{3.1}$$

Στην σχέση 3.1, το b_j αντιπροσοπεύει το bias, τα SxPxQxR είναι το μέγεθος της συνέλιξης των τεσσάρων διαστάσεων, το W_i δίνει το βάρος στην θέση s,p,q,r στην αντίστοιχη θέση του φίλτρου και το αποτέλεσμα της σχέσης είναι η τιμή του εικονοστοιχείου o.



Σχήμα 3.12: Μορφή συνέλιξης txtxtxt [56]

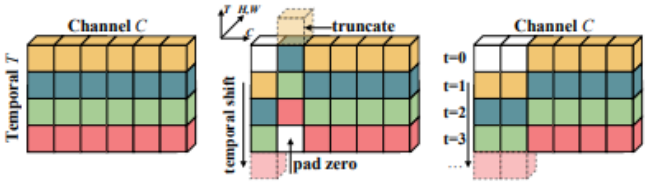
Στο Σχήμα 3.12 η διάσταση Action Unit απεικονίζεται με πράσινο χρώμα και έχουν παραληφθεί οι διαστάσεις των καναλιών εισόδου καθώς και ο αριθμός των παρτίδων.

3.2.3 Αναγνώριση Δράσεων και Υπολογιστικό Κόστος

Όλες οι παραπάνω τεχνικές αποτελούν νέες εξελίξεις στον τομέα της αναγνώρισης δράσεων μέχρι αυτήν την στιγμή, όμως με την εξέλιξη των Dataset ανακύπτει ένα σοβαρό πρόβλημα που οδηγησε τους ερευνητές να δώσουν μεγαλύτερη σημασία στην απαιτούμενη υπολογιστική ισχύ για την εκπαίδευση τέτοιων δικτύων. Επίσης, οι νέες απαιτήσεις υπό το πρίσμα του υλικού που είναι διαθέσιμο κάθε φορά για την εκτέλεση του εκάστοτε αλγόριθμου παίζουν καθοριστικό ρόλο για την ανάπτυξη νέων αρχιτεκτονικών με στόχο την αναγνώριση δράσεων. Κλείνοντας την βιβλιογραφική ανασκόπηση των αρχιτεκτονικών, παρακάτω αναφέρονται νέες τεχνικές που έχουν βασικό άξονα ανάπτυξής τους τα παραπάνω:

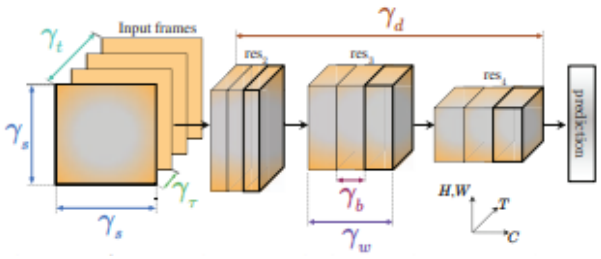
- Temporal Shift Module:Πρόκειται για μια τεχνική που στόχο έχει να μειώσει δραστικά το υπολογιστικό κόστος στην διαδικασία της αναγνώρισης δράσεων. Η ιδέα πίσω του είναι ότι

ενώ η πληροφορία εισέρχεται για επεξεργασία σε ένα τμήμα του δικτύου, το Temporal Shift Module αλλάζει την κατανομή της πληροφορίας στην χρονική διάσταση, απλά μετατοπίζοντας την είτε σε μία ή και σε 2 κατευθύνσεις, καθιστώντας το δίκτυο ικανό να διακρίνει χρονικά πρότυπα ευκολότερα και με μικρή υπολογιστική ισχύ.



Σχήμα 3.13: Σχηματική απεικόνιση του TSM.[43]

- X3D:Είναι ένα παράδειγμα αρχιτεκτονικής που βασίζεται στην ισορροπία ανάμεσα στην ακρίβεια της ταξινόμησης και της πολυπλοκότητας. Με την σταδιακή δοκιμή άλλων παραμέτρων στο δίκτυο, είναι δυνατή η βελτιστοποίησή του ώστε να μην ξεπερνά έναν συγκεκριμένο βαθμό πολυπλοκότητας αλλά και να είναι ταυτόχρονα αποτελεσματικό στον διαχωρισμό κλάσεων.



Σχήμα 3.14: Αρχιτεκτονική X3D,με τις παραμέτρους γ_i που προσαρμόζονται.[44]

Κεφάλαιο 4

Μεθοδολογία-Πειραματική Διαδικασία

Στο πρώτο μέρος της ενότητας αυτής, γίνεται περιγραφή της διαδικασίας που προηγήθηκε των πειραμάτων και αφορά την ταξινόμηση βίντεο με πειράματα προ-κλινικών μελετών σε επιμύες στην δοκιμασία εξαναγκασμένης κολύμβησης. Αυτό διότι όπως περιγράφηκε σε προηγούμενο κεφάλαιο, η επιβλεπόμενη ταξινόμηση επιτάσσει την ύπαρξη σετ δεδομένων και τα δεδομένα ground truth, δηλαδή την κατηγοριοποίηση των δειγμάτων στην κατηγορία που πραγματικά ανήκουν. Στην συνέχεια, γίνεται η αναφορά στην πειραματική διαδικασία για την αυτόματη αναγνώριση της δράσης σύμφωνα με τις κατηγορίες που έχουν βρεθεί από την χειροκίνητη ταξινόμηση με την χρήση νευρωνικών δικτύων.

4.1 Προεπεξεργασία Δεδομένων

Για την εκπαίδευση των τεχνητών νευρωνικών δικτύων, έγινε εκ νέου συλλογή δεδομένων από πειράματα εξαναγκασμένης κολύμβησης, που έλαβαν χώρα στη συνεργαζόμενη ομάδα Νευροψυχοφαρμακολογίας στο εργαστήριο Φαρμακολογίας της Ιατρικής Σχολής Αθηνών (ΕΚΠΑ). Σε προηγούμενη ΔΕ [52] είχε πραγματοποιηθεί συλλογή και ταξινόμηση βίντεο και ταξινόμηση από του ίδιους με αυτά να αφορούν πειράματα Forced Swim Test για την έρευνα [53]. Σε αυτό το σετ δεδομένων υπήρχαν συνολικά 100 βίντεο στα με 50 fps και η ταξινόμηση από τους ερευνητές πραγματοποιήθηκε με την χρήση του ελεύθερου λογισμικού Kinoscope [54]. Στο τελικό Dataset συγκεντρώθηκαν συνολικά 200 βίντεο με διάρκεια 16 ωρών από τα πειράματα που προαναφέρθηκαν. Τα βίντεο αυτά περιελάμβαναν κάθε ένα 2 επιμύες που εκτελούσαν την δοκιμασία ταυρόχρονα. Επίσης, η διάρκεια των βίντεο κυμαινόταν ανάμεσα σε 5 και 7 λεπτά. Τέλος, το framerate ήταν 25 ή 10 fps ανάλογα μα την ρύθμιση της συσκευής καταγραφής. Παρακάτω περιγράφεται η διαδικασία με την οποία έγινε η προεπεξεργασία των δεδομένων για την δημιουργία του τελικού Dataset που χρησιμοποιήθηκε για την διεξαγωγή των πειραμάτων.

4.1.1 Ταξινόμηση Βίντεο-Μορφή εκτιμήσεων

Τα επιπλέον βίντεο που συλλέχθηκαν στην δεύτερη φάση των πειραμάτων ήταν εν μέρει ταξινομημένα, για αυτό το λόγο σε συνεργασία με την Υποψήφια Διδάκτορα της Ιατρικής Σχολής Παυλίνα

Παυλίδη έγινε η ταξινόμηση και των υπολοίπων δεδομένων των πειραμάτων FST. Η διαδικασία έγινε στο γραφικό περιβάλλον του Kinoscope , με την διαδικασία να έχει ως εξής:

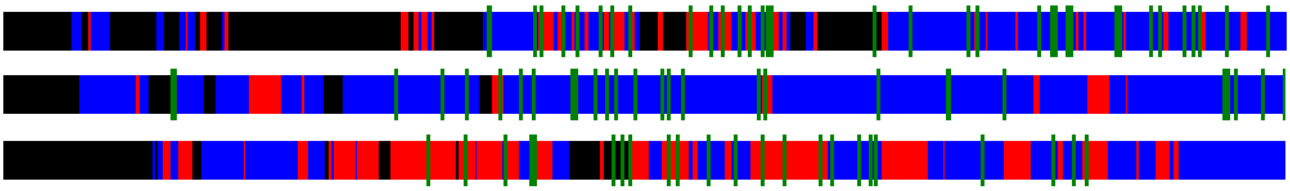
1. Εισαγωγή των στοιχείων για τον τύπο πειράματος που θα ακολουθήσει. Εδώ πρόκειται για πείραμα εξαναγκασμένης κολύμβησης FST .

2. Επιλογή του χρόνου διάρκειας του κάθε πειράματος. Ταξινομήθηκαν τα δεδομένα Test και η διάρκειά τους είναι 300 δευτερόλεπτα.

3. Εισαγωγή του ΑΑ του πειράματος στην βάση δεδομένων της εφαρμογής.

4. Ταξινόμηση της συμπεριφοράς του πειραματόζωου και καταγραφή του αποτελέσματος σε μορφή εικόνας με πλάτος 1000 εικονοστοιχεία.

Από την παραπάνω διαδικασία, εξάγεται η εικόνα που περιέχει την ταξινόμηση των κινήσεων και έχει την παρακάτω μορφή:



Σχήμα 4.1: Εικόνα με ταξινομήσεις δράσης επιμύων από λογισμικό Kinoscope

Όπως είναι φανερό η κάθε κατηγορία κίνησης του ζώου κωδικοποιείται με συγκεκριμένο χρώμα στην εικόνα που λαμβάνεται ως έξοδος. Πιο συγκεκριμένα, η αντιστοιχία έχει ως εξής:

Κατηγορία-Category	Χρώμα-Color	Κωδικός Αναγνώρισης Δράσης
Ακινησία-Immobility		0
Κολύμβηση-Swimming		1
Αναρρίχηση-Climbing		2
Τίναγμα Κεφαλής-Head Shake		3
Κατάδυση-Diving		4

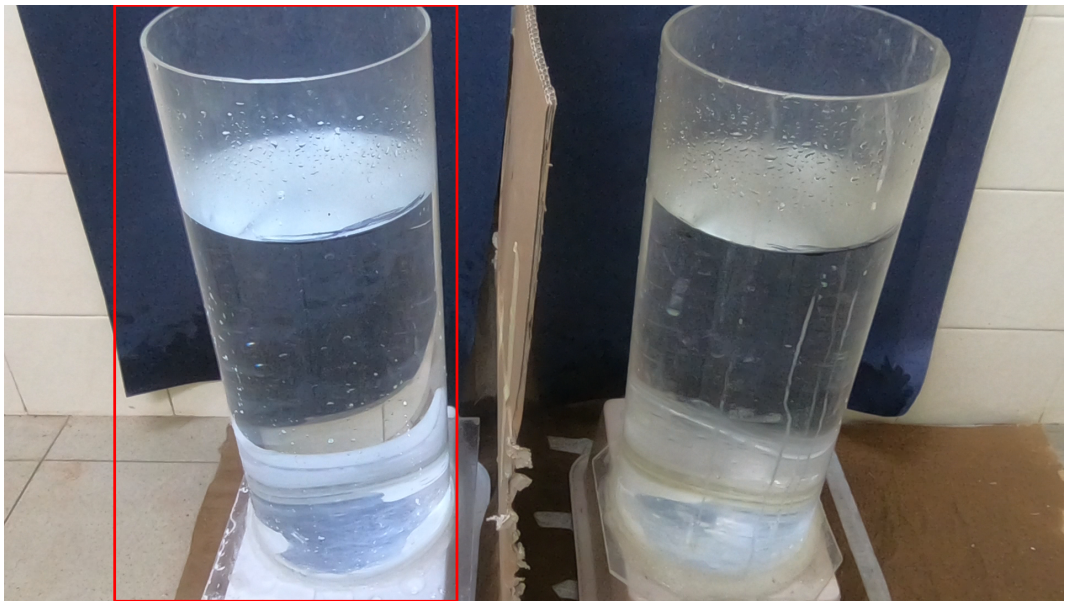
Πίνακας 4.1: Κατηγοριοποίηση κίνησης και χρώματος στην εικόνα εξόδου.

4.1.2 Εξαγωγή Περιοχής Ενδιαφέροντος

Το σύνολο των βίντεο που αποκτήθηκε περιελάμβανε 2 πειράματα τα οποία συμβαίνουν ταυτόχρονα. Για τον λόγο αυτό το επόμενο βήμα στην προεπεξεργασία των δεδομένων ήταν το κόψιμο της

περιοχής ενδιαφέροντος για κάθε πειραματόζωο. Επιλέχθηκε να γίνει προσεκτική επιλογή της περιοχής ενδιαφέροντος, έτσι ώστε αφενός να ελαχιστοποιηθεί το παρασκήνιο που στην περίπτωση της αναγνώρισης δράσης δεν θα περιείχε χρήσιμη πληροφορία, αλλά και να μικρύνει όσο είναι δυνατό το μέγεθος του τελικού Dataset .

Για τον λόγο αυτό δημιουργήθηκε script στην γλώσσα προγραμματισμού python για την κατάλληλη επιλογή της εκάστοτε περιοχής και με την βοήθεια της βιβλιοθήκης ffmpeg έγινε μαζικά το κόψιμο των επιμέρους πειραμάτων. Παρακάτω, δίνεται εικόνα με την διαδικασία της επιλογής της κατάλληλης περιοχής και με την εικόνα που δίνει ο κώδικας που συντάχθηκε σαν μια μορφή διεπαφής, ώστε να διευκολυνθεί ο χρήστης.



Σχήμα 4.2: Απόσπασμα από την διαδικασία κοψίματος των βίντεο. Φαίνεται η αρχική μορφή του βίντεο και η επιλεγμένη περιοχή σχεδιασμένη από τον κώδικα σαν διεπαφή.

Στην συνέχεια όλες οι περιοχές στην εικόνα περάστηκαν στο επόμενο μέρος του κώδικα που με την χρήση της ffmpeg και η διαδικασία αυτοματοποιήθηκε. Η παραπάνω διαδικασία είχε ως αποτέλεσμα την συλλογή τελικά 100 βίντεο από τα νέα δεδομένα που αναλογούν σε 7.7 Gigabytes δεδομένων.

4.1.3 Συγχρονισμός Εκτιμήσεων και Βίντεο

Κατά την διάρκεια της χειροκίνητης ταξινόμησης για την δημιουργία των δεδομένων ground truth το λογισμικό Kinoscope ξεκινά την καταγραφή από την στιγμή που ο χρήστης θα εισάγει την πρώτη κατηγορία κίνησης. Όμως αυτό συμβαίνει αρκετά δευτερόλεπτα μετά την έναρξη του βίντεο. Επίσης, δεν είναι σπάνιο φαινόμενο ο επιμύς που συμμετέχει στο πείραμα αντιδρά έντονα κατά την εισαγωγή του στο δοχείο με αποτέλεσμα να μην είναι σταθερή η χρονική στιγμή έναρξης της ταξινόμησης για κάθε πείραμα.

Για τον λόγο αυτό έγινε συγχρονισμός του κάθε βίντεο, με την ανάλογη εμφάνιση της ταξινομημένης κατηγορίας για τα βίντεο τα οποία ήταν ήδη ταξινομημένα και επαλήθευση της χρονικής στιγμής έναρξης της κατηγορίας δράσης σε όλα όσα ταξινομήθηκαν κατά την διάρκεια συγγραφής της παρούσας ΔΕ, ώστε να μην υπάρχει λανθασμένη αντιστοιχία κατηγορίας και βίντεο.

Τελευταία επέμβαση που έγινε στα δεδομένα βίντεο, ήταν η μετατροπή τους σε εικόνες, που αντιπροσωπεύουν το κάθε ένα frame του ώστε να είναι εφικτή η τροφοδότηση του δικτύου. Επίσης,

έγινε η κατάλληλη επαναδειγματοληψία ώστε όλα τα βίντεο να έχουν framerate, 25 fps.

4.1.4 Στατιστικά στοιχεία Dataset

Μετά την τελική συγκέντρωση των δεδομένων, εξήχθησαν τα στατιστικά για την κάθε μια κατηγορία δράσης των επιμύων. Η παραπάνω διαδικασία είναι σημαντική για την τελική αξιολόγηση του μοντέλου, αφού σε αυτό το βήμα θα γίνει φανερό εάν υπάρχει κάποια ανισορροπία στις κλάσεις των δεδομένων.

Για να εξαχθούν τα στατιστικά της κάθε κλάσης επί του συνόλου, εκτελέστηκε μέρος του αλγόριθμου που κατασκευάζει το αρχείο με όλα τα δείγματα και την κλάση στην οποία ανήκουν και τελικά πάρθηκαν τα παρακάτω αποτελέσματα σε μορφή πίνακα.

Κατηγορία-Category	Κωδικός Αναγνώρισης Δράσης	Ποσοστό Επί του Συνόλου
Ακινησία-Immobility	0	35%
Κολύμβηση-Swimming	1	27%
Αναρίχηση-Climbing	2	30%
Τίναγμα Κεφαλής-Head Shake	3	1%
Κατάδυση-Diving	4	7%

Πίνακας 4.2: Στατιστικά στοιχεία των κατηγοριών επί του συνόλου του σετ δεδομένων.

Από τα παραπάνω είναι εύκολο και ασφαλές να εξαχθεί το συμπέρασμα, ότι στο τελικό σετ δεδομένων υπάρχει σημαντική ανισορροπία ανάμεσα στις κλάσεις στα διαθέσιμα δεδομένα. Πιο συγκεκριμένα, η κατηγορία της Αναρίχησης,της Κολύμβησης και της Ακινησίας καταλαμβάνουν δυσανάλογα μεγάλο ποσοστό του Dataset σε σχέση με τις άλλες δύο κατηγορίες των Diving και Head Shake.Τα παραπάνω είναι σίγουρο ότι θα προκαλέσει πρόβλημα στην αναγνώριση αυτών των κατηγοριών κίνησης. Επιπλέον, είναι σημαντικό να τονιστεί ότι η μικρή διάρκεια της κίνησης Head Shake την καθιστά μια ποιοτική κίνηση και όχι ποσοτική όπως στις υπόλοιπες. Αυτό το χαρακτηριστικό την κάνει μια δράση δύσκολη στην αναγνώριση ιδιαίτερα με την έλλειψη ήχου στο Dataset.

4.1.5 Τελική μορφή Dataset

Τα βίντεο που αποκτήθηκαν, επεξεργάστηκαν με τέτοιο τρόπο ώστε να υπάρχει η μικρότερη δυνατή παρέμβαση σε αυτά. Επιλέχθηκε να μην γίνει κάποια περικοπή των βίντεο και να γίνει συγχρονισμός των εκτιμήσεων ώστε να είναι ευκολότερη η διόρθωση λαθών που τυχόν υπάρξουν και θα φανούν μετά από την εκπαίδευση και την αξιολόγηση του δικτύου. Στην [52], έχουν συμμετάσχει 2 ταξινομητές, και για αυτόν τον λόγο επιλέχθηκε να συμπεριληφθούν οι χειροκίνητες ταξινομήσεις και των 2 ώστε να φανεί, εάν η μοναδικότητα του ταξινομητή είναι πιθανό να επηρεάσει την διαδικασία. Γενικά, κατά την διαδικασία δημιουργίας του σετ δεδομένων εντοπίστηκαν προβλήματα που είναι πιθανό να επηρεάσουν την διαδικασία της αυτόματης ταξινόμησης. Αυτά είναι τα παρακάτω:

- Ασυμφωνία κατά την χειροκίνητη ταξινόμηση των κατηγοριών κίνησης ανάμεσα από τους ταξινομητές. Η πρόβλεψη για την επίδραση αυτής της ασυμφωνίας είναι ότι κατά την διάρκεια

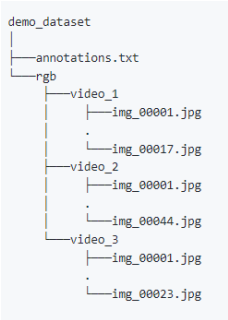
της εκπαίδευσης, το εκάστοτε δίκτυο είναι πολύ πιθανό να μην συγκλίνει διότι θα υπάρχουν μεγάλες διαφορές ανάμεσα σε βίντεο που έχουν ταξινομηθεί με υποκειμενικά κριτήρια. Επίσης, η εμπειρία του ταξινομητή στην διαδικασία της αναγνώρισης δράσης, αναμένεται να παίξει σημαντικό ρόλο στην τελική ακρίβεια του δικτύου.

- Έλλειψη ήχου από τα βίντεο. Το γεγονός ότι τα βίντεο απεικονίζουν 2 πειράματα εξαναγκασμένης κολύμβησης ταυτόχρονα, αποκλείει την χρήση του ήχου για την αναγνώριση δράσης. Αυτό τα είχε ιδιαίτερη σημασία στην περίπτωση της κατηγορίας Head Shake , που δίνει έναν χαρακτηριστικό ήχο και αποτελεί μια κίνηση με πολύ μικρή διάρκεια, της τάξεως του 0,5 sec.

- Κατά την καταγραφή των βίντεο υπάρχουν διαφορετικά φύλα ζώων τα οποία έχουν διαφορετικό μέγεθος. Αυτό σημαίνει ότι δεν είναι εύκολη η αναγνώριση της κίνησης όλων των άκρων τους, γεγονός που αναμένεται να επηρεάσει την διαδικασία.

- Η τοποθέτηση της κάμερας που καταγράφει τα πειράματα, βρίσκεται σε τέτοια θέση η οποία είναι πιθανό να μην αποτυπώνει με την μεγαλύτερη δυνατή ακρίβεια την κίνηση στον άξονα $y'y$. Το παραπάνω είναι πιθανό να δυσκολέψει την αναγνώριση της κατηγορίας Diving.

Το τελικό σετ δεδομένων με όλα τα βίντεο και μετά από την προ-επεξεργασία έχει την μορφή εικόνων από κάθε frame και κάθε ένα είναι σε διαφορετικό φάκελο.

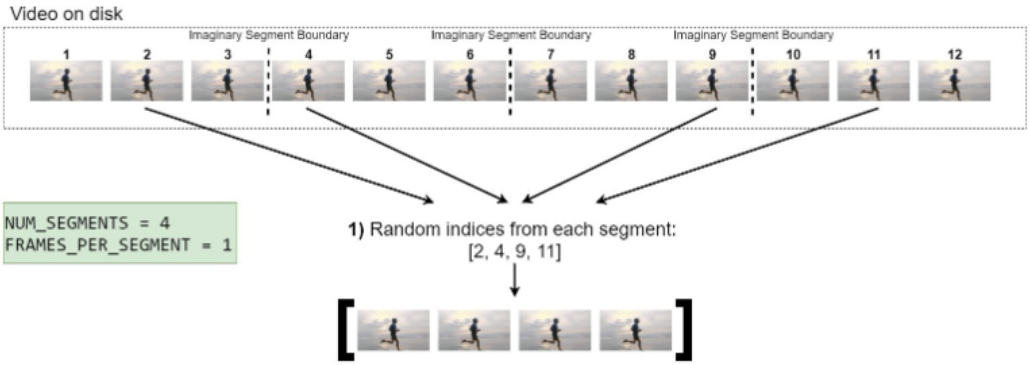


Σχήμα 4.3: Δομή φακέλων για την αποθήκευση των frames των βίντεο [55]

Για τον καλύτερο έλεγχο των παραμέτρων της δειγματοληψίας,επιλέχθηκε η κατάτμηση του δείγματος $N_{(frames)}$ σε επιμέρους κομμάτια και η τελική του επιλογή να γίνεται τυχαία μέσα από το κάθε κομμάτι αυτό (segment)

Μετά τα παραπάνω το τελικό σετ δεδομένων έχει τις παραμέτρους που δίνονται στον παρακάτω πίνακα.

Στην δεύτερη στήλη του πίνακα 4.3, δίνονται οι ενδεικτικές τιμές που είναι πιθανό να δώσουν την καλύτερη ακρίβεια σύμφωνα με το [40] χωρίς όμως να είναι δεσμευτικές και θα παρουσιαστούν ενδελεχώς τα αποτελέσματα των πειραμάτων με την τελική βελτιστοποίηση των παραμέτρων του καλύτερου μοντέλου που θα προκύψει.



Σχήμα 4.4: Μορφή τελικής δειγματοληψίας [55]

Παράμετρος	Ενδεικτικές τιμές καλύτερου μοντέλου	Δυνατότητα βελτιστοποίησης
Frames Per second	32	Ναι
Frames Per second	RGB	Όχι
Μέγεθος εικόνας	112x112	Ναι

Πίνακας 4.3: Πίνακας ενδεικτικών παραμέτρων του τελικού σετ δεδομένων.

