



ΠΑΝΕΠΙΣΤΗΜΙΟ ΔΥΤΙΚΗΣ ΑΤΤΙΚΗΣ

ΣΧΟΛΗ ΜΗΧΑΝΙΚΩΝ

ΤΜΗΜΑ ΗΛΕΚΤΡΟΛΟΓΩΝ & ΗΛΕΚΤΡΟΝΙΚΩΝ ΜΗΧΑΝΙΚΩΝ

**Πρόγραμμα Μεταπτυχιακών Σπουδών
«ΔΙΑΔΙΚΤΥΟ ΤΩΝ ΠΡΑΓΜΑΤΩΝ ΚΑΙ ΕΥΦΥΗ ΠΕΡΙΒΑΛΛΟΝΤΑ»**

ΜΕΤΑΠΤΥΧΙΑΚΗ ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

**“Αυτόματο Σύστημα Αναγνώρισης Τροφίμων με Χρήση Βαθιάς Μάθησης.
Περίπτωση Χρήσης: Αναγνώριση Περιεκτικότητας Συσκευασιών
Γάλακτος”**



Μεταπτυχιακός Φοιτητής: Όλσος Πάντος, ΑΜ msciot18005

Επιβλέπων: Χαράλαμπος Πατρικάκης, Καθηγητής

ΑΙΓΑΛΕΩ, ΣΕΠΤΕΜΒΡΙΟΣ 2022



UNIVERSITY OF WEST ATTICA

FACULTY OF ENGINEERING

DEPARTMENT OF ELECTRICAL & ELECTRONICS ENGINEERING

Master of Science in “INTERNET of THINGS AND INTELLIGENT ENVIRONMENTS”

MSc Thesis

«Automatic Food Detection System Using Deep Learning. Use Case: Detection of content levels in Milk Containers»



Student: Pantos, Olsis, Registration Number msciot18005

MSc Thesis Supervisor: Patrikakis, Charalampos, Professor

ATHENS-EGALEO, SEPTEMBER 2022

Η Διπλωματική Εργασία έγινε αποδεκτή και βαθμολογήθηκε από την εξής τριμελή επιτροπή:

Χαράλαμπος Πατρικάκης Καθηγητής (Επιβλέπων)	Γρηγόριος Καλτσάς Καθηγητής (Μέλος)	Αλέξανδρος Αλεξανδρίδης Καθηγητής (Μέλος)

Copyright © Με επιφύλαξη παντός δικαιώματος. All rights reserved.

ΠΑΝΕΠΙΣΤΗΜΙΟ ΔΥΤΙΚΗΣ ΑΤΤΙΚΗΣ και [Όλσος Πάντος],

[Σεπτέμβριος, 2022]

Απαγορεύεται η αντιγραφή, αποθήκευση και διανομή της παρούσας εργασίας, εξ ολοκλήρου ή τμήματος αυτής, για εμπορικό σκοπό. Επιτρέπεται η ανατύπωση, αποθήκευση και διανομή για σκοπό μη κερδοσκοπικό, εκπαιδευτικής ή ερευνητικής φύσης, υπό την προϋπόθεση να αναφέρεται η πηγή προέλευσης και να διατηρείται το παρόν μήνυμα. Ερωτήματα που αφορούν τη χρήση της εργασίας για κερδοσκοπικό σκοπό πρέπει να απευθύνονται προς τους συγγραφείς.

Οι απόψεις και τα συμπεράσματα που περιέχονται σε αυτό το έγγραφο εκφράζουν τον/την συγγραφέα του και δεν πρέπει να ερμηνευθεί ότι αντιπροσωπεύουν τις θέσεις του επιβλέποντος, της επιτροπής εξέτασης ή τις επίσημες θέσεις του Τμήματος και του Ιδρύματος.

ΔΗΛΩΣΗ ΣΥΓΓΡΑΦΕΑ ΔΙΠΛΩΜΑΤΙΚΗΣ ΔΙΑΤΡΙΒΗΣ

Ο κάτωθι υπογεγραμμένος Όλσος Πάντος του Πέτρου, με αριθμό μητρώου msciot18005, φοιτητής του Πανεπιστημίου Δυτικής Αττικής της Σχολής ΜΗΧΑΝΙΚΩΝ του Τμήματος ΗΛΕΚΤΡΟΛΟΓΩΝ ΚΑΙ ΗΛΕΚΤΡΟΝΙΚΩΝ ΜΗΧΑΝΙΚΩΝ,

δηλώνω υπεύθυνα ότι:

«Είμαι συγγραφέας αυτής της διπλωματικής εργασίας και ότι κάθε βοήθεια την οποία είχα για την προετοιμασία της είναι πλήρως αναγνωρισμένη και αναφέρεται στην εργασία. Επίσης, οι όποιες πηγές από τις οποίες έκανα χρήση δεδομένων, ιδεών ή λέξεων, είτε ακριβώς είτε παραφρασμένες, αναφέρονται στο σύνολό τους, με πλήρη αναφορά στους συγγραφείς, τον εκδοτικό οίκο ή το περιοδικό, συμπεριλαμβανομένων και των πηγών που ενδεχομένως χρησιμοποιήθηκαν από το διαδίκτυο. Επίσης, βεβαιώνω ότι αυτή η εργασία έχει συγγραφεί από μένα αποκλειστικά και αποτελεί προϊόν πνευματικής ιδιοκτησίας τόσο δικής μου, όσο και του Ιδρύματος.

Παράβαση της ανωτέρω ακαδημαϊκής μου ευθύνης αποτελεί ουσιώδη λόγο για την ανάκληση του διπλώματός μου».

Ο Δηλών



Όλσος Πάντος

ΠΕΡΙΛΗΨΗ

Τα συνελκτικά νευρωνικά δίκτυα και γενικότερα η βαθιά μάθηση έχουν οδηγήσει στην ανάπτυξη μοντέλων αναγνώρισης και ταξινόμησης αντικειμένων εφάμιλλες σε επιδόσεις με τις ανθρώπινες ικανότητες. Αυτές οι μέθοδοι όμως έχουν και μειονεκτήματα. Συγκεκριμένα, με την αύξηση του βάθους της αρχιτεκτονικής στα νευρωνικά δίκτυα αυξάνεται και η πολυπλοκότητα των υπολογισμών τα οποία χρειάζονται να εκτελεστούν. Επιπλέον, στα συνελκτικά νευρωνικά δίκτυα, με την αύξηση του βάθους της αρχιτεκτονικής δημιουργείται κορεσμός από ένα σημείο και μετά λόγω του φαινομένου του προβλήματος των εξαφανιζόμενων κλίσεων (Vanishing Gradient Problem).

Σκοπός της διπλωματικής εργασίας είναι η μελέτη και ανάπτυξη συστήματος αναγνώρισης τροφίμων το οποίο έχει τη δυνατότητα να αναγνωρίσει το γάλα σε διαυγείς φιάλες καθώς και την ποσότητα αυτού που περιέχουν.

Για την υλοποίηση της, αρχικά μελετήθηκε τι υπάρχει στην αγορά και στη διεθνή βιβλιογραφία για παρόμοια συστήματα και πως αυτά λειτουργούν. Στη συνέχεια, συγκεντρώθηκαν εικόνες μπουκαλιών με γάλα και ταξινομήθηκαν με βάση το περιεχόμενό τους. Τέλος, με τη βοήθεια της μεταφοράς μάθησης, που αυξάνει την αποδοτικότητα του εκπαιδευμένου νευρωνικού δικτύου, το νευρωνικό δίκτυο αρχιτεκτονικής MobileNet εκπαιδεύτηκε με δύο τρόπους. Με βάση τον πρώτο τρόπο, έγινε χρήση της πλατφόρμας Teachable Machine που περιείχε το εκπαιδευμένο νευρωνικό δίκτυο MobileNet το οποίο είχε προηγουμένως εκπαιδευτεί με την βάση δεδομένων ImageNet. Κατά τον δεύτερο τρόπο, δημιουργήθηκε πρόγραμμα το οποίο χρησιμοποιεί την αρχιτεκτονική MobileNet του οποίου έχει γίνει μείωση κατά 50% στο πλάτος κάθε στοιβάδας με νευρώνες.

Τα αποτελέσματα δείχνουν ότι τα δύο εκπαιδευμένα νευρωνικά δίκτυα πραγματοποίησαν επιτυχώς ταξινόμηση σε ποσοστό άνω του 95%. Τέλος, τα εκπαιδευμένα νευρωνικά δίκτυα μπορούν να ενσωματωθούν σε εφαρμογές ή να λειτουργήσουν σε προγράμματα πλοήγησης στο διαδίκτυο, όπως παρουσιάζεται στην εργασία.

ΛΕΞΕΙΣ – ΚΛΕΙΔΙΑ: Συνελκτικό Νευρωνικό Δίκτυο, Αναγνώριση Τροφίμων, Python, MobileNet, Ταξινόμηση Εικόνων.

ABSTRACT

Convolutional Neural Networks (CNN) and deep learning in general have led to the development of models for identifying and classifying objects comparable in performance to human capabilities. However, these methods also have disadvantages. As the depth of architecture in neural networks increases, so does the complexity of the calculations that need to be performed. In addition, in Convolutional Neural Networks, as the depth of the architecture increases, saturation is created from one point onwards due to the Vanishing Gradient Problem.

The purpose of the dissertation is the study and development of a food identification system that can identify milk in clear bottles and what quantities they contain.

For its implementation, it was first studied what is on the market and in the international literature for similar systems and how they work. Images containing bottles with milk were then collected and sorted according to their contents. Finally, with the help of transfer learning, which increases the efficiency of the trained neural network, the MobileNet architecture neural network was trained in two ways. In the first way, the Teachable Machine platform was used which contained the trained MobileNet neural network previously trained with the ImageNet database. In the second way, a program was created that uses the MobileNet architecture which has reduced by 50% the width of each layer of neurons.

The results show that the two trained neural networks successfully classify more than 95%. Finally, trained neural networks can be integrated into applications or operated in browsers as presented in the thesis.

KEYWORDS: Convolutional Neural Networks (CNN), Food Recognition, Python, MobileNet, Image Classification

ΕΥΧΑΡΙΣΤΙΕΣ

Στην μητέρα μου Βαλεντίνα,
στην μνήμη του πατέρα μου Πέτρο,
σε όσους με βοήθησαν.

ΠΙΝΑΚΑΣ ΠΕΡΙΕΧΟΜΕΝΩΝ

ΔΗΛΩΣΗ ΣΥΓΓΡΑΦΕΑ ΔΙΠΛΩΜΑΤΙΚΗΣ ΔΙΑΤΡΙΒΗΣ	5
ΠΕΡΙΛΗΨΗ	6
ABSTRACT	7
ΕΥΧΑΡΙΣΤΙΕΣ	8
ΠΙΝΑΚΑΣ ΠΕΡΙΕΧΟΜΕΝΩΝ	9
ΚΑΤΑΛΟΓΟΣ ΠΙΝΑΚΩΝ:	12
ΚΑΤΑΛΟΓΟΣ ΕΙΚΟΝΩΝ:	13
ΕΙΣΑΓΩΓΗ.....	15
ΠΑΡΑΔΟΣΙΑΚΕΣ ΜΕΘΟΔΟΙ ΠΡΟΓΡΑΜΜΑΤΙΣΜΟΥ ΔΙΑΤΡΟΦΗΣ.....	15
ΣΥΣΤΗΜΑΤΑ ΑΝΑΓΝΩΡΙΣΗΣ ΤΡΟΦΙΜΩΝ ΜΕ ΤΗ ΒΟΗΘΕΙΑ ΒΑΘΙΑΣ ΜΑΘΗΣΗΣ ΚΑΙ ΤΑ ΠΛΕΟΝΕΚΤΗΜΑΤΑ ΤΟΥΣ	15
ΘΕΜΑ ΤΗΣ ΕΡΓΑΣΙΑΣ.....	17
ΜΕΘΟΔΟΛΟΓΙΑ	17
ΔΟΜΗ ΕΡΓΑΣΙΑΣ	18
ΚΕΦΑΛΑΙΟ 1 ΓΕΝΙΚΗ ΘΕΩΡΙΑ ΒΑΘΙΑΣ ΜΑΘΗΣΗΣ	19
1.1 ΓΕΝΙΚΗ ΘΕΩΡΙΑ ΤΕΧΝΗΤΩΝ ΝΕΥΡΩΝΙΚΩΝ ΔΙΚΤΥΩΝ	19
1.1.1 Μοντέλα Νευρώνων	19
1.1.2 Αρχιτεκτονική Νευρωνικών Δικτύων.....	21
1.1.3 Backpropagation	23
1.1.4 Συναρτήσεις Ενεργοποίησης	23
1.1.5 Συναρτήσεις Απώλειας.....	25
1.1.6 Βελτιστοποίηση	25
1.2 ΣΥΝΕΛΙΚΤΙΚΑ ΝΕΥΡΩΝΙΚΑ ΔΙΚΤΥΑ	28
1.2.1 Αντίληψη Εικόνων από τα Συνελικτικά Νευρωνικά Δίκτυα	29
1.2.2 Λειτουργία Συνέλιξης	30
1.2.3 Βασικά Πλεονεκτήματα Συνελικτικών Νευρωνικών Δικτύων	32
1.2.4 Ομαδοποίηση	36
1.2.5 Παραλλαγές της Συνάρτησης Συνέλιξης	37
1.2.6 Τύποι Δεδομένων	39
1.2.7 Ταξινόμηση Εικόνων και Ανίχνευση Αντικειμένων σε Εικόνες.....	40
1.2.8 AlexNet.....	41
1.2.9 Google Inception V1, V2 και V3.....	41
1.2.10 VGG	46
1.2.11 MobileNet	47
1.2.12 ResNet	49
ΚΕΦΑΛΑΙΟ 2 ΜΕΘΟΔΟΙ ΑΝΑΓΝΩΡΙΣΗΣ ΤΡΟΦΙΜΩΝ ΚΑΙ ΒΑΣΕΙΣ ΔΕΔΟΜΕΝΩΝ	52
2.1 ΜΕΘΟΔΟΙ ΑΝΑΓΝΩΡΙΣΗΣ ΤΡΟΦΙΜΩΝ.....	52
2.1.1 Μεταφορά Μάθησης	52
2.1.2 Δημιουργία Νέας Αρχιτεκτονικής Νευρωνικού Δικτύου	53
2.1.3 Πλατφόρμες Βαθιάς Μάθησης	53
2.2 ΒΑΣΕΙΣ ΔΕΔΟΜΕΝΩΝ	55
2.2.1 Food-101	55
2.2.2 ImageNet	55
2.2.3 ChineseFoodNet	55
2.2.4 FoodX-251	55
2.2.5 UECFood100 και UECFood256	56
2.2.6 Food524DB	56

