

ΠΑΝΕΠΙΣΤΗΜΙΟ ΘΕΣΣΑΛΙΑΣ

ΠΟΛΥΤΕΧΝΙΚΗ ΣΧΟΛΗ



UNIVERSITY OF
THESSALY

ΤΜΗΜΑ ΜΗΧΑΝΙΚΩΝ Η/Υ, ΤΗΛΕΠΙΚΟΙΝΩΝΙΩΝ ΚΑΙ ΔΙΚΤΥΩΝ

ΑΝΑΓΝΩΡΙΣΗ ΠΕΖΩΝ ΣΕ ΣΥΝΘΗΚΕΣ ΣΚΟΤΟΥΣ ΜΕ ΣΥΝΕΛΕΚΤΙΚΑ ΝΕΥΡΩΝΙΚΑ ΔΙΚΤΥΑ

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

Αλέξανδρος Μαντσακώβ

Επιβλέποντες καθηγητές:

Δημήτριος Κατσαρός, Πανεπιστήμιο Θεσσαλίας

Ελευθέριος Τσουκαλάς Καθηγητής, Πανεπιστήμιο Θεσσαλίας

Βόλος, Οκτώβριος 2018

ΠΕΡΙΛΗΨΗ

Τα Συνελικτικά Νευρωνικά δίκτυα λόγω της ευρείας ανάπτυξης τους τα τελευταία χρόνια έχουν προτιμηθεί ιδιαίτερα για έργα όπως η ανίχνευση αντικείμενων και συγκεκριμένα για την ανίχνευση πεζών. Αυτή η έκρηξη ενδιαφέροντος, έχει υπάρξει κυρίως για την ανάπτυξη αυτο-οδηγούμενων αυτοκινήτων, μιας και η ανίχνευση πεζών, είτε την ημέρα είτε την νύχτα, έχει γίνει ένα από τα κύρια προβλήματα που δεν έχει αντιμετωπισθεί εξ' ολοκλήρου. Οι έρευνες γύρω από το θέμα αυτό αφορούν κυρίως συστήματα τα οποία έχουν αναπτυχθεί για την επιτυχή ανίχνευση πεζών κατά την διάρκεια της ημέρας όπου υπάρχει αρκετό φως και οι πεζοί είναι ευδιάκριτοι. Αντιθέτως έχουν γίνει ελάχιστες μελέτες που αφορούν την ανίχνευση κατά την περίοδο της νύχτας. Αυτή η διπλωματική εργασία στοχεύει στην εύρεση κατάλληλων Συνελικτικών Νευρωνικών δικτύων, δοκιμάζοντας έτοιμες αρχιτεκτονικές νευρωνικών δικτύων οι οποίες καταφέρνουν να αναγνωρίζουν την ύπαρξη πεζών σε εικόνες εν ώρα νυκτός. Για να επιτευχθεί αυτό το έργο χρησιμοποιήθηκαν τρία διαφορετικά σύνολα δεδομένων.

Αρχικά δοκιμάστηκε το σύνολο δεδομένων του Πανεπιστημίου του Caltech [51] για την προπόνηση των δικτύων, ωστόσο τα αποτελέσματα δεν ήταν ελπιδοφόρα και η έρευνα συνεχίστηκε με δύο ειδικά κατασκευασμένα σύνολα εικόνων, οι οποίες βρέθηκαν και συλλέχτηκαν από το διαδίκτυο. Από μία μεγάλη συλλογή αρχιτεκτονικών δικτύων που βρέθηκαν, μόνο τρεις ήταν συμβατές με το υπολογιστικό σύστημα που ήταν διαθέσιμο για τη διαδικασία της προπόνησης και αξιολόγησης των δικτύων. Σε αυτήν την έρευνα θα αναλυθεί η συμπεριφορά αυτών των αρχιτεκτονικών σε συνδυασμό με σύνολα δεδομένων που ελέγχουν την αναγνώριση πεζών εν ώρα νυκτός.

Λέξεις κλειδιά

Συνελικτικά νευρωνικά δίκτυα, Tensorflow, Αναγνώριση πεζών την νύχτα, Python, ειδικά σύνολα δεδομένων

ΛΙΣΤΑ ΕΙΚΟΝΩΝ

Εικόνα 1: Ένα παράδειγμα ενός πολυεπίπεδου Νευρωνικού Δικτύου	9
Εικόνα 2: Ένα παράδειγμα ενός απλού Νευρωνικού δικτύου	10
Εικόνα 3: feedforward ΝΔ.....	11
Εικόνα 4: Radial Basis function ΝΔ.....	12
Εικόνα 5: Επαναλαμβανόμενο ΝΔ	13
Εικόνα 6: Αρθρωτό ΝΔ	14
Εικόνα 7: Συνελικτικό ΝΔ	15
Εικόνα 8: Παράδειγμα μιας εικόνας σε κωδικοποίηση RGB (αριστερή εικόνα) και μίας σε κωδικοποίηση grayscale (δεξιά)	15
Εικόνες 9-12: Παραδείγματα εικόνων του Caltech Pedestrian Dataset. Είναι εμφανές ότι κάθε φορά που εμφανίζονται πεζοί, δεν αποτελούν και μεγάλο μέρος της εικόνας, δυσκολεύοντας έτσι την προπόνηση του ΣΝΔ.	20
Εικόνα 10: Αρχιτεκτονική του πρώτου δικτύου.....	26
Εικόνα 11: Αρχιτεκτονική του δεύτερου δικτύου.....	27
Εικόνα 12: Αρχιτεκτονική του τρίτου δικτύου	28
Εικόνα 13: Διάγραμμα μεγέθους παρτίδας και της αρχιτεκτονικής που χρησιμοποιήθηκε, για αποδεκτά μοντέλα	33
Εικόνα 14: Διάγραμμα μεγέθους παρτίδας και της αρχιτεκτονικής που χρησιμοποιήθηκε, για μη αποδεκτά μοντέλα	33
Εικόνα 15: Διάγραμμα μεγέθους παρτίδας και της αρχιτεκτονικής που χρησιμοποιήθηκε, για μη αποδεκτά μοντέλα	33
Εικόνα 16: Διάγραμμα ρυθμού μάθησης και μεγέθους παρτίδας, για αποδεκτά μοντέλα	35
Εικόνα 17: Διάγραμμα ρυθμού μάθησης και μεγέθους παρτίδας, για όλα τα μοντέλα	35
Εικόνα 18: Διάγραμμα ρυθμού μάθησης και μεγέθους παρτίδας, για μη αποδεκτά μοντέλα	35
Εικόνα 19: Διάγραμμα ρυθμού μάθησης και της αρχιτεκτονικής που χρησιμοποιήθηκε, για αποδεκτά μοντέλα	Error! Bookmark not defined.
Εικόνα 20: Διάγραμμα ρυθμού μάθησης και της αρχιτεκτονικής που χρησιμοποιήθηκε, για όλα τα μοντέλα	Error! Bookmark not defined.
Εικόνα 21: Διάγραμμα ρυθμού μάθησης και της αρχιτεκτονικής που χρησιμοποιήθηκε, για μη αποδεκτά μοντέλα	Error! Bookmark not defined.
Εικόνα 22: Διάγραμμα μεγέθους παρτίδας και της αρχιτεκτονικής που χρησιμοποιήθηκε, για αποδεκτά μοντέλα	38
Εικόνα 23: Διάγραμμα μεγέθους παρτίδας και της αρχιτεκτονικής που χρησιμοποιήθηκε, για όλα τα μοντέλα	38
Εικόνα 24: Διάγραμμα μεγέθους παρτίδας και της αρχιτεκτονικής που χρησιμοποιήθηκε, για μη αποδεκτά μοντέλα	38
Εικόνα 25: Διάγραμμα ρυθμού μάθησης και μεγέθους παρτίδας, για αποδεκτά μοντέλα	39
Εικόνα 26: Διάγραμμα ρυθμού μάθησης και μεγέθους παρτίδας, για όλα τα μοντέλα	39
Εικόνα 27: Διάγραμμα ρυθμού μάθησης και μεγέθους παρτίδας, για μη αποδεκτά μοντέλα	39
Εικόνα 28: Διάγραμμα ρυθμού μάθησης και της αρχιτεκτονικής που χρησιμοποιήθηκε, για αποδεκτά μοντέλα	40
Εικόνα 29: Διάγραμμα ρυθμού μάθησης και της αρχιτεκτονικής που χρησιμοποιήθηκε, για όλα τα μοντέλα	40

Εικόνα 30: Διάγραμμα ρυθμού μάθησης και της αρχιτεκτονικής που χρησιμοποιήθηκε, για μη αποδεκτά μοντέλα	40
Εικόνα 31: Διάγραμμα μεγέθους παρτίδας και της αρχιτεκτονικής που χρησιμοποιήθηκε, για μη αποδεκτά μοντέλα	43
Εικόνα 32: Διάγραμμα μεγέθους παρτίδας και της αρχιτεκτονικής που χρησιμοποιήθηκε, για αποδεκτά μοντέλα	43
Εικόνα 33: Διάγραμμα μεγέθους παρτίδας και της αρχιτεκτονικής που χρησιμοποιήθηκε, για όλα τα μοντέλα	43
Εικόνα 35: Διάγραμμα ρυθμού μάθησης και μεγέθους παρτίδας, για όλα τα μοντέλα	43
Εικόνα 36: Διάγραμμα ρυθμού μάθησης και μεγέθους παρτίδας, για αποδεκτά μοντέλα	43
Εικόνα 34: Διάγραμμα ρυθμού μάθησης και μεγέθους παρτίδας, για μη αποδεκτά μοντέλα	43
Εικόνα 37: Διάγραμμα ρυθμού μάθησης και της αρχιτεκτονικής που χρησιμοποιήθηκε, για μη αποδεκτά μοντέλα	43
Εικόνα 38: Διάγραμμα ρυθμού μάθησης και της αρχιτεκτονικής που χρησιμοποιήθηκε, για όλα τα μοντέλα	43
Εικόνα 39: Διάγραμμα ρυθμού μάθησης και της αρχιτεκτονικής που χρησιμοποιήθηκε, για όλα τα μοντέλα	43

ΠΙΝΑΚΑΣ ΠΕΡΙΕΧΟΜΕΝΩΝ

ΑΝΑΓΝΩΡΙΣΗ ΠΕΖΩΝ ΣΕ ΣΥΝΘΗΚΕΣ ΣΚΟΤΟΥΣ ΜΕ ΣΥΝΕΛΕΚΤΙΚΑ ΝΕΥΡΩΝΙΚΑ ΔΙΚΤΥΑ.....	1
ΠΕΡΙΛΗΨΗ	3
Λέξεις κλειδιά	3
ΛΙΣΤΑ ΕΙΚΟΝΩΝ	4
ΠΙΝΑΚΑΣ ΠΕΡΙΕΧΟΜΕΝΩΝ.....	6
1. ΕΙΣΑΓΩΓΗ	8
2. ΕΙΣΑΓΩΓΗ ΣΤΑ ΝΕΥΡΩΝΙΚΑ ΔΙΚΤΥΑ.....	9
2.1 Feedforward ΝΔ.....	11
2.2 ΝΔ με χρήση συνάρτησης Radial Basis	12
2.3 Επαναλαμβανόμενο ΝΔ (Recurrent Neutral Network).....	13
2.4 Αρθρωτό Νευρωνικό Δίκτυο.....	14
2.5 Συνελικτικά νευρωνικά δίκτυα (ΣΝΔ).....	15
3. ΒΙΒΛΙΟΓΡΑΦΙΚΗ ΑΝΑΣΚΟΠΗΣΗ.....	17
4. ΣΥΝΟΛΟ ΔΕΔΟΜΕΝΩΝ	20
4.1 Caltech Pedestrian Dataset.....	20
4.2 1 ^ο ειδικά κατασκευασμένο ΣΔ.....	21
4.3 2 ^ο ειδικά κατασκευασμένο ΣΔ.....	22
5. ΥΛΟΠΟΙΗΣΗ	24
5.1 Εισαγωγή στο εργαλείο TensorFlow.....	24
5.2 Ανάλυση αρχιτεκτονικών που χρησιμοποιήθηκαν	25
5.2.1 1 ^η δομή δικτύου.....	26
5.2.2 2 ^η δομή δικτύου	27
5.2.3 3^η δομή δικτύου	28
6. ΑΠΟΤΕΛΕΣΜΑΤΑ ΔΟΚΙΜΩΝ ΚΑΙ ΠΕΙΡΑΜΑΤΑ	30
6.1 1 ^ο Σύνολο Δεδομένων	33
6.2 2 ^ο Σύνολο Δεδομένων.....	37
6.3 1 ^ο σύνολο δεδομένων με διασταυρωμένη επικύρωση (10-fold cross validation).....	41
7. ΣΥΜΠΕΡΑΣΜΑΤΑ	45
8. ΜΕΛΛΟΝΤΙΚΗ ΕΡΓΑΣΙΑ.....	46
ΒΙΒΛΙΟΓΡΑΦΙΑ ΚΑΙ ΑΝΑΦΟΡΕΣ	48

1. ΕΙΣΑΓΩΓΗ

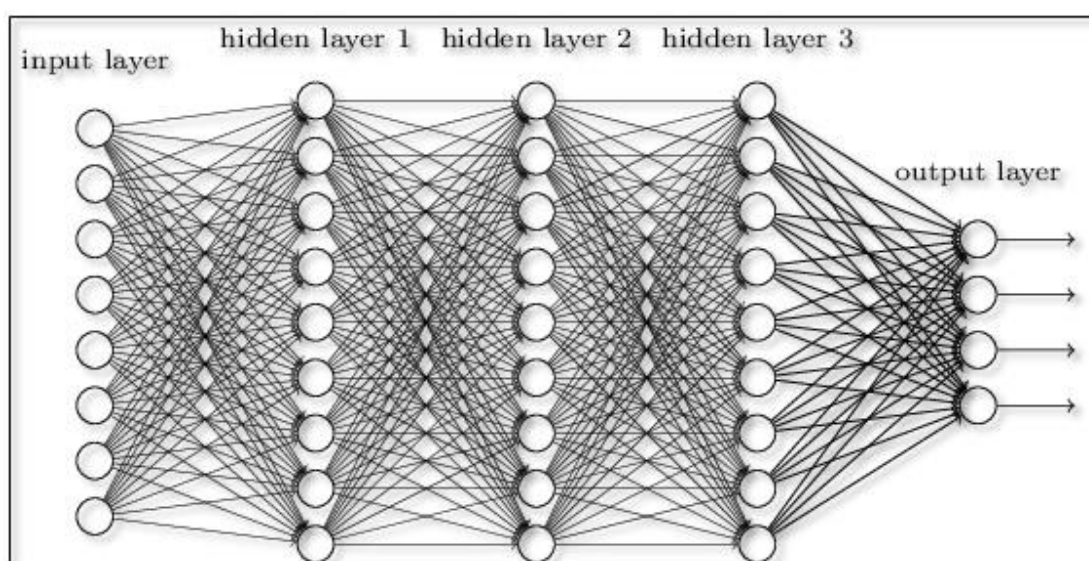
Λόγω της ραγδαίας αύξησης ενδιαφέροντος στα αυτό-οδηγούμενα αυτοκίνητα, η επιστήμη έχει αρχίσει να επιλύει τα προβλήματα που εμποδίζουν την επίτευξη αυτού του έργου. Η σημαντικότητα της τελειοποίησης των αυτό-οδηγούμενων αυτοκινήτων είναι μεγάλη, καθώς θα έχει ως αποτέλεσμα να μειωθούν τα αυτοκινητιστικά δυστυχήματα που συμβαίνουν ετησίως και παράλληλα να αυξηθεί η παραγωγικότητα του εμπορίου μέσω της αυτοματοποίησης της διαδικασίας της μεταφοράς προϊόντων. Η ανίχνευση πεζών, είναι ένας από τους κύριους λόγους που στέκονται εμπόδιο στην επίτευξη του έργου αυτού.

Οι πεζοί είναι ένα από τα πιο κοινά θύματα των αυτοκινητιστικών ατυχημάτων, ιδιαίτερα τη νύχτα. Για το λόγο αυτό η ανίχνευση πεζών είναι πια το μείζον θέμα προς επίλυση. Τα συνελκτικά νευρωνικά δίκτυα (ΣΝΔ) είναι μέχρι τώρα αυτά τα οποία έχουν δείξει τα καλύτερα αποτελέσματα και σε αναγνώριση αλλά και σε χρόνο. Μέχρι σήμερα έχουν αναπτυχθεί αρκετοί τρόποι οι οποίοι αξιοποιούν διαφορετικές τεχνικές ανάλυσης που είτε καταφέρνουν να αναγνωρίζουν πεζούς με μία καλή απόδοση αναγνώρισης αλλά όχι σε πραγματικό χρόνο, είτε με μέτρια απόδοση αναγνώρισης και σε πραγματικό χρόνο. Για να τεθεί όμως σε χρήση ένα τέτοιου είδους σύστημα θα χρειαστεί να υπάρχει άριστη απόδοση αναγνώρισης σε πραγματικό χρόνο διότι υπάρχουν αρκετές δυσκολίες που παρουσιάζονται. Αρχικά, η κίνηση της κάμερας που θα καταγράφει το δρόμο είναι ιδιαίτερα αυξημένη, λόγω της ύπαρξης αυτοκινήτου, με αποτέλεσμα η ανάλυση του βίντεο που θα χρησιμοποιηθεί σαν είσοδος στο νευρωνικό δίκτυο να επηρεάζεται αρνητικά (π.χ. θολούρα). Επιπρόσθετα τα ΣΝΔ εξέχουν στην αναγνώριση προτύπων μέσα σε μία εικόνα, στη συγκεκριμένη περίπτωση ο στόχος είναι να διδαχθεί το ΣΝΔ ώστε να αναγνωρίζει τη μορφή του ανθρώπινου σώματος. Ωστόσο το πρόβλημα που το αποτρέπει αυτό, είναι η στάση του σώματος η οποία μπορεί να πάρει πολλές διαφορετικές μορφές δυσκολεύοντας την προπόνηση του. Όλα αυτά τα προβλήματα χειροτερεύουν την επίδραση τους όταν η εικόνα δεν είναι κατάλληλα φωτισμένη, δηλαδή την νύχτα.