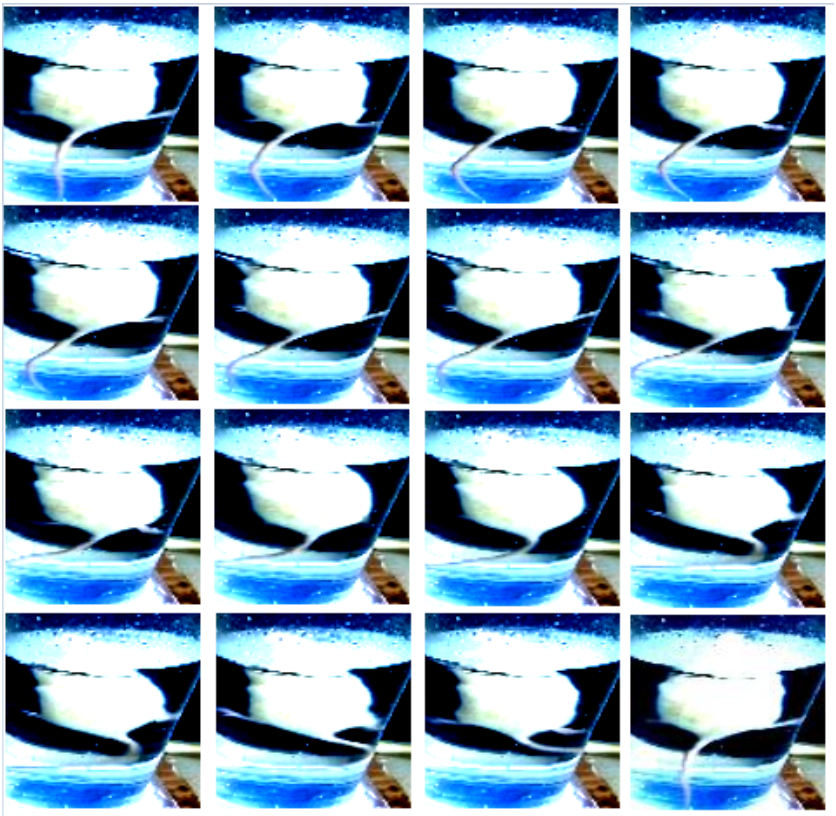
4.2 Προετοιμασία Πειραμάτων

4.2.1 Datalader

Όπως αναφέρθηκε και παραπάνω, οι εκτιμήσεις από τους παρατηρητές έχουν την μορφή εικόνας μεγέθους στο πλάτος 1000 εικονοστοιχείων. Για τον λόγο αυτό υλοποιήθηκε μέθοδος με την οποία λαμβάνεται η εικόνα της ταξινόμησης και ανάλογα με την χρονική διάρκεια των frames που αποτελεί το εκάστοτε δείγμα και τον παρατηρητή. Έτσι σχηματίζεται το αρχείο που περιέχει τα δείγματα με την κατάλληλη παραμετροποίηση και περιέχουν την χρονική διάρκεια του δείγματος, την διαδρομή για το ανάλογο Dataset και την εκτίμηση ground truth σε μορφή .txt.Για την εκπαίδευση του μοντέλου και για την αξιολόγησή του επιλέχθηκε να δημιουργηθούν 3 σετ, Training, Validation και Test.

Στην συνέχεια, δημιουργήθηκε κλάση Dataset η οποία δίνει την δυνατότητα για την φόρτωση των δειγμάτων από τα στιγμιότυπα που έχουν εξαχθεί και με την ground truth ταξινόμηση που έχει παρθεί από το μεσαίο frame της ομάδας. Η χρήση της κλάσης Dataset που δίνεται από την βιβλιοθήκη Pytorch παρέχει εκτός των άλλων και την δυνατότητα να εφαρμοστεί η τεχνική επαύξησης των δειγμάτων, Data Augmmentation . Όπως αναφέρθηκε και στην αντίστοιχη παράγραφο, τα δεδομένα περιέχουν 5 κλάσεις κίνησης, με τις κατηγορίες Climbing, Swimming και Immobility να αποτελούν το μεγαλύτερο μέρος του σετ δεδομένων. Για να αντιμετωπιστεί το πρόβλημα της ανισορροπίας των κλάσεων χρησιμοποιήθηκαν τεχνικές Data Augmentation που να προσομοιάζουν όσο καλύτερα γίνεται την κάθε πιθανή κίνηση του επιμύ.

Τέλος, για την τελική έξοδο δεδομένων από το Dataset γίνεται κατευθείαν μετατροπή σε μορφή κατάλληλη για είσοδο στο μοντέλο, Tensor με τις κατάλληλες διαστάσεις (C, B, H, W). Το Dataset δίνει εικόνες με την μορφή των παρακάτω:



Σχήμα 4.5: Δείγμα 16 frames όπως δίνεται από τυχαίο index του Dataset για την κατηγορία Swimming.

Όπως είναι φανερό και από τις παραπάνω εικόνες του σετ δεδομένων, κατά την διαδικασία των πειραμάτων, τα δίκτυα που θα δοκιμαστούν καλούνται να αναγνωρίσουν κινήσεις και τα πρότυπα αφορούν αποκλειστικά αυτές και όχι τα χαρακτηριστικά του παρασκηνίου, αφού στην αναγνώριση δράσης συμβάλλουν μόνο τα άκρα και η στάση του σώματος του επιμύ.

Το επόμενο βήμα για την προετοιμασία της εκπαίδευσης των δικτύων είναι η υλοποίηση Dataloader. Ο Dataloader δίνει την δυνατότητα να δοθούν στο δίκτυο ως είσοδοι οι αλληλουχίες εικόνων και οι ταξινομήσεις ground truth με το κατάλληλο μέγεθος παρτίδας (Batch Size). Έτσι είναι εφικτή η παραμετροποίηση ώστε να γίνει βελτιστοποίηση του μοντέλου μετά την επιλογή της βέλτιστης αρχιτεκτονικής. Το μέγεθος παρτίδας που επιλέχθηκε κυμαινόταν, ανάλογα με το μοντέλο που εκπαιδευόταν κάθε φορά, από 4 έως 32. Επιπλέον, κατά την διάρκεια της εκπαίδευσης η λήψη των δειγμάτων γίνεται τυχαία μέσα από το κομμάτι του Dataset που χωρίστηκε για εκπαίδευση. Όμως για την δοκιμή του μοντέλου από Validation και Testing ο Dataloader δίνει και την επιλογή τα στοιχεία να λαμβάνονται με την απλή προσπέλαση του σετ δεδομένων με την σειρά. Αυτό είναι εφικτό διότι κατά τις δύο παραπάνω φάσεις το μοντέλο δεν εκπαιδεύεται και κατά συνέπεια τα βάρη των φίλτρων του δεν αλλάζουν. Ακόμα το μέγεθος του σετ δεδομένων είναι τέτοιο που δεν θα ήταν δυνατή η φόρτωσή του στην μνήμη. Με την χρήση του Dataloader η φόρτωσή του γίνεται σταδιακά και όσο χρειάζεται ώστε να γίνει η εκπαίδευση αλλά και να μην εξαντληθεί η μνήμη. Η δημιουργία των δειγμάτων γίνεται από την CPU με Multiprocessing ώστε να μην επιβαρύνεται η GPU με την δημιουργία δειγμάτων.

Η εκπαίδευση συνελικτικών νευρωνικών δικτύων είναι μια ιδιαίτερα απαιτητική διαδικασία από πλευρά υπολογιστικής ισχύος. Για τον λόγο αυτό η εκπαίδευση και η αξιολόγηση έγιναν με την χρήση GPU με μνήμη 6 έως 24 Gigabyte σε υπολογιστή του Εργαστηρίου Τηλεπισκόπησης της

Σχολής Αγρονόμων Τοπογράφων Μηχανικών και Μηχανικών Γεωπληροφορικής του ΕΜΠ.

4.2.2 Επιλογή Συνάρτησης κόστους και Αλγόριθμου Βελτιστοποίησης

Η περίπτωση της αναγνώρισης δράσεων εμπίπτει στην κατηγορία ταξινομήσεων που αναζητείται η καλύτερη δυνατή αναγνώριση κλάσης ανάλογα με τα δεδομένα εισόδου. Αυτό οδήγησε στην επιλογή ως συνάρτηση κόστους της Cross Entropy Loss. Όπως ειπώθηκε και στο 2ο κεφάλαιο η συνάρτηση αυτή βασίζεται στην απόκλιση Kullback-Leibler. Κατά την φάση της εκπαίδευσης, μετά την κάθε παρτίδα, υπολογίζεται το κόστος και τα βάρη αλλάζουν ώστε να βελτιωθεί η επίδοση του μοντέλου. Επίσης, μετά το τέλος της κάθε εποχής υπολογίζεται το τελικό κόστος που υπήρξε και θα δίνεται διαγραμματικά μετά την κάθε εκπαίδευση στο επόμενο μέρος. Επιπλέον, κατά την διάρκεια της αξιολόγησης θα υπολογιστεί το κόστος με βάση την αξιολόγηση του δικτύου.

Ο αλγόριθμος βελτιστοποίησης που επιλέχθηκε για τα πειράματα είναι ο AdamW.Οι αλγόριθμοι της οικογένειας αυτής όπως αναφέρθηκε και στο Κεφάλαιο 2, χρησιμοποιούνται ευρέως σε εφαρμογές τεχνητών νευρωνικών δικτύων. Σημαντική υπερπαράμετρος για την βελτιστοποίηση του δικτύου είναι ο ρυθμός εκμάθησης. Ο ρυθμός εκμάθησης (Learning Rate) αρχικά επιλέχθηκε να έχει τιμή 10^{-4} και με τα αποτελέσματα που δίνονταν κατά την διάρκεια των πειραμάτων προσαρμόστηκε κατάλληλα ώστε να γίνει η βελτιστοποίηση.

Για την εκπαίδευση του δικτύου και για την αξιολόγηση από το σετ δεδομένων που χωρίστηκε για τον σκοπό αυτό, έγινε χρήση της βιβλιοθήκης Pytorch Lightning.Με την χρήση της, οι αρχιτεκτονικές είναι δυνατό να σχεδιαστούν με πολύ πιο γρήγορο τρόπο προσφέροντας ευελιξία κρατώντας τον κώδικα όσο πιο ευανάγνωστο γίνεται. Επιπλέον, χωρίζει τα βήματα των Train, Validation και Test σε χωριστές μεθόδους. Με τον κατακερματισμό αυτό ο κώδικας είναι απαλλαγμένος από επαναληπτικούς βρόγχους, γεγονός που μειώνει τις πιθανότητες για σφάλματα στον κώδικα. Ένα ακόμα σημαντικό πλεονέκτημα της χρήσης του Pytorch Lightning είναι ότι όλα τα παραπάνω βρίσκονται μέσα σε κλάση και ο κώδικας για βήμα μέσω των μεθόδων συντάσσεται αφαιρετικά, πάλι μειώνοντας την πιθανότητα σφαλμάτων ειδικά κατα την διαδικασία δημιουργίας νέων αρχιτεκτονικών.

Τέλος, η καταγραφή των πειραμάτων είναι εξαιρετικής σημασίας για την παρακολούθηση του μοντέλου και την βελτιστοποίηση. Τα πειράματα είχαν μεγάλη χρονική διάρκεια, δεδομένου ότι τα μοντέλα διέθεταν πολλές φορές εκατομμύρια παραμέτρους και το μέγεθος της παρτίδας ήταν πολλές φορές 4 ή 8. Για την καταγραφή των πειραμάτων επιλέχθηκε η βιβλιοθήκη mlflow, που δίνει και την δυνατότητα να αποθηκευτεί το checkpoint του καλύτερου μοντέλου από την διαδικασία της αξιολόγησης του δικτύου. Κατά την διαδικασία Validation και Test θα υπολογίζονται οι πίνακες σύγχυσης, γεγονός που θα δώσει μια αρκετά καλή εικόνα για την επίδοση του μοντέλου. Ιδιαίτερα σημαντική θα είναι η συμβολή του πίνακα σύγχυσης σε ό,τι αφορά την αξιολόγηση του μοντέλου στην ταξινόμηση κατηγοριών δράσης ανάμεσα σε αυτές που αποτελούν ιδιαίτερη περίπτωση (ποιοτική κίνηση) όπως το Head Shake.

4.3 Πειραματική Διαδικασία

Στο πρώτο μέρος των πειραμάτων, δοκιμάστηκαν αρχιτεκτονικές που περιλάμβαναν συνελίξεις τριών διαστάσεων, σε μια προσπάθεια να μοντελοποιηθούν τα χωρικά αλλά και τα χρονικά πρότυπα που εμφανίζονται στα δεδομένα εκπαίδευσης. Οι αρχιτεκτονικές αυτές σχεδιάστηκαν με βάση όσα έχουν δημοσιευτεί και με στόχο να διερευνηθούν οι δυνατότητες μικρών και ευέλικτων δικτύων με

λιγότερες από 10 εκατομμύρια παραμέτρους. Για τα δίκτυα αυτά έγιναν δοκιμές για την βελτιστοποίησή τους με τα αποτελέσματα να δίνονται στην αντίστοιχη παράγραφο. Για την καλύτερη ακρίβεια των δικτύων έγινε επίσης δοκιμή με διαφορετικό αριθμό συνελικτικών επιπέδων και μεγέθους εικόνας, με την συνάρτηση ενεργοποίησης να είναι η ReLU και τον αλγόριθμο βελτιστοποίησης AdamW . Για την καλύτερη μελέτη των δικτύων αυτών, θα υλοποιηθούν αρχιτεκτονικές που κατά την διάρκεια των πειραμάτων θα έχουν αυξανόμενο αριθμό συνελίξεων. Αναμένεται να υπάρξει ένα όριο στην βελτίωση της επίδοσης με την αύξηση των συνελικτικών επιπέδων. Κατά την διάρκεια της εκπαίδευσης και με την συνεχόμενη πρόσθεση συνελικτικών επιπέδων η συμπεριφορά του δικτύου γίνεται ασταθής και δυσχεραίνεται η εκπαίδευσή του. Η συνάρτηση ενεργοποίησης ReLU αναμένεται να βοηθήσει σημαντικά στην επίλυση του παραπάνω προβλήματος.

Στο δεύτερο μέρος της πειραματικής διαδικασίας, θα γίνει εκτενής διερεύνηση των αρχιτεκτονικών ResNet3D και R(2+1)D. Πρόκειται για αρχιτεκτονικές που έχουν πετύχει εξαιρετικά αποτελέσματα στην αναγνώριση δράσεων και ανήκουν στην κατηγορία των Deep Neural Networks. Σε συνέχεια της προηγούμενης παραγράφου, επιλύουν τα προβλήματα της αποσταθεροποιημένης εκπαίδευσης με την χρήση Skip Connections ώστε να μην παρατηρείται το vanishing gradient. Όπως και με τις αρχιτεκτονικές του πρώτου μέρους, θα γίνει συνολική παρουσίαση της πειραματικής διαδικασίας με τα αποτελέσματα του κάθε πειράματος κατά την διάρκεια της βελτιστοποίησης.

Στο πρώτο μέρος της πειραματικής διαδικασίας, θα χρησιμοποιηθούν δίκτυα με τα βάρη τους να είναι τυχαία αρχικοποιημένα, αλλά στο δεύτερο μέρος, τα μοντέλα θα είναι προεκπαιδευμένα στο dataset Kinetics 400 .

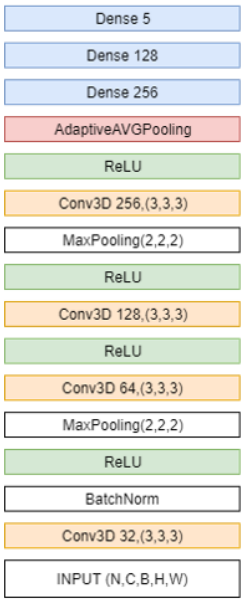
4.4 Simple 3D Convolution Layers

Τα πειράματα ξεκίνησαν με απλές αρχιτεκτονικές αρχικά 4 επιπέδων και στην συνέχεια αυξήθηκαν για να μελετηθούν οι δυνατότητές τους. Όπως αναφέρθηκε χρησιμοποιήθηκαν η συνάρτηση ενεργοποίησης ReLU και ο αλγόριθμος βελτιστοποίησης AdamW. Επιπλέον είναι σημαντικό να αναφερθεί ότι η χρήση της βιβλιοθήκης Pytorch Lightning που αναλύθηκε και σε προηγούμενη παράγραφο, έκανε την διαδικασία του Prototyping εξαιρετικά εύκολη και έδωσε την ευκαιρία για περαιτέρω ενασχόληση με την επίλυση των προβλημάτων που απορρέουν από την διαδικασία της βελτιστοποίησης των δικτύων.

4.4.1 Βασική Αρχιτεκτονική με 4 επίπεδα 3D Convolution Layers

Τα πειράματα ξεκίνησαν με την δημιουργία ενός σχετικά ρηχού δικτύου με 4 συνελικτικά επίπεδα τριών διαστάσεων. Η αρχιτεκτονική αυτή δημιουργήθηκε ώστε να διερευνηθεί η δυνατότητα αναγνώρισης δράσεων από τα τρισδιάστατα συνελικτικά επίπεδα από την βάση της. Κατά την διαδικασία της εκπαίδευσης γίνεται η προσπάθεια να αναγνωριστούν τα χρονικά πρότυπα μέσα από την χρήση των παραπάνω συνελίξεων. Παρακάτω δίνεται η αρχιτεκτονική και σχηματικά.

Στο Σχήμα 4.6 φαίνεται η αφετηρία των αρχιτεκτονικών που δοκιμάστικαν για την αναγνώριση των δράσεων στα πειράματα εξαναγκασμένης κολύμβησης. Ιδιαίτερο ενδιαφέρον έχει το επίπεδο Adaptive Average Pooling. Η χρήση του επιπέδου αυτού ήταν αναγκαία, διότι χωρίς αυτό θα χρειαζόταν ο υπολογισμός των παραμέτρων που εισάγονται κάθε φορά στα επίπεδα Dense N . Με την χρήση του η έξοδος είναι πάντα η ίδια χωρίς να χρειαστεί να υπολογιστεί οτιδήποτε άλλο, αφού πίσω από το επίπεδο αυτό υπολογίζεται αυτόματα το βήμα ώστε να δοθεί η επιθυμητή έξοδος.



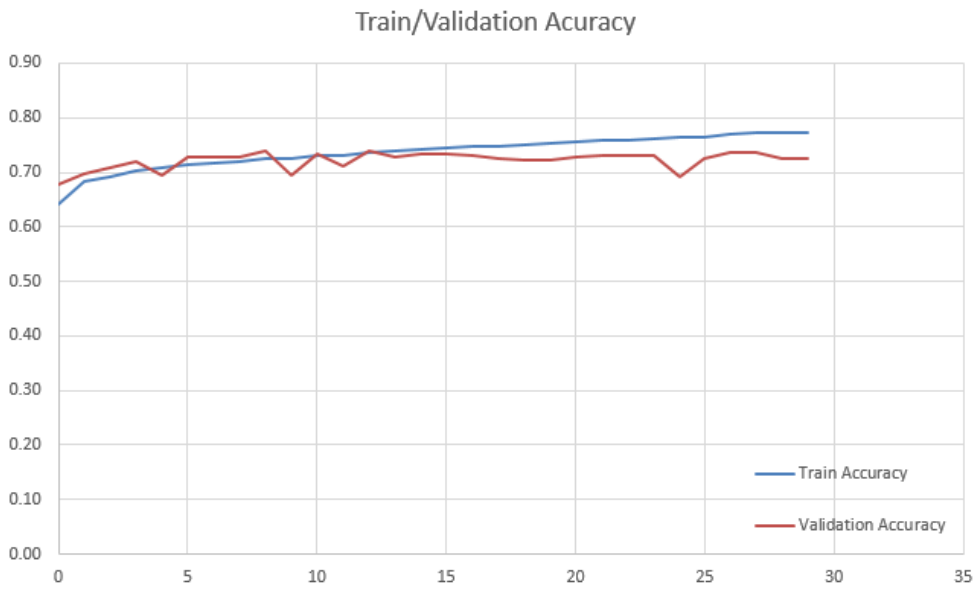
Σχήμα 4.6: Αρχιτεκτονική με 4 συνελικτικά επίπεδα τριών διαστάσεων.

Ο ρυθμός μάθησης επιλέχθηκε να είναι 10^{-4} ως αρχική τιμή. Στον πίνακα φαίνονται οι αρχικές υπερ-παράμετροι με τις οποίες εκπαιδεύτηκε το δίκτυο.

Υπερ-παράμετρος	**Τιμή**
Αριθμός συνελικτικών επιπέδων	4
Μέγεθος τυχαίου παραθύρου αποκοπής για εκπαίδευση	96X96
Ρυθμός εκμάθησης	10^{-4}
Μέγεθος εικόνας	112X112
Χρονική διάρκεια δείγματος	32 frames
Batch Size	32
fps	25
Number of Segments	16
Frames per Segment	1

Πίνακας 4.4: Πίνακας υπερπαραμέτρων βασικής αρχιτεκτονικής

Μετά την εκπαίδευση της βασικής αρχιτεκτονικής εξήχθησαν οι γραφικές παραστάσεις που αφορούν τα Train Accuracy και Validation Accuracy.

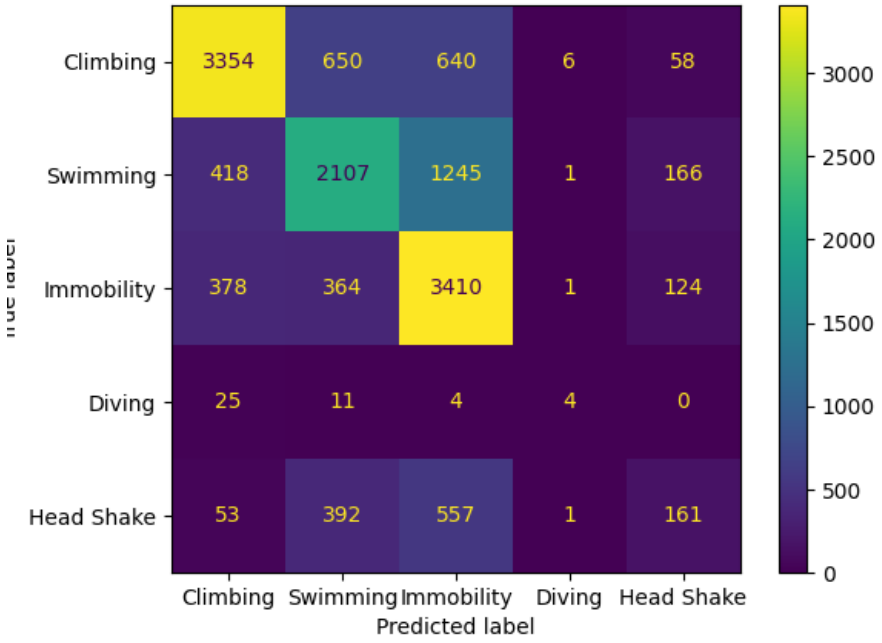


Σχήμα 4.7: Γραφική παράσταση Train Accuracy και Validation Accuracy.

Από το παραπάνω είναι φανερό ότι μετά την εποχή 27 η ακρίβεια δεν αυξάνεται και μένει σταθερή στο 0.77. Αυτό όμως που δίνει σημαντικά στοιχεία για την συμπεριφορά του μοντέλου είναι το γράφημα με την Validation Accuracy που δείχνει ότι αυτή μένει σταθερή και μετά από ένα σημείο το δίκτυο έχει σημάδια overfitting.Επίσης, είναι σημαντικό να δοθεί και η γραφική παράσταση Train/Validation Loss.Οι παραστάσεις αυτές είναι εξαιρετικά σημαντικές, διότι δίνουν σημαντικά στοιχεία για την πορεία της εκπαίδευσης και της συμπεριφοράς του μοντέλου.

Σχήμα 4.8: Γραφική παράσταση Validation Loss.

Τέλος, κρίνεται σημαντικό να δοθεί ο πίνακας σύγχυσης για την φάση Test του δικτύου. Από εδώ θα γίνει φανερό ποιές κατηγορίες κίνησης έχουν ταξινομηθεί σωστά και όπως είναι λογικό, οι προβλέψεις θα βρίσκονται επάνω στην διαγώνιο του πίνακα, αλλά και οι προβλέψεις στις οποίες υπήρξε σύγχυση σε ό,τι αφορά την ταξινόμηση.



Σχήμα 4.9: Πίνακας σύγχυσης Test.

Από τον παραπάνω πίνακα, είναι εφικτό να εξαχθούν και τα στατιστικά Precision,Recall και F1-Score. Όπως αναφέρθηκε και παραπάνω, τα στατιστικά αυτά δίνουν μια ξεκάθαρη εικόνα για την επίδοση του δικτύου για την κάθε κατηγορία. Σύμφωνα με αυτά θα γίνει ο σχολιασμός των παραπάνω και θα σχεδιαστούν με ορθότερο τρόπο οι μετέπειτα αρχιτεκτονικές.