2.2.7 <i>EgocentricFoods</i>	56
2.2.8 <i>Recipe1M+</i>	56
2.3 ΠΛΑΤΦΟΡΜΕΣ ΓΙΑ ΒΑΘΙΑ ΜΑΘΗΣΗ	56
2.3.1 <i>Tensorflow</i>	57
2.3.2 <i>Pytorch</i>	57
2.3.3 <i>Caffe</i> και <i>Caffe2</i>	57
2.3.4 <i>Keras</i>	57
2.3.5 <i>Theano</i>	58
2.3.6 <i>MXNet</i>	58
2.3.7 <i>Deeplearning4j</i>	58
2.4 ΒΙΒΛΙΟΓΡΑΦΙΑ ΜΕΛΕΤΩΝ	58
2.4.1 Μελέτη 1 – Αναγνώριση Τροφίμων με Βάση τη Βαθιά Μάθηση	59
2.4.2 Μελέτη 2 – Επισκόπηση Εφαρμογών Βαθιάς Μάθησης στα Τρόφιμα	60
2.4.3 Μελέτη 3 - Σύγκριση Διαφορετικών Πλατφορμών Αναγνώρισης Τροφίμων	63
2.4.4 Μελέτη 4- Αναγνώριση Τροφίμων για Εικόνες <i>Lifelog</i>	65
2.4.5 Μελέτη 5- <i>FoodAI</i>	65
2.4.6 Μελέτη 6 - <i>PureFoodNet</i>	68
2.4.7 Μελέτη 7 -Σύστημα Ταξινόμησης Τροφίμων με τη Χρήση Μεταφοράς Μάθησης στο <i>InceptionV3</i>	69
2.4.8 Μελέτη 8 <i>Logmeal</i>	69
2.5 ΔΙΑΔΙΚΤΥΑΚΕΣ ΕΦΑΡΜΟΓΕΣ ΓΙΑ ΑΡΧΑΡΙΟΥΣ ΣΤΗΝ ΤΝ ΠΟΥ ΜΠΟΡΟΥΝ ΝΑ ΧΡΗΣΙΜΟΠΟΙΗΘΟΥΝ ΓΙΑ ΕΚΠΑΙΔΕΥΣΗ ΣΥΣΤΗΜΑΤΟΣ ΑΝΑΓΝΩΡΙΣΗΣ ΤΡΟΦΙΜΩΝ	70
2.5.1 <i>Machine Learning For Kids</i>	70
2.5.2 <i>Teachable Machine</i>	71
3.1 ΣΥΛΛΟΓΗ ΕΙΚΟΝΩΝ ΓΙΑ ΤΗΝ ΕΚΠΑΙΔΕΥΣΗ ΤΟΥ ΝΕΥΡΩΝΙΚΟΥ ΔΙΚΤΥΟΥ	72
3.2 ΕΚΠΑΙΔΕΥΣΗ ΝΕΥΡΩΝΙΚΟΥ ΔΙΚΤΥΟΥ ΜΕ <i>TEACHABLE MACHINE</i>	73
3.2.1 Αποτελέσματα Εκπαίδευσης Νευρωνικού Δικτύου με <i>Teachable Machine</i>	74
3.3 ΕΚΠΑΙΔΕΥΣΗ ΝΕΥΡΩΝΙΚΟΥ ΔΙΚΤΥΟΥ <i>MOBILENET</i> ΜΕ ΧΡΗΣΗ ΜΕΤΑΦΟΡΑΣ ΜΑΘΗΣΗΣ	75
3.3.1 Πρόγραμμα Εκπαίδευσης Νευρωνικού Δικτύου με αρχιτεκτονική <i>MobileNet</i> και μεταφορά μάθησης	76
3.3.2 Αποτελέσματα Εκπαίδευσης Νευρωνικού Δικτύου Αρχιτεκτονικής <i>MobileNet</i> με μεταφορά μάθησης	77
3.4 ΕΦΑΡΜΟΓΕΣ ΜΕ ΤΟ ΕΚΠΑΙΔΕΥΜΕΝΟ ΝΕΥΡΩΝΙΚΟ ΔΙΚΤΥΟ ΑΠΟ ΤΗΝ ΠΛΑΤΦΟΡΜΑ <i>TEACHABLE MACHINE</i>	78
3.4.1 Εφαρμογή μέσω της <i>p5.js</i>	78
3.4.2 Εφαρμογή με <i>Keras.h5</i> και <i>Webhook</i>	79
3.4.3 Εφαρμογή με πρόγραμμα περιήγησης ιστού και <i>Javascript</i>	79
3.5 ΕΦΑΡΜΟΓΕΣ ΜΕ ΤΟ ΕΚΠΑΙΔΕΥΜΕΝΟ ΝΕΥΡΩΝΙΚΟ ΔΙΚΤΥΟ <i>MOBILENET</i>	80
3.6 ΠΑΡΑΤΗΡΗΣΕΙΣ	80
ΚΕΦΑΛΑΙΟ 4 ΣΥΜΠΕΡΑΣΜΑΤΑ – ΠΡΟΤΑΣΕΙΣ	81
ΒΙΒΛΙΟΓΡΑΦΙΑ-ΠΗΓΕΣ	84
ΠΑΡΑΡΤΗΜΑ 1: ΠΡΟΓΡΑΜΜΑ ΕΚΠΑΙΔΕΥΣΗΣ ΝΕΥΡΩΝΙΚΟΥ ΔΙΚΤΥΟΥ <i>MOBILENET</i> ΜΕ ΜΕΤΑΦΟΡΑ ΜΑΘΗΣΗΣ	93
ΠΑΡΑΡΤΗΜΑ 2: <i>KERAS.H5</i>.....	97
ΠΑΡΑΡΤΗΜΑΤΑ ΠΑΡΑΡΤΗΜΑ 3: ΠΡΟΓΡΑΜΜΑ .HTML ΑΡΧΕΙΟΥ ΜΕ ΕΝΣΩΜΑΤΩΜΕΝΟ ΠΡΟΓΡΑΜΜΑ JAVASCRIPT ΤΟ ΟΠΟΙΟ ΚΑΝΕΙ ΑΝΑΓΝΩΡΙΣΗ ΤΡΟΦΙΜΩΝ ΜΕΣΩ ΤΟΥ ΕΚΠΑΙΔΕΥΜΕΝΟΥ ΜΟΝΤΕΛΟΥ	99
ΠΑΡΑΡΤΗΜΑ 4: ΠΡΟΓΡΑΜΜΑ ΠΟΥ ΑΛΛΑΖΕΙ ΑΡΧΕΙΟ .H5 ΣΕ .JSON.....	102
ΠΑΡΑΡΤΗΜΑ 5: ΣΥΝΑΡΤΗΣΗ ΠΡΟΓΡΑΜΜΑΤΟΣ ΠΟΥ ΦΟΡΤΩΝΕΙ ΤΟ ΕΚΠΑΙΔΕΥΜΕΝΟ ΜΟΝΤΕΛΟ ΜΟΡΦΗΣ .JSON.....	103

Κατάλογος Πινάκων:

Πίνακας 1: Παραδείγματα δεδομένων που μπορούν να χρησιμοποιηθούν με συνελκτικά νευρωνικά δίκτυα	32
Πίνακας 2: Παρουσιάζονται τα χαρακτηριστικά κάθε επιπέδου του InceptionV2.....	46
Πίνακας 3: Αρχιτεκτονική MobiliNet[64].....	50
Πίνακας 4: Τίτλοι Μελετών.....	59
Πίνακας 5: Αποτελέσματα της Μελέτης 1.....	61
Πίνακας 6: Παρουσιάζεται η αποτελεσματικότητα μοντέλων αναγνώρισης τροφίμων με τη χρήση διαφορετικών βάσεων δεδομένων[106].....	63
Πίνακας 7: Σύγκριση αποτελεσματικότητας για τις σημαντικότερες πλατφόρμες αναγνώρισης τροφίμων υπό διαφορετικές συνθήκες όσον αφορά το φωτισμό και τη γωνία[108].....	64
Πίνακας 8: Σύγκριση της αποτελεσματικότητας που έχουν οι σημαντικότερες πλατφόρμες αναγνώρισης για διαφορετικού τύπου τροφές και ποτά[108].....	65
Πίνακας9: Παρουσιάζεται η ακρίβεια για κάθε αρχιτεκτονική που εξετάστηκε στην μελέτη με τη χρήση της βάσης δεδομένων FoodAI-756[110].....	68
Πίνακας 10: Ποσοστά ακρίβειας Top-1 και Top-5 για κάθε βάση δεδομένων[114].....	69
Πίνακας 11: Αποτελέσματα ακρίβειας Top-1 για διαφορετικά νευρωνικά δίκτυα[115].....	70
Πίνακας 12: Πίνακας 12: Περιεχόμενο της βάσης δεδομένων που χρησιμοποιήθηκε για την εκπαίδευση του νευρωνικού δικτύου.....	74
Πίνακας 13: Πίνακας σύγχυσης για τα δεδομένα επιβεβαίωσης της ορθής λειτουργίας του εκπαιδευμένου νευρωνικού δικτύου με Teachable Machine.....	76
Πίνακας 14: Πίνακας σύγχυσης για τα δεδομένα αξιολόγησης της ορθής λειτουργίας του εκπαιδευμένου νευρωνικού δικτύου με Teachable Machine.....	76
Πίνακας 15: Πίνακας σύγχυσης για τα δεδομένα επιβεβαίωσης της ορθής λειτουργίας του εκπαιδευμένου νευρωνικού δικτύου MobileNet.....	79
Πίνακας 16: Πίνακας σύγχυσης για τα δεδομένα αξιολόγησης της ορθής λειτουργίας του εκπαιδευμένου νευρωνικού δικτύου MobileNet.....	79

Κατάλογος Εικόνων:

<u>Εικόνα 1: Πάνω αριστερά παρουσιάζεται η έξυπνη κάμερα Smarter FridgeCam. Πάνω δεξιά παρουσιάζεται ψυγείο που διαθέτει ενσωματωμένο σύστημα αναγνώριση τροφίμων. Κάτω παρουσιάζεται η εφαρμογή FoodVisor.....</u>	17
<u>Εικόνα 2: Μη γραμμικό μοντέλο νευρώνων,[14].</u>	20
<u>Εικόνα 3: Αφινικός μετασχηματισμός λόγω της ύπαρξης πόλωσης. Σημειώνουμε ότι $U_k=U_b$ όταν $U_k=0$, [14].</u>	21
<u>Εικόνα 4: Δίκτυο πρόσθιας τροφοδότησης με ένα μεμονωμένο επίπεδο νευρώνων[14].</u>	22
<u>Εικόνα 5: Πλήρως συνδεδεμένο δίκτυο πρόσθιας τροφοδότησης με ένα κρυφό επίπεδο και ένα επίπεδο εξόδου, [16].</u>	23
<u>Εικόνα 6: Παρουσιάζεται συνάρτηση όπου δείχνει τη διαφορά μεταξύ τοπικού και ολικού ελαχίστου [32].</u>	26
<u>Εικόνα 7: Παρουσιάζεται πως η συνάρτηση βαθμωτής κατάβασης πλησιάζει προς το ολικό ελάχιστο σταδιακά [33].</u>	27
<u>Εικόνα 8: Τυπικό επίπεδο συνελκτικού δικτύου [18].</u>	29
<u>Εικόνα 9: Ένα παράδειγμα 2-D συνέλιξης χωρίς αναστροφή πυρήνα. Δείχνει πως σχηματίζεται το άνω αριστερά στοιχείο του τανυστή εξόδου εφαρμόζοντας τον πυρήνα στην αντίστοιχη άνω-αριστερά του τανυστή εισόδου [40].</u>	31
<u>Εικόνα 10: Συνέλιξη με 3-D τανυστή[41].</u>	32
<u>Εικόνα 11: Η μορφή της σποραδική συνδεσιμότητα πάνω και πλήρης κάτω[41].</u>	33
<u>Εικόνα 12: Σποραδική συνδεσιμότητα πάνω και πλήρης κάτω[40]. Επισημαίνονται οι συνδέσεις ως προς την μονάδα εξόδου s_3.</u>	34
<u>Εικόνα 13: Το δεκτικό πεδίο των μονάδων στα “βαθύτερα” στρώματα ενός συνελκτικού νευρωνικού δικτύου. Οι μονάδες στα βαθύτερα στρώματα συνδέονται έμμεσα με το σύνολο ή το μεγαλύτερο μέρος της εικόνας εισόδου[40].</u>	34
<u>Εικόνα 14: Κοινή χρήση παραμέτρων: Πάνω παρουσιάζεται ένα δίκτυο μερικής συνδεσιμότητας και κάτω πλήρους συνδεσιμότητας [40].</u>	35
<u>Εικόνα 15: Αποτελεσματικότητα ανίχνευσης άκρων (Efficiency of edge detection)[40].</u>	35
<u>Εικόνα 16: Μορφές ομαδοποίησης μέγιστου όρου και μέσο όρου [48].</u>	37
<u>Εικόνα 17: Συνέλιξη με δρασκέλισμα[47].</u>	38
<u>Εικόνα 18: Η επίδραση της μηδενικής επένδυσης(zero padding) στο μέγεθος του δικτύου: Δείχνει ένα συνελκτικό δίκτυο με πυρήνα πλάτους έξι σε κάθε επίπεδο[47].</u>	39
<u>Εικόνα 19: Αριστερά παρουσιάζεται μια εικόνα που έχει ταξινομηθεί ενώ δεξιά τα αντικείμενα της εικόνας έχουν εντοπιστεί και ταξινομηθεί [50].</u>	41
<u>Εικόνα 20: Η αρχιτεκτονική του AlexNet[54].</u>	41
<u>Εικόνα 21: Απλοϊκή βαθμίδα της αρχιτεκτονικής Inception [56].</u>	42
<u>Εικόνα 22: Βαθμίδα της αρχιτεκτονικής Inception με μειωμένες διαστάσεις[56].</u>	42
<u>Εικόνα 23: Αρχιτεκτονική GoogleNet[56].</u>	43
<u>Εικόνα 24: Παρουσιάζεται πως η συνέλιξη 5×5 αναπαρίσταται ισοδύναμα από 2 3×3 συνέλιξεις η μία μετά την άλλη[58].</u>	44
<u>Εικόνα 25: Παρουσιάζεται η αρχιτεκτονική μιας βαθμίδας Inception όταν η συνέλιξη 3×3 αντικαθίσταται ισοδύναμα από δύο συνέλιξεις 1×3 και 3×1 η μία μετά την άλλη[58].</u>	44
<u>Εικόνα 26: Παρουσιάζεται η αρχιτεκτονική της εικόνας 25 όταν το τελευταίο σκέλος της βαθμίδας το οποίο είναι συνέλιξη 3×3 αντικαθίσταται ισοδύναμα από δυο συνέλιξεις 1×3 και 3×1 σε παραλληλία[58].</u>	45
<u>Εικόνα 27: Αρχιτεκτονική του νευρωνικού δικτύου VGG-16[61].</u>	46
<u>Εικόνα 28: Εφαρμογές με την αρχιτεκτονική MobileNet[62].</u>	47
<u>Εικόνα 29: Πάνω, γίνεται συνέλιξη σε βάθος μιας εικόνας μεγέθους 12×12 με τρία κανάλια με 3 πυρήνες μεγέθους $5 \times 5 \times 1$ και παράγει μια νέα εικόνα με 8×8 με τρία κανάλια. Στη συνέχεια ενδιάμεσα παρουσιάζεται η συνέλιξη pointwise με ένα μόνο πυρήνα. Κάτω παρουσιάζεται το αποτέλεσμα αν γίνονταν συνέλιξη με 256 πυρήνες pointwise[63].</u>	48
<u>Εικόνα 30: Αριστερά είναι ένα επίπεδο συνέλιξης με batchnorm και ReLU ενώ δεξιά είναι μια σε βάθος διαχωριζόμενη συνέλιξη με επίπεδα για σε βάθος συνέλιξη και συνέλιξη pointwise ακολουθούμενο από batchnorm και ReLU[64].</u>	49
<u>Εικόνα 31: Μια βαθμίδα της αρχιτεκτονικής του νευρωνικού δικτύου ResNet[67].</u>	50
<u>Εικόνα 32: Αριστερά παρουσιάζεται η αρχιτεκτονική του νευρωνικού δικτύου VGG με 19 επίπεδα νευρώνων, στη μέση παρουσιάζεται η αρχιτεκτονική ενός συνελκτικού νευρωνικού δικτύου με 34 επίπεδα νευρώνων και δεξιά παρουσιάζεται το ίδιο νευρωνικό δίκτυο με τη χρήση βαθμίδων ResNet[67].</u>	51
<u>Εικόνα 33: Αρχιτεκτονική του νευρωνικού δικτύου FoodNet[103].</u>	61
<u>Εικόνα 34: Αρχιτεκτονική του νευρωνικού δικτύου WISer[104].</u>	61
<u>Εικόνα 35: Αρχιτεκτονική του νευρωνικού δικτύου LeNet-5[105].</u>	62
<u>Εικόνα 36: Σχεδιάγραμμα το οποίο δείχνει την αρχιτεκτονική του συστήματος αναγνώρισης[109].</u>	65