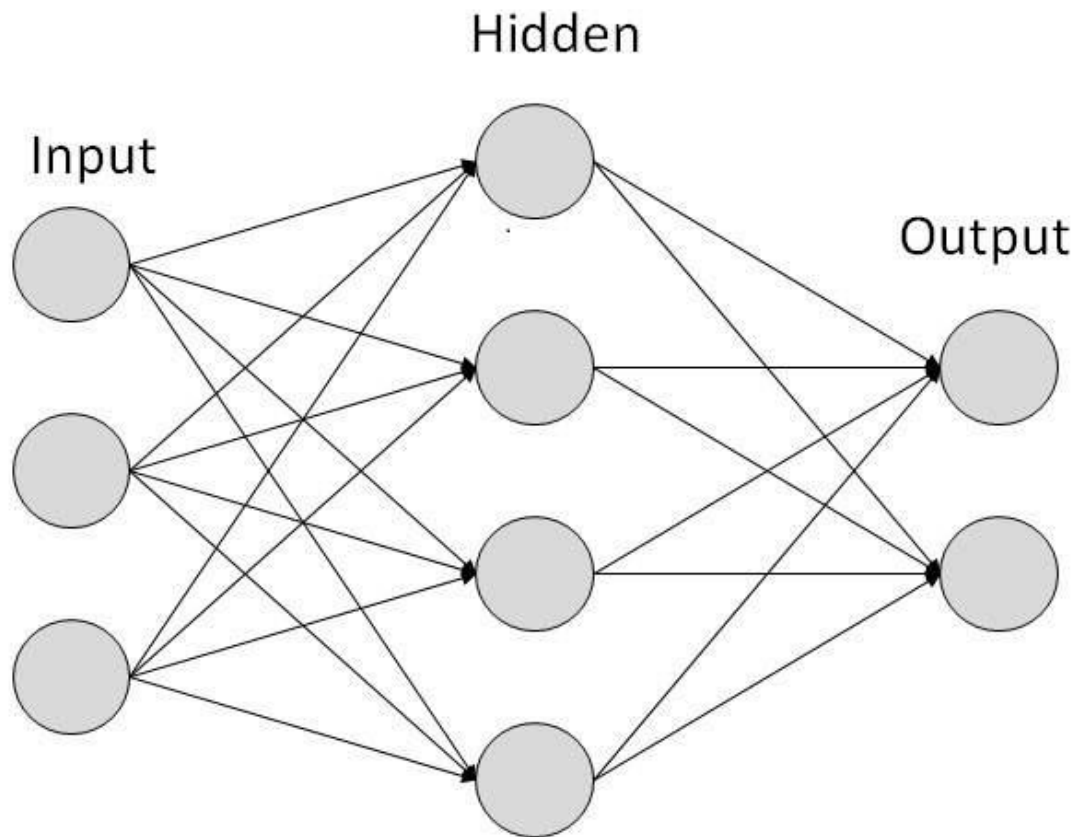
Σε αυτήν την διπλωματική εργασία παρουσιάζονται τα αποτελέσματα που βρέθηκαν εξετάζοντας ήδη υπάρχοντα ΣΝΔ που έχουν χρησιμοποιηθεί για αναγνώριση πεζών την ημέρα, αν λειτουργούν εξίσου καλά και για πεζούς εν ώρα νυκτός.

2. ΕΙΣΑΓΩΓΗ ΣΤΑ ΝΕΥΡΩΝΙΚΑ ΔΙΚΤΥΑ

Το πρώτο νευρωνικό δίκτυο (ΝΔ) εφευρέθηκε το 1943. Η ιδιότητα τους να αναγνωρίζουν μοτίβα έγινε άμεσα αντιληπτή, ωστόσο η ανάπτυξη τους αναστάθηκε λόγω έλλειψης υπολογιστικής ισχύος από τα τότε υπολογιστικά συστήματα. Ένα ΝΔ ορίζεται ως ένα δίκτυο από απλούς υπολογιστικούς κόμβους διασυνδεδεμένους μεταξύ τους. Οι νευρώνες χωρίζονται σε νευρώνες εισόδου, νευρώνες εξόδου και υπολογιστικούς νευρώνες ή κρυμμένους νευρώνες. Οι νευρώνες εισόδου δεν επιτελούν κανέναν υπολογισμό, μεσολαβούν απλώς ανάμεσα στις περιβαλλοντικές εισόδους του δικτύου και στους υπολογιστικούς νευρώνες. Οι νευρώνες εξόδου διοχετεύουν στο περιβάλλον τις τελικές αριθμητικές εξόδους του δικτύου. Οι υπολογιστικοί νευρώνες πολλαπλασιάζουν κάθε είσοδό τους με το αντίστοιχο *συναπτικό βάρος* και υπολογίζουν το ολικό άθροισμα των γινομένων. Το άθροισμα αυτό τροφοδοτείται ως όρισμα στη *συνάρτηση ενεργοποίησης*, την οποία υλοποιεί εσωτερικά κάθε κόμβος. Η τιμή που λαμβάνει η συνάρτηση για το εν λόγω όρισμα είναι και η έξοδος του νευρώνα για τις τρέχουσες εισόδους και βάρη [14]. Νευρωνικά δίκτυα με μόνο ένα επίπεδο υπολογιστικών νευρώνων ονομάζονται απλά νευρωνικά δίκτυα, ενώ ΝΔ με πολλαπλά επίπεδα θεωρούνται πολυεπίπεδα.



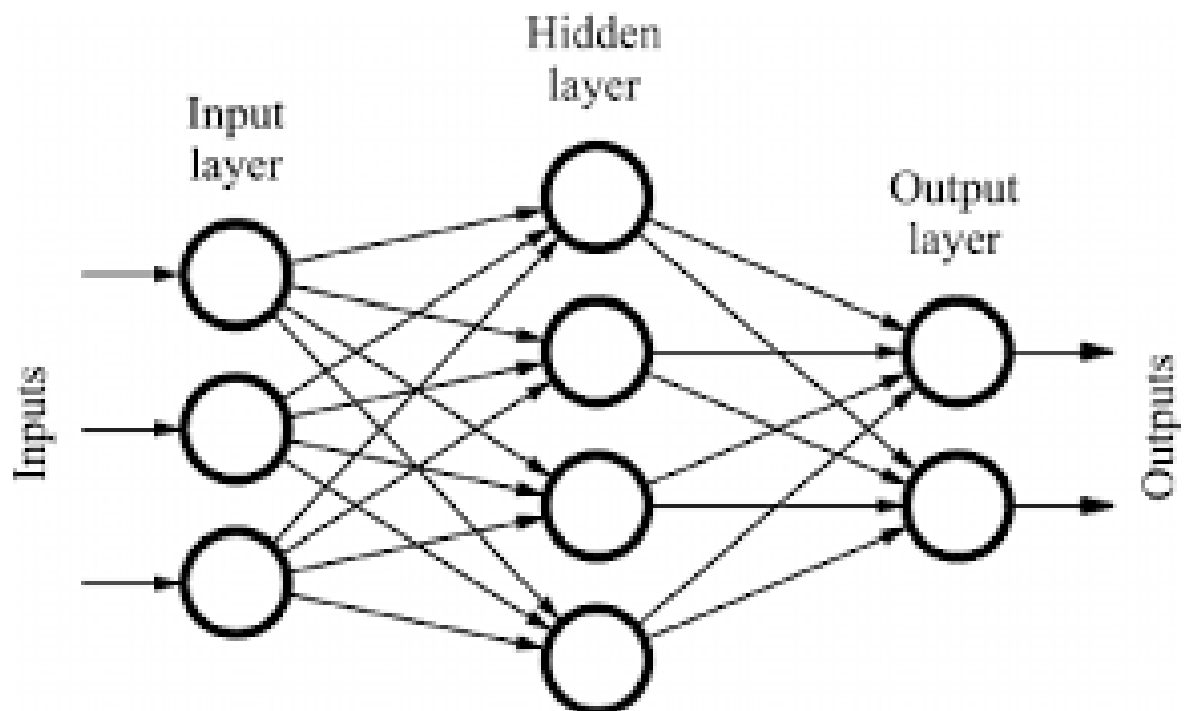
Εικόνα 1: Ένα παράδειγμα ενός πολυεπίπεδου Νευρωνικού Δικτύου



Εικόνα 2: Ένα παράδειγμα ενός απλού Νευρωνικού δικτύου

Για να αρχίσει το νευρωνικό δίκτυο να «μαθαίνει», θα χρειαστεί ένα μεγάλο αριθμό δεδομένων που θα είναι κατάλληλα επιλεγμένα έτσι ώστε μέσω της διαδικασίας της προπόνησης και οπισθοδρόμησης [16], να εξαχθούν τα κατάλληλα χαρακτηριστικά. Με αυτά το δίκτυο θα είναι σε θέση να εκτελέσει μία πιο πληροφορημένη διαδικασία αναγνώρισης. Για την βελτιστοποίηση αυτής της διαδικασίας υπάρχουν ορισμένες μεταβλητές οι οποίες αν τροποποιηθούν και θα λάβουν τις κατάλληλες τιμές επιτρέποντας στο δίκτυο να προπονηθεί πιο γρήγορα καθώς και να αναγνωρίζει με μεγαλύτερη ακρίβεια. Δύο από αυτές χρησιμοποιούνται και σε αυτήν την έρευνα και είναι ο ρυθμός μάθησης και το μέγεθος παρτίδας. Ο σκοπός τους είναι να καθορίζουν την ταχύτητα με την οποία αυξομειώνονται τα βάρη των φίλτρων και τον αριθμό των δειγμάτων που πρόκειται να διαδοθούν μέσω του δικτύου, αντίστοιχα. Υπάρχουν αρκετοί διαφορετικοί τύποι ΝΔ, και ο καθένας έχει αναπτυχθεί για να εξυπηρετεί και ένα διαφορετικό σκοπό, χρησιμοποιώντας διαφορετικές μαθηματικές λειτουργίες. Έξι από αυτούς τους τύπους είναι οι παρακάτω.

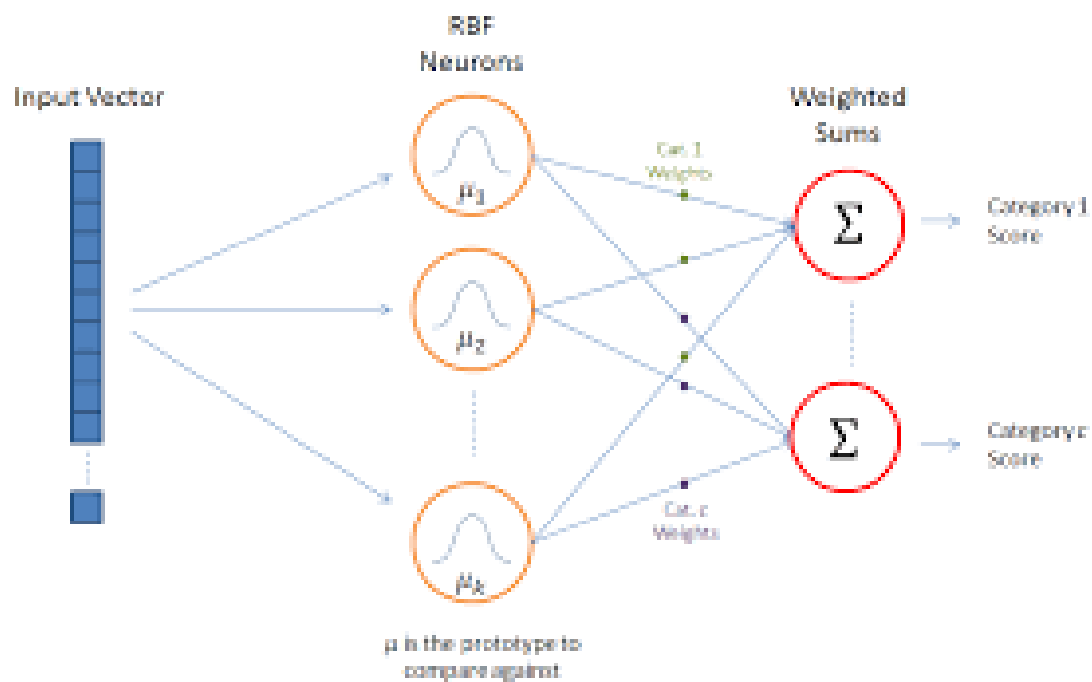
2.1 Feedforward ΝΔ



Εικόνα 3: feedforward ΝΔ

Αυτού του τύπου το ΝΔ έχει την απλούστερη μορφή που μπορεί να έχει ένα ΝΔ. Τα εισερχόμενα δεδομένα «ταξιδεύουν» μέσα στο δίκτυο προς μία μόνο κατεύθυνση. Αυτός ο τύπος δικτύου δεν κατέχει το χαρακτηριστικό της πίσω διάδοσης του λάθους. Τα feedforward ΝΔ βρίσκουν κυρίως εφαρμογή σε τομείς όπως το computer vision και η αναγνώριση ομιλίας.

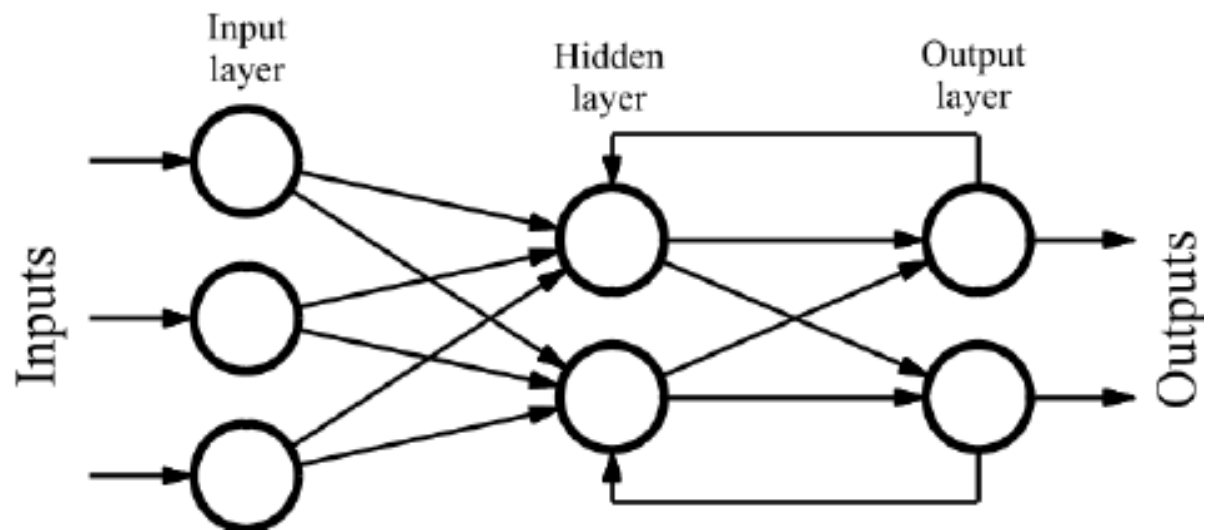
2.2 ΝΔ με χρήση συνάρτησης Radial Basis



Εικόνα 4: Radial Basis function ΝΔ

Η συνάρτηση Radial Basis εξετάζει την απόσταση μεταξύ ενός σημείου και ενός κέντρου. Το ΝΔ θα αποτελείται από δύο διαφορετικά επίπεδα. Στο πρώτο τα χαρακτηριστικά συνδυάζονται με την συνάρτηση Radial Basis και στο δεύτερο τα αποτελέσματα του πρώτου επιπέδου χρησιμοποιούνται για να επαναυπολογιστούν τα ίδια αποτελέσματα. Τέτοιου είδους συστήματα ΝΔ έχουν βρει χρήση σε συστήματα αποκατάσταση ισχύος.

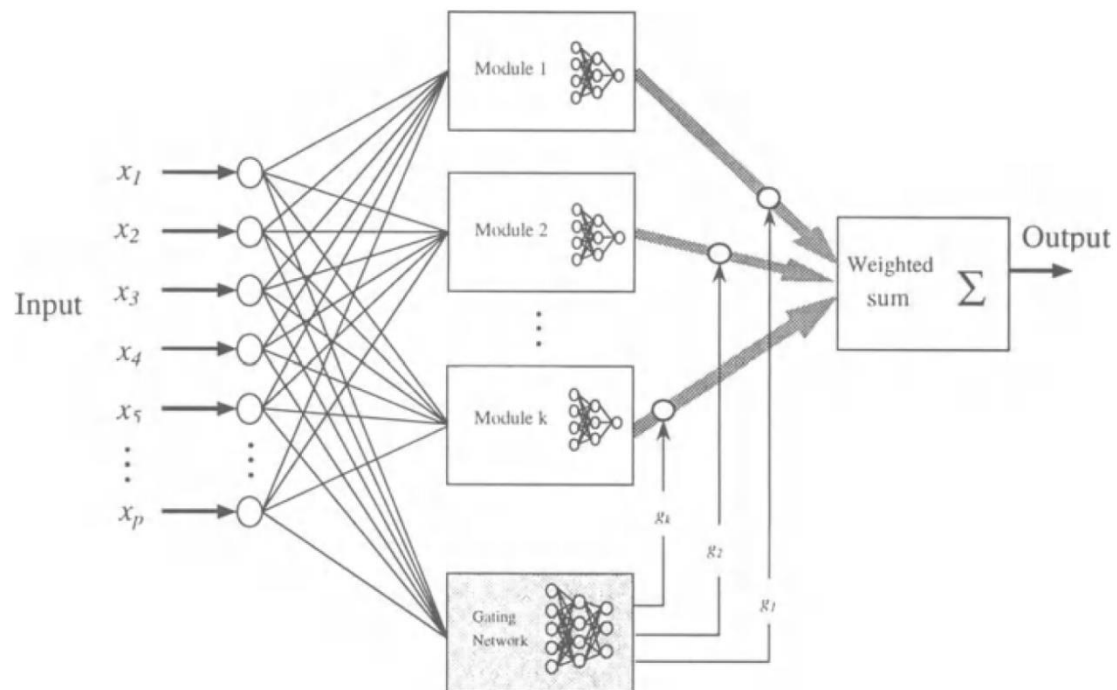
2.3 Επαναλαμβανόμενο ΝΔ (Recurrent Neutral Network)



Εικόνα 5: Επαναλαμβανόμενο ΝΔ

Η βασική ιδέα των επαναλαμβανόμενων ΝΔ είναι η αποθήκευση του αποτελέσματος ενός επιπέδου και η τροφοδότηση αυτού του αποτελέσματος στο ίδιο επίπεδο σαν είσοδο, βοηθώντας έτσι τη διαδικασία της πρόβλεψης του σωστού αποτελέσματος που πρέπει να υπολογίσει το συγκεκριμένο επίπεδο. Τέτοιου είδους τεχνικές χρησιμοποιούνται κυρίως για τον τομέα της μετατροπής γραπτού κειμένου σε προφορικό λόγο, αναγνώριση ομιλίας, πρόβλεψη χρονοσειρών και πολλών άλλων.