Κατηγορία	**Precision**	**Recall**	**F1-Score**	**Support**
Climbing	0.79	0.71	0.75	4708
Swimming	0.59	0.53	0.56	3937
Immobility	0.58	0.79	0.67	4277
Diving	0.3	0.1	0.14	44
Head-Shake	0.31	0.13	0.2	1164

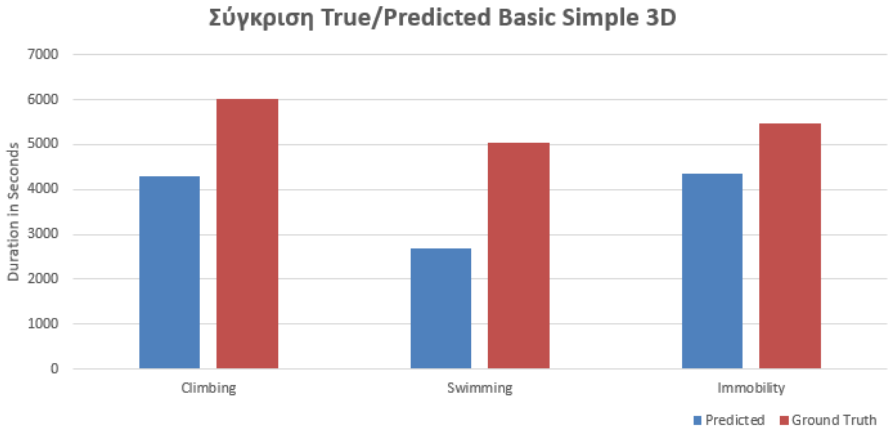
Πίνακας 4.5: Πίνακας στατιστικών πρώτης ταξινόμησης με 4 συνελικτικά επίπεδα.

	Precision	**Recall**	**F1-Score**
Macro Average	0.51	0.45	0.46
Weighted Average	0.63	0.63	0.62
Test Accuracy	63.2%		

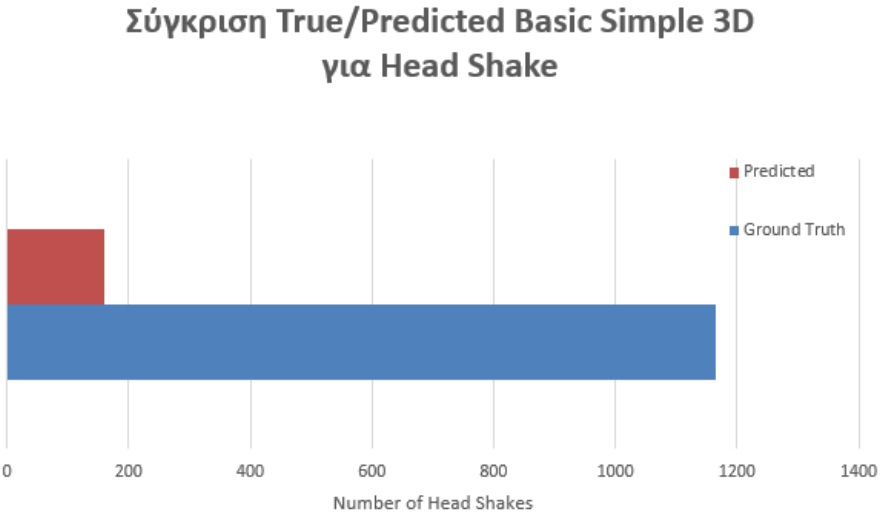
Πίνακας 4.6: Πίνακας μακρο-στατιστικών και σταθμισμένων .

Σχολιασμός

Μετά τα παραπάνω είναι εύκολη η εξαγωγή χρήσιμων συμπερασμάτων για την συμπεριφορά του μοντέλου. Η κατηγορίες με τα καλύτερα στατιστικά ήταν οι Climbing,Swimming και Immobility.



Σχήμα 4.10: Συγκριτικό διάγραμμα Pred/Ground Truth από Basic Simple 3D .



Σχήμα 4.11: Συγκριτικό διάγραμμα Pred/Ground Truth από Basic Simple 3D για τινάγματα κεφαλής.

Είναι φανερό ότι τα δεδομένα ήταν αρκετά ώστε να υπάρξει ικανοποιητική επίδοση του δικτύου σε αυτές τις κατηγορίες.

Στην συγκεκριμένη αρχιτεκτονική, το πρόβλημα εντοπίζεται στις άλλες δύο κατηγορίες, Diving και Head-Shake . Εδώ η επίδοση ήταν αρκετά χαμηλότερη από ότι στις τρεις παραπάνω και αυτό αποτελεί πεδίο βελτίωσης στις παρακάτω δοκιμές για τα απλά συνελικτικά δίκτυα.

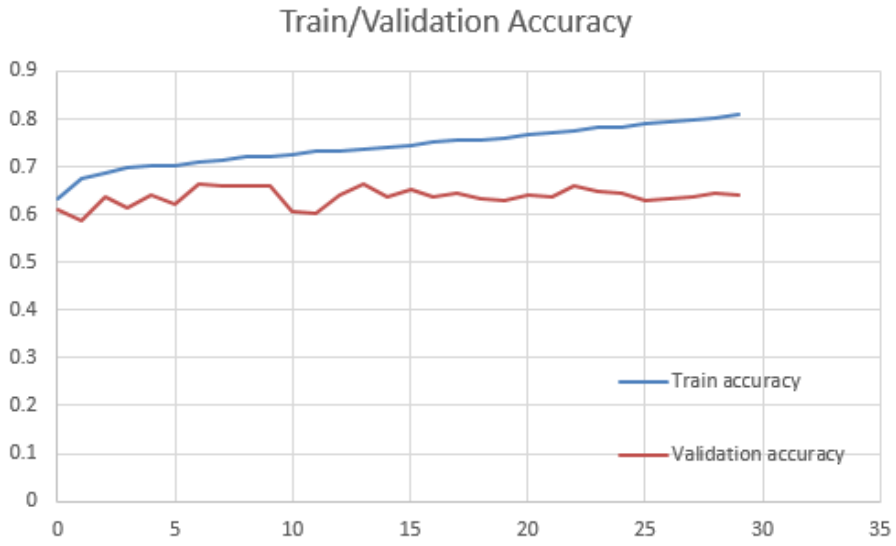
4.4.2 5 Συνελικτικά Επίπεδα Τριών Διαστάσεων και Βελτιστοποίηση Υπερπαραμέτρων Simple CNNs

Στην συνέχεια υλοποιήθηκε δίκτυο που περιλαμβάνει 5 συνελικτικά επίπεδα 3D . Στην περίπτωση αυτή, το δίκτυο περιλαμβάνει 5 επίπεδα τύπου Conv3D . Ο λόγος για τον οποίο επιλέχθηκε να προστεθούν παραπάνω επίπεδα είναι για να δοκιμαστεί η ικανότητα να μοντελοποιηθεί η κίνηση των επιμύων με παραπάνω επίπεδα και να εντοπιστούν πρότυπα στην χρονική διάσταση με καλύτερα αποτελέσματα από ότι στην προηγούμενη δοκιμή. Οι υπερπαράμετροι που χρησιμοποιήθηκαν είναι οι εξής:

Υπερ-παράμετρος	**Τιμή**
Αριθμός συνελικτικών επιπέδων	5
Μέγεθος τυχαίου παραθύρου αποκοπής για εκπαίδευση	112X112
Ρυθμός εκμάθησης	10^{-4}
Μέγεθος εικόνας	120X120
Χρονική διάρκεια δείγματος	32 frames
Batch Size	32
fps	25
Number of Segments	16
Frames per Segment	1

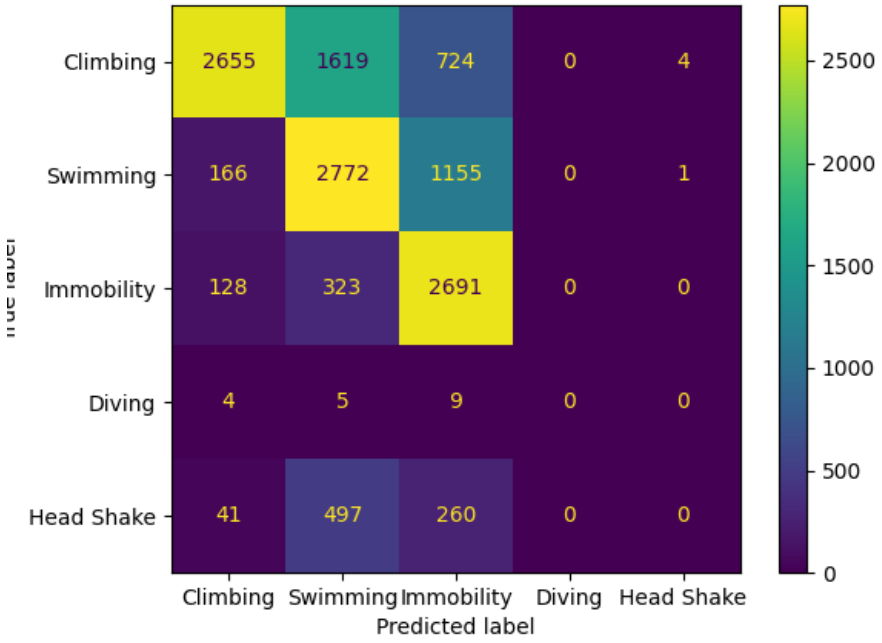
Πίνακας 4.7: Πίνακας υπερ-παραμέτρων αρχιτεκτονικής

Μετά από την εκπαίδευση δίνεται ο παρακάτω πίνακας σύγχυσης καθώς και η γραφική παράσταση της ακρίβειας Train και Validatinon :



Σχήμα 4.12: Γραφική παράσταση Train και Validation accuracy .

Στην συνέχεια και για την καλύτερη σύγκριση με το προηγούμενο δίνεται ο πίνακας με τα στατιστικά στοιχεία του δικτύου όπως και παραπάνω:



Σχήμα 4.13: Πίνακας Σύγχυσης δικτύου με 5 συνελικτικά επίπεδα (Conv3D).

Κατηγορία	**Precision**	**Recall**	**F1-Score**	**Support**
Climbing	0.88	0.53	0.66	5002
Swimming	0.53	0.67	0.6	4094
Immobility	0.55	0.85	0.67	3142
Diving	-	0	0	18
Head-Shake	0	0	0	798

Πίνακας 4.8: Πίνακας στατιστικών πρώτης ταξινόμησης με 5 συνελικτικά επίπεδα.

	Precision	**Recall**	**F1-Score**
Macro Average	-	0.41	0.38
Weighted Average	-	0.62	0.60
Test Accuracy	62.4%		

Πίνακας 4.9: Πίνακας μακρο-στατιστικών και σταθμισμένων .

Σχολιασμός

Από τα παραπάνω είναι φανερό ότι το δίκτυο που δόθηκε στην παραπάνω παράγραφο δεν είχε την ίδια επίδοση με το βασικό που προηγήθηκε. Υπάρχει βελτίωση ως προς την ακρίβεια στην κατηγορία Climbing , αλλά όχι ιδιαίτερη βελτίωση στις υπόλοιπες κατηγορίες. Αξιοσημείωτο είναι ότι το δίκτυο με τις υπερπαραμέτρους που δόθηκαν παραπάνω δεν ήταν σε θέση να αναγνωρίσει την κατηγορία της Κατάδυσης και του Τινάγματος Κεφαλής. Πιο συγκεκριμένα, υπήρχε σύγχυση της κατηγορίας Head Shake με τις τρεις κυρίαρχες κατηγορίες του σετ δεδομένων. Επιπλέον, το δίκτυο με τα 5 συνελικτικά επίπεδα, έδειξε σημάδια overfitting και αυτό αποτελεί σημαντικό στοιχείο ότι πρέπει να γίνει βελτιστοποίηση των υπερπαραμέτρων ώστε να υπάρξει βελτίωση της

συμπεριφοράς του.

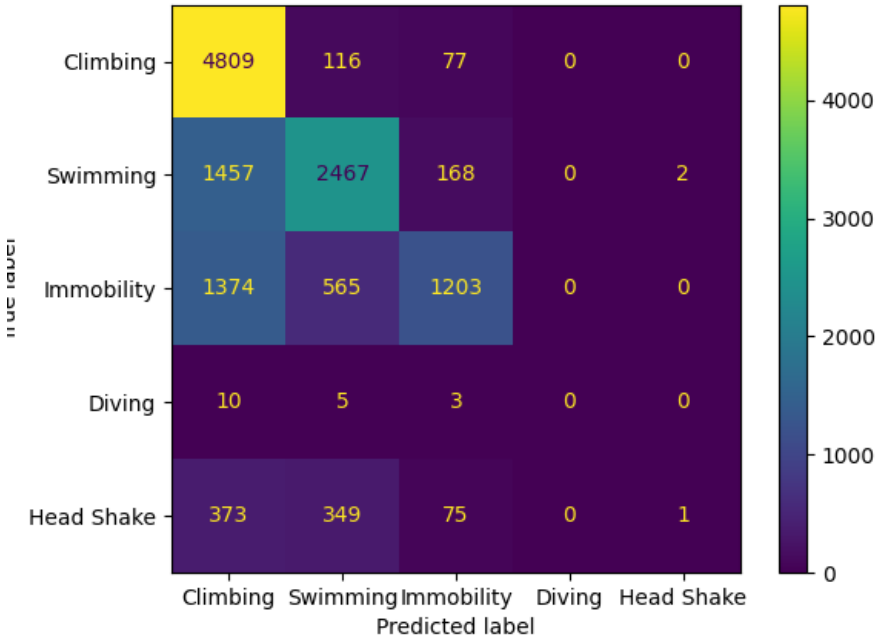
4.4.3 6 Συνελικτικά Επίπεδα Τριών Διαστάσεων

Στην συνέχεια, υλοποιήθηκε αρχιτεκτονική που περιλαμβάνει 6 συνελικτικά επίπεδα. Η συνάρτηση ενεργοποίησης επιλέχθηκε η ReLU και ο αλγόριθμος βελτιστοποίησης παρέμεινε ο ίδιος. Ο ρυθμός μάθησης για το αρχικό πείραμα παρέμεινε ο ίδιος και στην συνέχεια ρυθμίστηκε ώστε να βελτιστοποιηθεί το δίκτυο και να ληφθούν τα τελικά αποτελέσματα των αρχιτεκτονικών. Στην ομάδα πειραμάτων που αφορούν τα δίκτυα με 6 συνελικτικά επίπεδα έγινε η προσθήκη επιπέδου Dropout(p) . Το επίπεδο αυτό, είναι εξαιρετικής σημασίας για την αποφυγή της υπερπροσαρμογής στα τεχνητά νευρωνικά δίκτυα και ο ρόλος του είναι κατά την διάρκεια της εκπαίδευσης να βγάζει εκτός δικτύου κάποιες παραμέτρους τους, δίνοντάς τους μηδενική τιμή. Για το πρώτο πείραμα αυτής της σειράς, επιλέχθηκαν οι παρακάτω υπερπαράμετροι:

Υπερ-παράμετρος	**Τιμή**
Αριθμός συνελικτικών επιπέδων	6
Μέγεθος τυχαίου παραθύρου αποκοπής για εκπαίδευση	112X112
Ρυθμός εκμάθησης	10^{-4}
Μέγεθος εικόνας	120X120
Χρονική διάρκεια δείγματος	32 frames
Batch Size	16
fps	25
Number of Segments	16
Frames per Segment	1

Πίνακας 4.10: Πίνακας υπερ-παραμέτρων αρχιτεκτονικής

Μετά την εκπαίδευση του δικτύου προέκυψε ο παρακάτω πίνακας σύγχυσης:



Σχήμα 4.14: Πίνακας Σύγχυσης δικτύου με 6 συνελικτικά επίπεδα (Conv3D) και Dropout.

Με τον υπολογισμό των στατιστικών παρακάτω είναι δυνατόν να εξαχθεί μια ασφαλέστερη εικόνα για την επίδοση του δικτύου, ώστε να προσαρμοστούν καλυτερα οι υπερπαράμετροι.

Κατηγορία	**Precision**	**Recall**	**F1-Score**	**Support**
Climbing	0.6	0.96	0.73	5002
Swimming	0.7	0.60	0.64	4094
Immobility	0.78	0.38	0.51	3142
Diving	-	0	0	18
Head-Shake	0.33	0	0	798

Πίνακας 4.11: Πίνακας στατιστικών πρώτης ταξινόμησης με 6 συνελικτικά επίπεδα.

	Precision	**Recall**	**F1-Score**
Macro Average	-	0.38	0.38
Weighted Average	-	0.64	0.61
Test Accuracy	64.6%		

Πίνακας 4.12: Πίνακας μακρο-στατιστικών και σταθμισμένων .

Σχολιασμός

Από τα παραπάνω είναι εύκολο να εξαχθεί το συμπέρασμα ότι οι 2 κατηγορίες Diving και Head Shake απότελούν 2 περιπτώσεις που το μοντέλο δυσκολεύεται να αναγνωρίσει. Αυτό συμβαίνει διότι δεν υπάρχουν αρκετά δεδομένα εκπαίδευσης στις δύο αυτές κατηγορίες κίνησης.

Για τις υπόλοιπες τρεις κατηγορίες υπάρχει βελτίωση ως προς το Recall για την κίνηση Climbing και Swimming . Σε αυτές τις δύο κατηγορίες, υπάρχει σαφής βελτίωση του Recall και βελτίωση του

Precision στη κατηγορία Swimming. Μετά τα παραπάνω είναι σημαντικό να γίνει βελτιστοποίηση των υπερπαραμέτρων του παραπάνω δικτύου ώστε να επιλεχθουν οι βέλτιστες ρυθμίσεις που θα δώσουν την μεγαλύτερη δυνατή ακρίβεια και στατιστικά. Είναι σημαντικό να σημειωθεί ότι όπως αναφέρθηκε και παραπάνω, τα βίντεο αποτελούνται από πειράματα εξαναγκασμένης κολύμβησης που δεν περιέχουν κάποια λεπτομέρεια στο παρασκήνιο, ώστε να είναι δυνατή η σύνδεση της κίνησης με αυτό. Στην περίπτωση που υπήρχε κάποιο τέτοιο στοιχείο, η αναγνώριση θα ήταν αρκετά ευκολότερη αφού το πρόβλημα αναγνώρισης δράσεων θα είχε αναχθεί κατά ένα ποσοστό στην αναγνώριση ενός προτύπου του παρασκηνίου που χαρακτηρίζει την δράση.

4.4.4 Βελτιστοποίηση παραμέτρων δικτύου 6 επιπέδων

Όπως φάνηκε παραπάνω τα δίκτυα που σχεδιάστηκαν παρουσιάζουν το φαινόμενο της υπερπροσαρμογής, καθώ και την αδυναμία μετά από ένα σημείο να αναγνωρίσουν τις μικρές κινήσεις όπως το τίναγμα κεφαλής. Για να αποφευχθεί η υπερπροσαρμογή έγινε βελτιστοποίηση της παραμέτρου p=k του επιπέδου Dropout και έγινε επιπλέον χρήση του επιπέδου Batch Normalization μετά από κάθε συνελικτικό επίπεδο και πριν την συνάρτηση ενεργοποίησης.

Για την εύρεση της βέλτιστης επιλογής υπερπαραμέτρων και επιπέδων, έγιναν τα παρακάτω πειράματα με τις αρχιτεκτονικές να δίνονται σχηματικά στο Παράρτημα Α΄.

4.4.5 Αρχιτεκτονική 6 επιπέδων με επίπεδα Batch Normalization

Αφού σχεδιάστηκαν αρχιτεκτονικές που περιέχουν διάφορετικό αριθμό επιπέδων η κάθε μια, έγινε εκπαίδευση και πειραματισμός με τις παρακάτω υπερπαρμέτρους, ώστε να επιτευχθεί η καλύτερη δυνατή ακρίβεια.

4.4.5.1 Εύρεση βέλτιστου ρυθμού μάθησης

Στην παρακάτω σειρά πειραμάτων, δοκιμάστηκε η αρχιτεκτονική που αναφέρθηκε με όλες τις παραπάνω τιμές του ρυθμού μάθησης ώστε να βρεθεί η κατάλληλη και να συνεχιστούν τα πειράματα.

Learning Rate	**Ακρίβεια**
10^{-2}	57.2%
10^{-3}	62.3%
10^{-4}	64.8%

Πίνακας 4.13: Πίνακας Βελτιστοποίησης Learning Rate

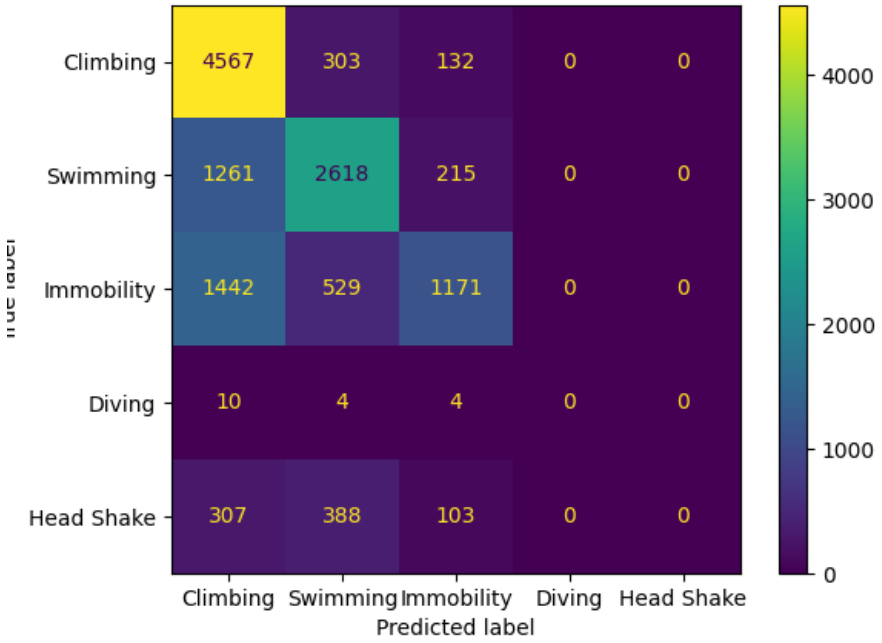
Η καλύτερη ακρίβεια δόθηκε με την επιλογή της τιμής 10^{-4} όπου η ακρίβεια στο Test Dataset ήταν 64.8%. Στην συνέχεια δίνεται ο πίνακας σύγχυσης και τα στατιστικά για το πείραμα αυτό.

Κατηγορία	Precision	Recall	F1-Score	Support
Climbing	0.60	0.91	0.72	5002
Swimming	0.68	0.63	0.32	4094
Immobility	0.72	0.59	0.5	3142
Diving	0	0	0	18
Head-Shake	0	0	0	798

Πίνακας 4.14: Πίνακας στατιστικών ταξινόμησης με 6 συνελικτικά επίπεδα και βελτιστοποιημένο learning rate .

	Precision	Recall	F1-Score
Macro Average	0	0.38	0.37
Weighted Average	0	0.69	0.60
Test Accuracy	64.8%		

Πίνακας 4.15: Πίνακας μακρο-στατιστικών και σταθμισμένων .



Σχήμα 4.15: Πίνακας σύγχυσης Testγια ρυθμό μάθησης 10^{-4}.

Σχολιασμός

Μετά την πρώτη σειρά πειραμάτων για την βελτιστοποίηση του δικτύου με απλές συνελίξεις τριών διαστάσεων είναι εμφανές ότι το βάθος του δικτύου παίζει σημαντικό ρόλο στην δυνατότητα ανίχνευσης προτύπων τόσο χωρικών αλλά και χρονικών. Από πλευρά ακρίβειας, είναι σημαντικό να τονιστεί ότι στα παραπάνω βρέθηκε ότι ο καλύτερος ρυθμός μάθησης ήταν η τιμή 10^{-4}. Μετά από προσεκτική εξέταση του πίνακα με τα στατιστικά του καλύτερου πειράματος, είναι φανερό ότι

το μοντέλο ήταν σε θέση να αναγνωρίσει σε ικανοποιητικό βαθμό τις τρεις κυρίαρχες κατηγορίες κίνησης του μοντέλου, αλλά όχι τόσο όσο στην προηγούμενη σειρά πειραμάτων όπου και τα στατιστικά από τον πίνακα 4.11 ήταν καλύτερα στις κατηγορίες αυτές. Επιπλέον η κατηγορία της Κολύμβησης ήταν βελτιωμένη σε σχέση με το αποτέλεσμα του δικτύου, αφού προστέθηκαν τα επίπεδα του Batch Normalization . Το ποσοστό επιτυχίας οφείλεται στην σταθερότερη επίδοση στην κατηγορία Climbing και είναι ίδιο με το προηγούμενο μοντέλο. Επίσης, δοκιμάστηκαν τα παραπάνω με διάφορους συνδυασμούς επιπέδων Batch Normalization τόσο σε όλα τα συνελικτικά επίπεδα, αλλά και σε κάποια από αυτά και τα καλύτερα αποτελέσματα δόθηκαν με την τοποθέτηση αυτού του είδους layer μετά από συνελικτικό επίπεδο(Convolution Layer) και πριν την συνάρτηση ενεργοποίησης. Έτσι από αυτό το σημείο και στο εξής, τα πειράματα θα αναφέρονται στην αρχιτεκτονική Batch Normalization σε όλα τα επίπεδα, αφού έδειξε τα περισσότερα σημάδια περαιτέρω βελτίωσης.