<u>Εικόνα 37: Παρουσιάζεται εικόνα όπου γίνεται επίδειξη της εφαρμογής FoodAI[12].</u>	66
<u>Εικόνα 38: Σχεδιάγραμμα λειτουργίας του συστήματος FoodAI[111].</u>	66
<u>Εικόνα 39: Ιστόγραμμα το οποίο παρουσιάζει τις κλάσεις της βάσης δεδομένων που χρησιμοποιήθηκε για την εκπαίδευση των νευρωνικών δικτύων FoodAI-756[110].</u>	67
<u>Εικόνα 40: Σχεδιάγραμμα της αρχιτεκτονικής του νευρωνικού δικτύου PureFoodNet[111].</u>	68
<u>Εικόνα 41: Αρχιτεκτονική συστήματος αναγνώρισης τροφίμων LogMeal[116].</u>	70
<u>Εικόνα 42: Επιφάνεια εργασιών όπου μπορεί να εκπαιδευτεί σύστημα αναγνώρισης μέσω της εφαρμογής Teachable Machine[121].</u>	71
<u>Εικόνα 43: Εικόνες από την βάση δεδομένων που χρησιμοποιήθηκε προκειμένου να εκπαιδευτεί το νευρωνικό δίκτυο.</u>	73
<u>Εικόνα 44: Η γραφική παράσταση παρουσιάζει την ακρίβεια ανά βήμα εκπαίδευσης (epochs).</u>	74
<u>Εικόνα 45: Στη συνάρτηση παρουσιάζεται η απώλεια (loss) σε συνάρτηση με το βήμα στο οποίο βρίσκεται η εκπαίδευση.</u>	74
<u>Εικόνα 46: Παρουσιάζεται ο αριθμός των εικόνων για κάθε βάση δεδομένων που είναι για εκπαίδευση.</u>	76
<u>Εικόνα 47: Παρουσιάζεται ο αριθμός των εικόνων για κάθε βάση δεδομένων που είναι για τη επιβεβαίωση της εκπαίδευσης του νευρωνικού δικτύου.</u>	77
<u>Εικόνα 48: Αριστερά παρουσιάζεται η πρόοδος της ακρίβειας κατά την διάρκεια της εκπαίδευσης ενώ δεξιά ο ρυθμός απώλειας. Με μπλε χρώμα παρουσιάζεται η πρόοδος για τα δεδομένα εκπαίδευσης ενώ με κόκκινο για τα δεδομένα που χρησιμοποιήθηκαν για επιβεβαίωση.</u>	77
<u>Εικόνα 49: Επιφάνεια εργασιών όπως εμφανίζεται στο χρήστη όταν συνδέεται μέσω προγράμματος περιήγησης.</u>	80

Η διατροφή αποτελεί ακρογωνιαίο λίθο της καθημερινής ζωής με πολυεπίπεδες επιπτώσεις στην υγεία, το πολιτισμό και τη συμπεριφορά των ανθρώπων. Συνήθειες που αφορούν την κατανάλωση τροφής αποτελούν πηγή πολλαπλών ασθενειών σε περιπτώσεις όπου οι τροφές είναι ποιοτικά ή ποσοτικά επιζήμιες για την υγεία [1]. Έχει αποδειχτεί ότι ασθένειες όπως διάφορες μορφές καρκίνου και καρδιακά προβλήματα συνδέονται άμεσα με το πόσες θερμίδες καταναλώνει ο άνθρωπος καθημερινά, ενώ η παχυσαρκία έχει ανακηρυχθεί το 2013 από το Ιατρικό Σωματείο της Αμερικής ως σοβαρή ασθένεια που πρέπει να αντιμετωπιστεί [2].

Παραδοσιακές Μέθοδοι Προγραμματισμού Διατροφής

Για την αντιμετώπιση των προβλημάτων που προέρχονται από την κακή διατροφή οι παραδοσιακές μέθοδοι συνήθως χρησιμοποιούν κάποια μορφή ημερολογιακής καταγραφής. Έτσι, ο χρήστης καταγράφει μόνος του τι έχει καταναλώσει και στη συνέχεια με τη βοήθεια διαιτολόγου, ο οποίος συμβουλεύει το χρήστη για τη σωστή διατροφή, ακολουθεί και την αντίστοιχη δίαιτα [3]. Αυτό ωστόσο έχει το μειονέκτημα ότι ο χρήστης μπορεί να μην καταγράψει με ακρίβεια τις τροφές που λαμβάνει, αφού ενδέχεται την ώρα που τις καταναλώνει να μην έχει τη δυνατότητα να καταγράψει τη δραστηριότητά του. Επιπλέον, ο διαιτολόγος δεν μπορεί να συμβουλέψει τον χρήστη σε ζωντανό χρόνο για το ποιες τροφές πρέπει να καταναλώσει και ποιες πρέπει να αποφύγει [4].

Συστήματα Αναγνώρισης Τροφίμων με τη Βοήθεια Βαθιάς Μάθησης και τα Πλεονεκτήματά τους

Με τη ανάπτυξη αυτόματων συστημάτων για την αναγνώριση τροφίμων διευκολύνεται η καταγραφή της τροφής που καταναλώνει ο χρήστης αφού το μόνο που χρειάζεται είναι μια φωτογραφία της. Με αυτό τον τρόπο, ο τελευταίος διευκολύνεται στο να διαχειριστεί καλύτερα ασθένειες που έχουν σχέση με την κατανάλωση τροφής όπως η παχυσαρκία ή ο διαβήτης, αφού δεν χρειάζεται να καταγράψει ο ίδιος τι έχει καταναλώσει και σε τι ποσότητες. Επιπλέον, με την εφαρμογή τέτοιων συστημάτων σε χώρους αποθήκευσης τροφίμων, ο χρήστης μπορεί να ενημερώνεται για τα διαθέσιμα ή μη τρόφιμα σε ζωντανό χρόνο προκειμένου να ρυθμίσει ευκολότερα τα ψώνια του [5,6], αλλά και να διευκολύνεται στην τήρηση μίας ισορροπημένης διαίτας και υγιεινής διατροφής. Η ποιότητα και η ασφάλεια των τροφίμων είναι ένα ιδιαίτερος σημαντικό κοινωνικό ζήτημα όντας στη βάση της ανθρώπινης υγείας, της κοινωνικής ανάπτυξης και της σταθερότητας, με τη διασφάλισή της να αποτελεί πολύπλοκη διαδικασία κατά την οποία πρέπει να λαμβάνονται υπόψη όλα τα στάδια της επεξεργασίας των τροφίμων, από την καλλιέργεια, τη συγκομιδή και την αποθήκευση έως την προετοιμασία και την κατανάλωση. Αυτές οι διαδικασίες, ωστόσο, συχνά απαιτούν εντατική εργασία. Σήμερα, η ανάπτυξη της μηχανικής όρασης μπορεί να βοηθήσει σημαντικά τους ερευνητές και τις βιομηχανίες στη βελτίωση της αποτελεσματικότητας της επεξεργασίας τροφίμων. Ως αποτέλεσμα, η μηχανική όραση έχει χρησιμοποιηθεί ευρέως σε όλες τις πτυχές της επεξεργασίας τροφίμων. Ταυτόχρονα, η επεξεργασία εικόνας, ένα σημαντικό συστατικό της μηχανικής όρασης, μπορεί να εκμεταλλευτεί μοντέλα μηχανικής και βαθιάς μάθησης προκειμένου να προσδιορίσει αποτελεσματικά τον τύπο και την ποιότητα του φαγητού. Ακολούθως, ο σχεδιασμός

Μεταπτυχιακή Διπλωματική Εργασία, Όλσους Πάντος, AM msciot18005

παρακολούθησης στο σύστημα μηχανικής όρασης μπορεί να εκτελέσει εργασίες όπως ταξινόμηση τροφίμων, εντοπισμό θέσεων ελαττωματικών σημείων ή ξένων αντικειμένων και αφαίρεση ακαθαρσιών. Η ευρεία εφαρμογή των παραπάνω μεθόδων στον τομέα της επεξεργασίας τροφίμων μπορεί να πιστοποιηθεί από πολλές μελέτες διεθνούς βιβλιογραφίας [5,6].

Η φύση και η ποιότητα των τροφών, καθώς και η ποσότητα που διαθέτουν οι χώροι αποθήκευσης και τα ψυγεία τροφίμων, έχουν γίνει αντικείμενο ακαδημαϊκής και βιομηχανικής έρευνας [5,6]. Συστήματα αυτοματισμού με τα οποία θα μπορούμε να γνωρίζουμε το είδος και τι ποσότητες τροφίμων που διαθέτουμε στο σπίτι μας μπορούν να βοηθήσουν ώστε αν υπάρξουν ελλείψεις, λόγου χάρη αν το μπουκάλι με το γάλα στο ψυγείο έχει χαμηλή ποσότητα γάλακτος, να λαμβάνει την ανάλογη ειδοποίηση ο ιδιοκτήτης ή ακόμα και να πραγματοποιείται αυτόματη παραγγελία για αγορά γάλακτος. Ένα σύστημα αυτοματισμού θα μπορούσε να ελέγχει πλήρως τον αποθηκευτικό χώρο ή το ψυγείο και να προτείνει τις αντίστοιχες συνταγές για μαγείρεμα μειώνοντας δραστικά το χρόνο που επενδύει ο άνθρωπος για την λήψη τροφής. Αυτό είναι ιδιαίτερα σημαντικό και για ανθρώπους που έχουν προβλήματα υγείας και αναγκάζονται να περιορίζουν το εύρος των διατροφικών επιλογών τους, όπως για παράδειγμα σε άτομα με αλλεργίες. Επιπλέον, με τη χρήση ενός αυτόματου συστήματος όρασης μπορούμε να μειώσουμε τη σπατάλη χρημάτων, καθώς ο χρήστης θα μπορεί να γνωρίζει τότε πλησιάζει η ημερομηνία λήξης των τροφίμων του και να προειδοποιείται να καταναλώσει άμεσα τη τροφή αυτή δεδομένου ότι είναι ακόμα βρώσιμη[7].

Για το σκοπό αυτό έχουν αναπτυχθεί στον εμπορικό τομέα συστήματα αυτόματης όρασης με τη χρήση βαθιάς μάθησης. Έξυπνες κάμερες, όπως το Smarter FridgeCam που έχει ενσωματωμένο σύστημα αναγνώρισης τροφίμων μπορούν να τοποθετηθούν σε αποθηκευτικούς χώρους και ψυγεία δίνοντας τη δυνατότητα ακριβούς ενημέρωσης και ελέγχου αναφορικά με το περιεχόμενο του ψυγείου του χρήστη[8].

Για την καλύτερη διαχείριση των τροφών και καλύτερο προγραμματισμό για την κατανάλωσή τους με υγιή τρόπο, εταιρείες όπως η Samsung και η LG κατασκευάζουν “έξυπνα ψυγεία” τα οποία έχουν ενσωματωμένες κάμερες με τη δυνατότητα αναγνώρισης τροφών μέσω συστήματος βαθιάς μάθησης. Τα συστήματα αυτά μπορούν επίσης να δουν ποιες τροφές είναι σε μικρές ποσότητες ή δεν υπάρχουν πλέον στο ψυγείο και στη συνέχεια να ειδοποιήσουν τον ιδιοκτήτη προκειμένου να κάνει τις απαραίτητες αγορές. Το ψυγείο, μάλιστα, της Samsung έχει τη δυνατότητα μέσω των εφαρμογών Whisk να ελέγξει και να συμβουλευτεί τον χρήστη σε ζωντανό χρόνο μέσω του κινητού του τηλεφώνου, καθώς και να προγραμματίσει γεύματα έως και μια εβδομάδα νωρίτερα με προτάσεις της εφαρμογής δημιουργώντας λίστες τροφών για αγορά συνδυαστικά με τις ήδη υπάρχουσες τροφές [9,10,11].

Πληθώρα δικτυακών εφαρμογών που μπορούν να εγκατασταθούν σε κινητά τηλέφωνα έχουν την δυνατότητα να αναγνωρίζουν τις τροφές από φωτογραφίες που λαμβάνονται αναγνωρίζοντας το περιεχόμενο τροφών και κάνοντας εκτίμηση της θρεπτικής τους αξίας. Τέτοιες εφαρμογές είναι το FoodAI το οποίο αναπτύχθηκε από ερευνητές του Πανεπιστημίου Διοίκησης Επιχειρήσεων της Σιγκαπούρης και η εφαρμογή για κινητές συσκευές FoodVisor, το οποίο αποτελεί ιδιωτική πρωτοβουλία. Με την βοήθεια τέτοιου τύπου εφαρμογών, ο χρήστης έχει την δυνατότητα να κρατάει ημερολόγιο με καταγραφές των τροφών που καταναλώθηκαν και να παίρνει συμβουλές ως προς την διατροφική αξία τους και για το σύνολο των θρεπτικών ουσιών που καταναλώθηκαν [12,13].



Εικόνα 1: Πάνω αριστερά παρουσιάζεται η έξυπνη κάμερα Smarter FridgeCam. Πάνω δεξιά παρουσιάζεται ψυγείο που διαθέτει ενσωματωμένο σύστημα αναγνώριση τροφίμων. Κάτω παρουσιάζεται η εφαρμογή FoodVisor.

Θέμα της Εργασίας

Η εργασία αυτή έχει ως θέμα την μελέτη και την ανάπτυξη συστήματος αναγνώρισης τροφίμων με τη βοήθεια της βαθιάς μάθησης. Για το σκοπό αυτό, αρχικά παρουσιάζεται η βασική δομή και τα χαρακτηριστικά ενός νευρωνικού δικτύου καθώς και ο τρόπος λειτουργίας τους. Εν συνεχεία, παρουσιάζονται και αναλύονται τα συνελκτικά νευρωνικά δίκτυα και τα πλεονεκτήματά τους, οι τύποι δεδομένων που χρησιμοποιούνται για την εκπαίδευση των νευρωνικών δικτύων και η διαφορά μεταξύ αναγνώρισης και εντοπισμού ενός αντικειμένου σε μια εικόνα. Επιπλέον παρουσιάζονται οι σημαντικότερες και πιο γνωστές αρχιτεκτονικές συνελκτικών νευρωνικών δικτύων πάνω στα οποία έχουν αναπτυχθεί εφαρμογές αναγνώρισης αντικειμένων και οι σημαντικότερες βιβλιοθήκες που υπάρχουν για τον προγραμματισμό νευρωνικών δικτύων. Ο τρόπος με τον οποίο μπορεί κάποιος να αναπτύξει μεθόδους για αναγνώριση τροφίμων κατηγοριοποιείται σε τρεις μεθόδους ενώ παρουσιάζονται οι πιο δημοφιλείς βάσεις δεδομένων που περιέχουν τροφές. Επιπλέον παρουσιάζεται βιβλιογραφία που αναλύει τις μεθόδους βαθιάς μάθησης και την αποτελεσματικότητά τους στην αναγνώριση τροφίμων. Τέλος, αναπτύσσονται συστήματα αναγνώρισης τροφίμων τα οποία έχουν τη δυνατότητα να αναγνωρίζουν την ποσότητα γάλακτος που περιέχεται σε μπουκάλια με την χρήση βαθιάς μάθησης.

Μεθοδολογία

Η διπλωματική εργασία έχει δύο σκέλη. Στο πρώτο σκέλος που αφορά την θεωρία, πραγματοποιείται βιβλιογραφική ανασκόπηση και γίνεται παρουσίαση πηγών και αναφορών που συλλέχθηκαν μέσω του Google Scholar και του Science Direct. Στο δεύτερο σκέλος της εργασίας που αφορά το πρακτικό μέρος, δημιουργείται βάση δεδομένων 5 κλάσεων. Στη συνέχεια με τη χρήση της βάσης δεδομένων αναπτύσσονται δύο μοντέλα νευρωνικών δικτύων που βασίζονται στην αρχιτεκτονική MobileNet και αναλύεται η

απόδοσή τους, ενώ επίσης παρουσιάζονται εφαρμογές στις οποίες μπορούν να εφαρμοστούν τα εκπαιδευμένα μοντέλα.

Δομή Εργασίας

Η δομή του κειμένου της εργασίας έχει χωριστεί σε 5 κεφάλαια. Στη συνέχεια παρουσιάζονται τα περιεχόμενα του κάθε κεφαλαίου.

Στο 1^ο κεφάλαιο με τίτλο “Γενική Θεωρία Βαθιάς Μάθησης” γίνεται αρχικά μια γενική εξήγηση του τι είναι νευρώνας και νευρωνικό δίκτυο. Στη συνέχεια παρουσιάζεται η λειτουργία και αρχιτεκτονική του συνελκτικού νευρωνικού δικτύου. Επεξηγείται ποια είναι η διαφορά μεταξύ ταξινόμησης και εντοπισμού αντικειμένων σε εικόνες. Στην ίδια ενότητα παρουσιάζονται και μερικές από τις πιο σημαντικές αρχιτεκτονικές που χρησιμοποιούν τα συνελκτικά νευρωνικά δίκτυα. Στην επόμενη ενότητα, γίνεται παρουσίαση των σημαντικότερων βιβλιοθηκών προγραμματισμού που αφορούν την βαθιά μάθηση.