2.4 Αρθρωτό Νευρωνικό Δίκτυο

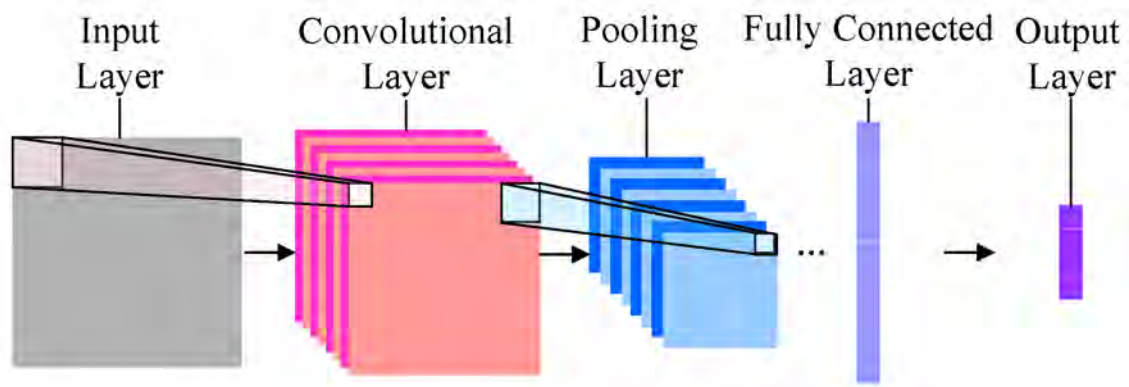


Εικόνα 6: Αρθρωτό ΝΔ

Το κύριο χαρακτηριστικό αυτού του τύπου δικτύων είναι ο συνδυασμός πολλών διαφορετικών ΝΔ, τα οποία θα εργάζονται ξεχωριστά, για την πρόβλεψη του επιθυμητού αποτελέσματος. Κάθε ΝΔ θα δέχεται μια ξεχωριστή είσοδο σε σύγκριση με τα υπόλοιπα ΝΔ, τα οποία θα διατελούν υπο-εργασίες, αποφεύγοντας την επικοινωνία μεταξύ τους.

Αυτά είναι μερικά παραδείγματα από υπάρχοντες τύπους ΝΔ τα οποία έχουν την δική τους χρήση και ο καθένας ειδικεύεται στην επίλυση συγκεκριμένης γκάμας προβλημάτων. Για το θέμα της αναγνώρισης πεζών χρησιμοποιείται ένας διαφορετικός τύπος ΝΔ, που εξειδικεύεται στην επεξεργασία εικόνας ή βίντεο με σκοπό την εύρεση κοινών χαρακτηριστικών. Η ονομασία αυτών των δικτύων είναι *Συνελικτικά νευρωνικά δίκτυα*.

2.5 Συνελικτικά νευρωνικά δίκτυα (ΣΝΔ)



Εικόνα 7: Συνελικτικό ΝΔ

Τα συνελικτικά νευρωνικά δίκτυα είναι όμοια με κανονικά Feedforward ΝΔ τα οποία δέχονται σαν είσοδο αρχεία εικόνας ή βίντεο. Ο κύριος σκοπός τέτοιων αρχιτεκτονικών είναι η αναγνώριση μοτίβων και αντικειμένων. Για τον άνθρωπο αυτή η διαδικασία είναι απλή, ωστόσο για έναν υπολογιστή όταν δέχεται μια εικόνα σαν είσοδο την αντιλαμβάνεται ως ένα πίνακα με αριθμούς. Αυτοί οι αριθμοί μπορούν να πάρουν αρκετές διαφορετικές τιμές που εξαρτώνται από την κωδικοποίηση της εικόνας.



Εικόνα 8: Παράδειγμα μιας εικόνας σε κωδικοποίηση RGB (αριστερή εικόνα) και μίας σε κωδικοποίηση grayscale (δεξιά)

Οι πιο κοινές τεχνικές κωδικοποίησης είναι η RGB και η grayscale. Η πρώτη έχει τη δυνατότητα να αναγνωρίζει το χρώμα που αναπαριστά κάθε pixel σαν συνδυασμός των τριών κύριων χρωμάτων (κόκκινο, πράσινο και μπλε), αλλά αναγκαστικά είναι πιο ακριβή σε θέμα μνήμης. Η δεύτερη παρόλο που αδυνατεί να αναγνωρίσει την έννοια του χρώματος, δίνει στο κάθε pixel μια αριθμητική τιμή που αντιστοιχίζεται σε μια απόχρωση του χρώματος γκρι, επιτρέποντας ταχύτερους χρόνους αναγνώρισης.

Τα ΣΝΔ διαθέτουν διαφορετικές αρχιτεκτονικές δικτύων από τα απλά ΝΔ. Το κάθε επίπεδο έχει τρεις διαστάσεις, ύψος, πλάτος, μήκος. Οι νευρώνες σε δύο διαφορετικά επίπεδα δεν είναι αναγκαστικό να είναι πλήρως συνδεδεμένοι μεταξύ τους. Το αποτέλεσμα του ΣΝΔ είναι ένα διάνυσμα αριθμών όπου η κάθε τιμή του, θα αναπαριστά μια πιθανότητα η εικόνα που λήφθηκε στην είσοδο να ανήκει σε μία κλάση (στην δική μας περίπτωση η κλάση είναι η ύπαρξη πεζού στην εικόνα).

Επιπρόσθετα η δομή των ΣΝΔ αποτελείται από δύο κύρια κομμάτια. Το ένα είναι η εξαγωγή χαρακτηριστικών, το οποίο αφιερώνεται στην εξαγωγή χαρακτηριστικών από την εικόνα, εφαρμόζοντας μια σειρά συνελίξεων (convolution) και συγκεντρώσεων (pooling). Το δεύτερο κομμάτι είναι η διαδικασία της κατάταξης της εικόνας με βάση τα χαρακτηριστικά που έχουν ήδη βρεθεί.

Ο τρόπος εξαγωγής δεδομένων βασίζεται κυρίως στις συνελίξεις που εκτελούνται πάνω στην εικόνα που θα δοθεί σαν είσοδος στο ΣΝΔ. Η συνέλιξη εκτελείται με την βοήθεια φίλτρων παράγοντας έτσι έναν πολυδιάστατο πίνακα, ο οποίος ονομάζεται χάρτης χαρακτηριστικών. Έπειτα αυτός ο χάρτης χαρακτηριστικών θα δεχτεί αρκετή επεξεργασία μέσω άλλων μαθηματικών συναρτήσεων όπως είναι οι συνελίξεις, οι συγκεντρώσεις και Rectified Linear Units (ReLU), οι οποίες θα εφαρμόζονται σε διαφορα επίπεδα του ΣΝΔ . Αφότου λάβει τέλος αυτή η διαδικασία ο τελικός χάρτης χαρακτηριστικών χρησιμοποιείται σαν είσοδος σε μια σειρά από πλήρως συνδεδεμένα στρώματα. Εντούτοις τέτοιου είδους επίπεδα μπορούν να λάβουν σαν είσοδο αποκλειστικά μονοδιάστατους πίνακες, αναγκάζοντας έτσι τον πίνακα χαρακτηριστικών από τρισδιάστατο να μετατραπεί σε μονοδιάστατο για να παραχθεί το επιθυμητό αποτέλεσμα.

3. ΒΙΒΛΙΟΓΡΑΦΙΚΗ ΑΝΑΣΚΟΠΗΣΗ

Η ήδη υπάρχουσα έρευνα που έχει γίνει μέχρι τώρα για την αναγνώριση πεζών την νύχτα αφορούν κυρίως μελέτες που χρησιμοποιούν ή ειδικές κάμερες ή θερμική ή υπέρυθρη όραση. Μικρός είναι ο αριθμός ερευνών που αφιερώνεται στην μελέτη με κανονικές κάμερες. Ένας από τους κύριους λόγους που συμβαίνει αυτό είναι λόγω της έλλειψης ορατότητας, τα χαρακτηριστικά ενός πεζού δεν είναι αρκετά ευδιάκριτα, έτσι ώστε ένα νευρωνικό δίκτυο να καταφέρει να τα αντιληφθεί.

Γενικότερα οι μελέτες που έχουν γίνει μέχρι και σήμερα, μπορούν να χωριστούν σε δύο κατηγορίες, αυτές που χρησιμοποιούν μία κάμερα και αυτές που χρησιμοποιούν περισσότερες από μια κάμερες. Αυτές οι κατηγορίες μπορούν να χρησιμοποιήσουν τα εξής εργαλεία για την αναγνώριση πεζών:

- Προσαρμοστική ώθηση (AdaBoost) [9], η οποία εφαρμόζεται συχνά για τη διαδικασία της αναγνώρισης προσώπου,
- Histogram of oriented gradients (HOG) [7] μαζί με την χρήση Support Vector Machines (SVM) [28],
- μέθοδοι βασισμένες στα Νευρωνικά Δίκτυα χρησιμοποιώντας την προσέγγιση ευαίσθητου πεδίου.

Έχοντας λοιπόν αποθαρρυνθεί από αυτό το γεγονός, οι προσπάθειες επίτευξης αυτού του έργου στράφηκαν στην χρήση των προαναφερθέντων τεχνικών όρασης. Ωστόσο παρουσιάστηκαν αρκετά τεχνικά προβλήματα, τα οποία καθιστούσαν την ολοκλήρωση του έργου όλο και πιο δύσκολη. Ένα παράδειγμα είναι η μελέτη [37] στην οποία χρησιμοποιήθηκε μία θερμική κάμερα και είχε σαν κύρια ιδέα ότι η θερμοκρασία του σώματος ενός ανθρώπου θα ξεπέρνα την θερμοκρασία του φόντου της υπόλοιπης εικόνας. Ωστόσο γρήγορα έγινε αντιληπτό πως αυτή η μέθοδος ήταν ελαττωματική για θερμές περιόδους του χρόνου όπως το καλοκαίρι, λόγω της αυξημένης θερμοκρασίας του περιβάλλοντος, δημιουργώντας ένα είδος καμουφλάζ γύρω από τον πεζό που θα υπάρχει στην εικόνα, καλύπτοντας έτσι την θερμοκρασία του σώματος του και κάνοντας έτσι πιο δύσκολη την εξαγωγή δεδομένων με οποιαδήποτε μέθοδο. Επιπρόσθετα μέθοδοι που απαιτούν την χρήση θερμικής κάμερας ή κάμερας νυχτερινής λήψης, δεν πρόκειται να είναι βιώσιμες ιδέες για μαζική παράγωγη λόγω του υψηλού κόστους που απαιτείται για την αγορά μίας τέτοιας κάμερας και για την χαμηλή ανάλυση που προσφέρει (640X480).

Ένα άλλο κριτήριο διαχωρισμού είναι το γεγονός αν η κάμερα θα τοποθετείται πάνω σε ένα σταθερό σημείο ή όχι. Στην περίπτωση χρήσης ενός Συνελικτικού Νευρωνικού Δικτύου αυτό παίζει σημαντικό ρόλο, διότι καθορίζει πόσο εύκολα θα καταφέρει το δίκτυο να αναγνωρίζει πεζούς μίας και αν η κάμερα βρίσκεται σε ένα σταθερό σημείο το φόντο της κάθε εικόνας θα παραμένει το ίδιο μην αλλάζοντας ραγδαία την πληροφορία που περιέχει η εικόνα. Μία πρόσφατη έρευνα που αξιοποίησε αυτήν την ιδέα, με την χρήση μίας Near Infrared (NIR) stereo network κάμερας και ένα νέο τρισδιάστατο μοντέλο ανίχνευσης πεζών [39], μέσω ελέγχου του μοντέλου για πολλές διαφορετικές καταστάσεις, τα αποτελέσματα που βρέθηκαν έδειξαν πως το μοντέλο είναι αποτελεσματικό και ανθεκτικό.

Υπάρχει και μία μελέτη όμως οι οποία αξιοποιεί τα Συνελικτικά Νευρωνικά Δίκτυα με το συνδυασμό ενός ασαφούς συστήματος συμπερασμάτων (fuzzy inference system ή FIS) [35] και την τεχνολογία δύο διαφορετικών ειδών κάμερας, Far-Infrared (FIR) και κανονικές. Το σύστημα έχοντας σαν είσοδο δύο τύπων εικόνας (μία κανονική και μία από την FIR κάμερα), διαλέγει την πιο κατάλληλη εικόνα με βάση την απόφαση του FIS και σε τελικό στάδιο αφήνει το δίκτυο να κρίνει για την ύπαρξη πεζών στην επιλεγμένη εικόνα.

Σε μεθόδους οι οποίες ενστερνίζονται την προσέγγιση της ανίχνευσης πεζών εν ώρα νυκτός με κανονική κάμερα και αξιοποιούν τον φωτισμό που παράγεται από το περιβάλλον (π.χ. φανάρια αμαξίου, φώτα από λάμπες στον δρόμο), χρησιμοποιούνται αρκετές μέθοδοι επεξεργασίας εικόνας. Οι δύο από τις πιο κλασικές μεθόδους είναι η histogram equalization (HE) [25] και η histogram specification (HS) [36], οι οποίες δυστυχώς, μετά την εφαρμογή τους παρουσιάζεται αύξηση θορύβου και μείωση συστατικών χαμηλής συχνότητας στην εικόνα. Έχει υπάρξει και μία γκάμα ερευνών που εφαρμόζουν αφαίρεση θορύβου και βελτιστοποίηση εικόνας, ωστόσο απαιτούν παραπάνω χρόνο υπολογισμού, καθυστερώντας αρκετά το σύστημα και μην επιτρέποντας το να δράσει σε πραγματικό χρόνο.

Κάθε έρευνα που εξετάστηκε χρησιμοποιούσε και ένα διαφορετικό συνδυασμό εργαλείων. Συχνά το Νευρωνικό Δίκτυο χρησιμοποιείται για τη δουλεία της κατάταξης, όπως και ένας Support Vector Machine (SVM) που προτιμάται στην πληθώρα των περιπτώσεων. Για την εργασία της εύρεσης χαρακτηριστικών χρησιμοποιούνται τεχνικές όπως Regions Of Interest (ROI) [34], HOG και Aggregated Channel Features (ACF) [26].

Σε αυτήν την έρευνα εξετάζουμε ορισμένες αρχιτεκτονικές που έχουν ήδη αναπτυχθεί και χρησιμοποιούνται σε συνεργασία με μερικές από τις προαναφερθείσες τεχνικές. Ωστόσο η εξέταση των δικτύων γίνεται χωρίς την χρήση περαιτέρω εργαλείων για να μελετηθεί το πόσο καλά μπορούν τα Συνελικτικά Νευρωνικά δίκτυα από μόνα τους να εκτελέσουν την εργασία της αναγνώρισης πεζών χωρίς καμία βοήθεια από κάποιον εξωτερικό παράγοντα.

4. ΣΥΝΟΛΟ ΔΕΔΟΜΕΝΩΝ

Το σύνολο δεδομένων(ΣΔ) είναι ένας από τους πιο σημαντικούς παράγοντες για την ανάπτυξη ενός αξιόπιστου ΝΔ. Κάθε ΣΔ, ή αλλιώς σύνολο εικόνων στην προκειμένη περίπτωση, είναι κρίσιμο να επιλέγεται προσεκτικά έτσι ώστε να περιέχει την κατάλληλη πληροφορία που θέλουμε να αναγνωρίζεται από το ΣΝΔ.

4.1 Caltech Pedestrian Dataset



Εικόνες 9-12: Παραδείγματα εικόνων του Caltech Pedestrian Dataset. Είναι εμφανές ότι κάθε φορά που εμφανίζονται πεζοί, δεν αποτελούν και μεγάλο μέρος της εικόνας, δυσκολεύοντας έτσι την προπόνηση του ΣΝΔ.

Κατά τη διάρκεια έρευνας θεωρίας για τα ΣΝΔ και τη στενή σχέση τους με τον τομέα της ανίχνευσης πεζών, αισθητή έκανε την παρουσία του ένα συγκεκριμένο σύνολο εικόνων, εν ονόματι Caltech Pedestrian Dataset [51], το οποίο ήταν ειδικά κατασκευασμένο για τέτοιου τύπου ΣΝΔ. Ποικίλες μελέτες βρέθηκαν οι οποίες με τον δικό τους τρόπο πετύχαιναν τον ίδιο στόχο, χρησιμοποιώντας αυτό το ΣΔ είτε για την προπόνηση του δικτύου, είτε για την εκτίμηση της τελικής του ακρίβειας. Το Caltech Pedestrian Dataset αποτελείται από μία συλλογή αρχείων βίντεο διάρκειας περίπου 10 ωρών και διαστάσεων 640x480 30Hz, το οποίο

τραβήχτηκε από μια κάμερα τοποθετημένη στο προστατευτικό τζάμι στο μπροστινό μέρος ενός αυτοκινήτου. Στο ΣΔ περιέχονται περίπου 250.000 frames τα οποία περιέχουν περίπου 350.000 οριοθετημένα κουτιά γύρω από τους υπάρχοντες πεζούς, οι οποίοι στο σύνολο είναι 2.300.

Το κύριο μειονέκτημα που παρατηρήθηκε κατά την διάρκεια χρήσης του για την διαδικασία της προπόνησης των δικτύων, ήταν πως για κάθε εικόνα που δεχόταν το ΝΔ σαν είσοδο, το ποσοστό πληροφορίας που αποτελούσε ο πεζός σε σύγκριση με το ποσοστό πληροφορίας που αποτελούσε η υπόλοιπη εικόνα(δρόμος, δέντρα, πεζοδρόμιο), έδειχνε να «μπερδεύει» τη διαδικασία αναγνώρισης του πεζού. Η μόνη λογική εξήγηση που μπορεί να δοθεί είναι ότι το ΣΝΔ δεν ήταν σε θέση να εστιάσει στον πεζό και να καταλάβει ότι αυτόν πρέπει να αναγνωρίζει, από την στιγμή που ο πεζός δεν καταλάμβανε αρκετό χώρο μέσα στην εικόνα. Εν τέλει, αυτός ήταν και ο κύριος λόγος που δεν χρησιμοποιήθηκε αυτό το ΣΔ για την ανάπτυξη της έρευνας.