4.4.5.2 Βελτιστοποίηση μεγέθους εικόνας

Κατά την διαδικασία της φόρτωσης των εικόνων γίνεται resize καθώς και τυχαία αποκοπή της κάθε εικόνας που αποτελεί την ομάδα του δείγματος. Αυτό συμβαίνει για τις ανάγκες της Επαύξησης Δεδομένων (Data Augmentation). Στο πρώτο μέρος που αφορά τα απλά συνελικτικά δίκτυα, λόγω των περιορισμών σε μνήμη το μέγεθος των δειγμάτων πριν την τυχαία αποκοπή κρατήθηκε στο 120X120. Για την καλύτερη εκπαίδευση και την διερεύνηση της επίδρασης του μεγέθους αποκοπής, επιλέχθηκαν τιμές που ήταν ακραίες αλλά και οι θεωρητικά βέλτιστες, ώστε να βρεθεί το ποσοστό αποκοπής που δίνει τα καλύτερα αποτελέσματα για τα δίκτυα.

Όπως και παραπάνω, τα πειράματα αυτής της σειράς ξεκίνησαν με τιμές εικόνας στην τυχαία αποκοπή και στην συνέχεια οι τιμές αυξήθηκαν ώστε να βρεθεί η καλύτερη ακρίβεια που μπορεί να προκύψει. Παρακάτω δίνεται ο πίνακας με τις τιμές του Random Crop και της ακρίβειας με αρχικό μέγεθος εικόνας 136X136.

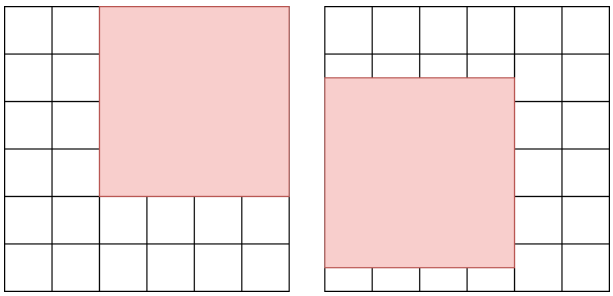
Random Crop Window Size	**Ακρίβεια**
62X72	61.3%
96X96	62.5%
120X120	70.1%

Πίνακας 4.16: Πίνακας Βελτιστοποίησης Random Crop Window

Σχολιασμός

Μετά το τέλος των πειραμάτων αυτής της σειράς για την βελτιστοποίηση του παραθύρου τυχαίας αποκοπής προέκυψαν οι ακρίβειες που δίνονται στον πίνακα 4.16. Κατά την διαδικασία της αποκοπής αυτής, γίνεται τυχαία επιλογή μιας περιοχής από την προεπιλεγμένη επιφάνεια και οι εικόνες έχουν αυτό το μέγεθος. Για την καλύτερη οπτικοποίηση της διαδικασίας, δίνεται παρακάτω παράδειγμα που περιγράφει την επικάλυψη του παραθύρου κατά την διαδικασία επιλογής και αποκοπής 2 δειγμάτων.

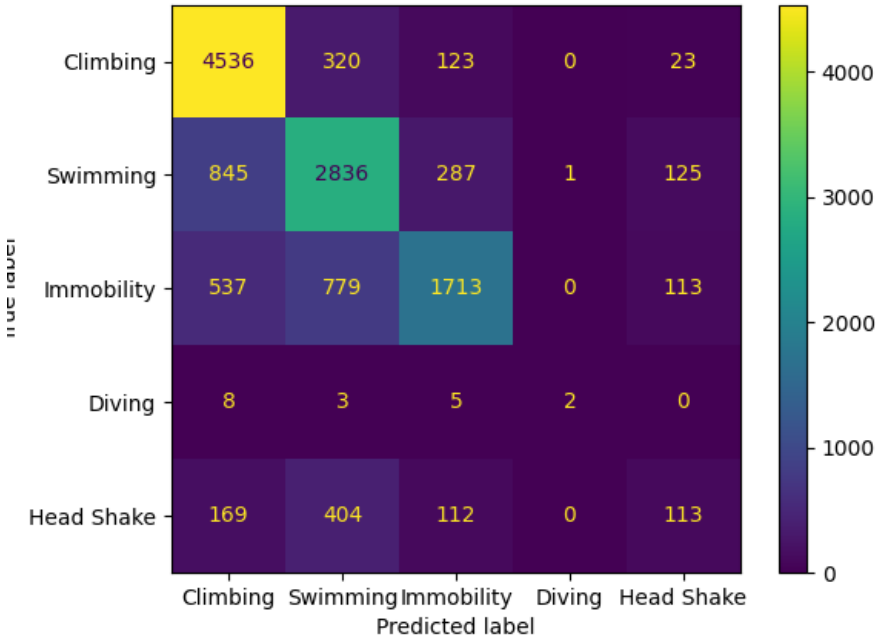
Από το σχήμα 4.16 είναι φανερό ότι κάθε φορά δεν κρατείται η ίδια περιοχή παρά ένα μόνο μέρος της. Τα πειράματα που πραγματοποιήθηκαν περιελάμβαναν δύο δεξαμενές με επιμύες και στην συνέχεια όπως αναφέρθηκε κατά την προετοιμασία του σετ δεδομένων, έγινε αποκοπή του κάθε πειράματος χωριστά. Αυτό οδήγησε στο να βρίσκεται ο επιμύς σε μια συγκεκριμένη περιοχή του βίντεο, την περιοχή ενδιαφέροντος, και σε όλη την υπόλοιπη εικόνα να περιέχεται μόνο πληροφορία που αφορά το δοχείο και δεν μπορεί να προσφέρει κάποια χρήσιμη πληροφορία για την αναγνώριση



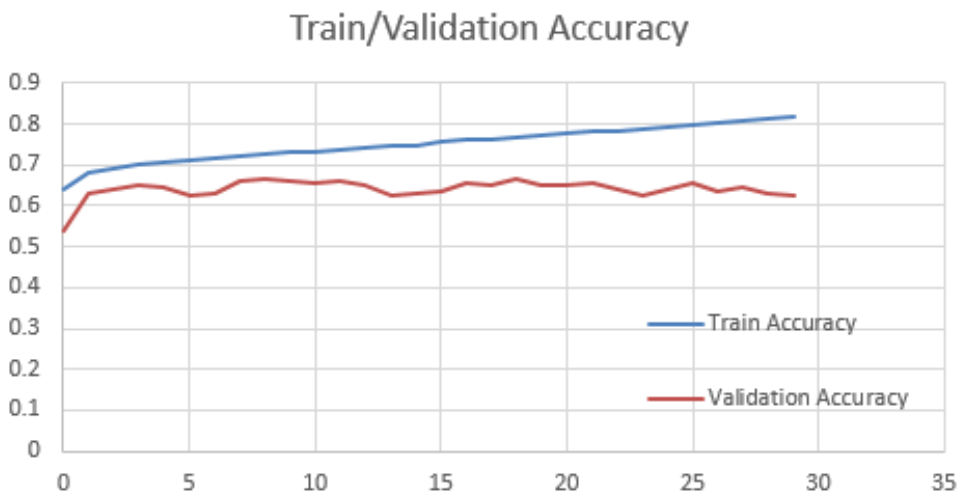
Σχήμα 4.16: Παράδειγμα random crop.

της δράσης του. Όταν γίνεται αποκοπή μεγάλης περιοχής των εικόνων που αποτελούν τα Frames του βίντεο, έχει ως αποτέλεσμα την αποκοπή περιοχής που περιλαμβάνει αναγκαία πληροφορία για την αναγνώριση της δράσης του πειραματόζωου.

Από τα πειράματα που έγιναν, βρέθηκε ότι η καταλληλότερη τιμή ήταν 120X120 εικονοστοιχεία για το παράθυρο τυχαίας αποκοπής. Αυτή η περιοχή σε σχέση με την αρχική εικόνα, αποτελεί περίπου το 90% της. Έτσι, μετά την βελτιστοποίηση αυτής της υπερπαραμέτρου του δικτύου είναι ασφαλές να εξαχθεί το συμπέρασμα, ότι για τα συγκεκριμένα δεδομένα θα πρέπει να χρησιμοποιηθεί αποκοπή του 10% της εικόνας για την καλύτερη εκπαίδευση του δικτύου. Παρακάτω δίνονται τα αποτελέσματα της καλύτερης τιμής random crop window με ακρίβεια 70.1%.



Σχήμα 4.17: Πίνακας σύγχυσης μετά από βελτιστοποίηση του random crop window .



Σχήμα 4.18: Γραφική παράσταση Train/Validation Accuracy μετά από βελτιστοποίηση του random crop window .

Κατηγορία	**Precision**	**Recall**	**F1-Score**	**Support**
Climbing	0.74	0.90	0.81	5002
Swimming	0.65	0.69	0.67	4094
Immobility	0.76	0.54	0.63	3142
Diving	0.66	0.11	0.19	18
Head-Shake	0.30	0.14	0.19	798

Πίνακας 4.17: Πίνακας στατιστικών μετά από βελτιστοποίηση του random crop window .

	Precision	**Recall**	**F1-Score**
Macro Average	0.62	0.47	0.50
Weighted Average	0.69	0.69	0.68
Test Accuracy	70.1%		

Πίνακας 4.18: Πίνακας μακρο-στατιστικών και σταθμισμένων μετά από βελτιστοποίηση του random crop window .

4.4.5.3 Βελτιστοποίηση επιπέδου Dropout

Στην παράγραφο αυτή θα παρουσιαστούν τα πειράματα που εκτελέστηκαν για την εύρεση της καλύτερης τιμής για την παράμετρο p του επιπέδου Dropout. Η χρήση αυτού του επιπέδου κρίνεται απαραίτητη ώστε να αντιμετωπιστεί το φαινόμενο της υπερ-προσαρμογής. Το επίπεδο αυτό όταν χρησιμοποιείται στο δίκτυο, στην φάση της εκπαίδευσης και ανάμεσα στα βήματα, βγάζει εκτός κάποιο ποσοστό νευρώνων με στόχο την καλύτερη εκπαίδευση του. Το ποσοστό που αποσυνδέεται κάθε φορά από το δίκτυο έως και αυτήν στιγμή, τα πειράματα, είχαν σταθερή τιμή και ίση με $p = 0.2$.

Στον παρακάτω πίνακα, δίνονται οι τιμές με τις οποίες εκπαιδεύτηκε ξανά το δίκτυο ώστε να βρεθεί η βέλτιστη τιμή p για την οποία το δίκτυο δίνει την καλύτερη ακρίβεια, κρατώντας όλες τις υπόλοιπες παραμέτρους σταθερές.

p	**Test Accuracy**
0.6	53.7%
0.5	68.7%
0.4	70.2%
0.3	67.1%

Πίνακας 4.19: Πίνακας Βελτιστοποίησης επιπέδου Dropout .

4.4.6 Σχολιασμός Πρώτης Σειράς Πειραμάτων

Από τα παραπάνω είναι εύκολο να εξαχθεί το συμπέρασμα ότι το μοντέλο βελτιώθηκε στην αναγνώριση όλων των κατηγοριών. Ιδιαίτερη βελτίωση έχει υπάρξει στην αναγνώριση των κινήσεων της Κατάδυσης και του Τινάγματος Κεφαλής. Όπως και στα παραπάνω πειράματα, παρατηρείται μεγάλη ανισορροπία ανάμεσα στις κλάσεις που το δίκτυο καλείται να αναγνωρίσει. Από τα μέχρι τώρα αποτελέσματα των πειραμάτων φαίνεται ότι το δίκτυο είναι σε θέση να αναγνωρίσει τις κυρίαρχες κατηγορίες δράσεων με σχετικά καλή ακρίβεια. Από τα πειράματα συμπεραίνεται ότι η πρόσθεση επιπέδων είναι σε θέση να βοηθήσουν στην αναγνώριση της κίνησης του επιμύ. Όμως όπως ήταν εμφανές, η ακρίβεια στο σύνολο των δεδομένων δοκιμής, μπορεί να αυξηθεί όταν γίνεται καλύτερη αναγνώριση στις κυρίαρχες κατηγορίες αλλά όταν η επίδοση σε αυτήν του Head-Shake αυξάνεται δεν αντικατοπτρίζεται άμεσα στην ακρίβεια αυτή. Οι δύο υπολοιπόμενες κατηγορίες της Κατάδυσης και του Τινάγματος Κεφαλής έχουν εμφανώς βελτιωθεί και θα αποτελέσουν αντικείμενο μελέτης στην επόμενη ενότητα.

4.5 Αρχιτεκτονικές Residual Neural Networks

Σε αυτήν την ενότητα θα εξεταστεί η δυνατότητα εφαρμογής αρχιτεκτονικών που δημοσιεύτηκαν στην [40]. Πρόκειται για τις αρχιτεκτονικές ResNet3D και R(2+1)D . Αυτές κάνουν χρήση τόσο συνελικτικών επιπέδων τριών διαστάσεων αλλά και συνδυασμού 2+1, δηλαδή 1xdxd ακολουθούμενες από μία ακόμα σε με μέγεθος tx1x1 . Επιπλέον τα δίκτυα αυτά κάνουν χρήση των Skip Connections . Αυτό όπως περιγράφτηκε και σε προηγούμενο, τους δίνει την δυνατότητα να αξιοποιούν περισσότερα επίπεδα και πλέον να γίνεται λόγος για Deep Neural Networks , και να είναι δυνατή η αντιμετώπιση του φαινομένου Vanishing Gradient . Από αυτές τις αρχιτεκτονικές που αναφέρθηκαν παραπάνω αναμένεται μεγαλύτερη ακρίβεια και είναι σημαντικό να διερευνηθεί η δυνατότητα τους να αναγνωρίσουν τόσο ποσοτικές κινήσεις,δηλαδή αυτές που χαρακτηρίζονται από την διάρκειά τους, αλλά και ποιοτικές όπως αυτή του Head Shake που αποτελεί μια καθαρά ποιοτική κίνηση και στόχος του πειράματος της εξαναγκασμένης κολύμβησης είναι η αναγνώριση και ταξινόμηση, μεταξύ άλλων, του αριθμού των μεμονωμένων στιγμιαίων τιναγμάτων.

4.5.1 Αρχιτεκτονική ResNet3D

Στην ενότητα αυτή πραγματοποιήθηκαν πειράματα με την χρήση της αρχιτεκτονικής ResNet3D . Τα συμπεράσματα που θα παρθούν από τα πειράματα που αφορούν το συγκεκριμένο δίκτυο, θα χρησιμοποιηθούν για την καλύτερη εκπαίδευση και επιλογή υπερ-παραμέτρων στην επόμενη αρχιτεκτονική R(2+1)D. Η αρχιτεκτονική ResNet3D αποτελείται από 5 ομάδες συνελικτικών επιπέδων σε τρεις διαστάσεις. Η [40] καθορίζει ότι το δίκτυο θα έχει στο σύνολό του 18 συνελικτικά επίπεδα και θα περιλαμβάνονται Skip Connections .

4.5.1.1 Pre-trained ResNet3D

Για την εκπαίδευση του δικτύου, χρησιμοποιήθηκαν οι βέλτιστες παράμετροι που δίνονται στην [40] και έδωσαν την καλύτερη δυνατή ακρίβεια σε βίντεο RGB .

Υπερ-παράμετρος	**Τιμή**
Μέγεθος τυχαίου παραθύρου αποκοπής για εκπαίδευση	116X116
Ρυθμός εκμάθησης	10^{-4}
Μέγεθος εικόνας	120X120
Χρονική διάρκεια δείγματος	32 frames
Batch Size	16
fps	25
Number of Segments	16
Frames per Segment	1

Πίνακας 4.20: Πίνακας υπερ-παραμέτρων αρχιτεκτονικής ResNet3D

Στον πίνακα 4.20 φαίνονται οι υπερπαράμετροι που χρησιμοποιήθηκαν για την εκπαίδευση του παραπάνω δικτύου. Στο πείραμα αυτό χρησιμοποιήθηκε το συγκεκριμένο δίκτυο ως προεκπαιδευμένο στο Dataset Kinetics 400 , με το σκεπτικό ότι είναι δυνατό το Transfer Learning και η αρχικοποίηση των βαρών να έχει θετικό αντίκτυπο στην ακρίβεια και να βοηθήσει στην μείωση του χρόνου εκπαίδευσης. Όπως βρέθηκε και στα παραπάνω, το μέγεθος της εικόνας σε σχέση με το παράθυρο της τυχαίας αποκοπής, πρέπει να διαφέρουν 10% για τα καλύτερα δυνατά αποτελέσματα στο σετ δεδομένων. Ο ρυθμός μάθησης επιλέχθηκε να είναι 10^{-4} καθώς και από την δημοσίευση που παρατέθηκε, τα αποτελέσματα ήταν καλύτερα όσο αυτός μειωνόταν, αλλά για να εξασφαλιστεί η επίδοση του δικτύου έγινε και σειρά πειραμάτων με ρυθμό 10^{-3}. Το Batch Size ήταν το βέλτιστο που μπορούσε να επιλεχθεί δεδομένων των περιορισμών σε μνήμη της GPU που ήταν διαθέσιμη την στιγμή των πειραμάτων. Τέλος, από τα δεδομένα επιλέχθηκε να χρησιμοποιηθούν 32 Frames και να γίνει επιλογή 16 από ισάριθμα segments κατά τον τρόπο που αναφέρθηκε στην αντίστοιχη ενότητα. Μετά την εκπαίδευση, δίνονται παρακάτω τα αποτελέσματα.

Κατηγορία	**Precision**	**Recall**	**F1-Score**	**Support**
Climbing	0.79	0.87	0.83	5002
Swimming	0.68	0.71	0.7	4094
Immobility	0.73	0.65	0.69	3142
Diving	0.6	0.16	0.26	18
Head-Shake	0.48	0.27	0.35	798

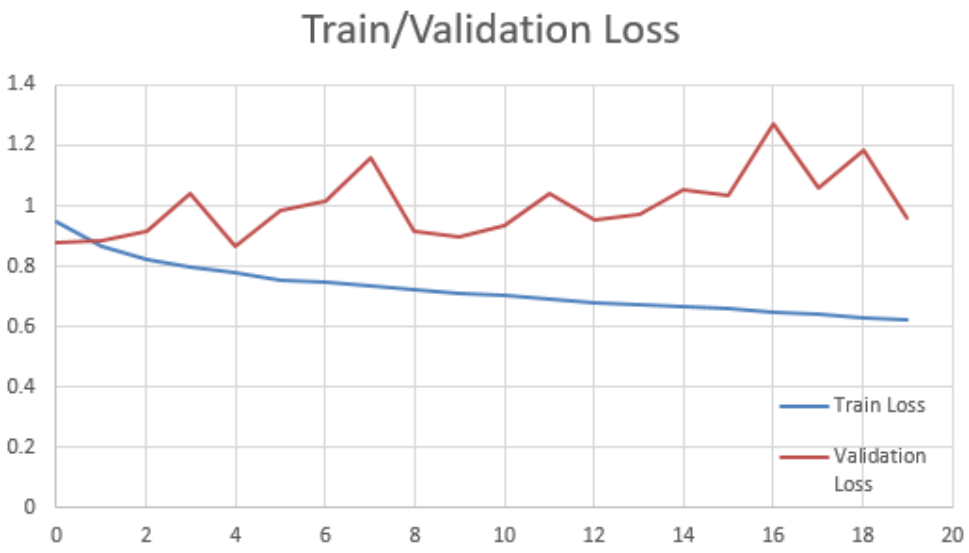
Πίνακας 4.21: Πίνακας στατιστικών Pretrained ResNet3D .

	Precision	**Recall**	**F1-Score**
Macro Average	0.65	0.54	0.56
Weighted Average	0.72	0.73	0.72
Test Accuracy	73.5%		

Πίνακας 4.22: Πίνακας μακρο-στατιστικών και σταθμισμένων ResNet3D .

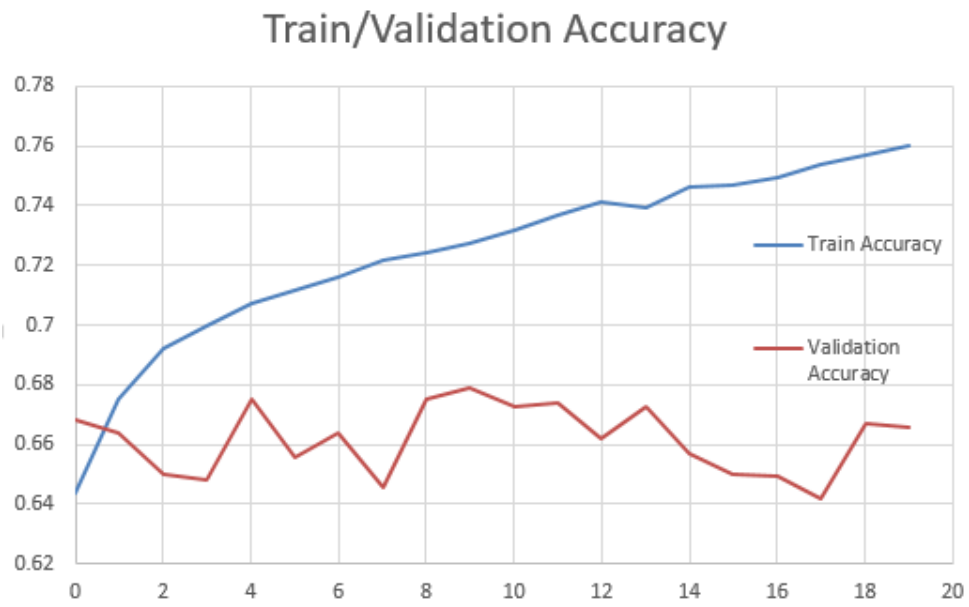
Από τα παραπάνω διαγράμματα είναι εμφανές ότι μετά από ένα σημείο το δίκτυο εμφανίζει υπερπροσαρμογή. Αυτό φαίνεται από το γεγονός ότι μετά από ένα σημείο στο Validation Dataset η

Σχήμα 4.19: Πίνακας σύγχυσης Pretrained ResNet3D .



Σχήμα 4.20: Διάγραμμα Train/Validation Accuracy Pretrained ResNet3D.

ακρίβεια δεν μεγαλώνει αλλά αντίθετα μειώνεται. Επίσης, η υπερπροσαρμογή φαίνεται από το ότι το κόστος μετά από ένα σημείο δεν μειώνεται αλλά αυξάνεται.



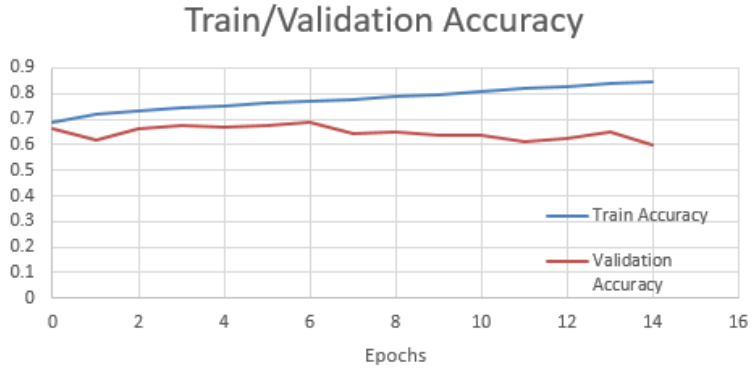
Σχήμα 4.21: Διάγραμμα Train/Validation Loss Pretrained ResNet3D.

Σχολιασμός

Τα πρώτα πειράματα με την χρήση Residual CNNs έδωσαν αρκετά καλά αποτελέσματα σε ό,τι αφορά την ταξινόμηση, ιδιαίτερα των κινήσεων που αφορούν τις κυρίαρχες κατηγορίες του σετ δεδομένων. Η κατηγορία Head Shake χωρίς κάποια βελτιστοποίηση βρίσκεται κοντά στο 30% recall.

4.5.1.2 Untrained ResNet3D

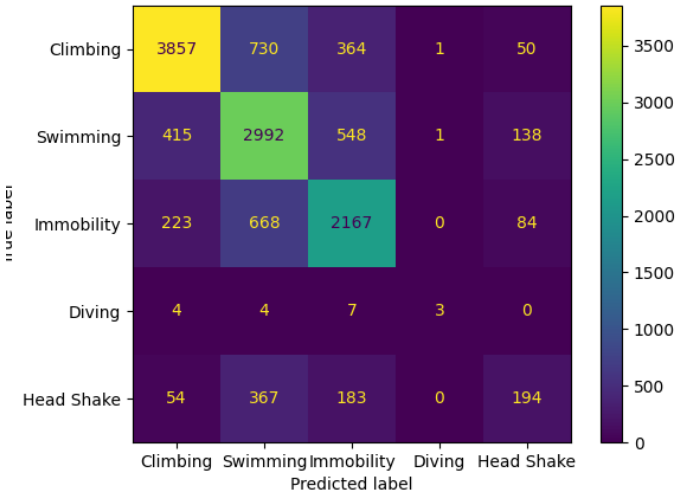
Στο επόμενο πείραμα, χρησιμοποιήθηκε το ίδιο δίκτυο και έγινε εκπαίδευση από την αρχή(From Scratch). Στόχος του πειράματος αυτού είναι να εξαχθούν χρήσιμα συμπεράσματα για την συμπεριφορά του δικτύου, αλλά και για το πόσο σταθερή είναι η διαδικασία την εκπαίδευσης μέσα από το διάγραμμα Train/Validation Loss ResNet3D .