Στο 2^ο κεφάλαιο με τίτλο “Μέθοδοι Αναγνώρισης Τροφίμων και Βάσεις Δεδομένων” παρουσιάζονται οι μέθοδοι με τις οποίες μπορούν να δημιουργηθούν συστήματα αναγνώρισης τροφίμων με βαθιά μάθηση και προβάλλονται οι πιο γνωστές και σημαντικές βάσεις δεδομένων που αφορούν τροφές. Τέλος, υπάρχει βιβλιογραφική ανασκόπηση όπου παρουσιάζονται μελέτες που αφορούν τον τομέα αναγνώρισης τροφίμων με τη χρήση βαθιάς μάθησης.

Στο 3^ο κεφάλαιο με τίτλο “Πλαίσιο Ανάπτυξης Εφαρμογών” γίνεται παρουσίαση της βάσης δεδομένων που δημιουργήθηκε για την εκπαίδευση των νευρωνικών δικτύων. Στην συνέχεια αναπτύσσονται με δύο τρόπους δύο μοντέλα αρχιτεκτονικής MobileNet με χρήση βαθιάς μάθησης. Στον πρώτο τρόπο, χρησιμοποιείται η διαδικτυακή πλατφόρμα Teachable Machine η οποία έχει ήδη ενσωματωμένη την αρχιτεκτονική MobileNet. Στο δεύτερο τρόπο δημιουργείται πρόγραμμα όπου η ίδια αρχιτεκτονική έχει μειωμένο το πλάτος κάθε στοιβάδας νευρώνων κατά 50%. Τέλος παρουσιάζονται εφαρμογές όπου μπορούν ενσωματωθούν τα εκπαιδευμένα μοντέλα.

Η διπλωματική εργασία ολοκληρώνεται με το 4^ο κεφάλαιο, με τίτλο “Επίλογος” όπου γίνεται ανασκόπηση της εργασίας. Στην ενότητα των παραρτημάτων υπάρχει ο κώδικας των προγραμμάτων που δημιουργήθηκαν για το πρακτικό μέρος της εργασίας.

ΚΕΦΑΛΑΙΟ 1

Γενική Θεωρία Βαθιάς Μάθησης

Σε αυτό το κεφάλαιο θα παρουσιαστεί το θεωρητικό τμήμα της διπλωματικής εργασίας. Αρχικά, θα εξεταστούν κάποιες βασικές έννοιες όπως τι είναι νευρώνας και τι τεχνητό νευρωνικό δίκτυο και τι σχέση έχουν με τη βαθιά μάθηση. Στη συνέχεια θα εξεταστούν τα Συνελικτικά Νευρωνικά Δίκτυα (Convolutional Neural Networks) και η λειτουργία ομαδοποίησης όπως και το πώς επεξεργάζεται το σύστημα τις εικόνες. Τέλος, θα παρουσιαστούν οι σημαντικότερες αρχιτεκτονικές που έχουν αναπτυχθεί για την πραγματοποίηση αναγνώρισεων σε εικόνες.

1.1 Γενική θεωρία Τεχνητών Νευρωνικών Δικτύων

Ένα νευρωνικό δίκτυο έχει κατανεμημένη αρχιτεκτονική και αποτελείται από απλές μονάδες επεξεργασίας οι οποίες ονομάζονται νευρώνες. Έχει τη δυνατότητα να αποθηκεύει εμπειρική γνώση και κατέχει δυο κοινά χαρακτηριστικά με τον ανθρώπινο εγκέφαλο[14]:

1. Έχει τη δυνατότητα να μαθαίνει λαμβάνοντας πληροφορίες από το περιβάλλον. Αυτό ονομάζεται διαδικασία μάθησης.

2. Η αποθήκευση της γνώσης πραγματοποιείται στους συνδέσμους (συνάψεις) μεταξύ των νευρώνων τα οποία ονομάζονται συναπτικά βάρη.

Εν συντομία, ένα τεχνητό νευρωνικό δίκτυο είναι ένα δίκτυο το οποίο μιμείται τη λειτουργία του νευρικού συστήματος των έμβιων οργανισμών, κυρίως του εγκεφάλου. Στη διεθνή βιβλιογραφία το σύνολο του πεδίου της μηχανική μάθησης που βασίζεται στα νευρωνικά δίκτυα ονομάζεται βαθιά μάθηση[14].

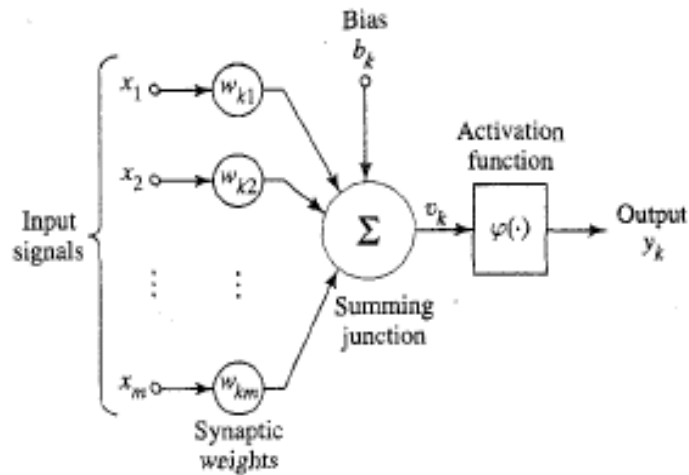
1.1.1 Μοντέλα Νευρώνων

Ο νευρώνας είναι μονάδα επεξεργασίας της πληροφορίας που αποτελεί τη βάση πάνω στην οποία φτιάχνουμε το νευρωνικό δίκτυο ενώ το βασικό σχηματικό διάγραμμα του παρουσιάζεται στην εικόνα 2. Τα κύρια στοιχεία αυτού του μοντέλου είναι τα ακόλουθα[14]:

1. Ένα σύνολο συνάψεων ή συνδέσμων που το καθένα έχει το δικό του βάρος. Συγκεκριμένα, το σήμα εισόδου x_j στην είσοδο της σύναψης j που συνδέεται με το νευρώνα k πολλαπλασιάζεται με το αντίστοιχο συναπτικό βάρος w_{kj} .

2. Έναν αθροιστή Σ που αθροίζει τα σήματα εισόδου τα οποία έχουν σταθμιστεί από τα αντίστοιχα συναπτικά βάρη του νευρώνα.

3. Μία συνάρτηση ενεργοποίησης $\varphi(\cdot)$ που περιορίζει το πλάτος του σήματος εξόδου του νευρώνα. Η συνάρτηση ενεργοποίησης ονομάζεται διαφορετικά συνάρτηση περιορισμού, επειδή περιορίζει το επιτρεπτό εύρος πλάτους του σήματος εξόδου σε συγκεκριμένες τιμές. Συνήθως το κανονικοποιημένο εύρος τιμών πλάτους της εξόδου του νευρώνα είναι ένα μοναδιαίο κλειστό διάστημα $[0,1]$ ή $[-1,1]$.



Εικόνα 2: Μη γραμμικό μοντέλο νευρώνα,[14].

Ο νευρώνας της εικόνας 2 περιέχει και μία εξωτερική διέγερση b_k που αυξάνει ή μειώνει τη δικτυακή διέγερση της συνάρτησης ενεργοποίησης ανάλογα με το εάν είναι θετική ή αρνητική, αντίστοιχα μετατοπίζοντας τη συνάρτηση (1) σε σχέση με την αρχή του άξονα όπως παρουσιάζεται στο σχήμα στην εικόνα 3 και αναγράφει η εξίσωση (3).

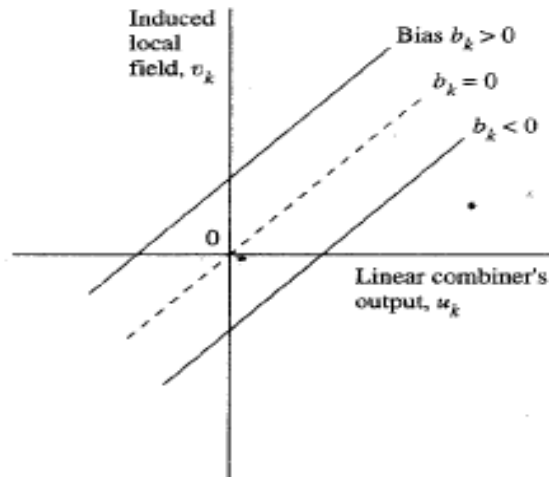
Το μοντέλο του νευρώνα k της εικόνας 2 απεικονίζεται μαθηματικά από το ακόλουθο ζεύγος εξισώσεων:

$$u_k = \sum_{j=1}^m w_{kj} * X_j \quad (1) \text{ και}$$

$$y_k = \phi(u_k + b_k) \quad (2)$$

όπου x_1, x_2, \dots, x_m είναι τα σήματα εισόδου, $w_{k1}, w_{k2}, \dots, w_{km}$ είναι τα αντίστοιχα συναπτικά βάρη του νευρώνα k , u_k είναι η έξοδος του γραμμικού συνδυαστή που οφείλεται στα σήματα εισόδου, b_k είναι η πόλωση (διέγερση), $\phi(\cdot)$ είναι η συνάρτηση ενεργοποίησης και y_k είναι το σήμα εξόδου του νευρώνα. Το αποτέλεσμα της διέγερσης από την πόλωση b_k είναι ότι εφαρμόζεται ένας αφινικός μετασχηματισμός στην έξοδο u_k όπως δείχνει η εξίσωση (3).

$$u_k = u_k + b_k \quad (3)$$



Εικόνα 3: Αφινικός μετασχηματισμός λόγω της ύπαρξης πόλωσης. Σημειώνουμε ότι $U_k = U_b$ όταν $U_k = 0$, [14].

Την εξωτερική διέγερση b_k του νευρώνα k μπορούμε να την υπολογίσουμε μέσω της εξίσωσης(2). Ισοδύναμα, μπορούμε τις εξισώσεις (1) και (3) να τις διατυπώσουμε ως εξής:

$$U_k = \sum_{j=0}^m W_{kj} * X_j \quad (4) \text{ και}$$

$$y_k = \Phi(u_k) \quad (5)$$

Στην (4) προσθέσαμε μια νέα σύναψη:

$$x_0 = +1 \quad (6)$$

και έχει βάρος

$$w_{k0} = b_k \quad (7).$$

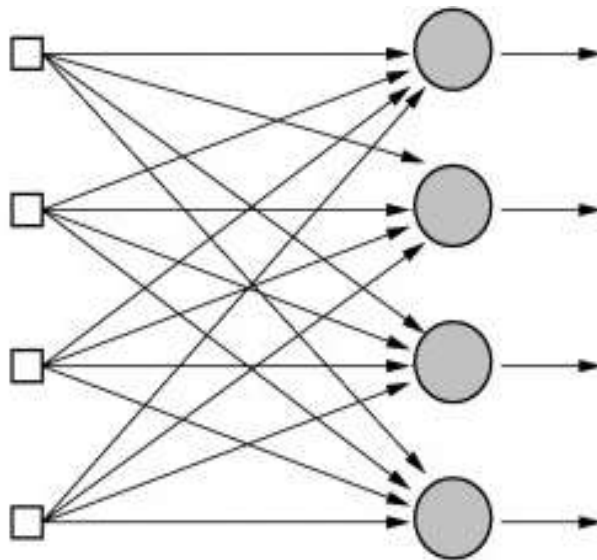
Ανάλογα με το συγκεκριμένο μοντέλο που χρησιμοποιείται, μπορεί να ονομάζονται ημι-γραμμική μονάδα (semi-linear unit), νευρώνας N_n (N_n neuron), δυαδικός νευρώνας (binary neuron), γραμμική συνάρτηση κατωφλίου (linear threshold function) ή νευρώνας McCulloch-Pitts (MCP).

Οι απλοί τεχνητοί νευρώνες, όπως το μοντέλο McCulloch-Pitts, μερικές φορές περιγράφονται ως "μοντέλα καρικατούρας" (caricature models), καθώς προορίζονται να αντανακλούν μία ή περισσότερες νευροφυσιολογικές παρατηρήσεις, αλλά χωρίς να λαμβάνονται υπόψη ο ρεαλισμός [15].

1.1.2 Αρχιτεκτονική Νευρωνικών Δικτύων

Ο αλγόριθμος μάθησης που χρησιμοποιούμε για την εκπαίδευση του νευρωνικού δικτύου σχετίζεται άμεσα με το πώς είναι δομημένοι οι νευρώνες του συστήματος. Οι βασικές δομές αρχιτεκτονικής των νευρωνικών δικτύων είναι οι ακόλουθες:

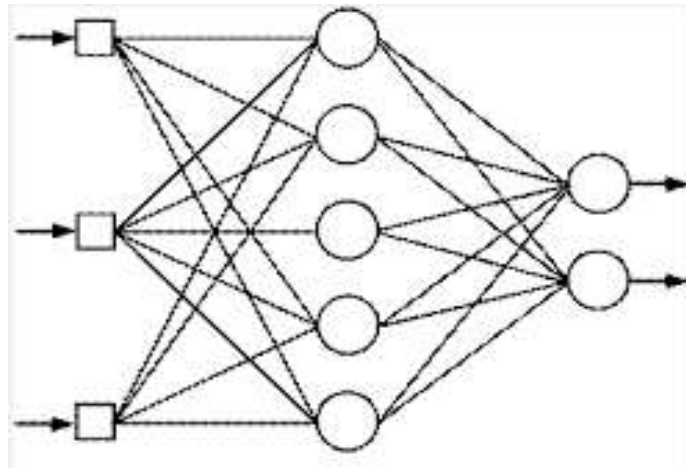
Δίκτυα Πρόσθιας Τροφοδότησης Ενός Επιπέδου: Είναι η απλούστερη μορφή ενός νευρωνικού δικτύου και αποτελείται από δύο επίπεδα (στρώματα), το στρώμα εισόδου, που ονομάζεται πηγαίος κόμβος, όπου εισέρχονται τα δεδομένα και το οποίο συνδέεται με το στρώμα νευρώνων που κάνουν υπολογισμούς και λειτουργούν ως έξοδος. Το στρώμα εισόδου στέλνει δεδομένα στο στρώμα εξόδου αλλά όχι αντίστροφα, όπως δείχνει η εικόνα 4. Αυτό το σύστημα είναι τύπου πρόσθιας τροφοδότησης (feedforward). Στην εικόνα 4 έχουμε ένα νευρωνικό δίκτυο ενός επιπέδου πρόσθιας τροφοδότησης με 4 κόμβους ως είσοδο και 4 νευρώνες ως έξοδο [14].



Εικόνα 4: Δίκτυο πρόσθιας τροφοδότησης με ένα μεμονωμένο επίπεδο νευρώνων[14].

Πολυεπίπεδα Δίκτυα Πρόσθιας Τροφοδότησης: Η δεύτερη κατηγορία πρόσθιας τροφοδότησης, η οποία λέγεται πολυεπίπεδο δίκτυο πρόσθιας τροφοδότησης (εμπροσθοδιάδοσης) ή, αλλιώς, νευρωνικά δίκτυα MLP (multilayer perceptron) έχει τουλάχιστον ένα κρυφό στρώμα με νευρώνες που λέγονται κρυφοί νευρώνες ή κρυφές μονάδες. Λέγονται κρυφές γιατί δεν γίνονται άμεσα αντιληπτές ούτε από το στρώμα εισόδου ούτε από το στρώμα εξόδου. Δουλειά τους είναι να παρεμβαίνουν θετικά μεταξύ της εξωτερικά προερχόμενης εισόδου και της εξόδου αυξάνοντας την αποτελεσματικότητα του νευρωνικού δικτύου. Όσο αυξάνουμε τα επίπεδα καθώς και τον αριθμό νευρώνων που έχουν, τόσο αυξάνεται η δυνατότητα στατιστικά υψηλότερης τάξης από την είσοδο του. Δηλαδή, το δίκτυο αποκτά μια γενική προοπτική λόγω του επιπλέον συνόλου συναπτικών συνδέσεων και της επιπλέον διάστασης νευρωνικών αλληλεπιδράσεων, παρά το τοπικό χαρακτήρα των συνδέσεων[14].