4.2 1^ο ειδικά κατασκευασμένο ΣΔ

Έπειτα από εκτεταμένη αναζήτηση και διερεύνηση αρκετών ΣΔ αποφασίστηκε ότι ένα ειδικά κατασκευασμένο ΣΔ θα ήταν η βέλτιστη λύση στο πρόβλημα της εύρεσης ενός κατάλληλου. Μέσω λοιπόν της βοήθειας της μηχανής αναζήτησης Google έγινε εφικτό η κατασκευή ενός τέτοιου ΣΔ, συλλέγοντας δηλαδή διάφορες εικόνες πεζών με ποικίλες στάσεις σώματος. Με αυτόν τον τρόπο σιγουρεύουμε ότι το ΝΔ, όταν χρειαστεί, θα καταφέρει να αναγνωρίσει παραπάνω παραδείγματα πεζών που θα του παρουσιαστούν κατά την διάρκεια της διαδικασίας αξιολόγησης. Το σετ εικόνων που αποθηκεύτηκε, αποτελείται από εικόνες πεζών στο φως της ημέρας. Αφότου σχηματίστηκε μία συλλογή επαρκούς αριθμού εικόνων, ήταν αναγκαίο να εφαρμοστεί αλλαγή μεγέθους σε κάθε εικόνα έτσι ώστε να συνάδουν με την αρχιτεκτονική των επιπέδων των ΣΝΔ.

Κατόπιν της διαδικασίας της αξιολόγησης των ΝΔ που είχαν κατασκευαστεί, έγινε γρήγορα εμφανές ότι το ΝΔ είχε να κάνει πάντα μία πρόβλεψη ανάμεσα σε δύο κλάσεις, αν δηλαδή η εικόνα περιέχει πεζό ή όχι. Εν ολίγοις κάθε πρόβλεψη που γινόταν είχε 50% πιθανότητα να είναι ορθή και 50% να είναι λανθασμένη. Μάλλον λόγω της έλλειψης επιλογών κλάσεων, το ΝΔ ανεξαρτήτως του αποτελέσματος που θα υπήρχε, προέβλεπε ως

αποτέλεσμα την ίδια κλάση μιας και το ρίσκο που έπαιρνε να είναι λανθασμένο ήταν 50%, ίσο με την πιθανότητα να είναι σωστή η πρόβλεψη. Δυστυχώς όμως αυτή η συμπεριφορά του δικτύου ακύρωνε την γενική ιδέα ότι απλά θα εξήγαγε χαρακτηριστικά από την εικόνα στην είσοδο και θα έκρινε αντικειμενικά την ύπαρξη πεζού σε αυτήν.

Η λύση που βρέθηκε στο συγκεκριμένο πρόβλημα ήταν να διευρυνθεί το πεδίο πρόβλεψης του δικτύου, ήτοι προσθέτοντας παραπάνω πιθανές κλάσεις πρόβλεψης. Για να εφαρμοστεί αυτή η μέθοδος χρειάστηκε να επεκτείνουμε το ΣΔ που ήδη υπήρχε και να προστεθούν εικόνες από άλλες δύο τυχαία επιλεγμένες κλάσεις. Φυσικά και αυτές οι εικόνες δέχτηκαν την ίδια προεπεξεργασία με τις εικόνες των πεζών, έτσι ώστε να είναι συμβατές με το δίκτυο. Οι τελικές εικόνες που περιέχονται στο ΣΔ είναι συνολικά 1756 τύπου RGB, όπου οι 684 είναι εικόνες πεζών, οι 632 ανήκουν στην πρώτη τυχαία κλάση και 440 στην δεύτερη τυχαία κλάση. Όλες έχουν τετραγωνικό μέγεθος, όπως προτείνεται σε ορισμένες μελέτες, $64 * 64$ (ύψος * φάρδος) αλλά αρκετά μικρό, λόγω της ανικανότητας του υπολογιστικού συστήματος να αντέξει κάποιο αρκετά μεγαλύτερο μέγεθος. Αυτό συμβαίνει διότι το μέγεθος των διαστάσεων κάθε επιπέδου μίας αρχιτεκτονικής συνδέεται στενά με το μέγεθος της εικόνας που θα λάβει σαν είσοδο. Δηλαδή όσο πιο πολλά τα pixel της εικόνας τόσο πιο μεγάλο το επίπεδο πάνω στο οποίο θα εκτελεστούν πράξεις (όπως συνέλιξη, μέγιστη συγκέντρωση).

4.3 2^ο ειδικά κατασκευασμένο ΣΔ

Το δεύτερο και τελευταίο ΣΔ που δημιουργήθηκε επικεντρώνεται σε μία διαφορετική προσέγγιση προπόνησης του δικτύου, τροφοδοτώντας όπως και στο προηγούμενο ΣΔ, εικόνες τριών διαφορετικών κλάσεων. Το μόνο χαρακτηριστικό που διαφοροποιεί αυτό το σύνολο εικόνων από το προηγούμενο είναι η αντικατάσταση του υποσυνόλου που περιέχει πεζούς εν ώρα ημέρας με πεζούς εν ώρα νυκτός. Οι εικόνες βρέθηκαν με παρόμοιο τρόπο όπως και οι εικόνες στο προηγούμενο ΣΔ και δέχτηκαν την ίδια προεπεξεργασία. Δυστυχώς δεν βρέθηκαν τόσες εικόνες πεζών, όσες υπάρχουν στο προηγούμενο ΝΔ, θέτοντας έτσι τα αποτελέσματα αυτού του σετ λιγότερο αξιόπιστα από τα αποτελέσματα του πρώτου σετ. Στο σύνολο του το ΣΔ αποτελείται από 1319 εικόνες, όπου 632 ανήκουν στην πρώτη τυχαία κλάση, 440 στην δεύτερη τυχαία κλάση και 247 στην κλάση των πεζών.

Πρέπει ακόμη να αναφερθεί ότι το υποσύνολο των πεζών αυτού του ΣΔ χρησιμοποιήθηκε και ως σετ αξιολόγησης των ΣΝΔ που προπονήθηκαν από τα δύο ειδικά κατασκευασμένα ΣΔ. Τοιουτοτρόπως όλα τα δίκτυα αξιολογούνται ορθολογικά ανεξαρτήτως του ΣΔ που χρησιμοποιείται για την προπόνηση και εξασφαλίζουν ότι η σύγκριση των αποτελεσμάτων θα γίνει με κοινό γνώμονα, το ίδιο σετ εικόνων.

5. ΥΛΟΠΟΙΗΣΗ

5.1 Εισαγωγή στο εργαλείο TensorFlow

Ένας από τους κύριους παράγοντες που διευκόλυναν αρκετά την ολοκλήρωση αυτής της έρευνας ήταν η βιβλιοθήκη λογισμικού ανοιχτού κώδικα με το όνομα TensorFlow, που χρησιμοποιήθηκε στη διαδικασία της συγγραφής κώδικα. Η βιβλιοθήκη που δημιουργήθηκε και δημοσιεύτηκε δωρεάν από την Google, εν ονόματι TensorFlow, προσφέρει μία τεράστια ποικιλία έτοιμων συναρτήσεων για μηχανική μάθηση (π.χ. Τεχνητά Νευρωνικά Δίκτυα, Βαθιά Μάθηση, Δίκτυα Bayes). Επίσης το TensorFlow χρησιμοποιεί γραφήματα ροής δεδομένων για να αναπαραστήσει υπολογισμούς, κοινές καταστάσεις και την λειτουργία που αλλάζει αυτές τις καταστάσεις. Οι κόμβοι που υπάρχουν σε αυτά τα γραφήματα ονομάζονται Tensors.

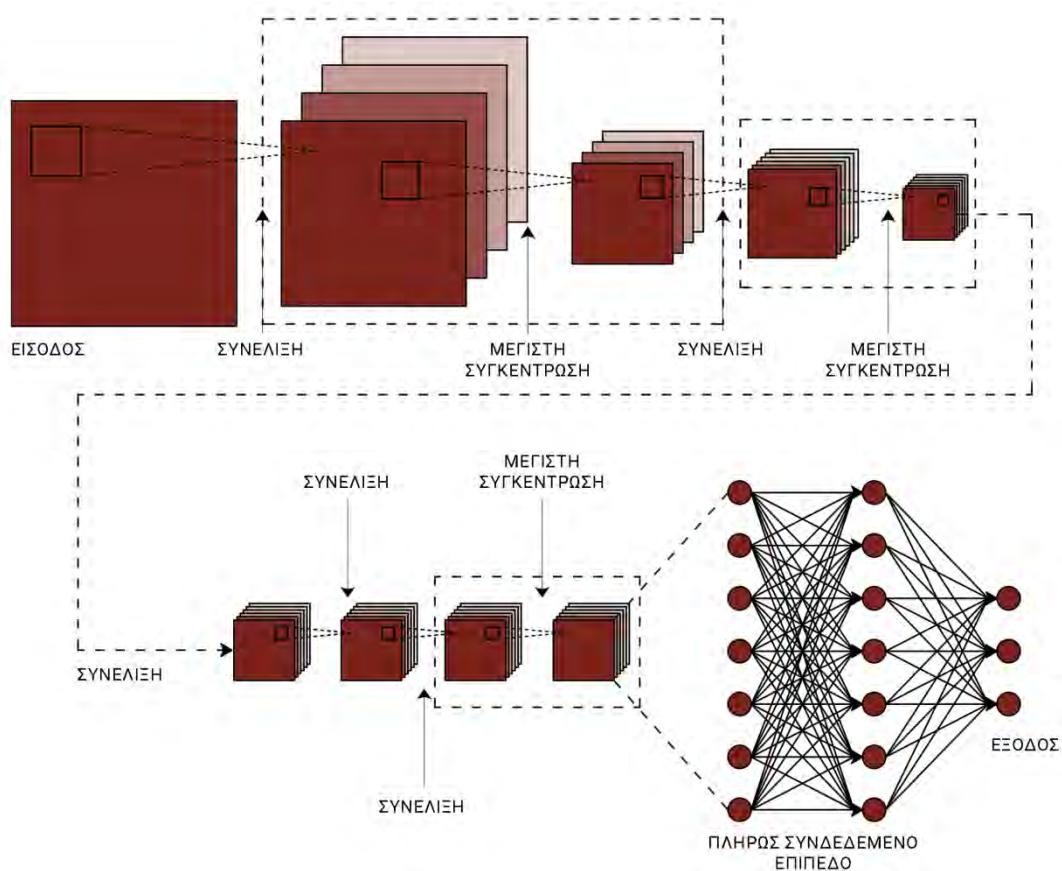
Το TensorFlow, όπως υποδηλώνει το όνομα, είναι ένα πλαίσιο για τον ορισμό και την εκτέλεση υπολογισμών που αφορούν τους Tensors. Ένας Tensor είναι μια γενίκευση των διανυσμάτων και των πολυδιάστατων πινάκων. Εσωτερικά το TensorFlow θεωρεί τους Tensors ως πολυδιάστατους πίνακες βασικών τύπων δεδομένων [\[40\]](#).

Σημαντικό χαρακτηριστικό του TensorFlow είναι ότι έχει δημιουργηθεί έτσι ώστε η εκτέλεση του κάθε προγράμματος να έχει τη δυνατότητα να εκμεταλλεύεται βέλτιστα κάθε υπολογιστικό πόρο(CPU, GPU, TPU) που μπορεί να διαθέσει το υπολογιστικό σύστημα, αλλά και να βοηθήσει στην διαδικασία συγγραφής κώδικα, μέσω διάφορων χρήσιμων API. Ένα από τα πιο χρήσιμα που αξιοποιηθήκαν σε αυτήν την έρευνα είναι το API των Estimators, το οποίο βοηθά στην αυτοματοποίηση της προπόνησης, αξιολόγησης και ανάπτυξης μοντέλων μηχανικής μάθησης [\[39\]](#). Σε αυτήν την έρευνα κάθε διαφορετική αρχιτεκτονική ΝΔ που χρησιμοποιήθηκε, αποτελεί και έναν ξεχωριστό Estimator.

5.2 Ανάλυση αρχιτεκτονικών που χρησιμοποιήθηκαν

Κάθε αρχιτεκτονική που εξετάστηκε για την συμβατότητα της με το έργο της αναγνώρισης πεζών εν ώρα νυκτός, έχει δομηθεί έτσι ώστε είτε να αναγνωρίζει αυτούσιες τις εικόνες πεζών, οι οποίες έχουν δεχτεί ορισμένες αλλαγές μέσω διάφορων περιγραφέντων εικόνων (αλλιώς image descriptor [\[52\]](#)), είτε να δέχεται κομμάτια εικόνων τα οποία έχουν επιλεγεί από προηγούμενες μεθόδους ως υποψήφια κομμάτια να περιέχουν πεζό. Οφείλει να ληφθεί υπόψη ότι όλες οι αρχιτεκτονικές ΣΝΔ είναι μόνο αυτοτελή κομμάτια μεθόδων αναγνώρισης πεζών και δεν δέχονται σαν είσοδο ακατέργαστες εικόνες τους. Επίσης έχουν εφαρμοστεί ορισμένες μετατροπές στις αρχιτεκτονικές δικτύων, με σκοπό να μπορούν να υποστηριχτούν από το σύστημα το οποίο θα υλοποιούσε τις δοκιμές. Λόγω της έλλειψης μνήμης που παρέχεται από την GPU (MSI GeForce GTX 750 Ti 2GB), το σύστημα αδυνατούσε να υποστηρίξει ΣΝΔ με πολλούς κόμβους, καθώς η GPU δεν είναι ικανή να τα αποθηκεύσει στην μνήμη όλα τα επίπεδα κόμβων. Για αυτό το λόγο ήταν εξίσου σημαντικό να μειωθεί το μέγεθος κάθε εικόνας που λαμβανόταν σαν είσοδος και το μέγεθος της εικόνας ήταν ένας καθοριστικός παράγοντας του μεγέθους και της περιπλοκότητας του δικτύου. Όσο μεγαλύτερη είναι η εικόνα, τόση παραπάνω πληροφορία περιέχει η εικόνα και τόσο μεγαλύτερη θα είναι η πολυπλοκότητα του δικτύου.

5.2.1 1^η δομή δικτύου



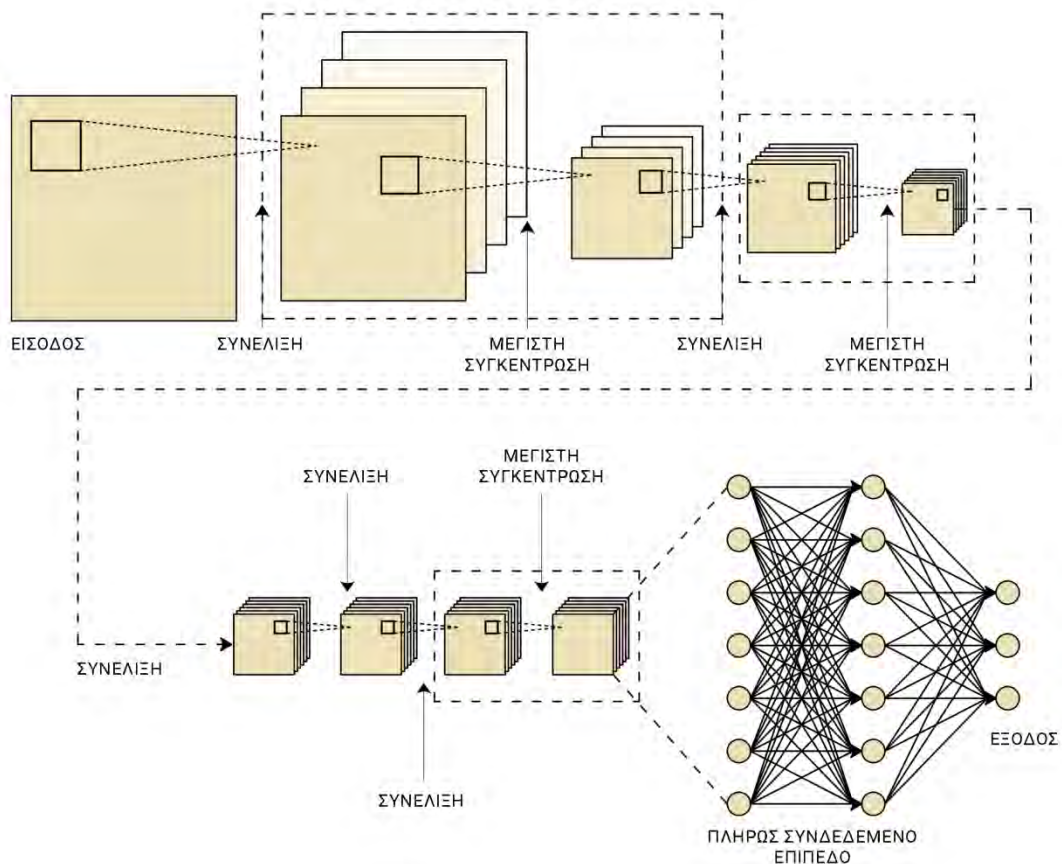
Εικόνα 10: Αρχιτεκτονική του πρώτου δικτύου

Το πρώτο δίκτυο βρέθηκε από την έρευνα [17] που διεξήχθη με κύριο γνώμονα την αναγνώριση πεζών τις νυκτερινές ώρες με χρήση κάμερας ορατού φωτός. Αρκετές έρευνες έχουν ήδη διεξαχθεί για την αναγνώριση πεζών χρησιμοποιώντας ειδικού τύπου τεχνολογίας λήψης εικόνας. Ένας τρόπος είναι μέσω υπέρυθρων φωτιστών (Near-infrared Illuminators) ή κάμερες θερμότητας (thermal camera). Εντούτοις τέτοιου είδους τεχνικές δεν αποτελούν βιώσιμες λύσεις για μαζική παράγωγή, λόγω αυξημένων οικονομικών απαιτήσεων.

Η αρχιτεκτονική αυτή αποτελείται από πέντε επίπεδα συνέλιξης, έξι επίπεδα ReLU, τρία επίπεδα μέγιστης συγκέντρωσης και δύο πλήρως συνδεδεμένα επίπεδα όπως φαίνεται και στην εικόνα 10. Σαν είσοδο λαμβάνει εικόνες μεγέθους 183*119 οι οποίες προτού εισαχθούν στο δίκτυο δέχονται κανονικοποίηση μεγέθους (size normalization) και ισοστάθμιση ιστογράμματος (histogram equalization).

Τα αποτελέσματα αυτής της μεθόδου αποδείχτηκαν εξαιρετικά σε σύγκριση με παρόμοιες μελέτες για αναγνώριση πεζών την νύχτα, σε πολλαπλά περιβάλλοντα και μάλιστα σε πραγματικό χρόνο. Παρατηρήθηκε επίσης ότι το δίκτυο απέδιδε καλύτερα όταν ως είσοδο δεχόταν τον συνδυασμό και των τριών ΣΔ.