Σχήμα 4.22: Διάγραμμα Train/Validation Accuracy Untrained ResNet3D .

Σχήμα 4.23: Διάγραμμα Train/Validation Loss Untrained ResNet3D .

Στα Σχήματα 4.22 και 4.23 φαίνονται τα διαγράμματα που προέκυψαν από την διαδικασία της εκπαίδευσης του μοντέλου. Γενικά η διαδικασία ήταν σχετικά σταθερή όπως φαίνεται και από τα παραπάνω. Η ακρίβεια στο σετ δεδομένων Test ήταν 70.5%.



Σχήμα 4.24: Πίνακας σύγχυσης Train/Validation Loss ResNet3D .

Ο πίνακας σύγχυσης δείχνει ότι υπάρχει σημαντική διαφορά ανάμεσα από τις δύο ταξινομήσεις που εκτελέστηκαν με την αρχιτεκτονική ResNet3D . Ήδη από το Σχήμα 4.24 φαίνεται ότι η κατηγορία του Head Shake δεν έχει ταξινομηθεί με την ίδια επιτυχία όπως και παραπάνω όπου το μοντέλο είναι προ-εκπαιδευμένο.

	Precision	**Recall**	**F1-Score**
Macro Average	0.63	0.52	0.54
Weighted Average	0.70	0.70	0.70
Test Accuracy	70.5%		

Πίνακας 4.24: Πίνακας μακρο-στατιστικών και σταθμισμένων Untrained ResNet3D.

Κατηγορία	Precision	Recall	F1-Score	Support
Climbing	0.84	0.77	0.80	5002
Swimming	0.62	0.73	0.67	4094
Immobility	0.66	0.68	0.67	3142
Diving	0.60	0.16	0.26	18
Head-Shake	0.41	0.24	0.30	798

Πίνακας 4.23: Πίνακας στατιστικών Untrained ResNet3D.

4.5.1.3 Βελτιστοποίηση ρυθμού μάθησης

Μια βασική υπερπαράμετρος του δικτύου είναι ο ρυθμός μάθησης. Όπως και παραπάνω, κρίνεται απαραίτητο να γίνει βελτιστοποίηση ώστε να βρεθεί ο καλύτερος ρυθμός, για τον οποίο το δίκτυο δίνει την βέλτιστη ακρίβεια. Τα πειράματα έγιναν με τις παρακάτω τιμές:

Learning Rate	Test Accuracy
10^{-3}	55.3%
10^{-4}	73.5%
10^{-5}	73.3%

Πίνακας 4.25: Πίνακας Βελτιστοποίησης ρυθμού μάθησης.

Τα πειράματα για την βελτιστοποίηση του ρυθμού μάθησης έδειξαν ότι για την καλύτερη εκδοχή του δικτύου ο βέλτιστος ρυθμός μάθησης ήταν 10^{-4}. Το παραπάνω ήταν αναμενόμενο διότι τα βάρη του προεκπαιδευμένου δικτύου δεν πρέπει να αλλάξουν δραστικά και να ακυρωθεί η προηγούμενη εκπαίδευση που το δίκτυο είχε λάβει. Αυτό θα χρησιμοποιηθεί και για τις δοκιμές της αρχιτεκτονικής R(2+1)D όπου και η ακρίβεια αναμένεται καλύτερη.

4.5.1.4 Βελτιστοποίηση μεγέθους εικόνας

Στην συνέχεια εκτελέστηκαν πειράματα για την εύρεση του βέλτιστου μεγέθους εικόνας για την αρχιτεκτονική αυτή. Τα αποτελέσματα είναι τα παρακάτω:

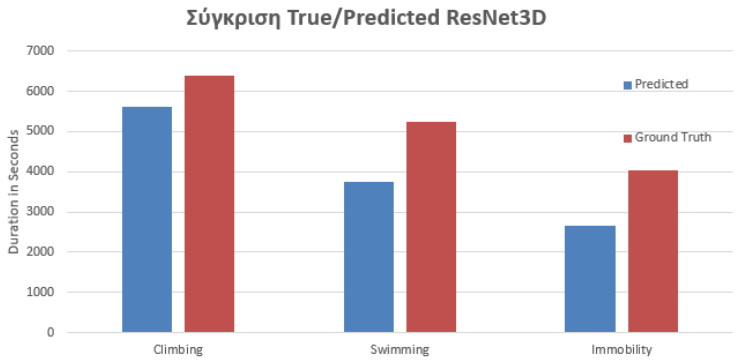
Frame Size	Test Accuracy
96X96	70.5%
112X112	70,8%
120X120	73.5%

Πίνακας 4.26: Πίνακας Βελτιστοποίησης μεγέθους εικόνας.

Από τα πειράματα που έγιναν για την βελτιστοποίηση του μεγέθους εικόνας, προέκυψε ότι η καλύτερη ακρίβεια για το δίκτυο αυτό δόθηκε με την υπερπαράμετρο μεγέθους εικόνας 120X120, όπως και στην αρχική μελέτη. Για τα πειράματα που θα ακολουθήσουν στο επόμενο δίκτυο, θα κρατηθεί αρχικά το ίδιο μέγεθος εικόνας ως αφετηρία για την βελτιστοποίηση του δικτύου.

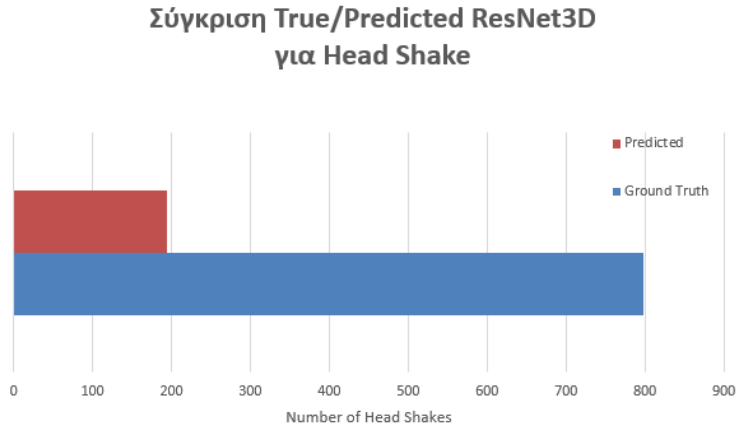
Σχολιασμός

Από τα παραπάνω είναι σαφής η βελτίωση των αποτελεσμάτων ειδικότερα στις 2 υπόλοιπες κατηγορίες. Πιο συγκεκριμένα τα Precision για όλες τις κατηγορίες είναι πάνω από το 45%. Η κατηγορία του Head Shake είναι ιδιαίτερα ενδιαφέρουσα, αφού η καλύτερη αναγνώρισή της δίνει την ικανότητα του δικτύου να ανιχνεύσει πρότυπα που έχουν μικρή χρονική διάρκεια, όπως αυτό της κίνησης του τινάγματος κεφαλής. Από τα αποτελέσματα η εικόνα για την επίδοση του μοντέλου είναι ότι είναι σε θέση να αναγνωρίσει καλύτερα από τα προηγούμενα τις κατηγορίες με μικρότερο αριθμό δειγμάτων, διατηρώντας την ικανοποιητική ακρίβεια για τις άλλες τρεις κατηγορίες των Climbing, Swimming και Immobility . Για την καλύτερη οπτικοποίηση των αποτελεσμάτων, είναι αναγκαία η μετατροπή, με βάση την επιλογή των υπερπαραμέτρων, των προβλέψεων σε πραγματικό χρόνο. Παρακάτω δίνεται το διάγραμμα με τις τρεις επικρατέστερες κατηγορίες:



Σχήμα 4.25: Συγκριτικό διάγραμμα Pred/Ground Truth από ResNet3D .

Στο σχήμα 4.25 φαίνονται συγκεντρωτικά τα αποτελέσματα των προβλέψεων για τις κατηγορίες Climbing, Swimming και Immobility , σε σύγκριση με τις χειροκίνητες ταξινομήσεις των παρατηρητών. Οι κατηγορίες αυτές απομονώθηκαν για δύο λόγους, αφενός γιατί είναι τρεις κατηγορίες που έχουν αρκετό ενδιαφέρον κατά την μελέτη της επίδρασης των ουσιών αλλά και γιατί χαρακτηρίζονται από την διάρκεια μέσα στο πείραμα. Επιπλέον, δίνεται αντίστοιχο διάγραμμα που απεικονίζει τα στοιχεία για την κατηγορία Head Shake .



Σχήμα 4.26: Συγκριτικό διάγραμμα Pred/Ground Truth από ResNet3D κατηγορίας Head Shake .

Στο παραπάνω διάγραμμα είναι εμφανής η αδυναμία του δικτύου να αναγνωρίσει σε βαθμό αντίστοιχο με τις άλλες τρεις κατηγορίες το τίναγμα κεφαλής του επιμύ. Αν και η επίδοση ήταν σημαντικά

βελτιωμένη σε σχέση με τα προηγούμενα, φαίνεται ότι υπήρξε σημαντική σύγχυση ειδικά σε αυτήν την κατηγορία. Η ταξινόμηση των κατηγοριών με το δίκτυο προεκπαιδευμένο κρίνεται ότι ήταν καλύτερη από όλα τα υπόλοιπα πειράματα. Ακόμα μετά από σειρά πειραμάτων επί της συγκεκριμένης τεχνικής δεν έδωσαν καλύτερα αποτελέσματα, οπότε θα χρησιμοποιηθούν αυτά τα αποτελέσματα για την παραπάνω διερεύνηση της αρχιτεκτονικής που ακολουθεί παρακάτω.

4.5.2 Αρχιτεκτονική R(2+1)D

Η αρχιτεκτονική αυτή ονομάζεται R(2+1)D , γιατί κάνει χρήση τρισδιάστατων συνελικτικών φίλτρων αλλάζοντας την χωρική και χρονική τους διάσταση για την αναγνώριση προτύπων στα βίντεο. Για την σειρά πειραμάτων που αφορούν αυτό το δίκτυο, χρησιμοποιήθηκαν αρχικά οι παράμετροι που δίνονται παρακάτω στην πρώτη σειρά πειραμάτων για την βελτιστοποίηση και στόχο είχαν την διερεύνηση των δυνατοτήτων του μοντέλου τόσο σε προεκπαιδευμένο αλλά και μη. Στην συνέχεια έγιναν πειράματα με τις αρχικές υπερπαραμέτρους που βρέθηκαν να δίνουν τα καλύτερα αποτελέσματα από το προηγούμενο (ResNet3D) και τα βάρη αρχικοποιήθηκαν με βάση την εκπαίδευση στο Dataset Kinetics 400 .

4.5.2.1 Αρχική εκπαίδευση δικτύου χωρίς αρχικοποιημένα βάρη(from scratch)

Στο πρώτο μέρος για την διερεύνηση των δυνατοτήτων του δικτύου, έγινε εκπαίδευση χωρίς να χρησιμοποιηθούν κάποια προηγούμενα βάρη από προηγούμενη εκπαίδευση. Στην περίπτωση αυτή, θα γίνει επιλογή των καλύτερων αρχικών υπερπαραμέτρων ώστε να διεξαχθεί η εκπαίδευση σε πρώτη φάση χωρίς την διαδικασία Transfer Learning . Οι υπερ-παράμετροι είναι οι εξής:

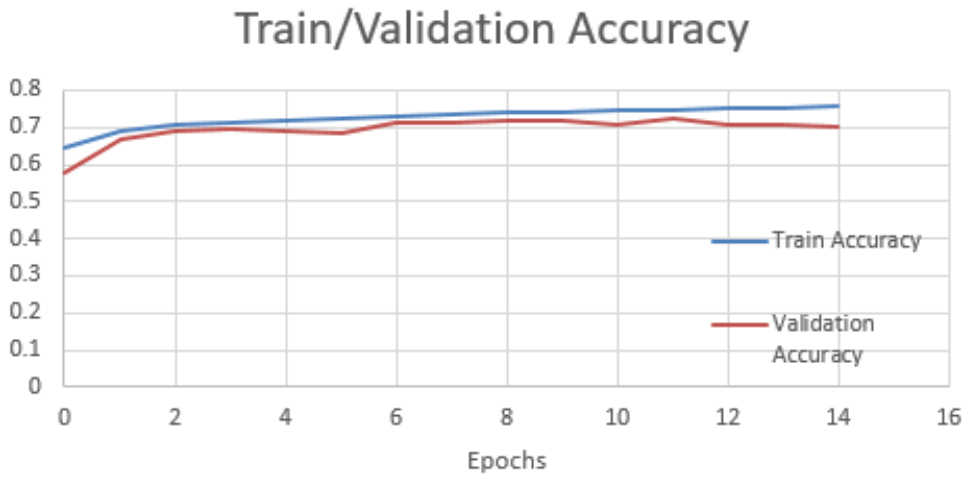
Υπερ-παράμετρος	**Τιμή**
Μέγεθος τυχαίου παραθύρου αποκοπής για εκπαίδευση	116X116
Ρυθμός εκμάθησης	10^{-3}
Μέγεθος εικόνας	120X120
Χρονική διάρκεια δείγματος	32 frames
Batch Size	16
fps	25
Number of Segments	16
Frames per Segment	1
Test Accuracy	63.5%

Πίνακας 4.27: Πίνακας υπερπαραμέτρων αρχιτεκτονικής R(2+1)D

Ο λόγος για τον οποίο το μέγεθος του δείγματος θα κρατηθεί ίδιος είναι γιατί μετά από πειραματισμό βρέθηκε πως η καλύτερη τιμή θα είναι τα 32 frames . Τόσο στην έρευνα που έχει δημοσιευτεί και αφορά τα συγκεκριμένα δίκτυα αλλά και από πειράματα που διεξήχθησαν και παραπάνω βρέθηκε ότι πρόκειται για την καλύτερη τιμή. Επιπλέον, οι κινήσεις που γίνεται προσπάθεια να αναγνωριστούν, αφορούν τόσο αυτές με σχετικά μεγάλη χρονική διάρκεια, αλλά και το τίναγμα κεφαλής που είναι μια στιγμιαία κίνηση και δεν θα ήταν δυνατό να αναγνωριστεί με μικρότερο μέγεθος δείγματος, δεδομένου και του resampling που έγινε κατά την μετατροπή των βίντεο σε frames .

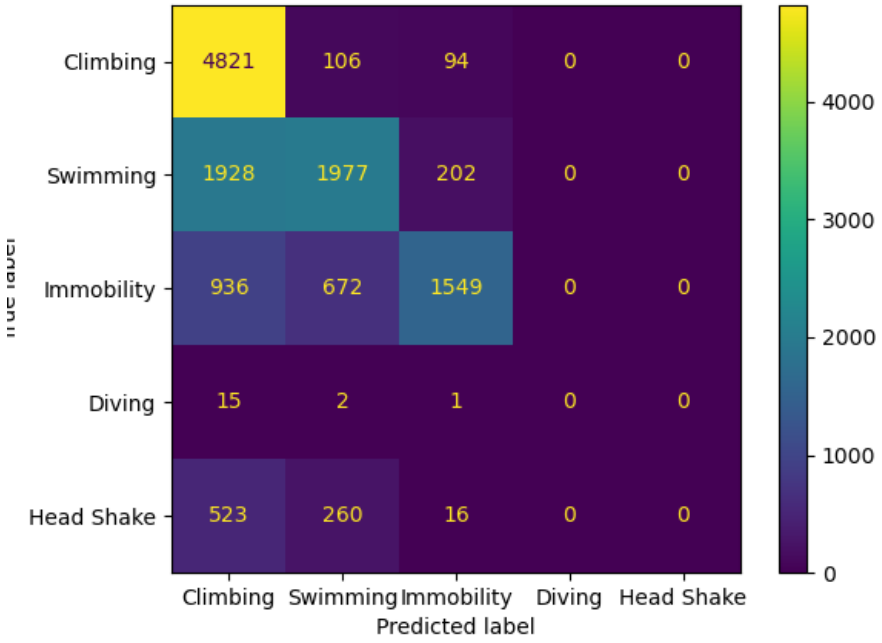
Τα αποτελέσματα από το πρώτο πείραμα, δίνονται στους παρακάτω πίνακες με τα στατιστικά και τον πίνακα σύγχυσης που προέκυψε από το Test Dataset .

Σχήμα 4.27: Γράφημα Train/Validation Loss για μοντέλο Untrained R(2+1)D.



Σχήμα 4.28: Γράφημα Train/Validation Accuracy για μοντέλο Untrained R(2+1)D.

Από τα αποτελέσματα είναι φανερό ότι το δίκτυο στην περίπτωση που πρόκειται για μη εκπαιδευμένο δεν έχει καταφέρει να αναγνωρίσει σε αποδεκτό βαθμό την κατηγορία του τινάγματος κεφαλής. Επίσης, το ίδιο συμβαίνει και για την κατηγορία της Κατάδυσης. Υπάρχει μεγάλη σύγχυση σε ότι αφορά αυτές τις 2 κατηγορίες με αυτές της Αναρρίχησης και της Κολύμβησης. Αυτό είναι δυνατό να ερμηνευτεί με βάση τα χαρακτηριστικά που μοιράζονται οι κατηγορίες αυτές και δεν διαχωρίστηκαν επαρκώς από το δίκτυο. Στις κατηγορίες Αναρρίχησης και Τινάγματος Κεφαλής, αλλά και κολύμβησης υπάρχει το στοιχείο της κίνησης. Το μοντέλο δεν ήταν σε θέση να αναγνωρίσει τις μικρές διαφορές ανάμεσα σε αυτές και να τις διαχωρίσει. Σε ό,τι αφορά την κατάδυση, είναι πιθανό να μην αναγνωρίστηκε σωστά η κάθετη κίνηση του επιμύ και για αυτόν τον λόγο να μην ήταν σε θέση να κατηγοριοποιήσει με σωστό τρόπο τις δύο αυτές κινήσεις. Στις επόμενες παραγράφους θα γίνει προσπάθεια για την βελτιστοποίηση του δικτύου σε ότι αφορά τις υπερπαραμέτρους ώστε να γίνει η καλύτερη προσαρμογή του δικτύου στις απαιτήσεις των πειραμάτων.



Σχήμα 4.29: Πίνακας σύγχυσης για μοντέλο Untrained R(2+1)D.

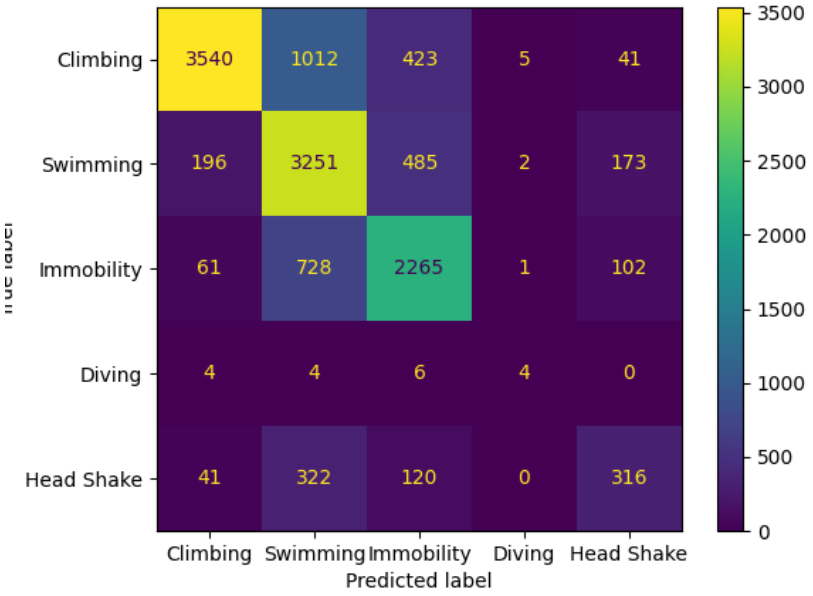
4.5.2.2 Βελτιστοποίηση ρυθμού μάθησης

Όπως ειπώθηκε παραπάνω τα πρώτα πειράματα έγιναν χρησιμοποιώντας το δίκτυο χωρίς προηγούμενη εκπαίδευση και συνεπώς δεν υπήρχε κάποια μεταφορά μάθησης από προηγούμενα σετ δεδομένων στο παρόν. Για τον λόγο αυτό το πρώτο πείραμα που δόθηκε παραπάνω έκανε χρήση ρυθμού μάθησης 10^{-3}. Στην συνέχεια, έγινε σειρά πειραμάτων, ώστε να βρεθεί ο βέλτιστος ρυθμός και να παρθούν τα καλύτερα δυνατά αποτελέσματα μετά από την βελτίωση αυτής της υπερ-παραμέτρου.

Learning Rate	**Test Accuracy %**
5^{-3}	71.5
10^{-3}	63.6
5^{-4}	62.9
10^{-4}	62.2

Πίνακας 4.28: Πίνακας βελτιστοποίησης ρυθμού μάθησης αρχιτεκτονικής R(2+1)D

Από τα παραπάνω επιβεβαιώνεται ότι κατά την εκπαίδευση του δικτύου από την αρχή είναι προτιμότερο να γίνεται επιλογή μεγαλύτερου ρυθμού από ότι όταν αυτό είναι προεκπαιδευμένο. Παρακάτω δίνεται ο πίνακας σύγχυσης για το καλύτερο πείραμα:



Σχήμα 4.30: Πίνακας σύγχυσης για μοντέλο Untrained R(2+1)D μετά την βελτιστοποίηση του ρυθμού μάθησης.

4.5.2.3 Βελτιστοποίηση χρονικού μεγέθους δείγματος

Όπως ειπώθηκε παραπάνω τα βίντεο περιλαμβάνουν πολλές κατηγορίες με διαφορετικά χαρακτηριστικά η κάθε μια σε ό,τι αφορά την χωρο-χρονική τους εξέλιξη. Για την επιλογή των δειγμάτων είναι σημαντικό να καθοριστεί η βέλτιστη επιλογή της χρονικής διάρκειας του δείγματος ώστε να επιτευχθεί η μέγιστη ακρίβεια στην ταξινόμηση.

Όλες οι κινήσεις του σετ δεδομένων εκτός του τινάγματος κεφαλής, αναμένεται να έχουν χρονική διάρκεια μεγαλύτερη από το 1,5 sec . Για τον λόγο αυτό επιλέχθηκε ως παράμετρος διάρκειας τα 1.3 sec ώστε να υπάρχει η δυνατότητα να εμπεριέχεται η κίνηση του τινάγματος κεφαλής, αλλά και να συνυπολογιστεί και ο χρόνος αντίδρασης του παρατηρητή. Για τις υπόλοιπες κινήσεις, η χρόνική διάρκεια, όπως αναφέρθηκε, εμπίπτει μέσα στα όρια του χρόνου αντίδρασης, σε προσπάθεια να αντιμετωπιστούν όποια σφάλματα προέρχονται από αυτόν. Για την επιβεβαίωση του παραπάνω μέσω πειραματικής διαδικασίας, πραγματοποιήθηκε σειρά πειραμάτων με στόχο την εύρεση του καλύτερου μεγέθους δείγματος:

Χρονικό μέγεθος Δείγματος (sec)	**Test Accuracy %**
2	68.2
1.3	71.5
1	67.5
0.8	64.1

Πίνακας 4.29: Πίνακας βελτιστοποίησης χρονικού δείγματος αρχιτεκτονικής R(2+1)D

Από τα παραπάνω εξάγεται το συμπέρασμα ότι πρέπει να κρατηθεί το δείγμα ως έχει και να συνεχιστούν τα πειράματα με αυτό το μέγεθος. Σε ό,τι αφορά την μη ικανοποιητική απόδοση με τις υπόλοιπες επιλογές, φαίνεται στην περίπτωση του μεγαλύτερου χρονικού δείγματος, υπάρχει μεγαλύτερη σύγχυση ανάμεσα στην κατηγορία που αναγνωρίζει το δίκτυο και στην εκτίμηση ground

truth . Επιπλέον, πιθανή είναι η ύπαρξη μικρής ποσότητας πληροφορίας, όταν το δείγμα είναι μικρότερο, ώστε να αναγνωριστούν τα υπάρχοντα πρότυπα και να ταξινομηθεί η δράση σωστά.