Οι πηγαίοι κόμβοι, δηλαδή οι κόμβοι του επιπέδου εισόδου, δίνουν τα στοιχεία του προτύπου ενεργοποίησης που στη συνέχεια γίνονται σήματα εισόδου για το πρώτο κρυφό στρώμα νευρώνων. Στη συνέχεια, τα σήματα εξόδου του πρώτου κρυφού επιπέδου νευρώνων περνάει στο επόμενο κρυφό, αν υπάρχει, μέχρι να φτάσει στο επίπεδο εξόδου το οποίο είναι η συνολική απόκριση του προτύπου ενεργοποίησης του δικτύου που προέρχεται από τους πηγαίους κόμβους. Οι νευρώνες κάθε στρώματος έχουν συνήθως ως σήματα εισόδου μόνο την έξοδο του προηγούμενου επιπέδου. Στη εικόνα 5 έχουμε ένα δίκτυο 3-5-2, δηλαδή 3 πηγαίους κόμβους, 5 κρυφοί νευρώνες του κρυφού στρώματος και 2 νευρώνες εξόδου[16].



Εικόνα 5: Πλήρως συνδεδεμένο δίκτυο πρόσθιας τροφοδότησης με ένα κρυφό επίπεδο και ένα επίπεδο εξόδου, [16].

Το νευρωνικό δίκτυο της εικόνα 5 λέγεται πλήρως συνδεδεμένο γιατί ο κάθε κόμβος σε όλα τα στρώματα του, συνδέεται με κάθε κόμβο του επόμενου. Αν λείπανε συνδέσεις επικοινωνίας τότε θα ονομαζόταν μερικώς συνδεδεμένο[17]. Μερικώς συνδεδεμένο είναι το συνελκτικό νευρωνικό δίκτυο που παρουσιάζεται στην ενότητα 1.2.

Να πούμε εδώ ότι υπάρχει και το νευρωνικό δίκτυο ανάδρασης που διαφέρει από τα νευρωνικά δίκτυα πρόσθιας τροφοδότησης στο γεγονός ότι έχει τουλάχιστον μία ανάδραση[18].

1.1.3 Backpropagation

Ο αλγόριθμος οπισθοδιάδοσης (backpropagation) είναι μια κοινή μέθοδος εκπαίδευσης ενός νευρωνικού δικτύου το οποίο χρησιμοποιείται σε συνδυασμό με μια μέθοδο βελτιστοποίησης. Η οπισθοδιάδοση απαιτεί να γνωρίζουμε εκ των προτέρων την επιθυμητή έξοδο για κάθε τιμή εισόδου προκειμένου να υπολογιστεί η συνάρτηση κόστους(error) [19].

Ο αλγόριθμος οπισθοδιάδοσης χωρίζεται σε δύο μέρη, διάδοση και βελτιστοποίηση. Η διάδοση περιλαμβάνει τη διαδικασία που οι τιμές εισόδου διαδίδονται προς τα μπροστά μέσω του δικτύου και παράγουν τις τιμές εξόδου. Στη συνέχεια υπολογίζεται το σφάλμα(κόστος) και τέλος διαδίδεται στην έξοδο προς τα πίσω μέσω του δικτύου χρησιμοποιώντας το στόχο του προτύπου εκπαίδευσης[20].

Στην ενότητα 1.1.4 παρουσιάζονται οι βασικές συναρτήσεις κόστους και η λογική πίσω από τη λειτουργία τους. Στο μέρος της βελτιστοποίησης αλλάζουν τα βάρη προκειμένου στην επόμενη διάδοση προς τα εμπρός να μειωθεί περαιτέρω το σφάλμα. Το ποσοστό της αλλαγής της τιμής των βαρών επηρεάζει τη ταχύτητα και τη ποιότητα της εκπαίδευσης. Εάν το ποσοστό είναι μεγάλο, τότε η εκπαίδευση του νευρωνικού δικτύου γίνεται πιο γρήγορα αλλά έχει ως επακόλουθο η ακρίβεια των αποτελεσμάτων να μειωθεί. Εάν το ποσοστό είναι μικρό, τότε αυξάνεται ο χρόνος εκπαίδευσης του νευρωνικού δικτύου, αλλά η ακρίβεια των αποτελεσμάτων είναι μεγαλύτερη [21].

1.1.4 Συναρτήσεις Ενεργοποίησης

Μετά από κάθε επίπεδο νευρώνων υπάρχει μια συνάρτηση ενεργοποίησης για κάθε νευρώνα το οποίο κανονίζει αν η έξοδος του νευρώνα θα περάσει ως είσοδος στο επόμενο επίπεδο και με ποια μορφή [22]. Στη συνέχεια παρουσιάζονται μερικές από τις σημαντικότερες συναρτήσεις ενεργοποίησης που υπάρχουν:

- Η συνάρτηση ενεργοποίησης δυαδικού βήματος (Binary Step) λαμβάνει ως έξοδο μόνο 2 τιμές, 0 και 1. Λειτουργεί ως ένα είδος κατωφλίου όπου όταν η είσοδος έχει αρνητική τιμή ($x < 0$) τότε η έξοδος είναι ίση με το 0 ($f(x)=0$). Εάν η είσοδος είναι 0 ή θετική ($x \geq 0$) τότε η έξοδος είναι ίση με το 1 ($f(x)=1$) [22].
- Η γραμμική συνάρτηση (Linear Function) έχει ως εξίσωση το $f(x)=a \cdot x$ όπου x η είσοδος, $f(x)$ η έξοδος και το a είναι σταθερά [22].
- Η σιγμοειδής συνάρτηση είναι μια μη γραμμική συνάρτηση ενεργοποίησης η οποία είναι ιδιαίτερα δημοφιλής. Έχει ως έξοδο τιμές οι οποίες κυμαίνονται από το 0 έως το 1. Η μαθηματική της μορφή είναι η εξής[23]:

$$f(x) = \frac{1}{1+e^{-x}} \quad (8)$$

- Η συνάρτηση tanh είναι παρόμοια με τη σιγμοειδή συνάρτηση. Είναι συμμετρική ως προς την αρχή των αξόνων και λαμβάνει τιμές εξόδου που κυμαίνονται από την τιμή -1 έως 1. Η tanh αποτυπώνεται μαθηματικά από την εξίσωση[23]:

$$f(x) = \frac{2}{1+e^{-2x}} \quad (9)$$

- Το ReLU (rectified linear activation function) είναι μια συνάρτηση ενεργοποίησης η οποία εάν λαμβάνει είσοδο που είναι θετική την εξάγει ως έχει ενώ σε κάθε άλλη περίπτωση εξάγει τον αριθμό 0. Το ReLU αποκαλείται συχνά φορέας ανίχνευσης (detector). Είναι δημοφιλής συνάρτηση ενεργοποίησης γιατί μπορεί να εκπαιδεύσει το νευρωνικό δίκτυο γρήγορα χωρίς ιδιαίτερα προβλήματα [24,25]. Αποτυπώνεται μαθηματικά από τη κάτωθι εξίσωση:

$$f(x)=x^+ = \max(0,x) \quad (10)$$

- Η Leaky ReLU συνάρτηση είναι ίδια με την απλή ReLU με τη διαφορά ότι όταν η έξοδος του νευρώνα είναι αρνητική τιμή, η έξοδος δεν είναι μηδενική όπως στη ReLU αλλά ένα γινόμενο μιας πολύ μικρής σταθεράς επί την τιμή της εισόδου. Αυτό συμβαίνει γιατί εάν η έξοδος από τη συνάρτηση ενεργοποίησης είναι 0 θα απενεργοποιούσε τους νευρώνες της περιοχής αυτής. Η συνάρτηση Leaky ReLU αποτυπώνεται μαθηματικά: [22]

$$\begin{aligned} f(x) &= 0.01x, \quad x < 0 \quad (11) \\ &= x, \quad x \geq 0 \end{aligned}$$

- Η παραμετροποιημένη ReLU συνάρτηση είναι ίδια με την απλή ReLU με τη διαφορά ότι όταν η έξοδος του νευρώνα είναι αρνητική τιμή, η έξοδος δεν είναι μηδενική όπως στη ReLU αλλά ένα γινόμενο μιας τιμής a επί την τιμή της εισόδου. Η παράμετρος a αλλάζει με βάση τη πορεία της εκπαίδευσης. Χρησιμοποιείται όταν η συνάρτηση Leaky ReLU αποτύχει στο στόχο της η οποία είναι να κρατήσει τους νευρώνες ενεργούς. Η παραμετροποιημένη ReLU αποτυπώνεται μαθηματικά [22]:

$$\begin{aligned} f(x) &= ax, \quad x < 0 \quad (12) \\ &= x, \quad x \geq 0 \end{aligned}$$

- Εκθετική γραμμική μονάδα (Exponential Linear Unit-ELU) αποτελεί και αυτή παραλλαγή της συνάρτησης ReLU. Σε αντίθεση με τη παραμετροποιημένη και Leaky ReLU όπου το αρνητικό μέρος της συνάρτησής είναι μια ευθεία γραμμή, η ELU

χρησιμοποιεί εκθετική καμπύλη για το ίδιο σκοπό. Αποτυπώνεται μαθηματικά από την ακόλουθη εξίσωση [22]:

$$f(x) = x, \quad x \geq 0 \quad (13)$$
$$= a(e^x - 1), \quad x < 0$$

- Η συνάρτηση Softmax είναι συνδυασμός πολλαπλών σιγμοειδών συναρτήσεων. Χρησιμοποιείται συνήθως σε προβλήματα ταξινόμησης πολλαπλώς κλάσεων. Η Softmax επιστρέφει την πιθανότητα για τα δεδομένα που ανήκουν σε συγκεκριμένη κλάση. Με δεδομένο ότι z_i είναι τα διανύσματα εισόδου, η μαθηματική διατύπωση της συνάρτησης Softmax είναι η εξής [26]:

$$\text{για } j=1, \dots, K. \quad (14)$$

1.1.5 Συναρτήσεις Απώλειας

Μετά την έξοδο από τη συνάρτηση ενεργοποίησης, εξετάζεται κατά πόσο η τιμή εξόδου απέχει από τη ζητούμενη τιμή έτσι ώστε να έχει το σύστημα μια εικόνα κατά πόσο απέχει από τη τιμή στόχο. Αυτό μπορεί να γίνει με τη συνάρτηση απώλειας, η οποία μετρά τη διαφορά μεταξύ της αναμενόμενης εξόδου στόχου t και της πραγματικής εξόδου y . Οι πιο γνωστές συναρτήσεις απώλειας παρουσιάζονται πιο κάτω:

Συνάρτηση τετραγωνικού σφάλματος ή κόστος είναι ίσως η πιο απλή αλλά ιδιαίτερα αποτελεσματική συνάρτηση απώλειας [27, 28]. Η εξίσωσή της είναι η ακόλουθη:

$$Cost = \sum_{i=1}^n (t(i) - y(i))^2 \quad (15)$$

Όπου το i αναφέρεται σε κάθε δείκτη του δείγματος που εξετάζεται. Το y στη πραγματική έξοδο και το t στη ζητούμενη τιμή στόχο. Επειδή η συνάρτηση θα παραγωγιστεί γράφεται επίσης ως:

$$Cost = \frac{1}{2} \sum_{i=1}^n (t(i) - y(i))^2 \quad (16)$$

Μια άλλη πολύ σημαντική συνάρτηση είναι εκείνη της απώλειας εντροπίας ή λογαρίθμου (logarithmic loss) η οποία μετρά τη διαφορά μεταξύ δύο κατανομών πιθανότητας [29]. Η κάτωθι συνάρτηση είναι η μαθηματική αποτύπωση της συνάρτησης εντροπίας:

$$L_{CE} = - \sum_{i=1}^n t_i \log(s_i) \quad (17)$$

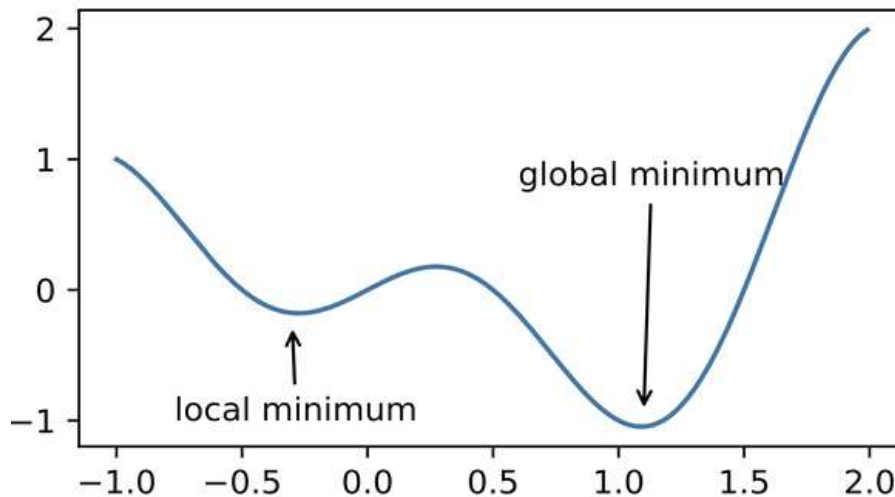
Όπου το s αντιπροσωπεύει τη κατανομή πιθανότητας που προβλέφθηκε από το μοντέλο. Το i είναι ο δείκτης κλάσης και το t παίρνει τη τιμή 1 για σωστή πρόβλεψη ή 0 για λάθος πρόβλεψη.

1.1.6 Βελτιστοποίηση

Σκοπός της διαδικασίας εκπαίδευσης είναι η εύρεση του ολικού ελαχίστου της συνάρτησης απώλειας. Δηλαδή, το σημείο εκείνο στη συνάρτηση απώλειας όπου η τιμή της είναι η μικρότερη δυνατή. Στη συνέχεια, γίνεται η οπισθοδιάδοση όπου υπολογίζεται η κλίση της συνάρτησης κόστους. Η κλίση της συνάρτησης κόστους είναι η μερική παράγωγος της συνάρτησης απώλειας προς τις παραμέτρους (gradient) του νευρωνικού δικτύου. Η κλίση συνάρτησης κόστους πολλαπλασιάζεται με το ρυθμό μάθησης που είναι μια πολύ μικρή τιμή που έχει κανονιστεί πριν από την εκκίνηση της εκπαίδευσης του νευρωνικού δικτύου. Εάν η τιμή του ρυθμού μάθησης είναι πολύ μεγάλη τότε ο χρόνος εκπαίδευσης μπορεί να μειωθεί αλλά υπάρχει η πιθανότητα να μην παρουσιάσει καλά αποτελέσματα, ενώ, εάν έχει πολύ μικρή τιμή τότε απαιτείται περισσότερος χρόνος για την εκπαίδευση. Πρέπει να τονιστεί εδώ ότι ακόμα και αν ο ρυθμός μάθησης είναι μικρός δεν σημαίνει απαραίτητα ότι θα πλησιάσει το ολικό ελάχιστο αφού μπορεί να συγκλίνει σε κάποιο τοπικό ελάχιστο της συνάρτησης

Μεταπτυχιακή Διπλωματική Εργασία, Όλσους Πάντος, AM msciot18005

απώλειας όπως φαίνεται στην εικόνα 6. Στη συνέχεια, το αποτέλεσμα του πολλαπλασιασμού δίνεται στο σύστημα βελτιστοποίησης το οποίο με τη σειρά του αλλάζει τα βάρη έτσι ώστε η συνάρτηση απώλειας να μειωθεί [30,31].