5.2.2 2^η δομή δικτύου



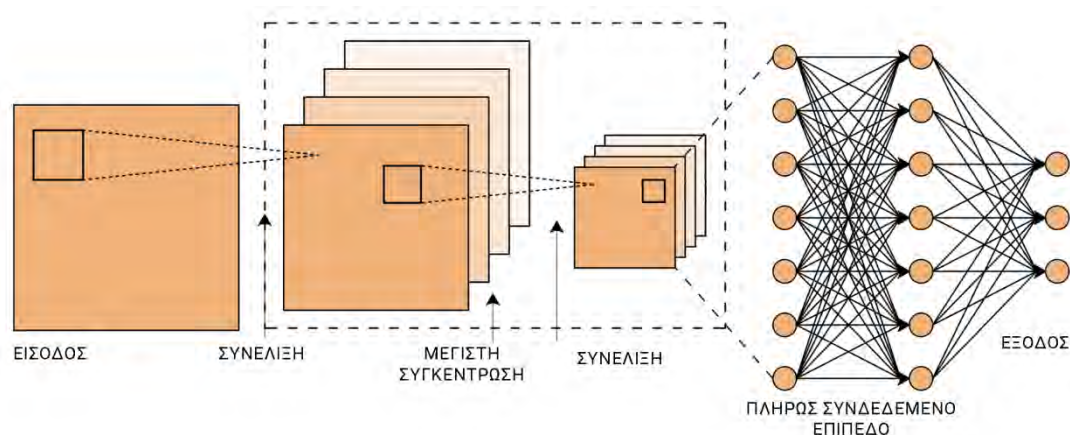
Εικόνα 11: Αρχιτεκτονική του δεύτερου δικτύου

Η δεύτερη έρευνα που μελετήθηκε κατάφερε να συνδυάσει αποτελεσματικά μία αλληλουχία εργαλείων τα οποία αποσκοπούν στην εξαγωγή χαρακτηριστικών από μία εικόνα. Αρχικά χρησιμοποιήθηκε ένας Aggregated Channel Features (ACF) [26] και έπειτα ένα πολυεπίπεδο ΣΝΔ. Το ΣΝΔ προτείνεται να έχει προπονηθεί με ένα αρκετά μεγάλο ΣΔ, έτσι ώστε να είναι χρήσιμο σε πιο περίπλοκες εικόνες. Σε τελικό στάδιο ένας απλός γραμμικός Support Vector Machine (SVM) [28] για το τελικό στάδιο της κατάταξης.

Το πολυεπίπεδο ΣΝΔ που μπορεί να φανεί στην εικόνα 11 αποτελείται από 7 επίπεδα και δέχεται σαν είσοδο εικόνες 227*227 pixels. Λόγω τεχνικών προβλημάτων όπως αναφέρθηκε και στην εισαγωγή αυτού του κεφαλαίου κρίθηκε αναγκαίο να αλλάξει το τελευταίο επίπεδο από 4096 κόμβους σε 1024.

Τα αποτελέσματα αυτής της μεθόδου αποδείχτηκαν πολλά υποσχόμενα. Στην διαδικασία αξιολόγησης έλαβαν μέρος τρία ΣΔ για ανίχνευση πεζών (INRIA, Caltech, ΕΤΗ).

5.2.3 3^η δομή δικτύου



Εικόνα 12: Αρχιτεκτονική του τρίτου δικτύου

Η Τρίτη και τελευταία αρχιτεκτονική ΣΝΔ που ήταν εφικτό να ενσωματώσουμε στην ερευνά, είναι ένα δίκτυο αρκετά μικρό σε μέγεθος το οποίο σαν στόχο δεν έχει τόσο την ακρίβεια, αλλά την ταχύτητα με την οποία επιτυγχάνεται η ανίχνευση. Η έρευνα στην οποία βρέθηκε το δίκτυο αυτό, εφαρμόζει πρώτα έναν καταρράκτη ταξινομητών [29] (VeryFast Cascade [30]) για να αποκλείσει κομμάτια εικόνας που γίνεται εύκολα κατανοητό ότι δεν περιέχουν πεζό και μετέπειτα προωθεί στο ΣΝΔ οτιδήποτε δεν κρίνει απορριπτέο.

Πρωτίστως το «μικρό» ΣΝΔ, όπως αναφέρεται και μέσα στην έρευνα, είναι βασισμένο σε ένα μεγαλύτερο ΣΝΔ, στηρίζεται στο ΣΝΔ του Krizhevsky [32]. Η αρχιτεκτονική του μικρού δικτύου είναι ορατή στην εικόνα 12 και όπως μπορεί να παρατηρηθεί από την έλλειψη αριθμού στρωμάτων συνελίξεων, το δίκτυο δεν είναι αρκετά ικανό να εξαγάγει πολλά χαρακτηριστικά από μία εικόνα.

Παρόλα αυτά όμως ο αλγόριθμος που παρουσιάζεται να συνδυάζει VeryFast Cascade και ΣΝΔ, επιτυγχάνει να εκτελείται σε πραγματικό χρόνο στα 15 frames ανά δευτερόλεπτο, έχοντας ένα μέσο ποσοστό απώλειας 26,2%. Το ποσοστό βρέθηκε χρησιμοποιώντας ως δείκτη αναφοράς το Caltech pedestrian dataset benchmark. Αξιοσημείωτο είναι ο συνδυασμός ταχύτητας και υψηλής ακρίβειας που πραγματοποιείται σε ένα από τα πιο δύσκολα και καλά ΣΔ που υπάρχουν μέχρι σήμερα.

6. ΑΠΟΤΕΛΕΣΜΑΤΑ ΔΟΚΙΜΩΝ ΚΑΙ ΠΕΙΡΑΜΑΤΑ

Συνολικά σε αυτήν την ενότητα ελέγχθηκαν τρεις διαφορετικές αρχιτεκτονικές ΣΝΔ ως προς την συμβατότητα τους με την αναγνώριση μη επεξεργασμένων εικόνων πεζών εν ώρα νυκτός, χρησιμοποιώντας δύο διαφορετικά ΣΔ. Είναι γνωστό ότι κάθε ΣΝΔ για να είναι σε θέση να αποδώσει βέλτιστα απαιτείται και η επιλογή του κατάλληλου συνδυασμού χαρακτηριστικών που θα χρησιμοποιηθούν κατά την διάρκεια της προπόνησης του δικτύου, όπως είναι ο ρυθμός μάθησης και το μέγεθος παρτίδας εικόνων που θα ληφθούν ως είσοδος σε κάθε επανάληψη.

Αρχικά για να γίνει εφικτό να πάρουμε βάσιμα και λεπτομερή αποτελέσματα για την ακρίβεια της πρόβλεψης κάθε αρχιτεκτονικής ΣΝΔ, δημιουργήθηκε μία επαναληπτική εργασία μέσα στην οποία επιλέγεται ποια αρχιτεκτονική θα χρησιμοποιηθεί και με τι χαρακτηριστικά. Αφότου οριστούν όλα τα προαπαιτούμενα χαρακτηριστικά (ρυθμός μάθησης και μέγεθος batch), τότε ξεκινά η διαδικασία της προπόνησης ενός ΣΝΔ και έπειτα αξιολογείται το δίκτυο από το σετ εικόνων που έχει επιλεγεί. Για τον ίδιο συνδυασμό χαρακτηριστικών που επιλέχτηκαν, η διαδικασία της προπόνησης θα επαναληφθεί είκοσι φορές, προπονώντας έτσι είκοσι διαφορετικά μοντέλα. Αυτή η διαδικασία θα διεξαχθεί για κάθε αρχιτεκτονική χρησιμοποιώντας όλους τους πιθανούς συνδυασμούς χαρακτηριστικών. Τα χαρακτηριστικά αυτά όπως παρουσιάστηκαν και στο κεφάλαιο της Εισαγωγής στα Νευρωνικά Δίκτυα είναι ο ρυθμός μάθησης και το μέγεθος παρτίδας τα οποία μπορούν να πάρουν τις τιμές [0.00001, 0.000025, 0.00005, 0.000075, 0.0001, 0.0005, 0.001] και [1, 50, 100, 150, 200] αντίστοιχα. Με βάση αυτά τα αποτελέσματα δημιουργήθηκαν και τα τρισδιάστατα ραβδογραφήματα που θα εμφανιστούν στην επόμενη υποενότητα, που δείχνουν τον μέσο όρο των μοντέλων που αναπαριστούν. Για παράδειγμα σε ένα γράφημα που έχει τρεις άξονες, τον άξονα του μήκους που αναπαριστά το μέγεθος παρτίδας, τον άξονα του πλάτους που αναπαριστά τον ρυθμό μάθησης και τον άξονα του ύψους που αναπαριστά την ακρίβεια των μοντέλων, κάθε ράβδος δείχνει τον μέσο όρο όλων των μοντέλων που έχουν εκπαιδευτεί με τις τιμές των χαρακτηριστικών που υπάρχουν στους άλλους δύο άξονες.

Σε όλες τις αρχιτεκτονικές εφαρμόστηκε ο ίδιος αλγόριθμος διόρθωσης λάθους κατά την πίσω διάδοση (ή αλλιώς back propagation) [2], ονόματι Adam (Adaptive Moment

Estimation) [3] και ο ίδιος τρόπος υπολογισμού απώλειας. Πρέπει επίσης να αναφερθεί πως τα βάρη των φίλτρων, που χρησιμοποιήθηκαν στα συνελκτικά επίπεδα των δικτύων, σε κάθε επανάληψη αρχικοποιούνται με τυχαίες τιμές, μιας και δεν βρέθηκαν ήδη προπονημένα δίκτυα, μην επιτρέποντας έτσι την χρήση τεχνικών μεταφοράς μάθησης [1], όπως η χρήση ήδη εκπαιδευμένων φίλτρων.

Εν τέλει για την καλύτερη προπόνηση του δικτύου, κάθε εικόνα επιλέγεται τυχαία για να τροφοδοτηθεί στο δίκτυο σαν είσοδος την στιγμή της προπόνησης του, αποφεύγοντας έτσι το φαινόμενο του overfitting [4]. Επειδή τα ΣΔ που χρησιμοποιούνται, περιέχουν εικόνες τριών κλάσεων όπως εξηγήθηκε και προηγουμένως, η τυχειότητα στην επιλογή της σειράς τροφοδότησης των εικόνων εξασφαλίζει ότι το δίκτυο που βρίσκεται ακόμα στο στάδιο της προπόνησης, δεν θα λαμβάνει παρτίδες εικόνων της ίδιας κλάσης σαν είσοδο. Παρακάτω θα αναλυθούν τα αποτελέσματα όλων των τεστ που τελέστηκαν σε όλες τις αρχιτεκτονικές που χρησιμοποιήθηκαν σε συνδυασμό με τα δύο ειδικώς κατασκευασμένα ΣΔ. Συνολικά εκπαιδεύτηκαν 2100 ΣΝΔ για κάθε ΣΔ, δηλαδή για κάθε διαφορετικό συνδυασμό χαρακτηριστικών προπονήθηκαν είκοσι ξεχωριστά ΣΝΔ, από τα οποία μόλις τελείωνε η διαδικασία της προπόνησης τους, πέραναν από ένα σετ αξιολόγησης που είχε τεθεί εξ αρχής και ήταν το ίδιο κάθε φορά για να εξασφαλιστεί πως θα υπάρχει ένας κοινός γνώμονας σύγκρισης των δικτύων. Ο λόγος που εκτελούνται είκοσι τεστ και το κάθε μοντέλο παρουσιάζει μια διαφορετική ακρίβεια, μολοντί δεν αλλάζει τίποτα άλλο, είναι ότι σε κάθε επανάληψη της προπόνησης του δικτύου, τα βάρη των φίλτρων στα επίπεδα των συνελίξεων, αρχικοποιούνται τυχαία και για αυτόν τον λόγο κάθε επανάληψη παράγει και ένα διαφορετικό μοντέλο.

Σε αυτό το κομμάτι της έρευνας με τη βοήθεια των είκοσι επτά εικόνων 13 έως και 42 γίνεται εμφανής η συμπεριφορά των τριών διαφορετικών αρχιτεκτονικών που χρησιμοποιήθηκαν και στις έρευνες [17, 21, 44]. Τα διαγράμματα που χρησιμοποιούν σαν άξονα μήκους τον αριθμό της αρχιτεκτονικής, περιγράφουν τα αποτελέσματα ακρίβειας για κάθε αρχιτεκτονική, συναρτήσεως της τιμής του χαρακτηριστικού, που μπορεί να εμφανιστεί στον άξονα του πλάτους. Μπορεί να παρατηρηθεί ότι υπάρχουν τριών ειδών ραβδοδιαγράμματα. Το πρώτο είδος προβάλλει στατιστικά για μοντέλα τα οποία με βάση την ακρίβεια τους θεωρούνται αποδεκτά (εικόνες 13, 16, 19, 23, 25, 28, 31, 34 και 37), δηλαδή παρουσιάζουν ποσοστά ακρίβειας μεγαλύτερα από 70%. Το δεύτερο είδος παρουσιάζει τα μη αποδεκτά μοντέλα που προπονήθηκαν (εικόνες 15, 18, 21, 24, 27, 30, 33, 36 και 40), ήτοι τα μοντέλα που παρουσιάζουν ποσοστά ακρίβειας μικρότερα από 70%. Το τρίτο και

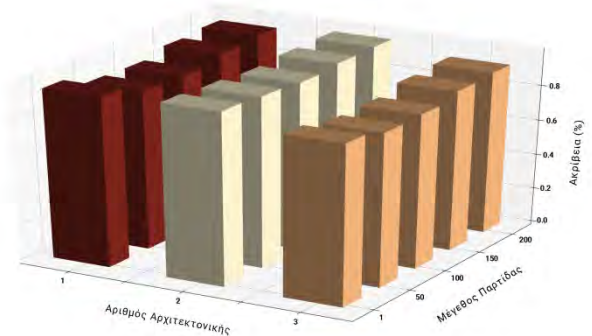
τελευταίο είδος απεικονίζει τα στατιστικά για όλα τα μοντέλα που έχουν αναπτυχθεί σχηματίζοντας μια γενική εικόνα των μοντέλων που προπονούνται με την βοήθεια του ΣΔ. Παρόμοια διαγράμματα θα παρατηρηθεί και στα επόμενα υποκεφάλαια που θα υποδεικνύονται οι συμπεριφορές των δικτύων με τη χρήση του δεύτερου ΣΔ και την βοήθεια μίας τεχνικής επαναδειγματοληψίας με το όνομα k-fold cross validation.

Κάτι τελευταίο που πρέπει να αναφερθεί είναι ότι οι τιμές 1, 2 και 3 που χρησιμοποιούνται για την αναπαράσταση τιμών αριθμού αρχιτεκτονικής ουσιαστικά αναφέρονται στις δομές δικτύων που έχουν ήδη προαναφερθεί στο κεφάλαιο 5. Αντίστοιχα οι τιμές 1, 2, 3 αναπαριστούν την πρώτη ([5.2.1](#)), δεύτερη ([5.2.2](#)) και τρίτη ([5.2.3](#)) δομή δικτύου

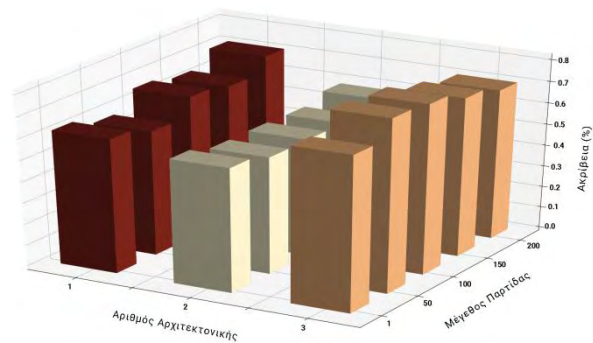
6.1 1^ο Σύνολο Δεδομένων

Από ό,τι μπορεί να φανεί από τις εικόνες 13 και 14, εφόσον ο μέσος όρος όλων των μοντέλων που φαίνονται στην εικόνα 13 είναι αρκετά υψηλός για κάθε συνδυασμό μπορούμε με ευκολία να συμπεράνουμε ότι το μέγεθος παρτίδας, δεν διαφοροποιεί έντονα τα αποτελέσματα και των τριών αρχιτεκτονικών. Αν όμως εστιάσουμε στο γράφημα της εικόνας 14 παρατηρούμε ότι η τρίτη αρχιτεκτονική είχε καλύτερη απόδοση από τις υπόλοιπες, μιας και ακόμα και τα μη αποδεκτά μοντέλα της, ξεπερνούσαν σε μέσο όρο ακρίβειας τα αποτελέσματα των άλλων δύο αρχιτεκτονικών.

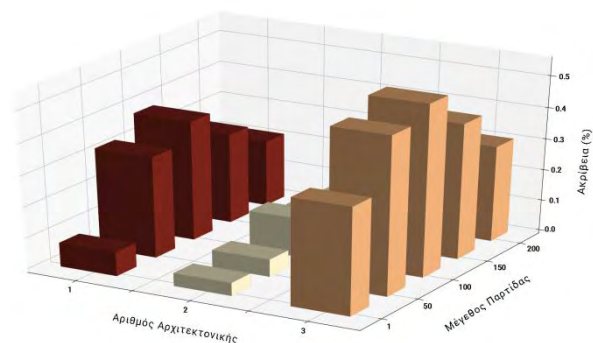
Μπορεί επίσης να παρατηρηθεί από τις εικόνες 14 και 15, ότι η δεύτερη αρχιτεκτονική δεν βγάζει καλύτερα αποτελέσματα από τις άλλες δυο. Αυτό μπορεί να συμβαίνει λόγω της πολυπλοκότητας που παρουσιάζει η δεύτερη αρχιτεκτονική του δικτύου και η αλλαγή που έχει γίνει στο δεύτερο πλήρως συνδεδεμένο επίπεδο που από 4096 κόμβους, εξαιτίας τεχνικών προβλημάτων (έλλειψη μνήμης στην GPU), μειώθηκε στους 1024 κόμβους.