4.5.2.4 Βελτιστοποίηση μεγέθους εικόνας

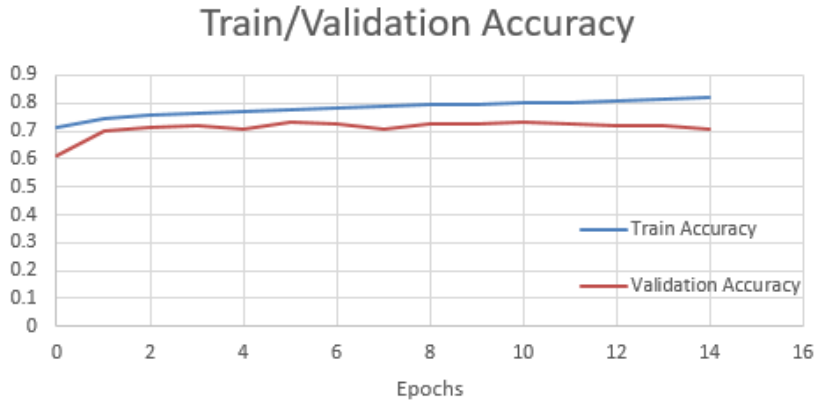
Στην συνέχεια έγινε διερεύνηση της επίδρασης του μεγέθους εικόνας στην δυνατότητα του δικτύου να αναγνωρίσει χωρο-χρονικά πρότυπα και να ταξινομήσει τις δράσεις στα σετ δεδομένων Validation/Test με την καλύτερη δυνατή ακρίβεια.

Τα πρώτα πειράματα που εκτελέστηκαν έκαναν χρήση μεγέθους εικόνας 120X120. Η παράμετρος αυτή επιλέχθηκε διότι μετά από μελέτη της αντίστοιχης δημοσίευσης που περιείχε την διαδικασία της βελτιστοποίησης, ήταν το πλέον κατάλληλο μέγεθος. Στην παράγραφο αυτή, θα εξεταστεί η πιθανότητα να επηρεάζεται η ικανότητα του μοντέλου να αναγνωρίζει τις δράσεις του επιμύ, ενώ είναι μη προεκπαιδευμένο, με διαφορετικά μεγέθη εικόνας και συγκεκριμένα μεγαλύτερα. Η αύξηση του μεγέθους της εικόνας είχε σημαντικό αντίκτυπο στην απαιτούμενη μνήμη της GPU. Για να μην υπάρχει σημαντική διαφοροποίηση και να εμφανιστεί η ανάγκη για επαναπροσαρμογή των υπερ-παραμέτρων για τα νέα μεγέθη έγινε χρήση κάρτας γραφικών Nvidia RTX 6000 24Gb που παραχωρήθηκε από το Εργαστήριο Τηλεπισκόπησης της ΣΑΤΜ-ΜΓ. Η χρήση της επέτρεψε την επιλογή μεγέθους παρτίδας έως και 32 για τα πειράματα που ακολούθησαν ώστε να γίνει μελέτη της επίδρασης του μεγέθους εικόνας στην ακρίβεια.

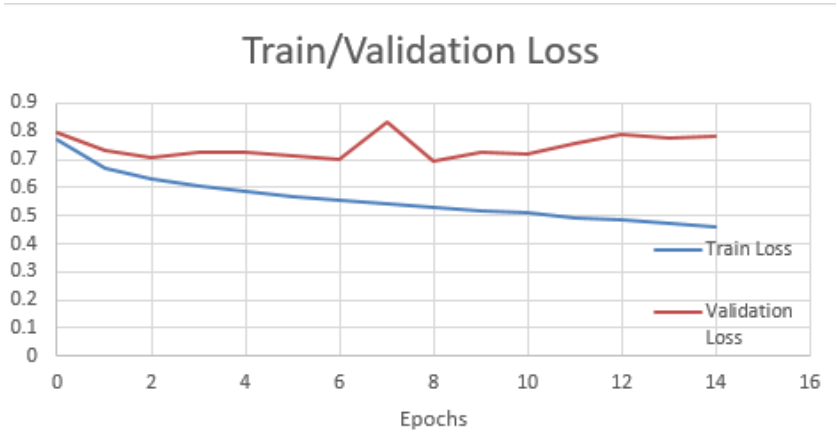
Μέγεθος Αρχικής εικόνας	**Test Accuracy %**
120X120	71.5
170X170	73.1
220X220	75.2
250X250	75.3

Πίνακας 4.30: Πίνακας βελτιστοποίησης μεγέθους εικόνας αρχιτεκτονικής Untrained R(2+1)D

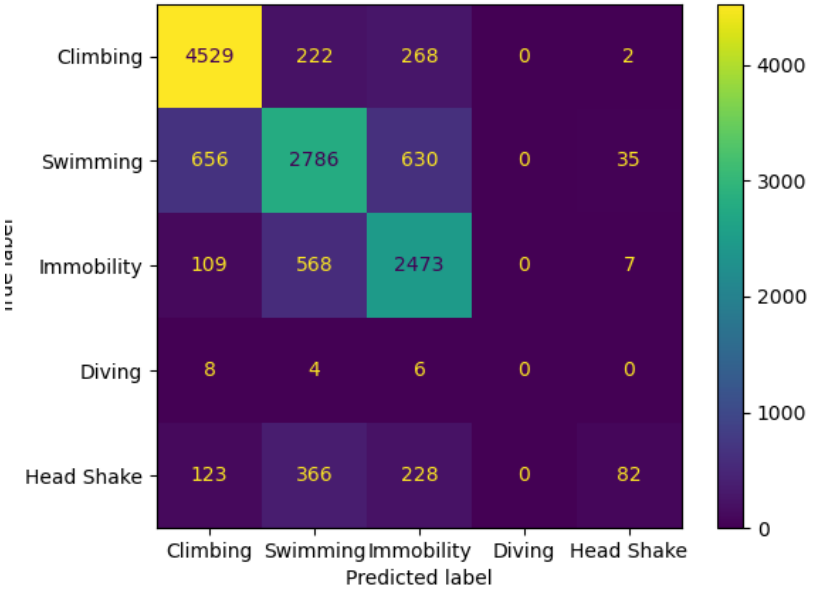
Στον παραπάνω πίνακα φαίνεται η διαφορά ανάμεσα στις εικόνες που δόθηκαν για την σειρά των πειραμάτων. Κατά την εκπαίδευση είναι πιθανό να υπάρχει διαφορά ανάμεσα στις κινήσεις που είναι εμφανείς ανάλογα με το μέγεθος εισόδου. Επίσης, ο συνδυασμός του μεγέθους και του μεγέθους παρτίδας κατά την εκπαίδευση είναι πιθανό να αλλάζουν την επίδοση του μοντέλου. Στο σημείο αυτό δίνονται τα στατιστικά μετά από το τέλος της σειράς πειραμάτων για το καλύτερο μοντέλο.



Σχήμα 4.31: Train/Validation Accuracy Untrained R(2+1)D μετά από βελτιστοποίηση μεγέθους εικόνας.



Σχήμα 4.32: Train/Validation Loss Untrained R(2+1)D μετά από βελτιστοποίηση μεγέθους εικόνας.



Σχήμα 4.33: Πίνακας σύγχυσης Untrained R(2+1)D μετά από βελτιστοποίηση μεγέθους εικόνας.

Επιπλέον δίνονται και οι πίνακες με τους σταθμισμένους μέσους όρους για τα στατιστικά και τα αντίστοιχα για κάθε κατηγορία.

Κατηγορία	Precision	Recall	F1-Score	Support
Climbing	0.83	0.90	0.86	5021
Swimming	0.70	0.67	0.69	4107
Immobility	0.68	0.78	0.73	3157
Diving	-	-	-	18
Head-Shake	0.65	0.10	0.17	799

Πίνακας 4.31: Πίνακας στατιστικών μετά από βελτιστοποίηση μεγέθους εικόνας.

	Precision	Recall	F1-Score
Macro Average	-	0.49	0.49
Weighted Average	-	0.75	0.73
Test Accuracy	75.3%		

Πίνακας 4.32: Πίνακας μακρο-στατιστικών και σταθμισμένων R(2+1)D μετά από βελτιστοποίηση μεγέθους εικόνας.

Σχολιασμός

Από τα παραπάνω είναι προφανές ότι η συνολική ακρίβεια στο Test Dataset δεν συνδέεται άμεσα με την ακρίβεια που μπορεί να επιτευχθεί στην κάθε μια κατηγορία χωριστά. Πιο συγκεκριμένα, στην περίπτωση που πάρθηκε ακρίβεια 75.3% η κατηγορία του Head Shake είχε recall 0.10, έναντι 0.39 και ακρίβεια 71.5%. Η απόδοση στις υπόλοιπες κατηγορίες κίνησης είναι εμφανώς βελτιωμένες και κοντά ή και ακόμα υψηλότερα από την ταξινόμηση που είναι πιθανό να πετύχουν αλλά και να συμφωνούν δύο παρατηρητές [52]. Οι κατηγορία της κατάδυσης δεν έδωσε αποτελέσματα και αυτό οφείλεται στην αδυναμία του μοντέλου να αναγνωρίσει επαρκώς τις διαφορές ανάμεσα από τις άλλες κατηγορίες και πιο συγκεκριμένα από τις Climbing, Swimming, Immobility.Οι κατηγορίες αυτές μοιράζονται χαρακτηριστικά όπως αναφέρθηκε και σε προηγούμενο και η αναγνώριση καθίσταται εξαιρετικά δύσκολη.

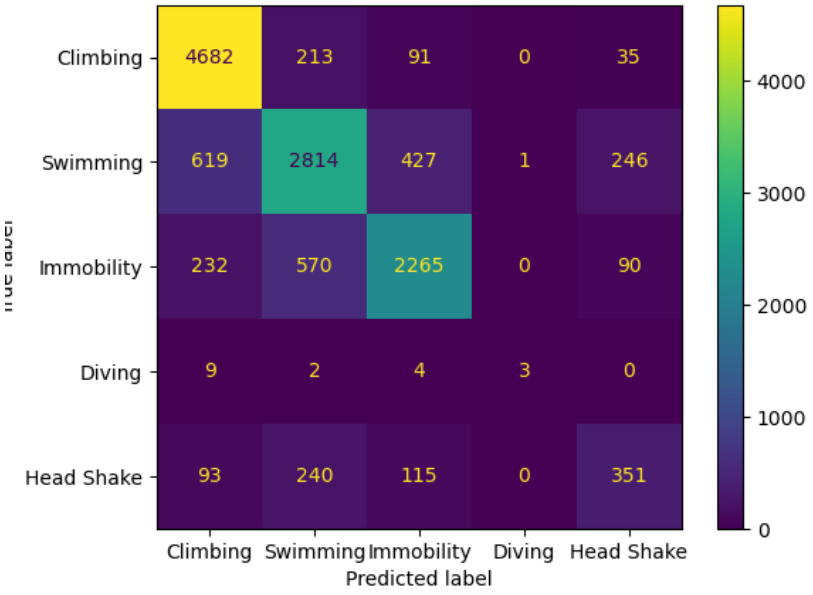
4.5.2.5 Transfer Learning

Στην συνέχεια θα διερευνηθεί εάν είναι εφικτή η εκμετάλλευση του δικτύου αυτού ως προεκπαιδευμένο σε άλλο σετ δεδομένων και η μεταφορά μάθησης στην περίπτωση των πειραμάτων της εξαναγκασμένης κολύμβησης για την αναγνώριση δράσεων των επιμύων. Το δίκτυο θα χρησιμοποιηθεί με αρχικοποιημένα τα βάρη του μετά από εκπαίδευση στο Dataset Kinetics 400. Το μοντέλο που θα χρησιμοποιηθεί όπως και παραπάνω έχει 18 επίπεδα όπως και παραπάνω, και η ακρίβεια που αναφέρθηκε στην αρχική μελέτη στο ίδιο σετ δεδομένων είναι 74.3%.

Για την σειρά πειραμάτων που θα ακολουθήσει θα χρησιμοποιηθούν οι ίδιες υπερ-παράμετροι που βρέθηκαν στο παραπάνω μέρος καθώς συμφωνούν αρχικά και με αυτές της αρχικής μελέτης αλλά είναι επιβεβλημένο να γίνει σωστό fine tuning του ρυθμού μάθησης, αφού πλέον γίνεται λόγος για ένα δίκτυο που έχει ήδη υποστεί εκπαίδευση και δεν πρέπει να γίνει δραστική αλλαγή των βαρών, παρά μόνο σωστή ρύθμιση ώστε να προσαρμοστεί στην ανίχνευση προτύπων που χαρακτηρίζουν το υπό μελέτη σετ δεδομένων.

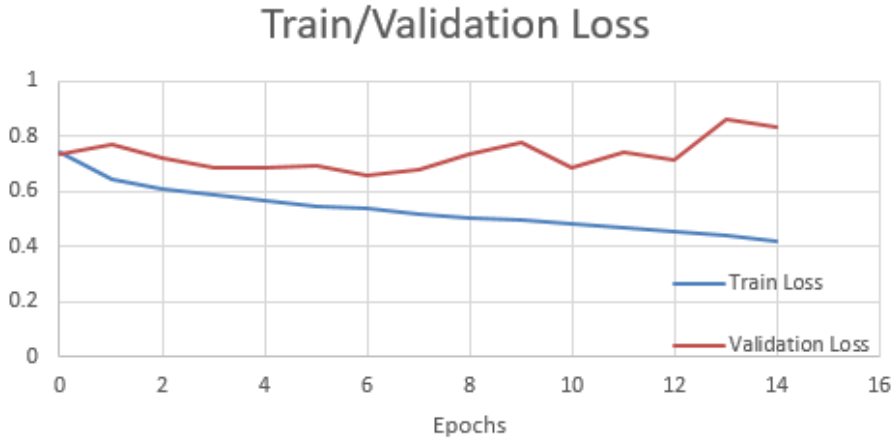
Learning Rate	Ακρίβεια
10^{-3}	75.6%
10^{-4}	76.4%
10^{-5}	77.2%

Πίνακας 4.33: Πίνακας Βελτιστοποίησης Learning Rate τελικού μοντέλου.

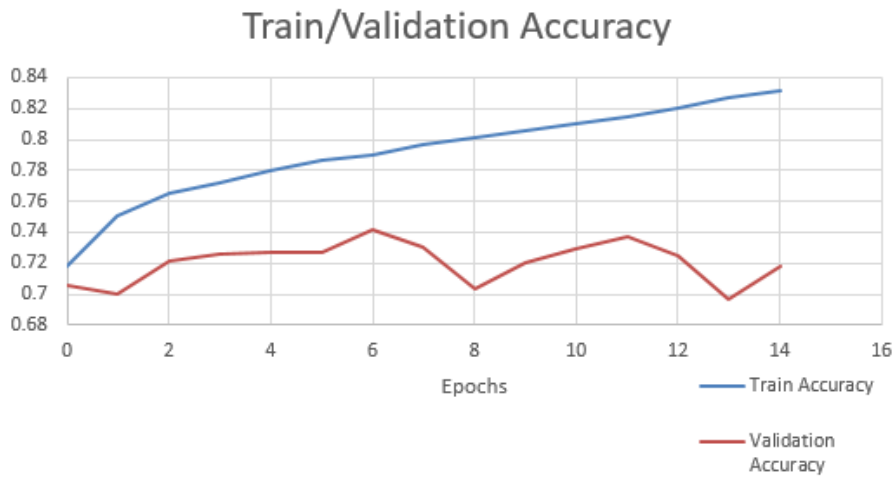


Σχήμα 4.34: Πίνακας σύγχυσης Train/Validation Accuracy για μοντέλο Pretrained R(2+1)D.

Ο παραπάνω πίνακας σύγχυσης, είναι το αποτέλεσμα του καλύτερου πειράματος που προέκυψε μετά από την επιλογή του καλύτερου ρυθμού μάθησης. Παρακάτω φαίνονται τα γραφήματα που πάρθηκαν από την εκπαίδευση του δικτύου.



Σχήμα 4.35: Διάγραμμα Train/Validation Loss για μοντέλο Pretrained R(2+1)D.



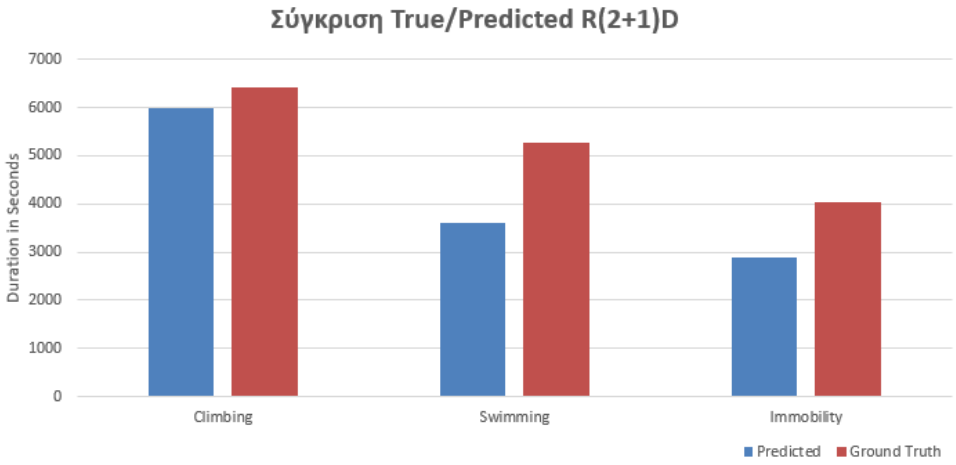
Σχήμα 4.36: Διάγραμμα Train/Validation Accuracy για μοντέλο Pretrained R(2+1)D.

Κατηγορία	**Precision**	**Recall**	**F1-Score**	**Support**
Climbing	0.83	0.93	0.87	5021
Swimming	0.73	0.68	0.70	4107
Immobility	0.78	0.71	0.74	3157
Diving	0.75	0.16	0.27	18
Head-Shake	0.48	0.43	0.46	799

Πίνακας 4.34: Πίνακας στατιστικών.

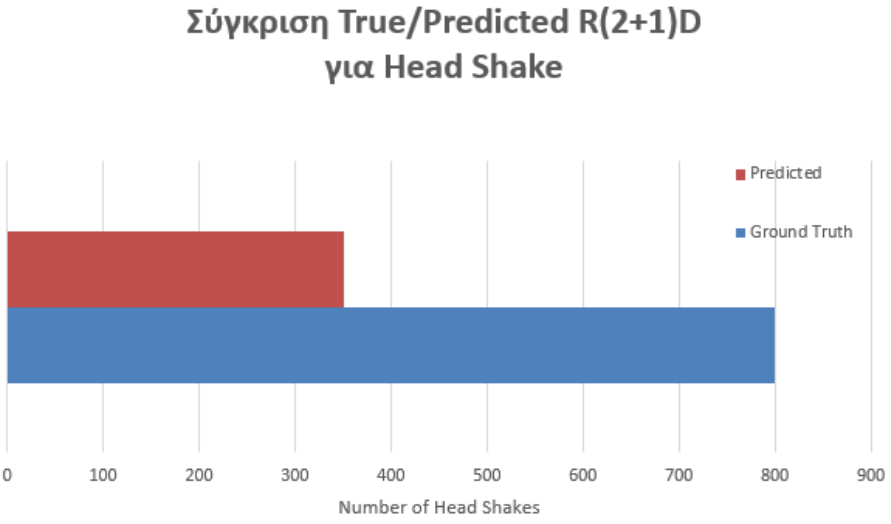
	Precision	**Recall**	**F1-Score**
Macro Average	0.71	0.58	0.61
Weighted Average	0.76	0.77	0.76
Test Accuracy	77.2%		

Πίνακας 4.35: Πίνακας μακρο-στατιστικών και σταθμισμένων



Σχήμα 4.37: Τελικά αποτελέσματα ταξινόμησης για μοντέλο Pretrained R(2+1)D.

Παραπάνω φαίνονται τα αποτελέσματα της τελικής ταξινόμησης με το προεκπαιδευμένο μοντέλο R(2+1)D. Το σχήμα 4.37 δείχνει τις τρεις κυρίαρχες κατηγορίες και την ταξινόμησή τους σε σχέση με τα δεδομένα Ground Truth. Η καλύτερη ακρίβεια όπως είναι φανερό και στον πίνακα 4.34 υπάρχει για την κατηγορία Climbing, όπου η ταξινόμηση είναι εξαιρετική και το recall στο 0.93. Ακολουθούν οι κατηγορίες κίνησης Swimming και Immobility με σχετικά καλή ταξινόμηση κοντά στα επίπεδα της συμφωνίας των παρατηρητών, σε περίπτωση που υπάρχουν τουλάχιστον 2. Η κατηγορία του Diving δεν δίνεται, διότι μετά από μελέτη των δεδομένων και πιο συγκεκριμένα του Support της δεν υπάρχουν αρκετά δείγματα ώστε να αναγνωριστεί σε σημαντικό βαθμό ακόμα και αν οι υπόλοιπες κατηγορίες έδειξαν βελτίωση στα στατιστικά τους. Εκτός αυτού, είναι μια κίνηση που δεν εμφανίζεται τόσο συχνά στο πείραμα της εξαναγκασμένης κολύμβησης και κατά την μελέτη των αποτελεσμάτων είναι δυνατό να παραληφθεί από την συνολική συμπεριφορά του ζώου, για την υπό μελέτη ουσία.



Σχήμα 4.38: Τελικά αποτελέσματα ταξινόμησης για Head Shakes, Pretrained R(2+1)D.

Σχολιασμός

Στα παραπάνω αποτελέσματα εξάγεται το συμπέρασμα ότι η ταξινόμηση ήταν επιτυχής για την πλειοψηφία των ταξινομήσεων μετά από την διαδικασία του Transfer Learning. Η διαδικασία αυτή προσέφερε την δυνατότητα να χρησιμοποιηθεί η εκπαίδευση σε ένα άλλο μεγάλο σετ δεδομένων και να γίνει αποτελεσματική αναγνώριση των δράσεων των επιμύων.

Πιο αναλυτικά, υπήρξε σημαντική βελτίωση στο recall ειδικά για την κατηγορία του Head Shake, γεγονός που δεν συνέβαινε στα υπόλοιπα πειράματα της σειράς. Γενικά, η εκπαίδευση του δικτύου σε εικόνες μεγαλύτερου μεγέθους, έδωσε την ικανότητα σε ένα βαθύτερο δίκτυο να αναγνωρίσει πρότυπα με μικρή χρονική διάρκεια. Η ικανότητα του δικτύου επηρεάστηκε σε μεγάλο βαθμό από την προηγούμενη εκπαίδευση σε σετ δεδομένων που έχει ήδη ταξινομηθεί με μεγάλη ακρίβεια ως προς τις δράσεις.

Η σημαντικότερη βελτίωση με την χρήση του προ-εκπαιδευμένου δικτύου ήταν στην αναγνώριση της κατηγορίας του Head Shake. Εδώ όπως αναφέρθηκε και παραπάνω, υπήρξε recall ίσο με 0.43. Αυτό, σε συνδυασμό με τα υψηλά ποσοστά στις υπόλοιπες 3 κατηγορίες των Swimming, Climbing και Immobility κάνουν το δίκτυο αυτό με την μορφή που δόθηκε το καλύτερο μέχρι στιγμής, μετά την βελτιστοποίηση και όλα τα πειράματα που εκτελέστηκαν. Επιπλέον, η ακρίβεια στο σετ δεδομένων Test ήταν στο ίδιο επίπεδο με την ακρίβεια που δόθηκε στην αρχική δημοσίευση και μετά την βελτιστοποίηση του δικτύου.

Κεφάλαιο 5

Συμπεράσματα-Μελλοντικές Επεκτάσεις

5.1 Συμπεράσματα

Στην παρούσα ΔΕ έγινε διερεύνηση των δυνατοτήτων των τεχνητών νευρωνικών δικτύων για την αναγνώριση δράσεων σε βίντεο, με την χρήση εικόνων RGB και τρισδιάστατων συνελικτικών φίλτρων. Η αναγνώριση δράσεων είναι ένα εξαιρετικά σημαντικό πεδίο έρευνας, με εφαρμογές στην ασφάλεια, στην Φαρμακολογία αλλά και την καθημερινότητα. Το πεδίο στο οποίο έγινε έρευνα στην ΔΕ αφορά την αναγνώριση δράσεων σε βίντεο εξαναγκασμένης κολύμβησης για την μελέτη αντικαταθλιπτικών φαρμάκων.

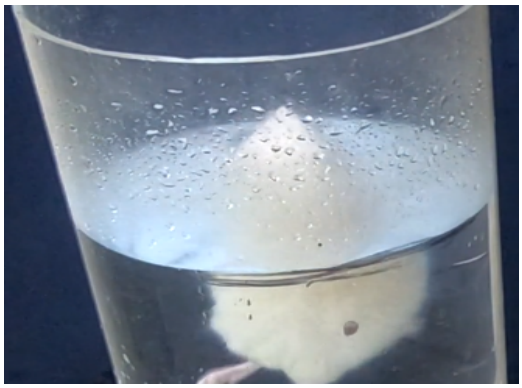
Η ταξινόμηση δράσεων των πειραματόζωων, είναι μια χρονοβόρα διαδικασία που απαιτεί γνώση των κινήσεων των επιμύων. Επιπλέον, τα δεδομένα συχνά αποτελούνται από βίντεο που ξεπερνούν σε διάρκεια τις 15 ώρες στο σύνολό τους ανά σειρά πειραμάτων εξαναγκασμένης κολύμβησης. Η αυτόματη αναγνώριση των δράσεων στα πειράματα εξαναγκασμένης κολύμβησης είναι σημαντική, διότι αφενός η διαδικασία επιταχύνεται σημαντικά συμπαρασύροντας την μελέτη των ουσιών με πολλαπλά ωφέλη για του μελετητές, αφετέρου η έρευνα συμβάλλει στην καλύτερη εξέλιξη του πεδίου της Αναγνώρισης Δράσεων εν γένει.