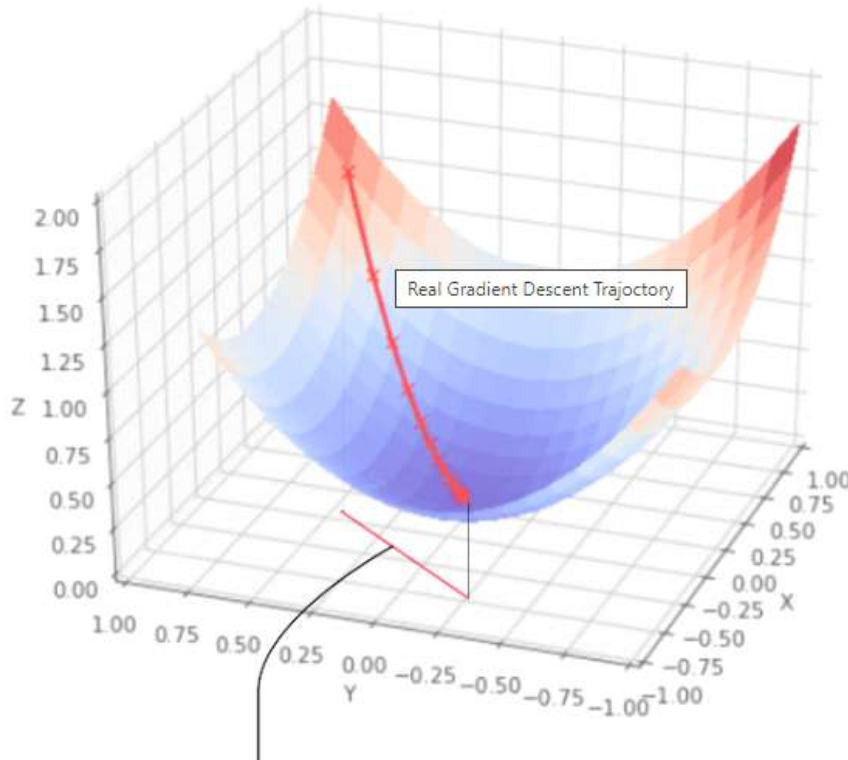
Εικόνα 6: Παρουσιάζεται συνάρτηση όπου δείχνει τη διαφορά μεταξύ τοπικού και ολικού ελαχίστου [32].

Πιο κάτω παρουσιάζονται μερικοί από τους σημαντικότερους αλγόριθμους βελτιστοποίησης που υπάρχουν:

- Ο αλγόριθμος βαθμωτής κατάβασης(Gradient Descent) είναι ο πιο συνηθισμένος αλγόριθμος βελτιστοποίησης από το οποίο έχουν αναπτυχθεί πολλές παραλλαγές. Κάτωθι αποτυπώνεται η μαθηματική του μορφή[31]:

$$a_{n+1} = a_n - \eta \nabla F(a_n) \quad (18)$$

όπου a_n είναι η παράμετρος του νευρωνικού δικτύου πριν την οπισθοδιάδοση, η ο ρυθμός εκπαίδευσης του δικτύου και $F(a_n)$ η συνάρτηση κόστους. Τα προτερήματα της είναι το γεγονός ότι είναι εύκολο να υπολογιστεί και να εφαρμοστεί καθώς επίσης είναι εύκολα κατανοητή σε σχέση με άλλους αλγόριθμους. Τα μειονεκτήματα της είναι το γεγονός ότι δεν μπορεί να ξεχωρίσει ένα τοπικό ελάχιστο από το ολικό με αποτέλεσμα να υπάρχει η περίπτωση το τελικό αποτέλεσμα να μην είναι το βέλτιστο. Επίσης, αν η βάση δεδομένων είναι μεγάλη, τότε χρειάζεται πολύ μνήμη αλλά και χρόνο προκειμένου να ολοκληρωθεί η διαδικασία της εκπαίδευσης.



Εικόνα 7: Παρουσιάζεται πως η συνάρτηση βαθμωτής κατάβασης πλησιάζει προς το ολικό ελάχιστο σταδιακά [33].

- Ο αλγόριθμος στοχαστικής βαθμωτής κατάβασης (Stochastic Gradient Descent) είναι παραλλαγή του αλγορίθμου βαθμωτής κατάβασης. Αντί να χρησιμοποιήσει την παράμετρο που δίνεται από το δίκτυο, χρησιμοποιεί μια εκτίμηση το οποίο παρήχθη από ένα τυχαίο υποσύνολο της βάσης δεδομένων. Έχει τη δυνατότητα να συγκλίνει πιο γρήγορα σε ελάχιστο αν και παρουσιάζει θόρυβο. Επίσης, παρουσιάζει το ίδιο πρόβλημα όπως στη βαθμωτή κατάβαση η οποία είναι η ικανότητα εντοπισμού του ολικού ελάχιστου. Τέλος, χρησιμοποιεί λιγότερη μνήμη[34].
- Ο αλγόριθμος στοχαστικής βαθμωτής κατάβασης με ορμή (Stochastic Gradient Descent with Momentum) είναι παραλλαγή του αλγορίθμου στοχαστικής βαθμωτής κατάβασης. Οι παράμετροι του δικτύου είναι λιγότερο θορυβώδη σε σχέση με τη στοχαστική πράγμα που οδηγεί σε μεγαλύτερες ταχύτητες εκπαίδευσης. Η συνάρτησή της χρησιμοποιεί μια επιπλέον υπερπαράμετρο σε σχέση με την απλή στοχαστική η οποία ονομάζεται Ορμή (Momentum) και συμβολίζεται με το γ . Κάτω παρουσιάζεται η συνάρτηση του αλγορίθμου στοχαστικής βαθμωτής κατάβασης με ορμή[35]:

$$a_{n+1} = a_n - \eta \nabla F(a_n) + \gamma \Delta(a_n) \quad (19)$$

- Η AdaGrad (Adaptive Gradient Algorithm) αποτελεί τροποποίηση του αλγορίθμου στοχαστικής βαθμωτής κατάβασης. Έχει το χαρακτηριστικό ότι έχει διαφορετικό ρυθμό εκπαίδευσης για τη κάθε παράμετρο. Όταν οι παράμετροι είναι περισσότερο σποραδικές τότε αυξάνεται ο ρυθμός εκμάθησης. Όταν οι παράμετροι είναι λιγότερο σποραδικές τότε ο ρυθμός εκπαίδευσης μειώνεται. Αυτό έχει ως επακόλουθο να

βελτιωθεί η επίδοση του μοντέλου σε σχέση με τη στοχαστική βαθμωτή κατάβαση. Η συνάρτησή της εξακολουθεί να έχει ένα αρχικό ρυθμό εκπαίδευσης η [31]. Η συνάρτηση της AdaGrad αποτυπώνεται κάτωθι :

$$\alpha_{t+1} = \alpha_t - \frac{\eta}{\sqrt{G_t + \epsilon}} g_t \quad (20)$$

όπου G_t είναι το άθροισμα των τετραγώνων των προηγούμενων παραμέτρων του δικτύου όπως παρουσιάζεται πιο κάτω:

$$G_t = \sum_{t=1}^T g^2(t) \quad (21)$$

το ϵ είναι μια σταθερά που έχει πολύ μικρή τιμή και σκοπός της είναι ο παρονομαστής να μην λαμβάνει ποτέ τη τιμή 0. Το g_t είναι η παράγωγος για συγκεκριμένες παραμέτρους τη συγκεκριμένη στιγμή t :

$$g_t = \nabla F(a_n) \quad (22)$$

- Η RMSProp (Root Mean Square Propagation) είναι μέθοδος που όπως και με τη AdaGrad ο ρυθμός εκπαίδευσης αλλάζει για κάθε παράμετρο[36]. Η συνάρτηση που υπολογίζει τις παραμέτρους είναι:

$$\alpha_{t+1} = \alpha_t - \frac{\eta}{\sqrt{E[g^2]_t + \epsilon}} g_t \quad (23)$$

όπου το β συνήθως παίρνει τη τιμή 0.9. Το $E[g^2]_t$ βγαίνει από την εξίσωση:

$$E[g^2]_t = E[g^2]_{t-1} + 0.1g^2_t \quad (24)$$

- Η Adam(Adaptive Moment Estimation) είναι μέθοδος η οποία υπολογίζει το προσαρμοσμένο ρυθμό μάθησης για κάθε παράμετρο. Ο τύπος της είναι[126]:

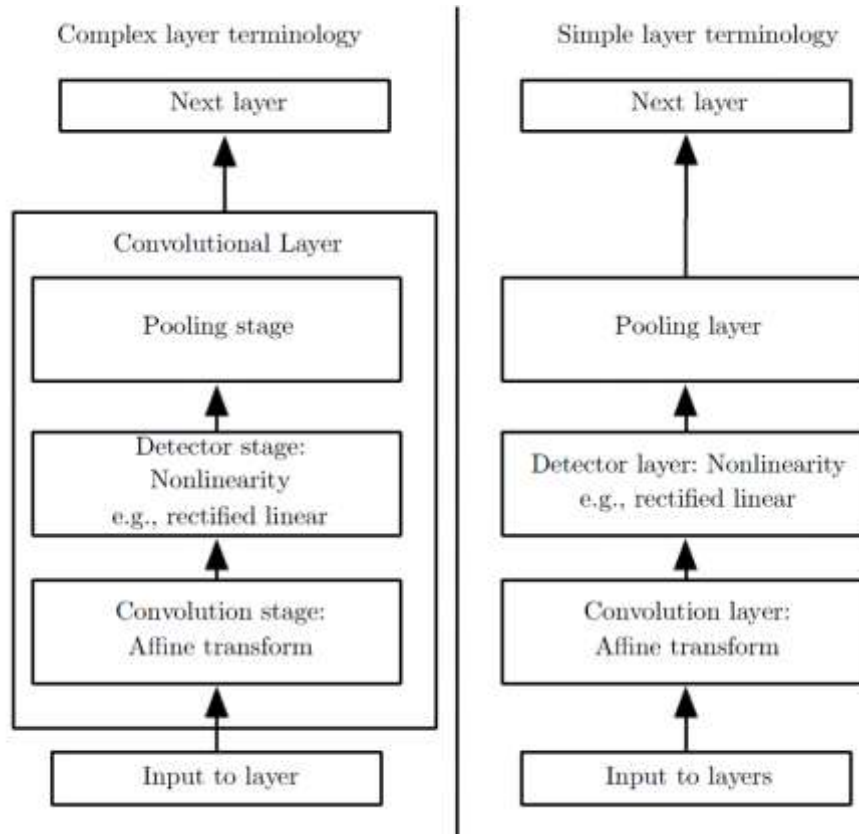
$$\alpha_{t+1} = \alpha_t - \frac{\eta}{\sqrt{v_t + \epsilon}} m_t \quad (25)$$

όπου m είναι ο μέσος όρος και v η διακύμανση.

1.2 Συνελκτικά Νευρωνικά Δίκτυα

Τα Συνελκτικά νευρωνικά δίκτυα, γνωστά στη διεθνή βιβλιογραφία ως Convolutional Neural Networks (CNN), είναι ένα εξειδικευμένο είδος νευρωνικού δικτύου για την επεξεργασία δεδομένων με τοπολογία μορφής πλέγματος. Χρησιμοποιούνται κυρίως για την ταξινόμηση εικόνων και την αναγνώριση αντικειμένων από εικόνες που δίνονται για ανάλυση. Παραδείγματα εφαρμογών με CNN είναι συστήματα αναγνώρισης προσώπων, οδών, όγκων και πολλών άλλων οπτικών δεδομένων (visual data). Η αποτελεσματικότητα των συνελκτικών νευρωνικών δικτύων (CNN) στην αναγνώριση εικόνων είναι ένας από τους κύριους λόγους για τους οποίους ο κόσμος έχει στραφεί στην μελέτη της βαθιάς μάθησης. Τα CNN τροφοδοτούν μείζονες εξελίξεις στην υπολογιστική όραση (Computer Vision=CV), η οποία έχει εφαρμογές στο πεδίο των αυτοκινούμενων οχημάτων, ρομποτικής, αεροσκαφών, ασφάλειας καθώς επίσης ιατρικές διαγνώσεις και θεραπείες για άτομα με προβλήματα όρασης. Τα συνελκτικά νευρωνικά δίκτυα μπορούν επίσης να εκτελούν πιο κοινότερες επιχειρηματικές δραστηριότητες όπως η οπτική αναγνώριση χαρακτήρων (OCR) για την ψηφιοποίηση κειμένου και την επεξεργασία φυσικής γλώσσας(Natural Language Processing = NLP) σε αναλογικά και χειρόγραφα έγγραφα, όπου οι εικόνες είναι σύμβολα προς μεταγραφή. Ωστόσο, τα CNN δεν περιορίζονται στην αναγνώριση εικόνων μόνο. Έχουν εφαρμοστεί απευθείας σε αναλυτικά κείμενα. Εφαρμόζονται επίσης στον ήχο όταν παρουσιάζεται οπτικά ως φασματογράφημα καθώς επίσης και για δεδομένα γραφημάτων με συνελκτικά δίκτυα[37].

Ένα απλό συνελκτικό νευρωνικό δίκτυο συνήθως αποτελείται από ένα επίπεδο εισόδου, από πολλαπλά επίπεδα νευρώνων όπου γίνεται η συνέλιξη, πολλαπλά επίπεδα νευρώνων ReLU, πολλαπλά επίπεδα νευρώνων ομαδοποίησης (pooling) και στο τέλος ένα επίπεδο νευρώνων πλήρως συνδεδεμένο. Η αρχιτεκτονική που ακολουθεί είναι το επίπεδο συνέλιξης, ακολουθούμενο από επίπεδο ReLU που με τη σειρά του ακολουθεί το επίπεδο ομαδοποίησης, επαναλαμβανόμενα συνέχεια έως ότου στο τέλος καταλήγουν στο τελευταίο επίπεδο με τους πλήρως συνδεδεμένους νευρώνες[38]. Η εικόνα 8 παρουσιάζει γραφικά ένα τυπικό επίπεδο συνελκτικού νευρωνικού δικτύου.



Εικόνα 8: Τυπικό επίπεδο συνελκτικού δικτύου [18].

1.2.1 Αντίληψη Εικόνων από τα Συνελκτικά Νευρωνικά Δίκτυα

Τα συνελκτικά νευρωνικά δίκτυα απορροφούν και επεξεργάζονται εικόνες ως τανυστές. Οι τανυστές είναι πίνακες αριθμών με πρόσθετες διαστάσεις.

Τα συνελκτικά νευρωνικά δίκτυα αντιλαμβάνονται εικόνες ως όγκους, δηλαδή ως τρισδιάστατα αντικείμενα και δεν βλέπουν μόνο το ύψος και το πλάτος των εικόνων. Αυτό οφείλεται στο γεγονός ότι οι ψηφιακές έγχρωμες εικόνες έχουν μια κόκκινη-μπλε-πράσινη (RGB) κωδικοποίηση. Αυτά τα τρία χρώματα αναμειγνύονται για να παράγουν το φάσμα χρωμάτων που μπορεί να γίνει αντιληπτό από τους ανθρώπους. Ένα συνελκτικό δίκτυο καταλαβαίνει τέτοιες εικόνες ως τρία ξεχωριστά στρώματα χρώματος στοιβαγμένα το ένα πάνω στο άλλο.

Επομένως, ένα συνελκτικό δίκτυο λαμβάνει μια κανονική έγχρωμη εικόνα ως ένα ορθογώνιο παραλληλόγραμμο του οποίου το πλάτος και το ύψος μετρούνται με τον αριθμό των εικονοστοιχείων (pixels) κατά μήκος αυτών των διαστάσεων και του οποίου το βάθος είναι τριών επιπέδων, ένα για κάθε χρώμα του RGB. Αυτά τα επίπεδα βάθους ονομάζονται κανάλια (channels) [39].

1.2.2 Λειτουργία Συνέλιξης

Για την λειτουργία του, το Συνελικτικό Νευρωνικό Δίκτυο ακολουθεί τη μαθηματική πράξη της Συνέλιξης (Convolution). Από τα μαθηματικά, η συνέλιξη είναι μια μαθηματική λειτουργία με την οποία δύο συναρτήσεις (x και w) παράγουν μια τρίτη συνάρτηση ($x*w$) η οποία εκφράζει πως το σχήμα της μίας συνάρτησης μετασχηματίζεται από την άλλη. Ο όρος Συνέλιξη αναφέρεται στη τελική συνάρτηση καθώς επίσης και στη διαδικασία υπολογισμού της. Ορίζεται ως το ολοκλήρωμα του αποτελέσματος των δύο συναρτήσεων μετά την αναστροφή και μετατόπιση της μίας εξ αυτών πάνω στην άλλη. Η συνάρτηση η οποία αναστρέφεται είναι η w στη συγκεκριμένη περίπτωση ενώ η συνάρτηση πάνω στη οποία θα ολισθαίνει η w είναι η x . Το αποτέλεσμα για κάθε μετατόπιση παράγει τη συνάρτηση της συνέλιξης. Η Συνέλιξη έχει εφαρμογές σε πεδία της επιστήμης όπως εκείνες των πιθανοτήτων και στατιστικής, ακουστικής, φασματοσκοπίας, επεξεργασία σήματος και εικόνας, μηχανικής, φυσικής, όρασης υπολογιστών και διαφορικές εξισώσεις. Η συνάρτηση της Συνέλιξης αποτυπώνεται μαθηματικά από την εξίσωση 26.

$$s(t) = \int x(a)w(t-a)da \quad (26)$$

Συνήθως η πράξη της συνέλιξης συμβολίζεται με το σύμβολο $*$ όπως δείχνει η συνάρτηση (27):

$$s(t) = (x * w)(t) \quad (27)$$

Το σύμβολο t που χρησιμοποιείται παραπάνω, δεν αντιπροσωπεύει απαραίτητα το χρόνο. Σε αυτή τη περίπτωση όμως, ο τύπος της συνέλιξης θα μπορούσε να περιγραφεί ως ο σταθμισμένος μέσος όρος της συνάρτησης $x(a)$ την χρονική στιγμή t όπου η στάθμιση δίνεται από $w(-a)$ μετατοπιζόμενο κατά t . Όσο το t αλλάζει, η σταθμισμένη συνάρτηση δίνει έμφαση σε διαφορετικά μέρη της συνάρτησης εισόδου.