Εικόνα 13: Διάγραμμα μεγέθους παρτίδας και της αρχιτεκτονικής που χρησιμοποιήθηκε, για αποδεκτά μοντέλα



Εικόνα 14: Διάγραμμα μεγέθους παρτίδας και της αρχιτεκτονικής που χρησιμοποιήθηκε, για μη αποδεκτά μοντέλα

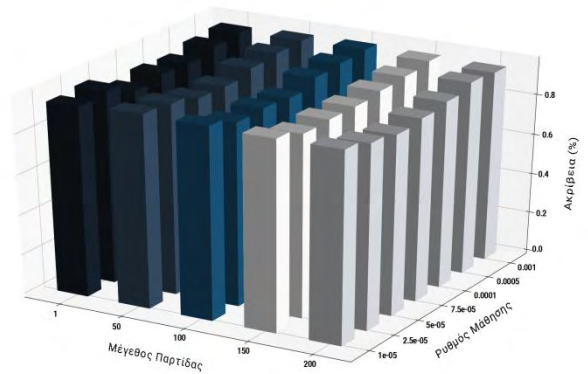


Εικόνα 15: Διάγραμμα μεγέθους παρτίδας και της αρχιτεκτονικής που χρησιμοποιήθηκε, για μη αποδεκτά μοντέλα

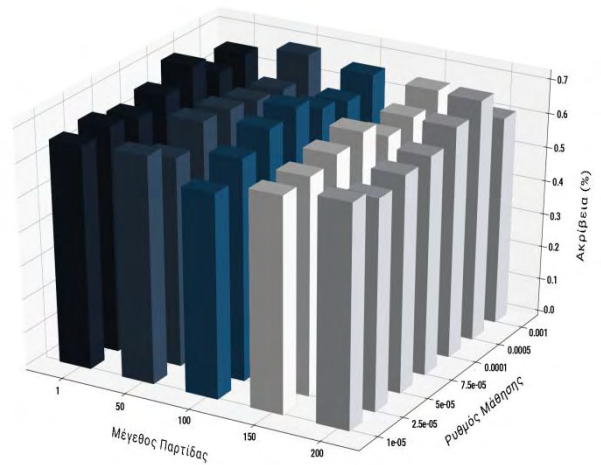
Επιπρόσθετα από ότι γίνεται γνωστό και από την έρευνα [\[44\]](#), όπου παρουσιάζεται η χρήση της αρχιτεκτονικής, χρησιμοποιείται μόνο για την εξαγωγή χαρακτηριστικών ενώ η πρόβλεψη της κλάσης γίνεται από ένα SVM (Support Vector Machine). Προσέχοντας την εικόνα 15, τα αποτελέσματα που αποφέρει η δεύτερη αρχιτεκτονική, είναι υπερβολικά χαμηλές τιμές ακρίβειας. Έπειτα κοιτάζοντας και το διάγραμμα στην εικόνα 14 και παρατηρώντας τις εξίσου χαμηλές τιμές ακρίβειας όλων των μοντέλων που χρησιμοποιούν την δεύτερη αρχιτεκτονική, συμπεραίνεται εύκολα ότι τα μοντέλα που δεν παρουσιάζουν αποδεκτές τιμές ακρίβειας αποτελούν την πλειονότητα των μοντέλων που προπονούνται με την χρήση της δεύτερης αρχιτεκτονικής.

Αντίθετα με την πρώτη αρχιτεκτονική, η τρίτη φαίνεται και στα τρία διαγράμματα να παρουσιάζει τα καλύτερα αποτελέσματα. Πολύ πιθανόν να οφείλεται στην απλότητα που παρουσιάζει η αρχιτεκτονική, καθώς το δίκτυο απαρτίζεται από 7 επίπεδα τα οποία φαίνονται ελάχιστα σε σύγκριση με τα 13 επίπεδα που έχει η δεύτερη αρχιτεκτονική. Βεβαίως το «βάθος» ενός ΣΝΔ δεν είναι αναγκαστικά καθοριστικός παράγοντας της ακρίβειας του, μιας και ένα πιο βαθύ ΣΝΔ σημαίνει πως μπορεί να εκτελέσει μία πιο εξειδικευμένη εξαγωγή χαρακτηριστικών. Το μόνο μειονέκτημα ωστόσο που θεωρείται ότι επηρεάζει αρνητικά την ακρίβεια του δικτύου είναι ένα φαινόμενο ονόματι υπερπροσαρμογή (ή αλλιώς overfitting) [\[52\]](#). Μεγαλύτερα ΣΝΔ είναι πιο επιρρεπή σε υπερπροσαρμογή και ένας τρόπος αντιμετώπισης τους είναι τα επίπεδα dropout που μηδενίζουν τυχαία κάποια βάρη (ποια βάρη θα είναι αυτά, αποφασίζεται από την θέση που θα τοποθετηθεί επίπεδο dropout). Από ότι φαίνεται το ήδη υπάρχων dropout επίπεδο δεν ήταν αρκετό στην περίπτωση μας, στην περίπτωση που όντως το δίκτυο παρουσίαζε υπερπροσαρμογή.

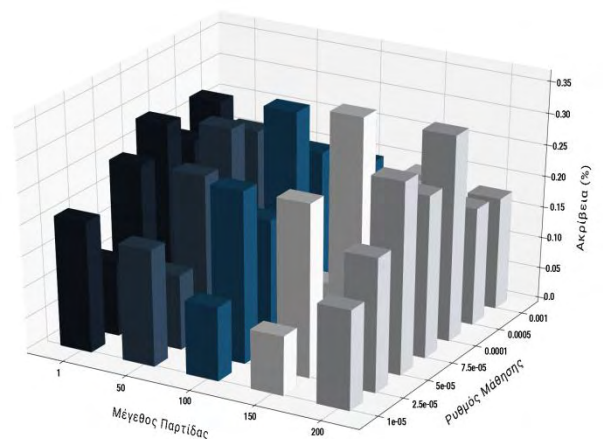
Τα διαγράμματα στις εικόνες 16 έως και 18 αποσκοπούν στην εύρεση ενός συνδυασμού χαρακτηριστικών (ρυθμός μάθησης και μέγεθος παρτίδας) που να ξεχωρίζει σε οποιαδήποτε αρχιτεκτονική και να χρησιμοποιηθεί. Από τα σχήματα 16 και 17 φαίνεται πως οποιοσδήποτε συνδυασμός των δυο αυτών χαρακτηριστικών δεν έχει κάνει κάποια μεγάλη διαφορά. Αυτή η πληροφορία, κάνει πιο εμφανές το γεγονός ότι κάθε μοντέλο χρειάζεται λειτουργεί βέλτιστα με έναν συγκεκριμένο συνδυασμό χαρακτηριστικών που έχει βρεθεί έπειτα από μία εξονυχιστική έρευνα με τη μέθοδο του *brute force*.



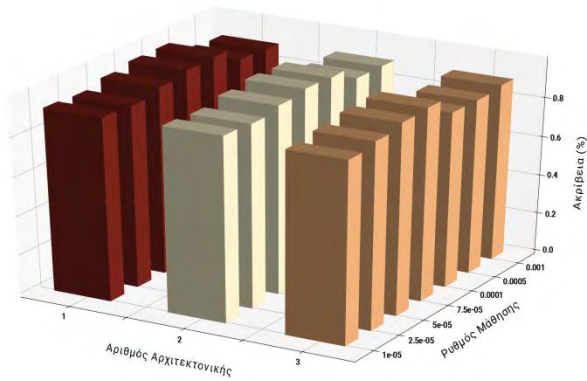
Εικόνα 16: Διάγραμμα ρυθμού μάθησης και μεγέθους παρτίδας για αποδεκτά μοντέλα



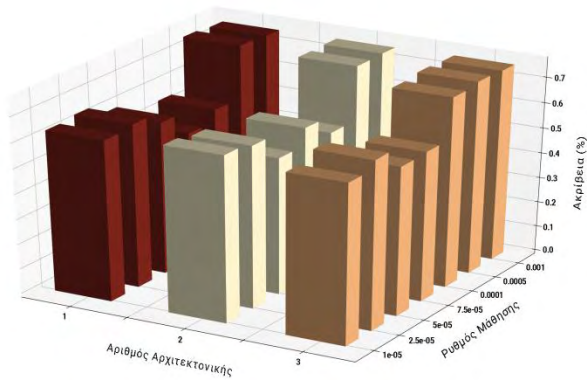
Εικόνα 17: Διάγραμμα ρυθμού μάθησης και μεγέθους παρτίδας, για όλα τα μοντέλα



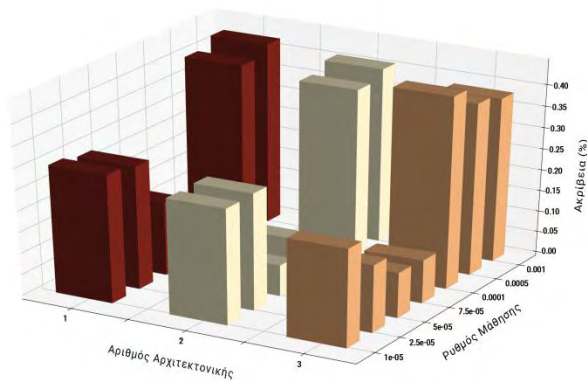
Εικόνα 18: Διάγραμμα ρυθμού μάθησης και μεγέθους παρτίδας, για μη αποδεκτά μοντέλα



Εικόνα 21: Διάγραμμα ρυθμού μάθησης και αρχιτεκτονικής που χρησιμοποιήθηκε, για αποδεκτά μοντέλα



Εικόνα 19: Διάγραμμα ρυθμού εκμάθησης και αρχιτεκτονικής που χρησιμοποιήθηκε για όλα τα μοντέλα



Εικόνα 20: Διάγραμμα ρυθμού εκμάθησης και αρχιτεκτονικής που χρησιμοποιήθηκε για μη αποδεκτά μοντέλα

Το χαρακτηριστικό του ρυθμού μάθησης φαίνεται να επηρεάζει σοβαρά την ακρίβεια των μοντέλων που αναπτύσσονται. Στην εικόνα 20 φαίνεται πως οποιαδήποτε διαφορετική τιμή, εκτός από 0.0005 και 0.001, δεν είναι αρκετή για να προπονήσει ένα αξιόλογο δίκτυο αναγνώρισης πεζών εν ώρα νυκτός, κάνοντας έτσι την τιμή που θα πάρει ο ρυθμός μάθησης έναν καθοριστικό παράγοντα στην δημιουργία ενός αξιόπιστου ΣΝΔ για την εκτέλεση του έργου. Γίνεται επίσης και στο διάγραμμα 21 ότι οι ρυθμοί μάθησης 0.0005 και 0.001 είναι οι καλύτεροι μιας και ακόμα και στα χειρότερα μοντέλα μπορεί να παρατηρηθεί η αλλαγή στην ένδειξη

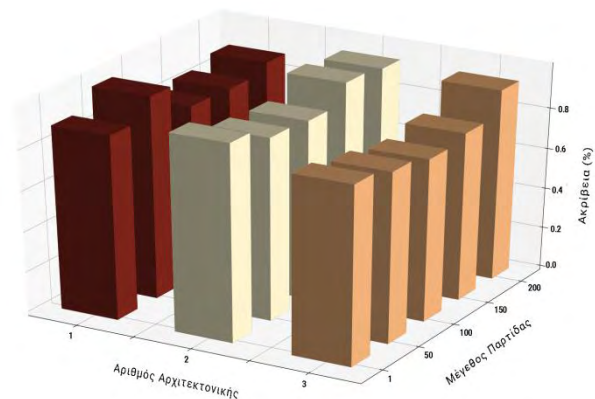
6.2 2° Σύνολο Δεδομένων

Προτού εξετάσουμε τα αποτελέσματα που δείχνουν τα διαγράμματα στις εικόνες 22 έως και 30, πρέπει να αναφερθεί ότι το υποσύνολο των πεζών εν ώρα νυκτός του δεύτερου ΣΔ, είναι εξαιρετικά μικρότερο σε μέγεθος από αυτό που υπάρχει στο πρώτο ΣΔ. Το μέγεθος του ΣΔ που θα χρησιμοποιηθεί είναι ύψιστης σημασίας. Σε αυτήν την περίπτωση καθώς θέλουμε το ΣΝΔ να αναγνωρίζει πεζούς, το υποσύνολο των πεζών την νύχτα οφείλει να είναι μεγαλύτερο ή ίσο με τα άλλα δύο υποσύνολα που βρίσκονται μέσα στο πρώτο ΣΔ.

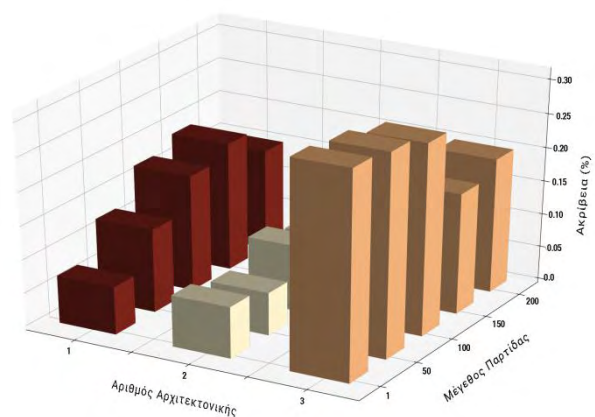
Δυστυχώς λόγω έλλειψης υλικού δεν βρέθηκαν παραπάνω από 247 εικόνες που να είναι ευδιάκριτοι οι πεζοί και να υπάρχει ο κατάλληλος φωτισμός. Επιπρόσθετα όπως αναφέρθηκε και στην υποενότητα [4.3](#), το ΣΔ που χρησιμοποιείται για τον έλεγχο του μοντέλου που αναπτύσσεται, απαρτίζεται αποκλειστικά από εικόνες του υποσυνόλου πεζών αυτού του ΣΔ. Εξαιτίας λοιπόν αυτής της έλλειψης, είναι εμφανές και στα διαγράμματα ότι τα αποτελέσματα δεν είναι καθόλου ελπιδοφόρα, αλλά ούτε και αρκετά αντιπροσωπευτικά των δυνατοτήτων των ΣΝΔ.

Εστιάζοντας στα γραφήματα που παρουσιάζονται στην αρχή του υποκεφαλαίου και δείχνουν αποτελέσματα για όλα τα μοντέλα (Εικόνες 23, 26 και 29), παρατηρούνται μικρά ποσοστά ακρίβειας στην πληθώρα των μοντέλων που αναπτύσσονται. Μολονότι υπάρχουν αρκετά μοντέλα για κάθε αρχιτεκτονική που θεωρούνται αποδεκτά και πετυχαίνουν κατά μέσο όρο ίδιες τιμές ακρίβειας, όπως φαίνεται και στο σχήμα στην εικόνα 22, ρίχνοντας μία ματιά στο διάγραμμα της εικόνας 23 που δείχνει τα αποτελέσματα για όλα τα μοντέλα που έχουν προπονηθεί, ο μέσος όρος των αποτελεσμάτων ακρίβειας δείχνει μία ραγδαία πτώση ακρίβειας. Με αυτό το δεδομένο γίνεται εύκολα κατανοητό ότι τα μοντέλα με αποδεκτή ακρίβεια είναι η μειονότητα των μοντέλων που προπονούνται με το δεύτερο ΣΔ. Συγκρίνοντας λοιπόν τα διαγράμματα στις εικόνες 23 και 24, παρατηρείται ότι όλες οι τιμές είναι παραπλήσιες αν όχι σχεδόν ίδιες

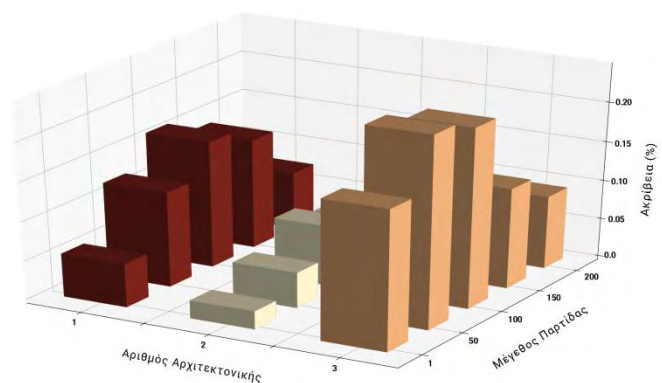
Επιπλέον όλες οι αρχιτεκτονικές φαίνεται να έχουν παρόμοιες συμπεριφορές με αυτές που είχαν και στην υποενότητα [6.1](#). Βλέποντας και συγκρίνοντας τα αντίστοιχα διαγράμματα (Εικόνες 22 έως και 30) αυτής της υποενότητας με την προηγούμενη, γίνεται αντιληπτό πως η μορφή των διαγραμμάτων μένει ίδια, αλλά οι τιμές ακρίβειας μειώνονται δραματικά. Μπορεί να μην φαίνεται με την πρώτη μάτια, αλλά αν μετατρέψουμε τα διαγράμματα σε ίδια κλίμακα τότε είναι πιο εύκολο. Τα μοντέλα έχουν τόσο χαμηλή ακρίβεια, που σε όλα τα γραφήματα που δείχνουν τα αποτελέσματα όλων των μοντέλων η μέγιστη τιμή που παρουσιάζει κάθε αρχιτεκτονική κυμαίνεται από 20% μέχρι περίπου 25%.



Εικόνα 22: Διάγραμμα μεγέθους παρτίδας και της αρχιτεκτονικής που χρησιμοποιήθηκε, για αποδεκτά μοντέλα

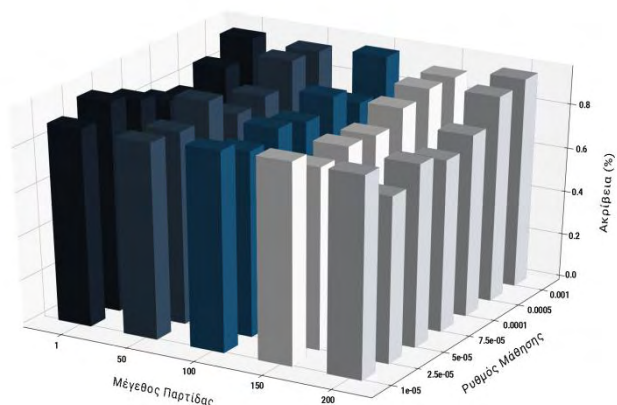


Εικόνα 23: Διάγραμμα μεγέθους παρτίδας και της αρχιτεκτονικής που χρησιμοποιήθηκε, για όλα τα μοντέλα

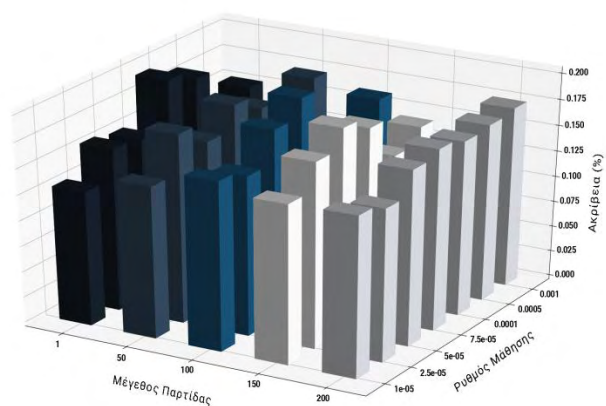


Εικόνα 24: Διάγραμμα μεγέθους παρτίδας και της αρχιτεκτονικής που χρησιμοποιήθηκε, για μη αποδεκτά μοντέλα

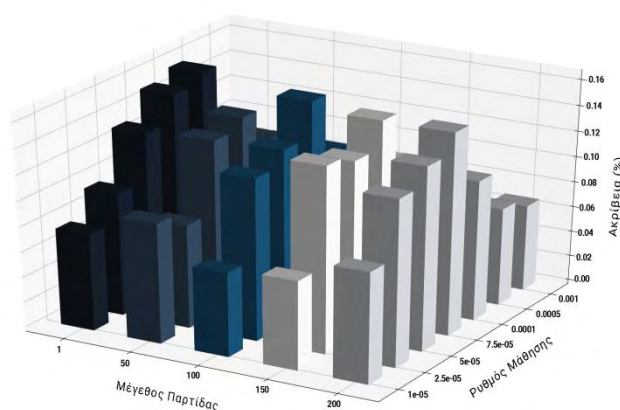
Ακριβώς το ίδιο συμβαίνει και για τα διαγράμματα στις εικόνες 25, 26 και 27. Όπως και στην υποενότητα [6.1](#), δεν παρατηρείται κάποιο συγκεκριμένο ζεύγος χαρακτηριστικών που να ξεχωρίζει από τα υπόλοιπα. Αλλά δεν βρίσκεται και κάποιο ζεύγος που να προπονει αποδεκτά μοντέλα, αντίθετα δηλαδή με τα δεδομένα που φαίνονται στο διάγραμμα της εικόνας 17, όπου κάθε ζεύγος χαρακτηριστικών αποδίδει πολύ καλά, μιας και το διάγραμμα της εικόνας 26 είναι σχεδόν ίδιο με το διάγραμμα της εικόνας 27. Αυτό σημαίνει ότι σχεδόν όλα τα μοντέλα που έχουν προπονηθεί, ανεξάρτητα της αρχιτεκτονικής που έχει χρησιμοποιηθεί, κατά μέσο όρο, επιδεικνύουν ποσοστά ακρίβειας μικρότερα του 20%.