Η συλλογή και επεξεργασία των δεδομένων, ανέδειξε χαρακτηριστικά του σετ δεδομένων που έπαιξαν σημαντικό ρόλο στην ακρίβεια της ταξινόμησης, όπως δίνονται παρακάτω:

1. Ανισορροπία κλάσεων: Μετά την προετοιμασία του Dataset εξήχθησαν τα στατιστικά για κάθε κατηγορία που περιλαμβάνεται σε αυτό. Από τον πίνακα 4.2 φαίνονται αναλυτικά τα ποσοστά της κάθε κίνησης που το μοντέλο θα κληθεί να ταξινομήσει. Είναι φανερό ότι το μεγαλύτερο μέρος του σετ περιλαμβάνει δείγματα από τις κατηγορίες Climbing, Swimming και Immobility . Οι δύο υπολοιπόμενες κατηγορίες καλύπτουν μικρό μέρος των δεδομένων και ακόμα και μετά την τυχαία επιλογή τους για την καλύτερη κατανομή στα τρία Split ο αριθμός των δειγμάτων ήταν και πάλι μη ικανοποιητικός. Το γεγονός αυτό είναι καταλυτικής σημασίας για την εκπαίδευση και την τελική επίδοση των δικτύων που σχεδιάστηκαν, αφού σε όλες τις περιπτώσεις, που αφορούν τα βελτιστοποιημένα μοντέλα, παρατηρήθηκαν στατιστικά όπως η ακρίβεια πάνω του 70%.
2. Η χειροκίνητη ταξινόμηση αποτελεί σημαντικό παράγοντα για την επιτυχία της ταξινόμησης από το μοντέλο που βρίσκεται κάθε φορά υπό δοκιμή. Αρχικά, μελετήθηκε ο χρόνος

αντίδρασης του παρατηρητή με μελέτη δημοσιεύσεων που αφορούν το θέμα αυτό. Η ταξινόμηση που λαμβάνει χώρα κατά την φάση της δημιουργίας των δεδομένων εκπαίδευσης από τους παρατηρητές, επηρεάζεται από παράγοντες όπως:

- Χρόνος αντίδρασης: Τα δεδομένα αποτελούνται από ακολουθίες frames και όχι από μεμονωμένες εικόνες όπως σε άλλα προβλήματα. Αυτό σημαίνει ότι για μεγαλύτερο αριθμό fps ο χρόνος αντίδρασης του παρατηρητή μικραίνει και υπάρχει η πιθανότητα λανθασμένης ταξινόμησης κατά της διαδικασία δημιουργίας των δειγμάτων για την εκπαίδευση του μοντέλου. Για τον λόγο αυτό έγινε πειραματισμός και σε συνάρτηση με τα δεδομένα των μελετών που παραθέτονται και των αποτελεσμάτων των ταξινομήσεων επιλέχθηκε ως το καλύτρο μέγεθος του δείγματος να είναι 32 frames.
- Κατά την διαδικασία της δημιουργίας του σετ δεδομένων που θα χρησιμοποιηθεί για την εκπαίδευση του δικτύου, τα δείγματα περιέχονται σε βίντεο με μεγάλη διάρκεια. Η χειροκίνητη ταξινόμηση απαιτεί σημαντική προσπάθεια από τους ταξινομητές, γεγονός που προσθέτει τον παράγοντα της κόπωσης στην διαδικασία με αποτέλεσμα την πιθανή συσσώρευση λανθασμένων εκτιμήσεων στα δεδομένα ground truth.
- Καθώς εκτυλίσσεται το πείραμα εξαναγκασμένης κολύμβησης, οι συνθήκες παίζουν σημαντικό ρόλο για την αναγνώριση της δράσης του επιμύ. Αρχικά, παρατηρήθηκε ότι κατά την διάρκεια της κολύμβησης το πειραματόζωο με τις κινήσεις που εκτελεί πετάει νερό στα τοιχώματα του δοχείου. Αυτό συμβαίνει μετά από την πάροδο κάποιων λεπτών του πειράματος, με αποτέλεσμα να προσθέτει αλλοιώσεις στην τελική εικόνα. Αυτό αποκτά ιδιαίτερη σημασία αφού το τίναγμα της κεφαλής συμβαίνει μετά την πάροδο κάποιου χρονικού διαστήματος από την έναρξη του κάθε μεμονωμένου πειράματος. Στα



Σχήμα 5.1: Δείγμα εικόνας με αλλοιωμένα χαρακτηριστικά των άνω άκρων του επιμύ.

Σχήμα 5.2: Δείγμα εικόνας με φανερά όλα τα άκρα του επιμύ από τα αρχικά λεπτά του πειράματος.

παραπάνω σχήματα είναι εμφανής η μεγάλη διαφοροποίηση των κινήσεων του επιμύ κατά την διάρκεια του πειράματος. Τα σχήματα 5.1 και 5.2 απεικονίζουν απόσπασμα κίνησης που αναφέρεται στην κατηγορία Swimming και είναι φανερό ότι κατά την διαδικασία του πειράματος πολλές φορές δεν είναι φανερά όλα τα άκρα του πειραματόζωου με αποτέλεσμα να δυσκολεύει η διαδικασία της αναγνώρισης της κίνησης.

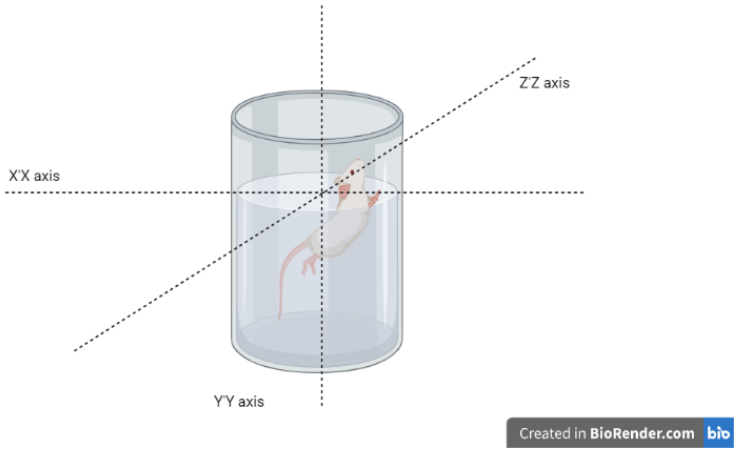
Μετά την ολοκλήρωση της προετοιμασίας των δεδομένων, έγινε σχεδιασμός και υλοποίηση αρχιτεκτονικών με συνελικτικά επίπεδα τριών διαστάσεων για την πρώτη σειρά πειραμάτων. Αρχικά, εξετάστηκε το πόσο επηρεάζει την επίδοση του δικτύου το βάθος του. Η μελέτη αφορούσε στο πρώτο μέρος της την διερεύνηση των δυνατοτήτων να αναγνωρίσουν δράσεις με την χρήση απλών δικτύων με μικρό υπολογιστικό κόστος. Τα δίκτυα αυτά περιείχαν έως και 2.1 εκατομμύρια παραμέτρους και τα πρώτα πειράματα της σειράς αυτής έδειξαν ότι ακόμα και με χρήση ενός σχετικά ρηχού δικτύου είναι εφικτή η αναγνώριση δράσεων σε σημαντικό βαθμό. Ειδικότερα, παρατηρήθηκε μια σχέση ανάμεσα στη ακρίβεια που επιτεύχθηκε και τον αριθμό των επιπέδων του δικτύου, ιδιαίτερα στην κατηγορία του Head Shake. Η κίνηση αυτή αναγνωρίστηκε καλύτερα με μικρό αριθμό συνελικτικών επιπέδων, ιδιαίτερα μετά την μελέτη και προσαρμογή των υπερ-παραμέτρων του δικτύου.

Η διαδικασία της βελτιστοποίησης των υπερ-παραμέτρων του δικτύου, έδειξε ότι τόσο ο ρυθμός μάθησης αλλά και το μέγεθος της εικόνας είχαν μεγάλη σημασία για την αναγνώριση προτύπων στα βίντεο. Τα πρώτα πειράματα είχαν ακραία μικρές τιμές μεγέθους εικόνας, ώστε να μελετηθεί όλο το πιθανό εύρος τιμών που μπορούν να χρησιμοποιηθούν. Μια μικρότερου μεγέθους εικόνα,56X56, βοήθησε στην δραστική μείωση του χρόνου εκπαίδευσης, αλλά είχε σημαντικό αντίκτυπο στην τελική ακρίβεια επί του σετ δεδομένων Test. Με την αύξηση του μεγέθους πάρθηκαν αποτελέσματα που έδειξαν ότι υπάρχει σύνδεση μεταξύ του μεγέθους εικόνας και της ικανότητας του δικτύου να εντοπίσει χωροχρονικά πρότυπα και να τα ταξινομήσει τελικά με σωστό τρόπο στις ζητούμενες κατηγορίες.

Σημαντική βελτίωση παρατηρήθηκε μετά την προσαρμογή του επιπέδου Dropout, όπου η συμβολή του ήταν καθοριστική για την καλύτερη εκπαίδευση του δικτύου και την αποφυγή της υπερπροσαρμογής (Overfitting). Η χρήση περισσότερων συνελικτικών επιπέδων έδωσε καλύτερη απόδοση στην αναγνώριση προτύπων που συνδέονται με τις κινήσεις που έχουν σχετικά μεγαλύτερη χρονική διάρκεια. Το πρόβλημα του Overfitting ήταν σημαντικό αντικείμενο πειραματισμού καθ΄ όλη την διάρκεια των πειραμάτων που αφορούσαν την διερεύνηση των δυνατοτήτων των απλών δικτύων που παρουσιάστηκαν παραπάνω.

Η ταχύτητα εκπαίδευσης των απλών δικτύων ήταν της τάξης των 10 έως 20 ωρών, ανάλογα το μέγεθος της εικόνας και των αριθμό δειγμάτων ανά παρτίδα και την GPU που χρησιμοποιήθηκε. Επιπλέον, μετά το τέλος των πειραμάτων για τα δίκτυα αυτά το μέγεθος του τελικού αποθηκευμένου μοντέλου ήταν 2 τάξεις μικρότερο από αυτό των υπολοίπων όπως το ResNet3D και το R(2+1)D.

Παρατηρήθηκε ότι η επίδοση τους ήταν στενά συνδεδεμένη με την χρήση των τεχνικών επαύξησης δεδομένων(Data Augmentation). Αυτό φαίνεται να απορρέει από το γεγονός ότι κίνηση του πειραματόζωου εκτελείται με μεγάλο εύρος και γύρω από τους 3 άξονες που φαίνονται στο Σχήμα 5.3. Για την καλύτερη εκπαίδευση και την αποφυγή της υπερπροσαρμογής, επιλέχθηκαν τεχνικές επαύξησης που να είναι όσο το δυνατό πιο κοντά στις κινήσεις που δύναται να εκτελέσει ο επιμύς κατά την διάρκεια του πειράματος της εξαναγκασμένης κολύμβησης. Μετά από όλα τα πειράματα είναι προφανές ότι οι τεχνικές αυτές πρέπει να προσαρμόζονται στις διαφορετικές συνθήκες που υπάρχουν, όταν γίνεται προσπάθεια γενίκευσης του μοντέλου. Για παράδειγμα, σε μεγάλο εύρος δεδομένων βίντεο η κάμερα ήταν τοποθετημένη σε γωνία τέτοια που ήταν πιθανή η απόκρυψη λεπτομερειών λόγω του φαινομένου της ανάκλασης.



Σχήμα 5.3: Απεικόνιση των αξόνων κίνησης του πειραματόζωου.

Η διαδικασία μελέτης της χρήσης Βαθιάς Μάθησης για την αναγνώριση δράσεων στα δεδομένα βίντεο συνεχίστηκε με την εφαρμογή Residual Neural Networks. Η σημαντική διαφορά με τα προηγούμενα μοντέλα που δοκιμάστηκαν, είναι η ύπαρξη συνδέσεων ανάμεσα από τα επίπεδα. Αυτό δίνει την δυνατότητα να κατασκευαστούν βαθύτερες αρχιτεκτονικές, που περιέχουν πάνω από 30 εκατομμύρια εκπαιδεύσιμες παραμέτρους.

Από τα πρώτα πειράματα που εκτελέστηκαν με την αρχιτεκτονική ResNet3D τα αποτελέσματα ήταν εμφανώς βελτιωμένα. Στα πειράματα αυτά, που ήταν τα πρώτα με χρήση αυτού του είδους δικτύου, έγιναν δοκιμές και βρέθηκε ότι τα καλύτερα αποτελέσματα πάρθηκαν με χρήση προ-εκπαιδευμένου δικτύου αφού η ακρίβεια στο σετ δεδομένων Test βελτιώθηκε κατά +3%. Η συγκεκριμένη αρχιτεκτονική, δοκιμάστηκε με τις καλύτερες παραμέτρους που έχουν δοθεί [40] και ο σκοπός των πρώτων πειραμάτων ήταν να εξεταστεί εάν είναι δυνατό να παρθούν με μια out of the box αρχιτεκτονική αποτελέσματα που να προσεγγίζουν αυτά των Benchmark Datasets, ώστε να γίνει αποσφαλμάτωση του σετ δεδομένων. Οι δοκιμές αυτές κρίνονται απαραίτητες, διότι στην περίπτωση των δεδομένων βίντεο που χρησιμοποιήθηκαν, χρειάστηκε να γίνει συγχρονισμός των χειροκίνητων ταξινομήσεων με τις αναπαριστάμενες δράσεις. Επιπλέον, οι δοκιμές είχαν ως στόχο τον εντοπισμό προβλημάτων που συνδέονται με την ταξινόμηση των δεδομένων από τους παρατηρητές, που στην περίπτωση αυτή ήταν τρεις. Από τα πειράματα που διεξήχθησαν με χρήση του δικτύου ResNet3D βγήκαν αποτελέσματα που δεδομένου του μικρού μεγέθους του σετ δεδομένων είναι εφάμιλλα με αυτά των Benchmark Datasets. Τα αποτελέσματα έδειξαν ότι ο συγχρονισμός ήταν σωστός στα δεδομένα των βίντεο καθώς και ότι η ταξινόμηση ήταν σωστή στα αρχικά βίντεο από τους παρατηρητές. Έτσι, τα αποτελέσματα των πρώτων πειραμάτων με απλά συνελικτικά δίκτυα αποκτούν εγκυρότητα και δείχνουν ότι με μικρό αριθμό συνελικτικών επιπέδων είναι δυνατή η αναγνώριση δράσεων στα πειράματα εξαναγκασμένης κολύμβησης, κρατώντας παράλληλα χαμηλά το υπολογιστικό κόστος.

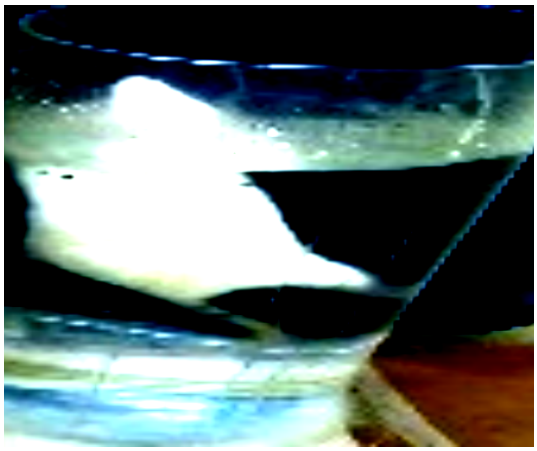
Στην συνέχεια, ξεκίνησαν οι σειρές πειραμάτων με την τελική αρχιτεκτονική που θα χρησιμοποιηθεί, την R(2+1)D . Η επιλογή του δικτύου αυτού έγινε με γνώμονα το ότι μέχρι και την στιγμή δημοσίευσης του είχε πετύχει την καλύτερη ακρίβεια 91.5% (top5) και 73.3% (top1) για RGB στο σετ δεδομένων Sports 1-M.

Στα πειράματα που έγιναν με την παραπάνω αρχιτεκτονική έγινε προσπάθεια να διερευνηθεί η ικανότητα αυτού του είδους δικτύου να αναγνωρίσει χωρο-χρονικά πρότυπα σε δεδομένα βίντεο RGB . Η μετά τις δοκιμές μεγεθών και την βελτιστοποίηση των παραμέτρων του πάρθηκε ακρίβεια

77.2% επί του Test Dataset. Η επίδοση κρίνεται εξαιρετική, αφού είναι μέσα στις top1 επιδόσεις για το είδος αυτό.

Μετά από προσεκτική μελέτη των δεδομένων που τροφοδότησαν το δίκτυο βγήκαν σημαντικά συμπεράσματα για τους παράγοντες που είτε βοήθησαν στην αναγνώριση των δράσεων ή την δυσκόλεψαν έως έναν βαθμό.

- Οι ενδιάμεσες συνδέσεις που χαρακτηρίζουν αυτήν την ομάδα δικτύων είναι καθοριστικής σημασίας για την επιτυχία της ταξινόμησης. Βρέθηκε ότι ανάμεσα στις δύο κατηγορίες δικτύων που δοκιμάστηκαν, τα απλά και τα Residual Neural Networks , υπάρχει μια ζώνη που τα μεν πρώτα αδυνατούν να αντεπεξέλθουν στην αναγνώριση σύντομων κινήσεων όπως το τίναγμα κεφαλής. Πιο αναλυτικά, όσο προσθέτονται συνελικτικά επίπεδα με εκπαιδεύσιμες παραμέτρους και χωρίς Skip Connections, τόσο πιο δύσκολη γίνεται η αναγνώριση των κινήσεων. Άρα με την πρόσθεση επιπέδων διαφαίνεται η ανάγκη των συνδέσεων ώστε να είναι δυνατή η αποτελεσματική εκπαίδευση του δικτύου.
- Από την πειραματική διαδικασία, φάνηκε να υπάρχει σύνδεση ανάμεσα στην τελική ακρίβεια και το μέγεθος των εικόνων που κάθε φορά τροφοδοτούν το δίκτυο. Η καλύτερη ακρίβεια δόθηκε με μέγεθος εικόνας 250X250,με την αρχική να είναι κατά μέσο όρο 800X1080 στα αρχικά βίντεο. Σημαντικό στοιχείο στα πειράματα αυτά είναι η απαίτηση για μνήμη που κατά την σειρά πειραμάτων με εικόνες του παραπάνω μεγέθους και Batch Size 16, ανήλθε στα 23Gb . Επίσης, τα τελικά αποθηκευμένα μοντέλα όπως αυτά πάρθηκαν κατά την διαδικασία της εκπαίδευσης, κρατώντας τα 3 καλύτερα για αποφυγή της υπερπροσαρμογής, είχαν μέγεθος της τάξης των 300 Mb έναντι 2.5 Mb για τα απλά δίκτυα που προαναφέρθηκαν.
- Η κινήσεις των πειραματόζωων είναι τέτοιες που λόγω της θέσης της συσκευής καταγραφής, δυσκολεύουν την αποτύπωση όλων των κινήσεων που χαρακτηρίζουν την δράση. Γενικά ο επιμύς, κινείται σε όλη την περιοχή του δοχείου και σε συνδυασμό με την ύπαρξη του φαινομένου της ανάκλασης, τα άκρα του δεν είναι πάντα εμφανή. Στην παρακάτω εικόνα είναι εμφανές το προαναφερθέν: Στο Σχήμα 5.4 φαίνεται το φαινόμενο που αναφέρθηκε. Η



Σχήμα 5.4: Απεικόνιση πειραματόζωου με κρυμμένα τα άνω άκρα.

κίνηση του επιμύ στην περίπτωση που φαίνεται είναι πιθανό να συγχυστεί με την κατηγορία της Αναρρίχησης ενώ ανήκει στην κατηγορία της Κολύμβησης. Ο λόγος είναι ότι τα άκρα του δεν είναι φανερά και η κίνηση προσομοιάζει την κάθετη τοποθέτηση του σώματος όπως συμβαίνει και με την Αναρρίχηση.

- Η χρήση προ-εκπαιδευμένου δικτύου βοήθησε αρκετά, τόσο στην εκπαίδευση αλλά και στην τελική επίδοση του δικτύου. Από τα αποτελέσματα των πειραμάτων είναι φανερή η συνεισφορά της χρήσης βαρών από άλλα σετ δεδομένων. Υπάρχουν χαρακτηριστικά σε αυτά που βοηθούν στην αναγνώριση προτύπων που συνδέονται με μικρές κινήσεις όπως αυτή του Τινάγματος Κεφαλής. Στην πορεία της πειραματικής διαδικασίας, δοκιμάστηκαν όλοι οι συνδυασμοί που θα μπορούσαν να δώσουν καλύτερα αποτελέσματα, όμως φάνηκε ότι τουλάχιστον στα δεδομένα που συγκεντρώθηκαν για την παρούσα ΔΕ υπάρχουν κοινά στοιχεία με αυτά των σετ δεδομένων που έγινε η αρχική εκπαίδευση.

Συνοψίζοντας, στην παρούσα ΔΕ έγινε ενδελεχής πειραματισμός με δίκτυα συνελικτικών επιπέδων τριών διαστάσεων (3D Convolutions). Αρχικά, δοκιμάστηκαν απλά δίκτυα και δείχτηκε ότι είναι δυνατή η αναγνώριση των δράσεων των πειραματόζωων με ικανοποιητικό βαθμό, ίσο σε αρκετές περιπτώσεις με το ποσοστό συμφωνίας δύο διαφορετικών παρατηρητών-ταξινομητών. Στην συνέχεια έγινε έλεγχος των αποτελεσμάτων και ακολούθησε σειρά πειραμάτων για τον έλεγχο του σετ δεδομένων με την αρχιτεκτονική ResNet3D, όπου και επιβεβαιώθηκε η ακεραιότητά του και η δυνατότητα βελτίωσης της ακρίβειας με την χρήση αρχιτεκτονικών που διαθέτουν Skip Connections. Στο τελευταίο μέρος της πειραματικής διαδικασίας, έγινε πειραματισμός και βελτιστοποίηση με την αρχιτεκτονική R(2+1)D και στην συνέχεια χρήση αρχικοποιημένων βαρών από το Kinetics 400. Όλα έδειξαν ότι τα δίκτυα αυτά είναι σε θέση να δώσουν εξαιρετικά αποτελέσματα στην αναγνώριση δράσεων για τα δεδομένα που αφορούν το πείραμα εξαναγκασμένης κολύμβησης. Η χρήση των δικτύων αυτών είναι δυνατό να αυτοματοποιήσει την διαδικασία της ταξινόμησης δράσεων για τα πειράματα αυτά, δίνοντας την δυνατότητα στους ερευνητές να επιταχύνουν δραστικά την διαδικασία μελέτης νέων αντικαταθλιπτικών φαρμάκων.

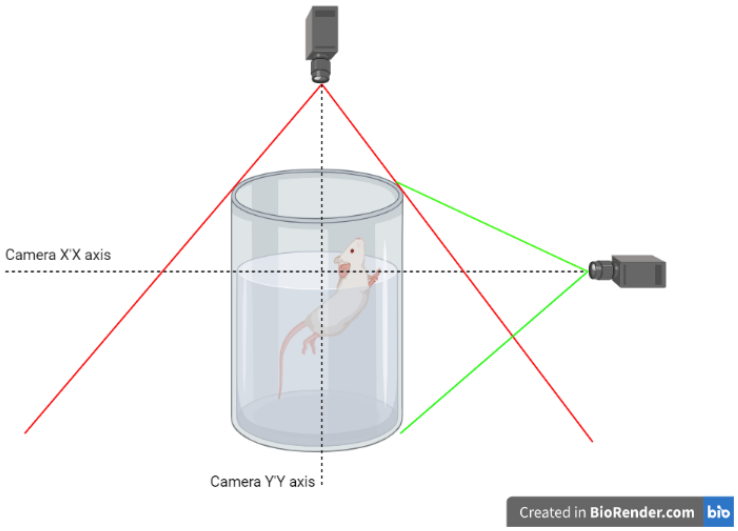
5.2 Μελλοντικές Επεκτάσεις

Μέσα από την διαδικασία που ακολουθήθηκε βγήκαν χρήσιμα συμπεράσματα για την περαιτέρω βελτίωση της διαδικασίας αναγνώρισης δράσεων σε δεδομένα βίντεο. Η μελέτη επικεντρώθηκε σε δεδομένα RGB όπως θα ήταν εύκολη η διάθεσή τους από το εκάστοτε εργαστήριο που εκτελεί τα εν λόγω πειράματα.