Στην ορολογία των Συνελικτικών Νευρωνικών Δικτύων, το πρώτο όρισμα (σε αυτό το παράδειγμα, η συνάρτηση x) στην συνέλιξη αναφέρεται συχνά ως είσοδος και το δεύτερο όρισμα (σε αυτό το παράδειγμα, η συνάρτηση w) ως kernel ή φίλτρο. Η έξοδος αναφέρεται συνήθως ως ο χάρτης χαρακτηριστικών (feature map)[40].

Στις εφαρμογές μηχανικής μάθησης, η είσοδος είναι συνήθως ένας πολυδιάστατος πίνακας δεδομένων και το φίλτρο που λέγεται επίσης και πυρήνας είναι συνήθως ένας πολυδιάστατος πίνακας παραμέτρων που προσαρμόζονται από τον αλγόριθμο εκμάθησης. Αυτοί οι πολυδιάστατοι πίνακες όταν είναι τριών ή περισσότερων διαστάσεων θα αναφέρονται ως τανυστές (tensors).

Η εξίσωση 26 δείχνει το τύπο του μηχανισμού της συνέλιξης για συνεχής συναρτήσεις. Για διακριτές συναρτήσεις ο τύπος της συνέλιξης ορίζεται ως:

$$s(t) = (x * w)(t) = \sum_{a=-\infty}^{\infty} x(a)w(t-a) \quad (28)$$

Εάν χρησιμοποιήσουμε ένα διδιάστατο χάρτη χαρακτηριστικών I ως είσοδο μας και χρησιμοποιήσουμε έναν διδιάστατο πυρήνα K τότε ο τύπος της συνέλιξης έχει ως εξής:

$$S(i, j) = (I * K)(i, j) = \sum_m \sum_n I(m, n)K(i-m, j-n) \quad (29)$$

Μέσω της αντιμεταθετικής ιδιότητας η εξίσωση (29) μπορεί να διατυπωθεί ισοδύναμα ως:

$$S(i, j) = (I * K)(i, j) = \sum_m \sum_n I(i-m, j-n)K(m, n) \quad (30)$$

Συνήθως ο τύπος (30) είναι πιο εύκολος να εφαρμοστεί σε μια βιβλιοθήκη μηχανικής μάθησης, επειδή υπάρχουν λιγότερες διακυμάνσεις στο εύρος των έγκυρων τιμών των m και n .

Πρέπει να τονιστεί εδώ ότι η συνέλιξη έχει τις ίδιες ιδιότητες με το πολλαπλασιασμό. Πιο συγκεκριμένα την αντιμεταθετική, την προσεταιριστική και επιμεριστική ιδιότητα[41, 42].

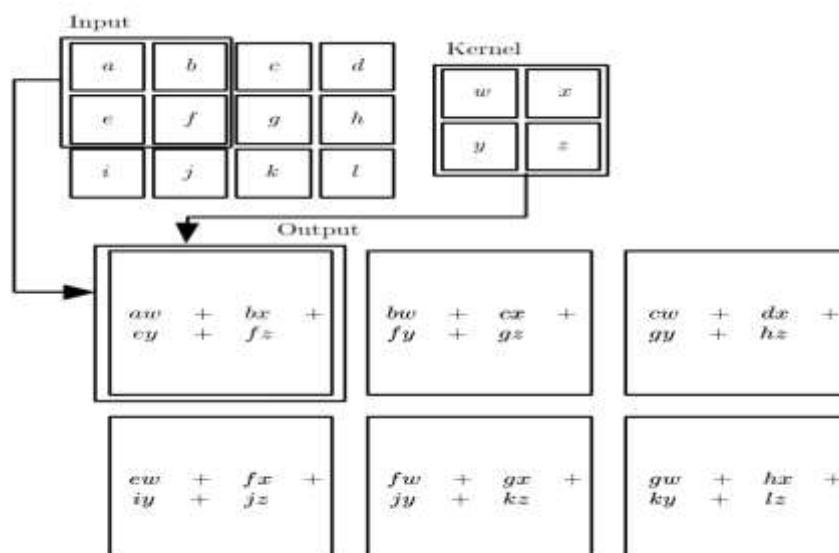
Η αντιμεταθετική ιδιότητα της συνέλιξης προκύπτει επειδή έχουμε αναστροφή του πυρήνα σε σχέση με την είσοδο, με την έννοια ότι όσο το m αυξάνεται, ο δείκτης στην είσοδο αυξάνεται, αλλά ο δείκτης στον πυρήνα(kernel) μειώνεται. Ο μόνος λόγος για την αντιστροφή στον πυρήνα είναι να αποκτηθεί η αντιμεταθετική ιδιότητα. Ενώ η αντιμεταθετική ιδιότητα είναι χρήσιμη για τη σύνταξη μαθηματικών αποδείξεων, δεν είναι σημαντική για την υλοποίηση ενός νευρωνικού δικτύου. Αντ' αυτού, πολλές βιβλιοθήκες νευρωνικών δικτύων εφαρμόζουν μια σχετική συνάρτηση που ονομάζεται συνάρτηση αλληλοσυσχέτισης (cross-correlation function), η οποία είναι η ίδια με την συνέλιξη αλλά χωρίς να αντιστρέφει τον πυρήνα:

$$S(i, j) = (I * K)(i, j) = \sum_m \sum_n I(i + m, j + n)K(m, n) \quad (31)$$

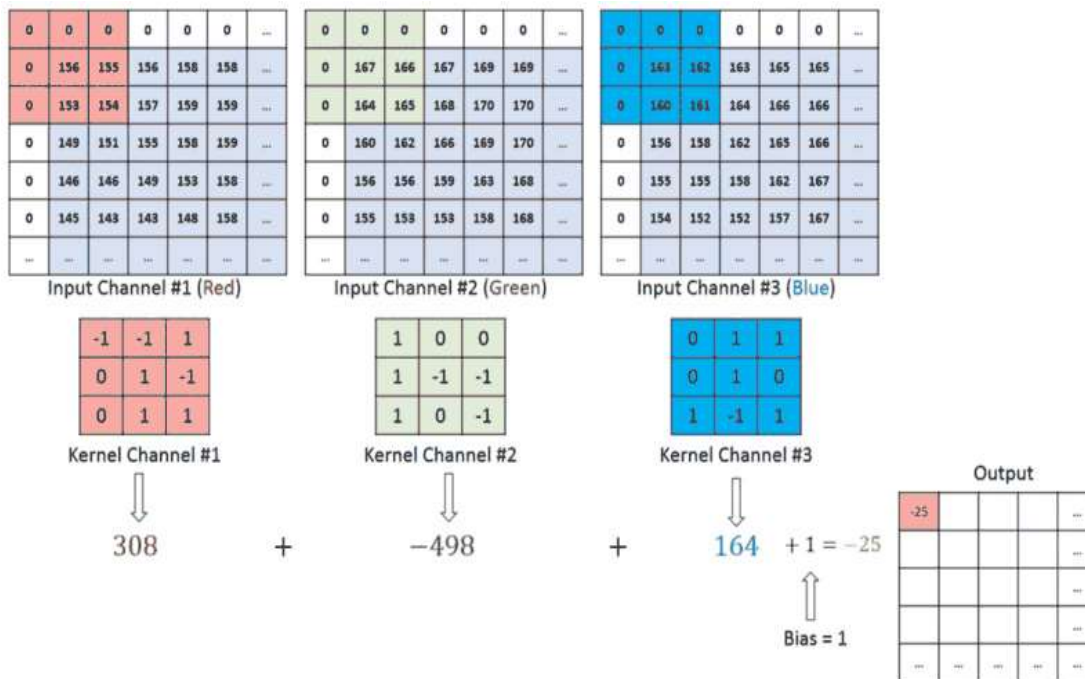
Πολλές βιβλιοθήκες μηχανικής μάθησης εφαρμόζουν αλληλοσυσχέτιση, αλλά το ονομάζουν συνέλιξη.

Το σχήμα της εικόνας 9 δείχνει ένα παράδειγμα συνέλιξης (χωρίς αναστροφή του πυρήνα) που εφαρμόζεται σε ένα 2-D τανυστή.

Η διακριτή(discrete) συνέλιξη μπορεί να θεωρηθεί ως πολλαπλασιασμός με πίνακα. Ωστόσο, ο πίνακας έχει αρκετές καταχωρήσεις(entries) που έχουν περιορισμό ώστε να είναι ίσες με άλλες καταχωρήσεις. Η συνέλιξη συνήθως αντιστοιχεί σε πολύ σποραδικούς(sparse) πίνακες, δηλαδή πίνακες των οποίων οι καταχωρήσεις είναι ως επί τω πλείστων ίσες με το μηδέν. Αυτό συμβαίνει επειδή ο πυρήνας είναι συνήθως πολύ μικρότερος από τον χάρτη χαρακτηριστικών που έχει εισέλθει για τη συνέλιξη. Κάθε αλγόριθμος νευρωνικού δικτύου που λειτουργεί με πολλαπλασιασμό πινάκων και δεν εξαρτάται από συγκεκριμένες ιδιότητες της δομής τους, μπορούν να λειτουργήσουν με συνέλιξη χωρίς να απαιτούνται περαιτέρω αλλαγές στο νευρωνικό δίκτυο. Τυπικά συνελκτικά νευρωνικά δίκτυα κάνουν χρήση περαιτέρω εξειδικεύσεων προκειμένου να αντιμετωπιστούν αποτελεσματικά οι μεγάλες εισροές εισόδων, αλλά αυτές δεν είναι απολύτως αναγκαίες από θεωρητική άποψη[43].



Εικόνα 9: Ένα παράδειγμα 2-D συνέλιξης χωρίς αναστροφή πυρήνα. Δείχνει πως σχηματίζεται το άνω αριστερά στοιχείο του τανυστή εξόδου εφαρμόζοντας τον πυρήνα στην αντίστοιχη άνω-αριστερά τανυστή εισόδου [40].



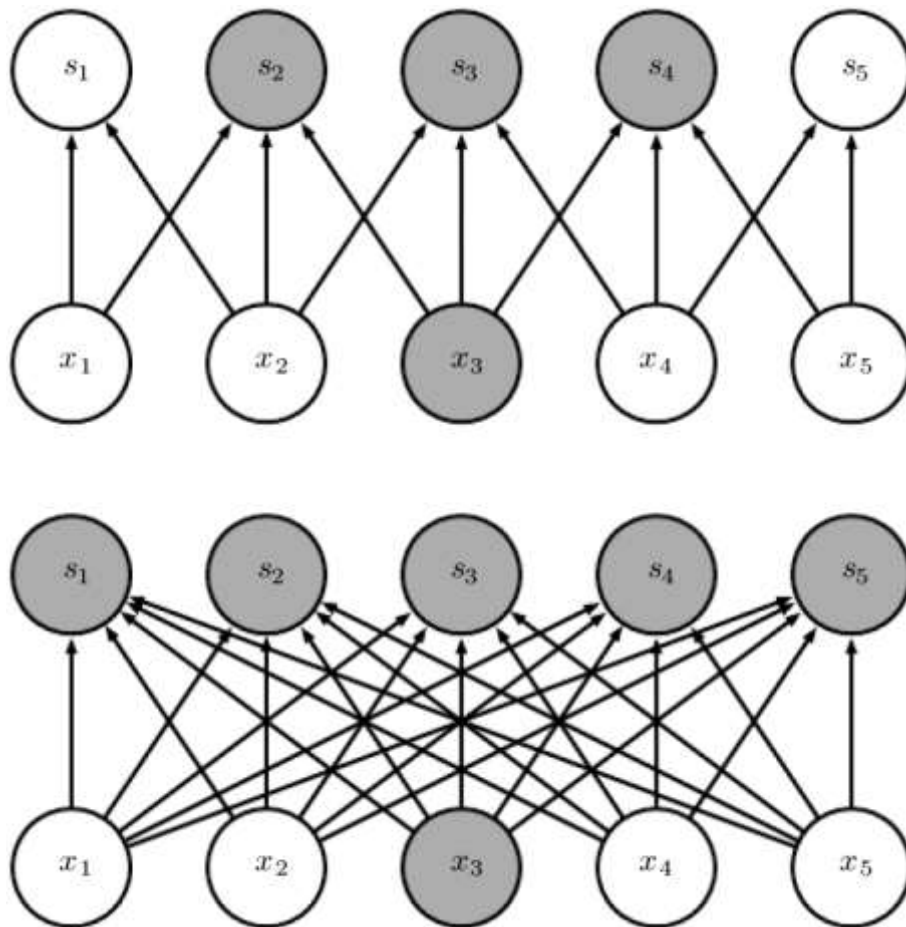
Εικόνα 10: Συνέλιξη με 3-D ταυυστή[41].

1.2.3 Βασικά Πλεονεκτήματα Συνελικτικών Νευρωνικών Δικτύων

Η συνέλιξη αξιοποιεί τρεις σημαντικές ιδιότητες που μπορούν να βοηθήσουν στη βελτίωση ενός συστήματος μηχανικής μάθησης: σποραδικές αλληλεπιδράσεις (sparse interactions), κοινή χρήση παραμέτρων (parameter sharing) και ισοδύναμες αναπαραστάσεις (equivariant representations). Επιπλέον, η συνέλιξη παρέχει ένα μέσο για εργασία με εισόδους μεταβλητού μεγέθους. Θα περιγραφεί στη συνέχεια κάθε μια από αυτές τις ιδιότητες με τη σειρά[44].