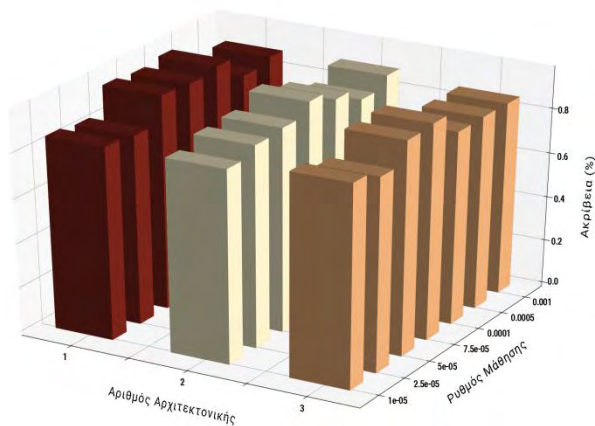
Εικόνα 25: Διάγραμμα ρυθμού μάθησης και μεγέθους παρτίδας, για αποδεκτά μοντέλα



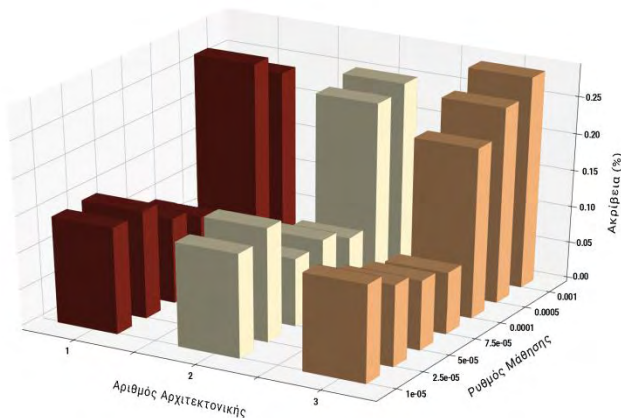
Εικόνα 26: Διάγραμμα ρυθμού μάθησης και μεγέθους παρτίδας, για όλα τα μοντέλα



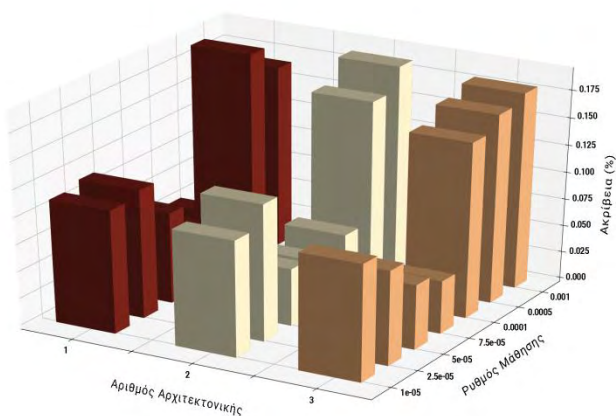
Εικόνα 27: Διάγραμμα ρυθμού μάθησης και μεγέθους παρτίδας, για μη αποδεκτά μοντέλα



Εικόνα 29: Διάγραμμα ρυθμού μάθησης και της αρχιτεκτονικής που χρησιμοποιήθηκε, για αποδεκτά μοντέλα



Εικόνα 30: Διάγραμμα ρυθμού μάθησης και της αρχιτεκτονικής που χρησιμοποιήθηκε, για όλα τα μοντέλα



Εικόνα 28: Διάγραμμα ρυθμού μάθησης και της αρχιτεκτονικής που χρησιμοποιήθηκε, για μη αποδεκτά μοντέλα

Εν τέλει στις εικόνες 28 έως και 30 αναδεικνύονται διαγράμματα που εξισώνουν τον ρυθμό μάθησης και τον αριθμό αρχιτεκτονικής. Στην εικόνα 30, όπως περιγράφεται και στην λεζάντα, μπορεί να φανεί η συμπεριφορά των αρχιτεκτονικών σε συνδυασμό με κάθε ρυθμό μάθησης. Όπως και στην πρώτη υποενότητα, οι μόνοι ρυθμοί μάθησης που ξεχωρίζουν είναι αυτοί που έχουν την τιμή 0.001 και 0.0005. Όπως και οι εικόνες 19, 20 και 21, οι εικόνες 28, 29 και 30 έχουν ακριβώς την ίδια μορφή με τις εικόνες 19, 20 και 21 με την μόνη διαφορά στα ποσοστά ακρίβειας. Η διαφορά των δύο ποσοστών που βλέπουμε στις εικόνες 29 και 20, ξεπερνάει το 50% μιας και το καλύτερο μοντέλο κάθε αρχιτεκτονικής ίσα ίσα φτάνει κοντά στο 25%. Όπως εξηγήθηκε και στα προηγούμενα 3 διαγράμματα, παρόλο που στην εικόνα 28 φαίνονται τα αποδεκτά μοντέλα, κοιτάζοντας την εικόνα 29, συμπεραίνεται ότι δεν υπάρχουν αρκετά μοντέλα που να τα καταφέρνουν να προπονηθούν κατάλληλα με το δεύτερο ΣΔ. Καταλήγοντας φαίνεται ότι το ΣΔ που παρουσιάζεται δεν τα καταφέρνει σε μεγάλο βαθμό να προπονήσει ένα αξιόπιστο δίκτυο, είτε για τους λόγους που αναφέρθηκαν στην αρχή της υποενότητας είτε διότι απλά δεν είναι συμβατό με την μελέτη.

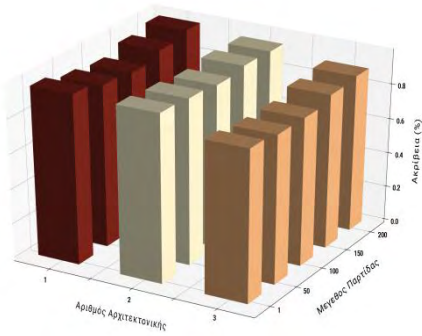
6.3 1° σύνολο δεδομένων με διασταυρωμένη επικύρωση (10-fold cross validation)

Σε αυτό το κομμάτι της ανάλυσης των πειραμάτων, θα παρουσιαστεί πως, με την χρήση του πρώτου ΣΔ και της τεχνικής του 10-fold cross validation¹ [49], διαφοροποιήθηκαν τα αποτελέσματα που βρέθηκαν στην πρώτη υποενότητα.

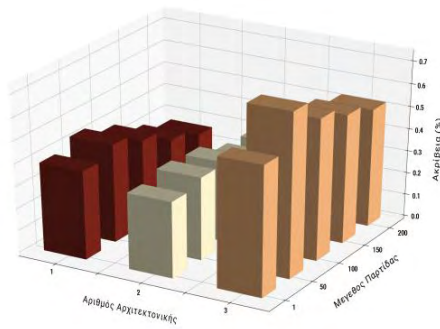
Πριν υπάρξει συνέχεια, είναι απαραίτητο να σημειωθεί ότι υπάρχει μία επιφύλαξη ως προς τα αποτελέσματα που θα προσφέρει αυτή η τεχνική επικύρωσης μοντέλου. Επειδή ο τρόπος που επιλέγονται τα στοιχεία που θα εισαχθούν στο σετ εικόνων προπόνησης και στο σετ εικόνων αξιολόγησης είναι τυχαίος. Επειδή όμως το ΣΔ περιέχει τριών τύπου εικόνες αυτό σημαίνει πως το σετ εικόνων αξιολόγησης δεν θα περιέχει αποκλειστικά εικόνες πεζών αλλά και άλλου τύπου. Έτσι το σετ υπάρχει περίπτωση να είναι αναξιόπιστο λόγω της έλλειψης εικόνων πεζών. Αυτή η έλλειψη δείχνει ότι το σετ αξιολόγησης δεν αντιπροσωπεύει πόσο καλά το ΣΝΔ αξιολογεί την αναγνώριση πεζών, αλλά το πόσο αποτελεσματικά αξιολογείται η αναγνώριση και των τριών κλάσεων που περιέχονται στο ΣΔ. Επιπρόσθετα από τον διαχωρισμό του αρχικού ΣΔ σε δύο υποσύνολα του, δεν θα είναι γνωστό αν ο αριθμός εικόνων που περιέχουν πεζούς, θα είναι αρκετός, έτσι ώστε να προπονηθεί κατάλληλα το ΣΝΔ. Έχοντας αναφέρει τις αμφιβολίες που υπάρχουν για την αξιοπιστία των τεστ που εκτελέστηκαν, σε αυτό το κομμάτι της έρευνας θα αναλυθούν τα αποτελέσματα αυτής της μεθόδου.

¹Το *k-cross fold validation* είναι η διασταυρούμενη επικύρωση (*cross validation*) είναι μία διαδικασία επαναδειγματοληψίας που χρησιμοποιείται για την αξιολόγηση μοντέλων μηχανικής μάθησης σε περιορισμένο δείγμα δεδομένων. Η μεταβλητή *k* αντιπροσωπεύει τον αριθμό ζευγαριών υποσυνόλων δεδομένων που θα κοπούν από το αρχικό ΣΔ. Ο ψευδοκώδικας του αλγορίθμου μπορεί να εμφανιστεί στον παρακάτω σύνδεσμο [49]

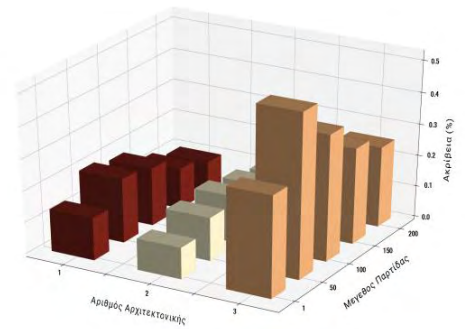
Συγκρίνοντας τα διαγράμματα της πρώτης υποενότητας [6.1](#) με τα διαγράμματα στις εικόνες 31 έως και 39, γίνεται ολοφάνερο ότι κάθε διάγραμμα που αναφέρεται σε αυτήν την τεχνική παρουσιάζει μικρότερες τιμές ακρίβειας. Ακόμα και τα διαγράμματα που αφορούν μόνο τα μοντέλα που δείχνουν να έχουν καλά ποσοστά ακρίβειας (Εικόνες 31, 34 και 37), δεν δείχνουν καλύτερα αποτελέσματα από τα διαγράμματα της πρώτης υποενότητας. Παρατηρώντας στο διάγραμμα στην εικόνα 35, τις μπάρες που δείχνουν τις ακρίβειες των μοντέλων που έχουν προπονηθεί με ένα συγκεκριμένο ζεύγος χαρακτηριστικών, υπάρχει μία δραματική πτώση της ακρίβειας σε κάθε μπάρα του διαγράμματος. Σε αυτήν την εικόνα αναλύονται όλα τα αποτελέσματα, ανεξαρτήτως της αρχιτεκτονικής που χρησιμοποιείται, δίνοντας μας έτσι μία γενική εικόνα ως προς την επίδραση της τεχνικής του 10-fold cross validation.



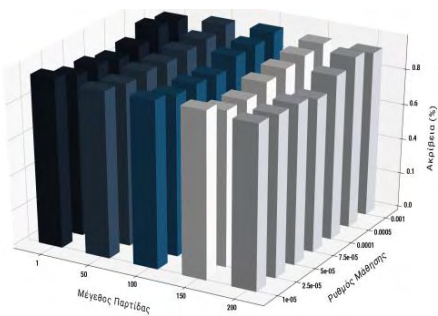
Εικόνα 33: Διάγραμμα μεγέθους παρτίδας και της αρχιτεκτονικής που χρησιμοποιήθηκε, για αποδεκτά μοντέλα



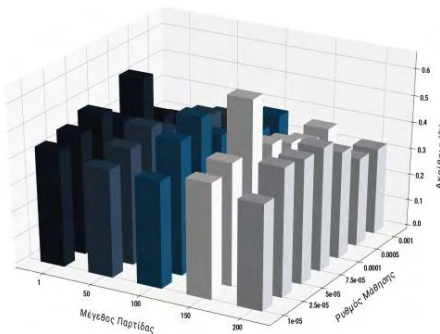
Εικόνα 31: Διάγραμμα μεγέθους παρτίδας και της αρχιτεκτονικής που χρησιμοποιήθηκε, για όλα τα μοντέλα



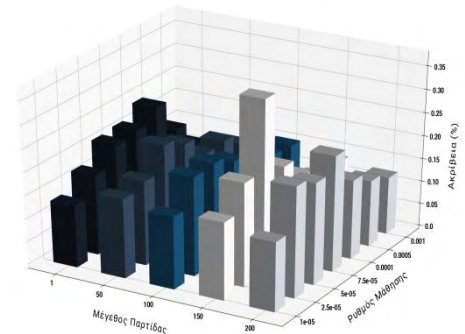
Εικόνα 32: Διάγραμμα μεγέθους παρτίδας και της αρχιτεκτονικής που χρησιμοποιήθηκε, για μη αποδεκτά μοντέλα



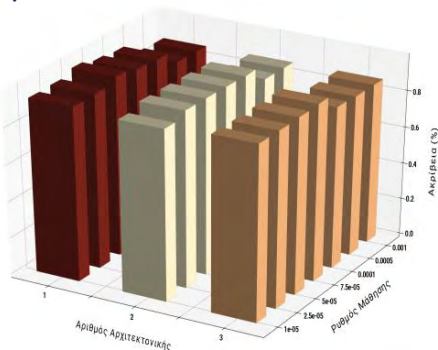
Εικόνα 36: Διάγραμμα ρυθμού μάθησης και μεγέθους παρτίδας, για αποδεκτά μοντέλα



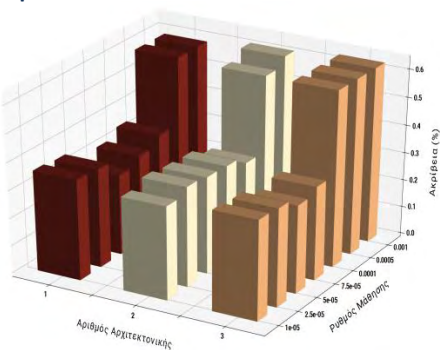
Εικόνα 35: Διάγραμμα ρυθμού μάθησης και μεγέθους παρτίδας, για όλα τα μοντέλα



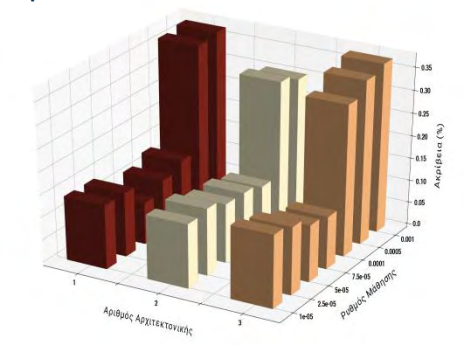
Εικόνα 34: Διάγραμμα ρυθμού μάθησης και μεγέθους παρτίδας, για μη αποδεκτά μοντέλα



Εικόνα 37: Διάγραμμα ρυθμού μάθησης και της αρχιτεκτονικής που χρησιμοποιήθηκε, για όλα τα μοντέλα



Εικόνα 38: Διάγραμμα ρυθμού μάθησης και της αρχιτεκτονικής που χρησιμοποιήθηκε, για όλα τα μοντέλα



Εικόνα 39: Διάγραμμα ρυθμού μάθησης και της αρχιτεκτονικής που χρησιμοποιήθηκε, για μη αποδεκτά μοντέλα

Συγκρίνοντας τα διαγράμματα της πρώτης υποενότητας [6.1](#) με τα διαγράμματα στις εικόνες 31 έως και 39, γίνεται ολοφάνερο ότι κάθε διάγραμμα που αναφέρεται σε αυτήν την τεχνική παρουσιάζει μικρότερες τιμές ακρίβειας. Ακόμα και τα διαγράμματα που αφορούν μόνο τα μοντέλα που δείχνουν να έχουν καλά ποσοστά ακρίβειας (Εικόνες 31, 34 και 37), δεν δείχνουν καλύτερα αποτελέσματα από τα διαγράμματα της πρώτης υποενότητας. Παρατηρώντας στο διάγραμμα στην εικόνα 35, τις μπάρες που δείχνουν τις ακρίβειες των μοντέλων που έχουν προπονηθεί με ένα συγκεκριμένο ζεύγος χαρακτηριστικών, υπάρχει μία δραματική πτώση της ακρίβειας σε κάθε μπάρα του διαγράμματος. Σε αυτήν την εικόνα

αναλύονται όλα τα αποτελέσματα, ανεξαρτήτως της αρχιτεκτονικής που χρησιμοποιείται, δίνοντας μας έτσι μία γενική εικόνα ως προς την επίδραση της τεχνικής του 10-fold cross validation.