Αρχικά, παρατηρήθηκε ότι τα βίντεο τα οποία διατέθηκαν περιείχαν δύο πειράματα με επιμύες που λάμβαναν χώρα την ίδια χρονική περίοδο στον ίδιο χώρο. Αυτό είχε ως αποτέλεσμα την ύπαρξη ήχου και από τα δύο πειράματα στο βίντεο και δεν υπήρχε η δυνατότητα διαχωρισμού του κατά την διαδικασία της προ-επεξεργασίας. Η ύπαρξη ήχου θα είχε ως αποτέλεσμα την πολύ καλύτερη αναγνώριση της κίνησης Head Shake , αφού πάντα συνοδεύεται από χαρακτηριστικό ήχο. Για την καλύτερη καταγραφή του πειράματος, θα μπορούσε να προταθεί η καταγραφή κάθε πειράματος χωριστά ή και δύο ταυτόχρονα, σε διαφορετικό χώρο ή σε ειδικά διαμορφωμένο δοχείο που θα επιτρέπει την καταγραφή ήχου μόνο από ένα πειραματόζωο. Η καταγραφή τέτοιων δεδομένων και η εκπαίδευση ενός δικτύου θα είχε πολλαπλά θετικά αποτελέσματα για την ορθή ταξινόμηση δράσεων.

Η μελέτη των δεδομένων κατά την προ-επεξεργασία, έδειξε ότι πολλές φορές η κίνηση των άκρων του επιμύ αποκρύπτεται είται λόγω της ανάκλασης ή λόγω των σταγόνων του νερού πάνω στα τοιχώματα του δοχείου που βρίσκεται το πειραματόζωο. Για τον λόγο αυτό θα ήταν χρήσιμη η τοποθέτηση δύο καμερών ώστε να είναι δυνατή η καταγραφή του πειράματος από γωνίες τέτοιες ώστε να είναι ορατές οι λεπτομέρειες αυτές που θα δώσουν την μέγιστη πληροφορία για την κίνησή του και κατά συνέπεια θα συμβάλλουν στην αναγνώριση των δράσεων.

Στο Σχήμα 5.5 φαίνεται μια πιθανή τοποθέτηση των συσκευών εγγραφής. Με τον τρόπο αυτό θα



Σχήμα 5.5: Απεικόνιση προτεινόμενου τρόπου τοποθέτησης συσκευών καταγραφής.

καταγραφούν δύο σετ εικόνων. Το πρώτο θα αφορά τον κατακόρυφο άξονα, που στόχο θα έχει να αποτυπώσει όλες τις κινήσεις που:

1. Εκτελούνται στην επιφάνεια της δεξαμενής, όπως αυτή της κολύμβησης και το πειραματόζωο κάνει κυκλικές κινήσεις σε όλη την ελεύθερη επιφάνεια και χαρακτηρίζουν την κίνηση Swimming.
2. Την σταθερή στάση του επιμύ όταν εκτελεί την κίνηση Immobility.
3. Τις κινήσεις των άκρων του επιμύ όταν αυτός είναι σε οριζόντια κίνηση και τα άκρα δεν θα ήταν δυνατό να φανούν από κάποια άλλη γωνία με τα βέλτιστα αποτελέσματα.

Επιπλέον, κατά την καταγραφή του πειράματος στον οριζόντιο άξονα, θα υπάρξουν τα παρακάτω θετικά αποτελέσματα:

1. Θα γίνει καταγραφή του πειραματόζωου στην κατακόρυφη θέση, ώστε όλα τα άκρα να είναι εμφανή κατά την διάρκεια της κίνησης Climbing.
2. Με την καταγραφή σε θέση κάθετα στον κατακόρυφο άξονα που αναφέρθηκε και παραπάνω θα εξαλειφθεί κατά το δυνατό το φαινόμενο της ανάκλασης που πολλές φορές αποκρύπτει χρήσιμα χαρακτηριστικά του επιμύ.
3. Την καλύτερη καταγραφή των κινήσεων του Head Shake και των Climbing .

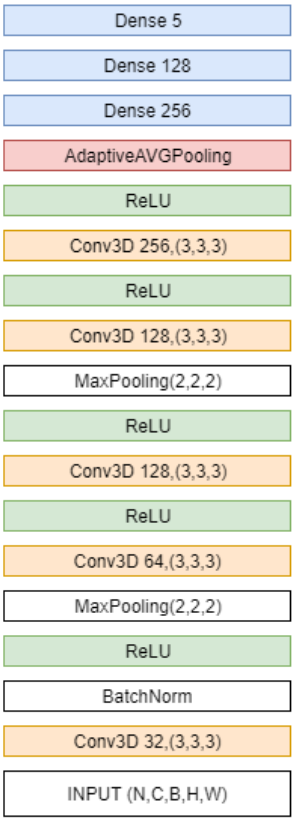
Από τα παραπάνω είναι δυνατό να υπάρξει καλύτερη επίδοση του δικτύου, αφού θα υπάρχουν περισσότερες λεπτομέρειες για την ταξινόμηση της κάθε δράσης.

Τέλος, θα μπορούσε να γίνει ανίχνευση των περιοχών που υπάρχει η κίνηση κάθε φορά και αποκοπή της υπόλοιπης περιοχής. Η Region Of Interest(ROI) θα περιλαμβάνει τα άκρα του πειραματόζωου και η ανίχνευση θα γίνεται χωριστά πριν την ταξινόμηση των δράσεων. Η έρευνα έχει δείξει ότι δεν συμβάλλουν όλα τα frames του βίντεο στην ανίχνευση της δράσης, παρά μόνο κάποια από αυτά. Στην περίπτωση που γινόταν μια ανίχνευση των περιοχών αυτών, θα ήταν δυνατή η μείωση της μνήμης στην τελική ταξινόμηση εάν γινόταν περικοπή του παρασκηνίου. Επιπλέον, εάν γινόταν ανίχνευση της περιοχής που περιέχει την σημαντική για την ταξινόμηση πληροφορία, θα ήταν δυνατό να χρησιμοποιηθεί εικόνα μεγαλύτερης ανάλυσης μόνο για αυτήν την περιοχή και όχι για όλη την εικόνα [58]

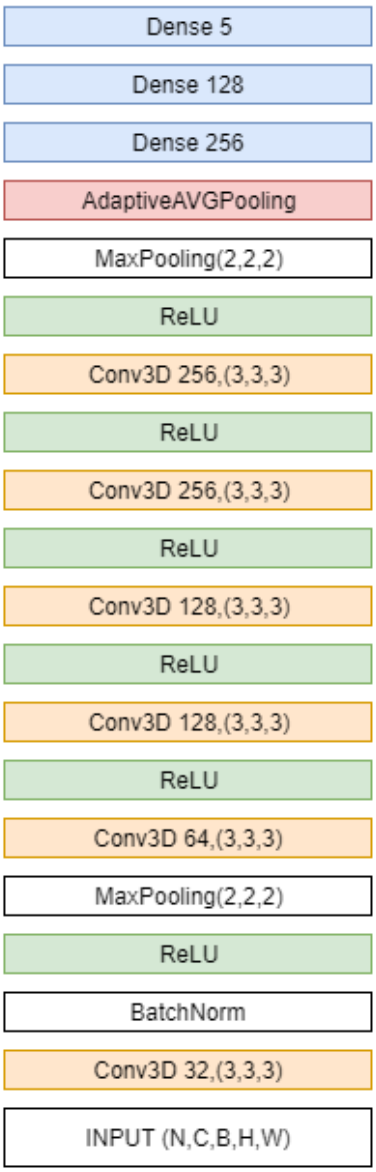
Παράρτημα Αρχιτεκτονικών

Παρουσίαση Αρχιτεκτονικών

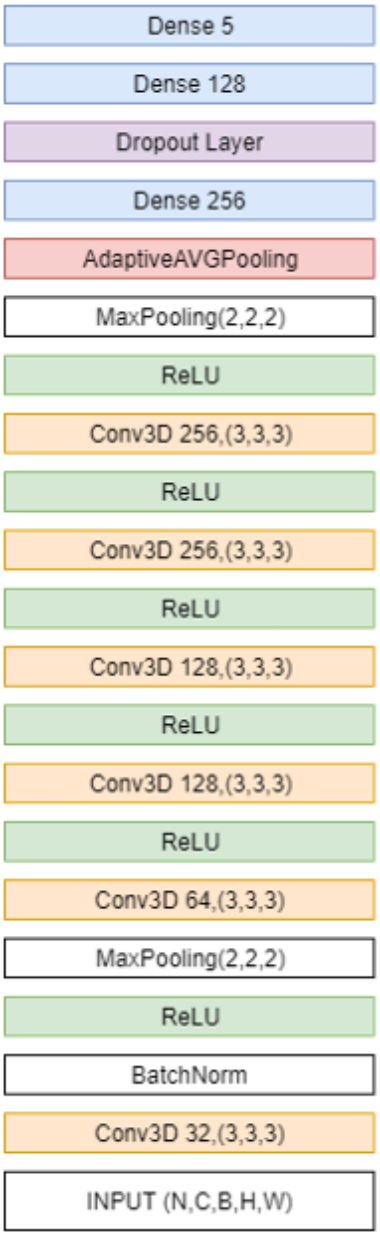
Στο παράστημα Α΄ θα παρουσιαστούν όλες οι αρχιτεκτονικές που δοκιμάστηκαν και αφορούν τα απλά συνελικτικά δίκτυα με όλους τους συνδυασμούς που κατασκευάστηκαν για τις ανάγκες των πειραμάτων.



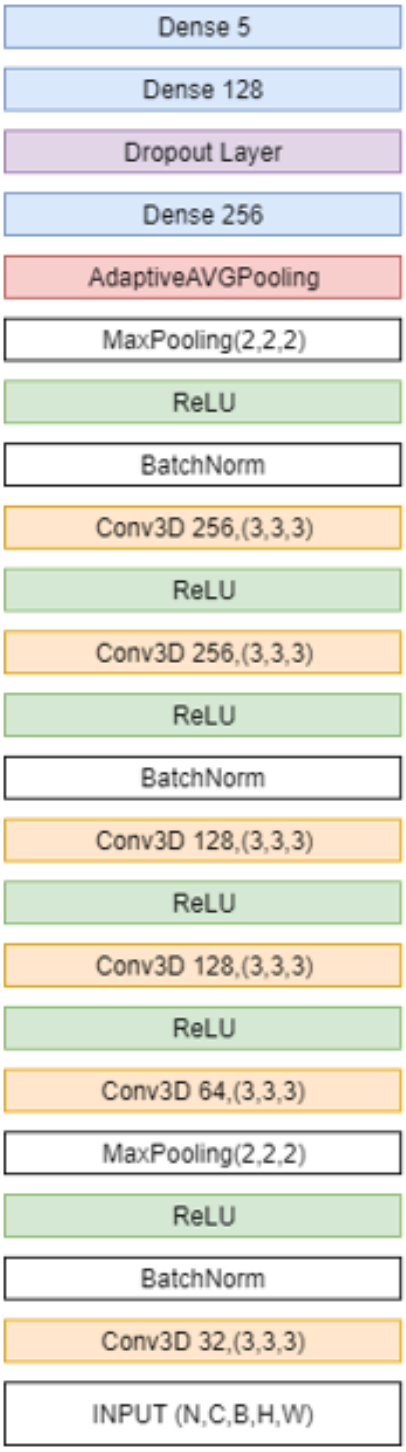
Σχήμα 5.6: Αρχιτεκτονική 5-Conv3D .



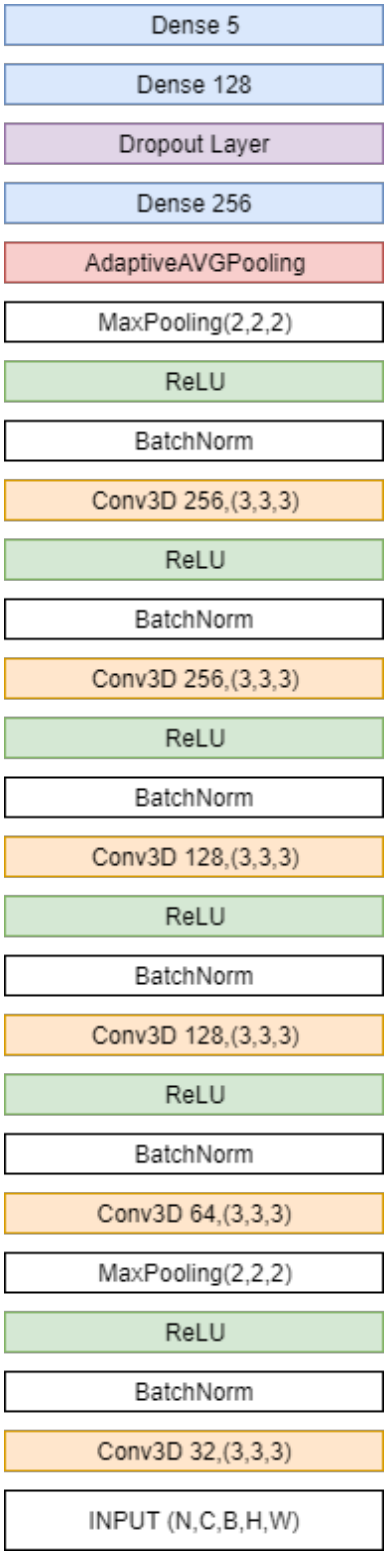
Σχήμα 5.7: Αρχιτεκτονική 6-Conv3D .



Σχήμα 5.8: Αρχιτεκτονική 6-Conv3D-Dropout .



Σχήμα 5.9: Αρχιτεκτονική 6-Conv3D-Dropout-Batch Normalization .



Σχήμα 5.10: Αρχιτεκτονική 6-Conv3D-Dropout-Full Batch Normalization .

Κατάλογος Σχημάτων

Κατάλογος Πινάκων

Βιβλιογραφία

[1] An Introduction to Convolutional Neural Networks,Keiron O'Shea and Ryan Nash, 2015, ArXiV 1511.08458

[2] Huang, Thomas S.: Computer vision: Evolution and promise. 1996.

[3] Rosenblatt, F. (1958). The perceptron: a probabilistic model for information storage and organization in the brain. Psychological Review, 65 6, 386–408.

[4] Gu J., Wang Z., Kuen J., Ma L., Shahroudy A., Shua, B., Liu T., Wang X., Wang G. (2015). Recent Advances in Convolutional Neural Networks. CoRR, abs/1512.07108. arxiv.org/abs/1512.0710

[5] Huang, T. S. (1996). Computer Vision: Evolution And Promise.

[6] Zhu Y., Liu, Zolfaghari, Xiong, Wu, Zhang, Tighe, Manmatha, Li, (2020). A Comprehensive Study of Deep Video Action Recognition. CoRR, abs/2012.06567. arxiv.org/abs/2012.06567

[7] Popescu, M.-C., Balas, Perescu-Popescu, Mastorakis(2009). Multilayer perceptron and neural networks. WSEAS Transactions on Circuits and Systems.

[8] Nwankpa, Ijomah, Gachagan,Marshall(2018). Activation Functions: Comparison of trends in Practice and Research for Deep Learning. arXiv. doi.org/10.48550/ARXIV.1811.03378

[9] Hui Liu, 2020, in Robot Systems for Rail Transit Applications. (2020). Rail transit channel robot systems, 7.3.4.2.1 Sigmoid activation function.

[10] RADU,COSTEA,Stan,V.(2020).Automatic Traffic Sign Recognition Artificial Inteligence-Deep Learning Algorithm. doi.org/10.1109/ECAI50035.2020.9223186

[11] Scabini L. F. S., Bruno O. M. (2021). Structure and Performance of Fully Connected Neural Networks: Emerging Complex Network Properties. arXiv.doi.org/10.48550/ARXIV.2107.14062

[12] Vaswani, Shazeer N., Parmar N., Uszkoreit J., Jones, L., Gomez A. N., Kaiser L., Polosukhin I. (2017). Attention Is All You Need. arXiv.doi.org/10.48550/ARXIV.1706.03762

[13] Albert-Laszlo B., Albert R. (1999). Emergence of Scaling in Random Networks. Science, 286(5439), 509–512. https://doi.org/10.1126/science.286.5439.509

[14] O'Shea K., Nash R. (2015). An Introduction to Convolutional Neural Networks.

[15] Chen, T. (2017). What is "padding" in Convolutional Neural Network?. Retrieved from https://medium.com/machine-learning-algorithms/what-is-padding-in-convolutional-neuralnetwork-c120077469cc

[16] Albawi S., Mohammed, T. A., Al-Zawi, S. Understanding of a convolutional neural network. 2017 International Conference on Engineering and Technology (ICET), 2017. doi: 10.1109/ICEngTechnol.2017.8308186

[17] Yessou H., Sumbul G., Demir B. (2020). A Comparative Study of Deep Learning Loss Functions for Multi-Label Remote Sensing Image Classification.

[18] Ma J. (2020). Segmentation Loss Odyssey. arXiv. https://doi.org/10.48550/ARXIV.2005.13449

[19] Zhang J. (2019). Gradient Descent based Optimization Algorithms for Deep Learning Models Training.

[20] Salman S., Liu X. (2019). Overfitting Mechanism and Avoidance in Deep Neural Networks. arXiv. https://doi.org/10.48550/ARXIV.1901.06566

[21] Rao M. R., Prasad V., Teja P., Zindavali M., Reddy O. (2018). A Survey on Prevention of Overfitting in Convolution Neural Networks Using Machine Learning Techniques. International Journal of Engineering and Technology(UAE), 7, 177–180. https://doi.org/10.14419/ijet.v7i2.32.15399

[22] P. Thanapol K. Lavangnananda P. Bouvry, F. Pinel and F. Leprévost, "Reducing Overfitting and Improving Generalization in Training Convolutional Neural Network (CNN) under Limited Sample Sizes in Image Recognition," 2020 - 5th International Conference on Information Technology (InCIT), 2020, pp. 300-305, doi: 10.1109/InCIT50588.2020.9310787.

[23] Srivastava N., Hinton G., Krizhevsky A., Sutskever I., Salakhutdinov, R. Dropout: a simple way to prevent neural networks from overfitting. The Journal of Machine Learning Research, vol. 15, no. 1, pp. 1929–1958, 2014.

[24] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in 32nd International Conference on Machine Learning, ICML 2015, 2015, pp. 448–456.

[25] Zhu Y., Li X., Liu C., Zolfaghari M., Xiong Y., Wu C., Zhang Z., Tighe J., Manmatha R., Li M. (2020). A Comprehensive Study of Deep Video Action Recognition. CoRR, abs/2012.06567. https://arxiv.org/abs/2012.06567

[26] Karen Simonyan and Andrew Zisserman. Two-Stream Convolutional Networks for Action Recognition in Videos. In Advances in Neural Information Processing Systems (NeurIPS), 2014.

[27] Limin Wang, Zhe Wang, Yuanjun Xiong, and Yu Qiao.CUHK and SIAT Submission for THUMOS15 Action Recognition Challenge. THUMOS'15 Action Recognition Challenge, 2015.

[28] Limin Wang, Yuanjun Xiong, Zhe Wang, and Yu Qiao. Towards Good Practices for Very Deep Two-Stream ConvNets. arXiv preprint arXiv:1507.02159, 2015.

[29] Limin Wang, Yuanjun Xiong, Zhe Wang, Yu Qiao, Dahua Lin, Xiaoou Tang, and Luc Van Gool. Temporal Segment Networks: Towards Good Practices for Deep Action Recognition. In The European Conference on Computer Vision (ECCV), 2016.

[30] Wang Yifan, Jie Song, Limin Wang, Luc Van Gool, and Otmar Hilliges. Two-Stream SR-CNNs for Action Recognition in Videos. In The British Machine Vision Conference (BMVC), 2016.

[31] Christoph Feichtenhofer, Axel Pinz, and Andrew Zisserman. Convolutional Two-Stream Network Fusion for Video Action Recognition. In The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016.

[32] Limin Wang, Yuanjun Xiong, Zhe Wang, and Yu Qiao. Towards Good Practices for Very Deep Two-Stream ConvNets. arXiv preprint arXiv:1507.02159, 2015.

[33] Karen Simonyan and Andrew Zisserman. Two-Stream Convolutional Networks for Action Recognition in Videos. In Advances in Neural Information Processing Systems (NeurIPS), 2014.

[34] Christoph Feichtenhofer, Axel Pinz, and Richard P. Wildes. Spatiotemporal Residual Networks for Video Action Recognition. In Advances in Neural Information Processing Systems (NeurIPS), 2016.

[35] A. Ullah, J. Ahmad, K. Muhammad, M. Sajjad, and S. W. Baik. Action Recognition in Video Sequences using Deep Bi-Directional LSTM With CNN Features. IEEE Access, 2017.

[36] Christoph Feichtenhofer, Haoqi Fan, Jitendra Malik, and Kaiming He. SlowFast Networks for Video Recognition. In The IEEE International Conference on Computer Vision (ICCV), 2019.

[37] Joao Carreira and Andrew Zisserman. Quo Vadis, Action Recognition? A New Model and the Kinetics Dataset. In The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017.

[38] Kensho Hara, Hirokatsu Kataoka, and Yutaka Satoh. Can Spatiotemporal 3D CNNs Retrace the History of 2D CNNs and ImageNet? In The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2018.

[39] Ali Diba, Mohsen Fayyaz, Vivek Sharma, M. Mahdi Arzani, Rahman Yousefzadeh, Juergen Gall, and Luc Van Gool. Spatio-Temporal Channel Correlation Networks for Action Classification. In The European Conference on Computer Vision (ECCV), 2018.

[40] Du Tran, Heng Wang, Lorenzo Torresani, Jamie Ray, Yann LeCun, and Manohar Paluri. A Closer Look at Spatiotemporal Convolutions for Action Recognition. In The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2018.

[41] Du Tran, Heng Wang, Lorenzo Torresani, and Matt Feiszli. Video Classification With Channel-Separated Convolutional Networks. In The IEEE International Conference on Computer Vision (ICCV), 2019.

[42] Yan Li, Bin Ji, Xintian Shi, Jianguo Zhang, Bin Kang, and Limin Wang. TEA: Temporal Excitation and Aggregation for Action Recognition. In The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2020.

[43] Ji Lin, Chuang Gan, and Song Han. TSM: Temporal Shift Module for Efficient Video Understanding. In The IEEE International Conference on Computer Vision (ICCV), 2019

[44] Christoph Feichtenhofer. X3D: Expanding Architectures for Efficient Video Recognition. In The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2020.

[45] Sah S. (2020). Machine Learning: A Review of Learning Types. https://doi.org/10.20944/preprints202007.0230.v1

[46] Zhuang F., Qi Z., Duan K., Xi D., Zhu Y., Zhu H., Xiong H., He Q. (2019). A Comprehensive Survey on Transfer Learning. arXiv. https://doi.org/10.48550/ARXIV.1911.02685

[47] https://towardsdatascience.com/activation-functions-neural-networks-1cbd9f8d91d6

[48] Kuteeva E., Hökfelt T. G. M., Wardi T., Ogren S. O. (2010). Galanin, galanin receptor subtypes and depression-like behaviour. Experientia Supplementum, 102, 163–181.

[49] R. D. Porsolt, A. Bertin and M. Jalfre, "Behavioral despair in mice: A primary screening test for antidepressants.," Archives internationales de pharmacodynamie et de therapie, vol. 229, pp. 327–336, Oct. 1977.

[50] R. D. Porsolt, G. Anton, N. Blavet, and M. Jalfre, "Behavioural despair in rats: A new model sensitive to antidepressant treatments," European Journal of Pharmacology, vol. 47, pp. 379–391, Feb. 1978.

[51] M. J. Detke, M. Rickels and I. Lucki, "Active behaviors in the rat forced swimming test differentially produced by serotonergic and noradrenergic antidepressants," Psychopharmacology, vol. 121, pp. 66–72, Sept. 1995.

[52] Alexandros Vythoulkas, Action Recognition and Classification in Pre-Clinical Experiment Videos with Machine Learning Techniques, National Technical University of Athens, Thesis Project, Athens July 2019.

[53] N. Kokras, K. Antoniou, H. G. Mikail, V. Kafetzopoulos, Z. Papadopoulou-Daifoti and C. Dalla, "Forced swim test: What about females?," Neuropharmacology, vol. 99, pp. 408–421, Dec. 2015.

[54] N. Kokras, D. Baltas, F. Theocharis, and C. Dalla, "Kinoscope: An Open-Source Computer Program for Behavioral Pharmacologists," Frontiers in Behavioral Neuroscience, vol. 11, May 2017

[55] https://github.com/RaivoKoot/Video-Dataset-Loading-Pytorch

[56] Zhang S., Guo S., Huang W., Scott M. R., Wang L. (2020). V4D:4D Convolutional Neural Networks for Video-level Representation Learning. arXiv. doi.org/10.48550/ARXIV.2002.07442

[57] Woods D. L., Wyma J. M., Yund E. W., Herron T. J., Reed B. (2015). Factors influencing the latency of simple reaction time. Frontiers in Human Neuroscience, 9. doi:10.3389/fnhum.2015.00131 10.3389/fnhum.2015.00131

[58] Kumar A. R., Ravindran B., Raghunathan A. (2019, January). Pack and Detect. Proceedings of the ACM India Joint International Conference on Data Science and Management of Data. doi.org/10.1145/3297001.3297020

[59] Massoudi, Massoud Verma, Siddhant Jain, Riddhima. (2021). Urban Sound Classification using CNN. 583-589. 10.1109/ICICT50816.2021.9358621.