Πιο αναλυτικά, είναι φανερό από τις εικόνες 14, 20, 32 και 38 ότι οι πρώτες δύο αρχιτεκτονικές δεν αντιδρούν θετικά με την είσοδο αυτής της τεχνικής στον τρόπο προπόνησης και αξιολόγησης του δικτύου. Για την ακρίβεια παρατηρείται μία μείωση περίπου 20% στην ακρίβεια των μοντέλων που προπονούσαν, αλλά στην εικόνα 38 φαίνεται ότι υπάρχει μια μείωση μόνο κατά 10% αντί για 20% όπως αναφέραμε προηγουμένως, για μοντέλα που έχουν προπονηθεί με ρυθμό μάθησης ίσο με 0.0005 ή 0.001.

7. ΣΥΜΠΕΡΑΣΜΑΤΑ

Εν κατακλείδι, σε αυτή την έρευνα αναπτύχθηκαν δύο ειδικά σύνολα δεδομένων τα οποία χρησιμοποιήθηκαν για την προπόνηση τριών διαφορετικών αρχιτεκτονικών ΣΝΔ. Το κάθε ΣΔ περιέχει τρεις κλάσεις εικόνων, των οποίων η μία αντιπροσωπεύει την κλάση πεζών (εν ώρα ημέρας ή νυκτός). Η κάθε δομή δικτύου προπονείται με σκοπό να παρατηρηθεί η ικανότητα στην αναγνώριση πεζών τη νύχτα.

Με σκοπό να γίνει μία εξονυχιστική έρευνα της συμβατότητας των αρχιτεκτονικών εκτελέστηκε μία σειρά αξιολογήσεων. Κάθε αξιολόγηση χρησιμοποιούσε και διαφορετικό συνδυασμό χαρακτηριστικών, διεξάγοντας με αυτόν τον τρόπο μια μορφή fine tuning σε κάθε αρχιτεκτονική ΣΝΔ. Με τη μέθοδο αυτή, στα διαγράμματα που παρουσιάζονται στην ενότητα [6.1](#) είναι ευδιάκριτο για ποιόν συνδυασμό χαρακτηριστικών η κάθε αρχιτεκτονική, δρα με το βέλτιστο δυνατό τρόπο.

Το 2^ο ΣΔ δυστυχώς λόγω της έλλειψης υλικού όπως προαναφέρθηκε δεν απέδωσε καθόλου καλά. Για αυτόν τον λόγο οι σειρές αξιολογήσεων που χρησιμοποίησαν το πρώτο ΣΔ [6.1](#) και [6.3](#), είναι αυτές οι οποίες θεωρήθηκαν πιο αξιόπιστες. Παρατηρώντας τα ευρήματα της έρευνας, το μικρό ΣΝΔ που βρέθηκε στην έρευνα [\[21\]](#) έδειξε να έχει τα καλύτερα αποτελέσματα ακρίβειας σε μέσο όρο από τα άλλα δύο που δοκιμάστηκαν. Εστιάζοντας στα γραφήματα 14, 20, 32 και 38 της πρώτης και τρίτης έρευνας, διακρίνεται η μικρή ανωτερότητα των τιμών της τρίτης αρχιτεκτονικής από των υπόλοιπων δύο. Στις εικόνες 14 και 32, τα διαγράμματα που εξισώνουν, την αρχιτεκτονική που χρησιμοποιείται και το χαρακτηριστικό του μεγέθους παρτίδας, η τρίτη αρχιτεκτονική παρουσιάζει αρκετά καλύτερα δεδομένα από τις υπόλοιπες αρχιτεκτονικές, για κάθε τιμή που μπορεί να πάρει η μεταβλητή του μεγέθους παρτίδας. Επιπρόσθετα στις εικόνες 20 και 38, παρόλο που τα διαγράμματα μοιάζουν σε μεγάλο βαθμό στη μορφή τους, παρατηρείται ότι όλες οι αρχιτεκτονικές δείχνουν την ίδια συμπεριφορά για κάθε ρυθμό μάθησης.

Γενικότερα τα καλύτερα αποτελέσματα φάνηκαν από τα ΣΝΔ που προπονήθηκαν με τη βοήθεια του πρώτου ΣΔ στην πρώτη σειρά τεστ και κυριότερα το τρίτο ΣΝΔ που ήταν και το λιγότερο περίπλοκο δίκτυο. Ωστόσο πιστεύουμε ότι αν επεκταθεί κατάλληλα το υποσύνολο των πεζών στο δεύτερο ΣΔ τότε τα αποτελέσματα των ίδιων ερευνών που εκτελέστηκαν στην υποενότητα [6.2](#) θα διαφοροποιηθούν ραγδαία προς το καλύτερο.

8. ΜΕΛΛΟΝΤΙΚΗ ΕΡΓΑΣΙΑ

Ο καθοριστικός παράγοντας που απέτρεψε αυτήν την έρευνα από το να ενσωματώσει περισσότερες και πιο πολύπλοκες αρχιτεκτονικές, ήταν η έλλειψη πιο προηγμένου εξοπλισμού. Πιο συγκεκριμένα η χρήση της κάρτας γραφικών NVIDIA MSI GeForce GTX 750 Ti 2GB (N750Ti-2GD5TLP), η οποία κατέχει μόνο 2GB μνήμης, δεν επέτρεπε την ύπαρξη πολλαπλών πλήρως συνδεδεμένων δικτύων, καθώς εμφανιζόταν συνεχώς το σφάλμα εξάντλησης πόρων. Ως αποτέλεσμα με έναν καλύτερο εξοπλισμό θα μπορούσε να υπάρξει μια καλύτερη ανάλυση των ίδιων ΣΝΔ αλλά και άλλων αρχιτεκτονικών.

Ο ίδιος περιορισμός δεν επιτρέπει την εισαγωγή εικόνων μεγαλύτερων διαστάσεων και καλύτερης ανάλυσης, κάνοντας με αυτόν τον τρόπο πιο εύκολο το έργο της αναγνώρισης πεζών. Ιδανικά θα μπορούσαν να χρησιμοποιηθούν εικόνες τυπικών διαστάσεων 640X480, για καλύτερη ανάλυση εικόνας. Αυτές οι διαστάσεις τυχαίνει να είναι και η ελάχιστη ανάλυση που έχει κάθε κάμερα αυτοκινήτου που τοποθετείται στο μπροστινό τζάμι του αμαξιού (ή αλλιώς dashcam).

Ένας άλλος δρόμος που θα μπορούσε να ακολουθηθεί για την βελτιστοποίηση της ταχύτητας των δικτύων θα ήταν η χρήση half-precision μεταβλητών και πινάκων (όπως οι Tensors), μιας και θα «ελάφρυνε» αρκετά το κάθε δίκτυο από την διαδικασία των υπολογισμών, λιγοστεύοντας την πολυπλοκότητα των πράξεων. Ένας άλλος τρόπος να υπήρχε το ίδιο αποτέλεσμα θα ήταν η αλλαγή της κωδικοποίησης των εικόνων εισόδου από RGB σε YCbCR [\[50\]](#). Θα ήταν καλύτερο βέβαια να μην γίνεται η μετατροπή της εικόνας στο υπολογιστικό σύστημα, αλλά να γίνεται άμεσα από τον φακό της κάμερας.

BIBΛΙΟΓΡΑΦΙΑ ΚΑΙ ΑΝΑΦΟΡΕΣ

- [1] S. Chang, F. Yang, W. Wu, Y. Cho and S. Chen, "Nighttime pedestrian detection using thermal imaging based on HOG feature," Proceedings 2011 International Conference on System Science and Engineering, Macao, 2011, pp. 694-698.
- [2] J. Ge, Y. Luo and G. Tei, "Real-Time Pedestrian Detection and Tracking at Nighttime for Driver-Assistance Systems," in IEEE Transactions on Intelligent Transportation Systems, vol. 10, no. 2, pp. 283-298, June 2009.
- [3] I. Parra Alonso *et al.*, "Combination of Feature Extraction Methods for SVM Pedestrian Detection," in IEEE Transactions on Intelligent Transportation Systems, vol. 8, no. 2, pp. 292-307, June 2007.
- [4] J. Dong, J. Ge and Y. Luo, "Nighttime Pedestrian Detection with Near Infrared using Cascaded Classifiers," 2007 IEEE International Conference on Image Processing, San Antonio, TX, 2007, pp. VI - 185-VI - 188.
- [5] D. Geronimo, A. M. Lopez, A. D. Sappa and T. Graf, "Survey of Pedestrian Detection for Advanced Driver Assistance Systems," in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 32, no. 7, pp. 1239-1258, July 2010.
- [6] V. Sze, Y. Chen, T. Yang and J. S. Emer, "Efficient Processing of Deep Neural Networks: A Tutorial and Survey," in Proceedings of the IEEE, vol. 105, no. 12, pp. 2295-2329, Dec. 2017.
- [7] Histogram Oriented Gradients(HOG):
<https://www.learnopencv.com/histogram-of-oriented-gradients>

- [8] Srinivas Suraj, Sarvadevabhatla Ravi Kiran, Mopuri Konda Reddy, Prabhu Nikita, Kruthiventi Srinivas S. S., Babu R. Venkatesh
A Taxonomy of Deep Convolutional Neural Nets for Computer Vision
- [9] Adaptive Boosting:
<https://en.wikipedia.org/wiki/AdaBoost>
- [10] S. Shi, Q. Wang, P. Xu and X. Chu, "Benchmarking State-of-the-Art Deep Learning Software Tools," *2016 7th International Conference on Cloud Computing and Big Data (CCBD)*, Macau, 2016, pp. 99-104.
- [11] Gu, Jiuxiang, Zhenhua Wang, Jason Kuen, Lianyang Ma, Amir Shahroudy, Bing Shuai, Ting Liu, Xingxing Wang, Gang Wang, Jianfei Cai and Tsuhan Chen. "Recent advances in convolutional neural networks." *Pattern Recognition* 77 (2018): 354-377.
- [12] Benenson R., Omran M., Hosang J., Schiele B. (2015) Ten Years of Pedestrian Detection, What Have We Learned?. In: Agapito L., Bronstein M., Rother C. (eds) *Computer Vision - ECCV 2014 Workshops*. ECCV 2014. Lecture Notes in Computer Science, vol 8926. Springer, Cham
- [13] 6 Types of Artificial Neural Networks Currently Being Used in Machine Learning
<https://analyticsindiamag.com/6-types-of-artificial-neural-networks-currently-being-used-in-todays-technology/>
- [14] Νευρωνικό δίκτυο Wikipedia:
https://el.wikipedia.org/wiki/%CE%9D%CE%B5%CF%85%CF%81%CF%89%CE%BD%CE%B9%CE%BA%CF%8C_%CE%B4%CE%AF%CE%BA%CF%84%CF%85%CE%BF

- [15] An intuitive guide to Convolutional Neural Networks: <https://medium.freecodecamp.org/an-intuitive-guide-to-convolutional-neural-networks-260c2de0a050>
- [16] what is back propagation: <https://en.wikipedia.org/wiki/Backpropagation>
- [17] Kim JH, Hong HG, Park KR. Convolutional Neural Network-Based Human Detection in Nighttime Images Using Visible Light Camera Sensors. Passaro VMN, ed. Sensors (Basel, Switzerland).
- [18] Jiang, Xiaoheng & Pang, Yanwei & Li, Xuelong & Pan, Jing. (2015). Speed up deep neural network based pedestrian detection by sharing features across multi-scale models. Neurocomputing.
- [19] Y. Tian, P. Luo, X. Wang and X. Tang, "Deep Learning Strong Parts for Pedestrian Detection," *2015 IEEE International Conference on Computer Vision (ICCV)*, Santiago, 2015, pp. 1904-1912.
- [20] H. Fukui, T. Yamashita, Y. Yamauchi, H. Fujiyoshi and H. Murase, "Pedestrian detection based on deep convolutional neural network with ensemble inference network," *2015 IEEE Intelligent Vehicles Symposium (IV)*, Seoul, 2015, pp. 223-228.
- [21] Angelova, Anelia & Krizhevsky, Alex & Vanhoucke, Vincent & Ogale, Abhijit & Ferguson, Dave. (2015). Real-Time Pedestrian Detection with Deep Network Cascades. 32.1-32.12. 10.5244/C.29.32.
- [22] S. Zha, F. Luisier, W. Andrews, N. Srivastava, and R. Salakhutdinov, "[Exploiting Image-trained CNN Architectures for Unconstrained Video Classification](#)," 26th British Machine Vision Conference (BMVC'15). pp. 60.1-60.13, 2015.

- [23] Sze, Vivienne, Yu-Hsin Chen, Tien-Ju Yang and Joel S. Emer. "Efficient Processing of Deep Neural Networks: A Tutorial and Survey." *Proceedings of the IEEE 105 (2017): 2295-2329.*
- [24] V. John, S. Mita, Z. Liu and B. Qi, "Pedestrian detection in thermal images using adaptive fuzzy C-means clustering and convolutional neural networks," *2015 14th IAPR International Conference on Machine Vision Applications (MVA)*, Tokyo, 2015, pp. 246-249.
- [25] Histogram Equalization:
https://en.wikipedia.org/wiki/Histogram_equalization
- [26] Piotr Dollar, Zhuowen Tu, Pietro Perona and Serge Belongie. Integral Channel Features. In A. Cavallaro, S. Prince and D. Alexander, editors, *Proceedings of the British Machine Conference*, pages 91.1-91.11. BMVA Press, September 2009. doi:10.5244/C.23.91.
- [27] Cudaconvnet library:
<https://code.google.com/p/cuda-convnet/> A. Krizhevsky, Cudaconvnet.
- [28] Support Vector Machine:
<https://medium.com/machine-learning-101/chapter-2-svm-support-vector-machine-theory-f0812effc72>
- [29] Cascading classifiers:
https://en.wikipedia.org/wiki/Cascading_classifiers
- [30] R. Benenson, M. Matthias, R. Tomofte, and L. Van Gool. Pedestrian detection at 100 frames per second. CVPR, 2012.

- [31] 31) R. Benenson, M. Omran, J. Hosang, and B. Schiele. Ten years of pedestrian detection, what have we learned? 2nd Workshop on Road scene understanding and Autonomous driving, ECCV, 2014.
- [32] A. Krizhevsky, I. Sutskever, and G. Hinton. Imagenet classification with deep convolutional neural networks. NIPS, 2012.
- [33] Softmax function:
https://en.wikipedia.org/wiki/Softmax_function
- [34] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, San Diego, CA, USA, 2005, pp. 886-893 vol. 1. doi: 10.1109/CVPR.2005.177
- [35] Region Of Interest (ROI):
https://en.wikipedia.org/wiki/Region_of_interest
- [36] Fuzzy Inference System:
<http://www.cs.princeton.edu/courses/archive/fall07/cos436/HIDDEN/Knapp/fuzzy004.htm>
- [37] Histogram Specification:
http://fourier.eng.hmc.edu/e161/lectures/contrast_transform/node3.html
- [38] Yuxi Chen and Chongzhao Han, "Night-time pedestrian detection by visual-infrared video fusion," 2008 7th World Congress on Intelligent Control and Automation, Chongqing, 2008, pp. 5079-5084.

- [39] Li J, Zhang F, Wei L, Yang T, Lu Z. Nighttime Foreground Pedestrian Detection Based on Three-Dimensional Voxel Surface Model. *Sensors* (Basel, Switzerland). 2017;17(10):2354. doi:10.3390/s17102354.
- [40] TensorFlow Estimators: Managing Simplicity vs. Flexibility in High-Level Machine Learning Frameworks, Cassandra Xia and Clemens Mewald and D. Sculley and David Soergel and George Roumpos and Heng-Tze Cheng and Illia Polosukhin and Jamie Alexander Smith and Jianwei Xie and Lichan Hong and Martin Wicke and Mustafa Ispir and Philip Daniel Tucker and Yuan Tang and Zakaria Haque, 2017
- [41] Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., Devin, M., Ghemawat, S., Irving, G., Isard, M. and Kudlur, M., 2016, November. Tensorflow: a system for large-scale machine learning. In *OSDI* (Vol. 16, pp. 265-283).
- [42] M. Hassaballah, Aly Amin Abdelmegeid and Hammam A. Alshazly
“ Image Features Detection, Description and Matching”
- [43] “Pedestrian Detection with Deep Convolutional Neural Network” Xiaogang Chen, Pengxu Wei, Wei Ke, Qixiang Ye, Jianbin Jiao School of Electronic,Electrical and Communication Engineering, University of Chinese Academy of Science, Beijing, China
- [44] Transfer learning:
<https://data-flair.training/blogs/transfer-learning/>
- [45] Back propagation:
<https://medium.com/datathings/neural-networks-and-backpropagation-explained-in-a-simple-way-f540a3611f5e>

- [46] Adam optimizer:
<https://machinelearningmastery.com/adam-optimization-algorithm-for-deep-learning/>
- [47] Overfitting:
<https://en.wikipedia.org/wiki/Overfitting>
- [48] K-fold cross validation:
<https://machinelearningmastery.com/k-fold-cross-validation/>
- [49] YCbCR coding:
<https://www.roman10.net/2011/08/18/ycbcr-color-spacean-intro-and-its-applications/>
- [50] Caltech Pedestrian Dataset:
http://www.vision.caltech.edu/Image_Datasets/CaltechPedestrians/
- [51] Image descriptor:
https://en.wikipedia.org/wiki/Visual_descriptor
- [52] Overfitting:
<https://en.wikipedia.org/wiki/Overfitting>

ΑΝΑΦΟΡΑ ΕΙΚΟΝΩΝ

Εικόνα 1:

https://www.tutorialspoint.com/artificial_intelligence/images/atypical_ann.jpg

Εικόνα 2:

<http://180.179.138.23/images/news/1442/cover.png>

Εικόνα 3:

<https://msatechnosoft.in/blog/wp-content/uploads/2018/05/feed-forward-neural-network-MSA-Technosoft.png>

Εικόνα 4:

https://chrisjmccormick.files.wordpress.com/2013/08/architecture_simple2.png

Εικόνα 5:

https://www.researchgate.net/profile/Ramon_Quiza/publication/234055140/figure/fig3/AS:299964002521099@1448528399394/Graph-of-a-recurrent-neural-network.png

Εικόνα 6:

https://www.mathworks.com/matlabcentral/answers/uploaded_files/17966/Sin%20t%C3%ADtulo.png

Εικόνα 7:

https://res.mdpi.com/information/information-07-00061/article_deploy/html/images/information-07-00061-g001.png

Εικόνα 8:

<https://helpx.adobe.com/content/dam/help/en/photoshop/ps-key-concepts/grayscale.jpg>