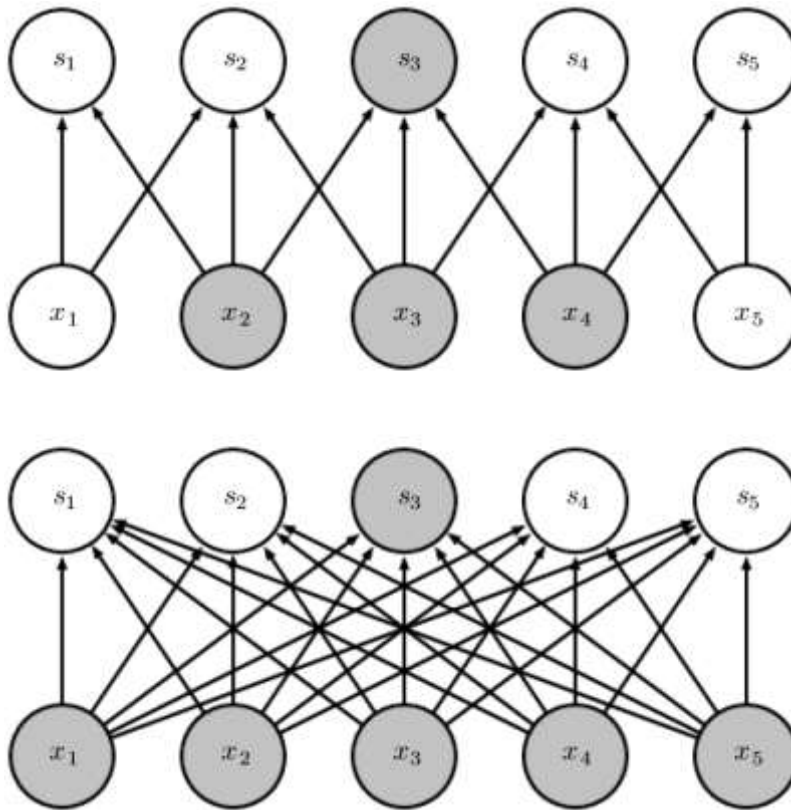
Σποραδικές Αλληλεπιδράσεις: Σε μια κλασική στοιβάδα ενός νευρωνικού δικτύου ο πίνακας εισόδων πολλαπλασιάζεται με ένα πίνακα παραμέτρων (βάρη) και παράγουν εξόδους που περιγράφουν την αλληλεπίδραση μεταξύ κάθε εισόδου και κάθε εξόδου. Αυτό σημαίνει ότι κάθε έξοδος αλληλοεπιδρά με κάθε είσοδο. Τα συνελικτικά δίκτυα, ωστόσο, έχουν συνήθως πιο σποραδικές αλληλεπιδράσεις (sparse interaction or sparse connectivity or sparse weights). Αυτό επιτυγχάνεται κάνοντας τον πυρήνα (kernel) μικρότερο από την είσοδο. Για παράδειγμα, κατά την επεξεργασία μιας εικόνας η εικόνα που έχει εισαχθεί, μπορεί να έχει χιλιάδες ή εκατομμύρια εικονοστοιχεία, αλλά μπορούμε να εντοπίσουμε μικρά, σημαντικά χαρακτηριστικά, όπως άκρες που έχουν εμβαδό μερικών δεκάδων εικονοστοιχείων (pixels). Αυτό μας επιτρέπει να αποθηκεύσουμε λιγότερες παραμέτρους, οι οποίες μειώνουν τις απαιτήσεις σε μνήμη και βελτιώνουν τη στατιστική τους απόδοση. Σημαίνει επίσης ότι ο υπολογισμός της εξόδου απαιτεί λιγότερες πράξεις. Αυτές οι βελτιώσεις στην απόδοση είναι συνήθως αρκετά μεγάλες. Εάν υπάρχουν m εισοδοί και n εξοδοί, τότε ο πολλαπλασιασμός πινάκων παράγει $m \times n$ παραμέτρους και οι αλγόριθμοι που χρησιμοποιούνται στην πράξη έχουν χρόνο εκτέλεσης $O(m \times n)$ (ανά παράδειγμα). Εάν περιορίσουμε τον αριθμό των συνδέσεων που μπορεί να έχει κάθε έξοδος σε k όπως μας επιτρέπει η μέθοδος των συνελικτικών νευρωνικών δικτύων, οι παράμετροι που χρειάζονται μειώνονται σε $k \times n$ και χρόνο εκτέλεσης σε $O(k \times n)$. Το αποτέλεσμα είναι ότι μπορούμε να χρησιμοποιήσουμε πολύ μικρότερο αριθμό συνδέσεων και να έχουμε καλά αποτελέσματα στη μηχανική μάθηση. Σε ένα βαθύ συνελικτικό νευρωνικό δίκτυο, οι μονάδες στα βαθύτερα στρώματα μπορούν έμμεσα να αλληλοεπιδράσουν με ένα μεγαλύτερο μέρος της εισόδου,

όπως φαίνεται στο σχήμα 13, αυτό με τη σειρά του επιτρέπει στο δίκτυο να περιγράψει αποτελεσματικά περίπλοκες αλληλεπιδράσεις[45].

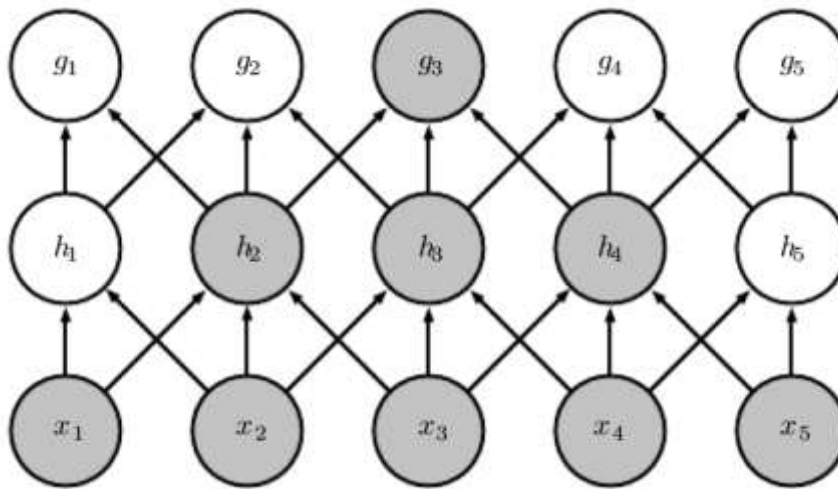


Εικόνα 11: Η μορφή της σποραδική συνδεσιμότητα πάνω και πλήρης κάτω[41].

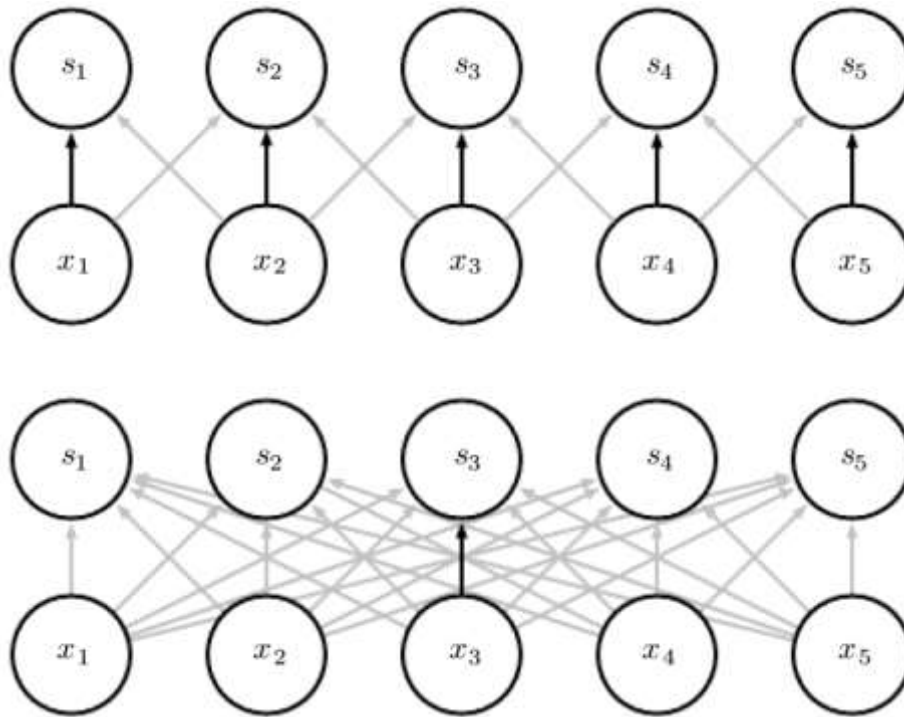
Κοινή Χρήση Παραμέτρων: Η κοινή χρήση παραμέτρων (Parameter Sharing) αναφέρεται στη χρήση της ίδιας παραμέτρου για περισσότερες από μία λειτουργίες σε ένα μοντέλο. Σε ένα παραδοσιακό νευρωνικό δίκτυο, κάθε στοιχείο του πίνακα με τα βάρη χρησιμοποιείται ακριβώς μία φορά κατά τον υπολογισμό της εξόδου ενός στρώματος. Πολλαπλασιάζεται με ένα στοιχείο της εισόδου και στη συνέχεια δεν επανεξετάζεται. Ως συνώνυμο για την κοινή χρήση παραμέτρων, μπορεί κανείς να πει ότι ένα δίκτυο έχει δεσμευμένα βάρη (tied weights), επειδή η τιμή του βάρους που εφαρμόζεται σε μία είσοδο δεσμεύεται με την τιμή ενός βάρους που εφαρμόζεται αλλού. Σε ένα συνελκτικό νευρωνικό δίκτυο, κάθε μέλος του πυρήνα χρησιμοποιείται για κάθε θέση της εισόδου (εκτός ίσως από τα οριακά εικονοστοιχεία, ανάλογα με τις σχεδιαστικές αποφάσεις σχετικά με το όριο). Δεν επηρεάζει τον χρόνο εκτέλεσης της πρόσθιας διάδοσης αφού εξακολουθεί να είναι $O(k \times n)$, ωστόσο, μειώνει ακόμα περισσότερο τις αποθηκευτικές απαιτήσεις του μοντέλου σε αριθμό k παραμέτρων που είναι πολύ μικρότερο από τις m παραμέτρους του κλασσικού νευρωνικού δικτύου. Δεδομένου ότι τα m και n έχουν συνήθως περίπου το ίδιο μέγεθος, το k είναι ουσιαστικά ασήμαντο σε σχέση με το $m \times n$. Συνεπώς, η συνέλιξη είναι δραματικά πιο αποτελεσματική από τον πολλαπλασιασμό πινάκων όπως είναι στα κλασσικά νευρωνικά δίκτυα όσον αφορά τις απαιτήσεις μνήμης και τη στατιστική απόδοση[46].



Εικόνα 12: Σποραδική συνδεσιμότητα πάνω και πλήρης κάτω[40]. Επισημαίνονται οι συνδέσεις ως προς την μονάδα εξόδου s_3 .



Εικόνα 13: Το δεκτικό πεδίο των μονάδων στα “βαθύτερα” στρώματα ενός συνελκτικού νευρωνικού δικτύου. Οι μονάδες στα βαθύτερα στρώματα συνδέονται έμμεσα με το σύνολο ή το μεγαλύτερο μέρος της εικόνας εισόδου[40].



Εικόνα 14: Κοινή χρήση παραμέτρων: Πάνω παρουσιάζεται ένα δίκτυο μερικής συνδεσιμότητας και κάτω πλήρους συνδεσιμότητας [40].

Ως παράδειγμα των δύο πρώτων αρχών στη πράξη, το σχήμα 15 δείχνει πώς η σποραδική συνδεσιμότητα και η κοινή χρήση παραμέτρων μπορούν να βελτιώσουν δραματικά την αποτελεσματικότητα μιας γραμμικής συνάρτησης για την ανίχνευση άκρων σε μια εικόνα.



Εικόνα 15: Αποτελεσματικότητα ανίχνευσης άκρων (Efficiency of edge detection)[40].

Ισοδύναμες Αναπαραστάσεις: Στην περίπτωση της συνέλιξης, η συγκεκριμένη μορφή κοινής χρήσης παραμέτρων προκαλεί στο επίπεδο να έχει μια ιδιότητα που ονομάζεται ισοδυναμία (equivariance) με το μηχανισμό μετατροπής. Το να πούμε ότι μια συνάρτηση είναι ισοδύναμη σημαίνει ότι εάν αλλάξει η είσοδος, η έξοδος αλλάζει με τον ίδιο τρόπο. Συγκεκριμένα, μια συνάρτηση $f(x)$ ισοδυναμεί με μια συνάρτηση g εάν $f(g(x)) = g(f(x))$. Στην περίπτωση της συνέλιξης, εάν αφήσουμε το g να είναι οποιαδήποτε συνάρτηση που μεταφράζει την είσοδο, δηλαδή την αλλάζει, τότε η συνάρτηση της συνέλιξης είναι ισοδύναμη με το g [40].

Η Συνέλιξη δεν είναι φυσικό ισοδύναμο για κάποιους άλλους μετασχηματισμούς, όπως αλλαγές στην κλίμακα ή περιστροφή μιας εικόνας. Για αυτό το σκοπό υπάρχουν άλλοι μηχανισμοί. Τέλος, ορισμένα είδη δεδομένων δεν μπορούν να υποβληθούν σε επεξεργασία από νευρωνικά δίκτυα[40].

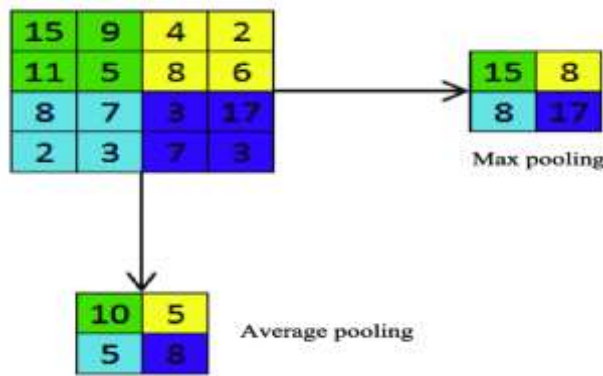
1.2.4 Ομαδοποίηση

Όπως και με το επίπεδο της συνέλιξης (Convolutional Layer), το επίπεδο ομαδοποίησης μειώνει την επιφάνεια του αντικειμένου που γίνεται η συνέλιξη (Convolved Feature). Χρησιμοποιείται γιατί με τη χρήση της μειώνεται η υπολογιστική ισχύς που απαιτείται για την επεξεργασία των δεδομένων αφού μειώνονται οι διαστάσεις. Επιπλέον, είναι χρήσιμη για την εξαγωγή κυρίαρχων χαρακτηριστικών τα οποία είναι αμετάβλητα ως προς τη θέση και τον άξονά τους, διατηρώντας έτσι τη διαδικασία με την οποία εκπαιδεύεται αποτελεσματικά το μοντέλο[47].

Η συνάρτηση ομαδοποίησης αντικαθιστά την έξοδο του δικτύου σε μια συγκεκριμένη θέση με μια συνοπτική στατιστική (summary statistic) των κοντινών εξόδων. Για παράδειγμα, η λειτουργία μέγιστης ομαδοποίησης (max pooling operation) αναφέρει τη μέγιστη απόδοση σε περιοχή σχήματος ορθογώνιου. Άλλες δημοφιλείς λειτουργίες ομαδοποίησης περιλαμβάνουν τον μέσο όρο σε μια περιοχή ορθογώνιου σχήματος, το πρότυπο L^2 μιας ορθογώνιας περιοχής ή έναν σταθμισμένο μέσο όρο που βασίζεται στην απόσταση από το κεντρικό εικονοστοιχείο[47].

Σε όλες τις περιπτώσεις, η ομαδοποίηση βοηθά να γίνει η αναπαράσταση σχεδόν αμετάβλητη στις μικρές διαφοροποιήσεις της εισόδου. Αμεταβλητότητα του μηχανισμού μετατροπής σημαίνει ότι εάν διαφοροποιηθεί η είσοδος ελάχιστα, οι τιμές των περισσότερων ομαδοποιημένων εξόδων δεν αλλάζουν. Η αμεταβλητότητα σε τοπικό μηχανισμό μετατροπής μπορεί να είναι μια πολύ χρήσιμη ιδιότητα αν μας ενδιαφέρει περισσότερο για το αν υπάρχει κάποιο στοιχείο από το τι ακριβώς είναι. Για παράδειγμα, όταν καθορίζεται αν μια εικόνα περιέχει ένα πρόσωπο, δεν χρειάζεται να γνωρίζουμε την τοποθεσία των ματιών με ακρίβεια επιπέδου εικονοστοιχείων, απλά πρέπει να γνωρίζουμε ότι υπάρχει ένα μάτι στην αριστερή πλευρά του προσώπου και ένα μάτι στη δεξιά πλευρά του προσώπου. Σε άλλες περιπτώσεις, είναι πιο σημαντικό να διατηρηθεί η θέση ενός χαρακτηριστικού. Για παράδειγμα, αν θέλουμε να βρούμε μια γωνία που ορίζεται από δύο άκρες που συναντώνται σε έναν συγκεκριμένο προσανατολισμό, πρέπει να διατηρήσουμε την θέση των άκρων αρκετά καλά για να ελέγξουμε αν συναντώνται. Το επίπεδο ομαδοποίησης δεν χάνει τα σημαντικότερα χαρακτηριστικά και ιδιότητες μιας εικόνας. Αντ' αυτού, εξάγει τις πιο κυρίαρχες πληροφορίες και ως εκ τούτου διατηρεί τη διαδικασία της αποτελεσματικής εκπαίδευσης του μοντέλου[47].

Οι συνηθέστεροι τύποι ομαδοποίησης είναι οι ομαδοποίηση μέγιστης τιμής (**Max Pooling**) και ομαδοποίηση μέσου όρου (**Average Pooling**). Στην ομαδοποίηση μέγιστης τιμής, η μέγιστη τιμή σε έναν επιλεγμένο πυρήνα διατηρείται και όλες οι άλλες τιμές απορρίπτονται. Στην ομαδοποίηση μέσου όρου ο μέσος όρος όλων των τιμών που υπάρχουν σε έναν επιλεγμένο πυρήνα είναι εκείνος που αποθηκεύεται[48].



Εικόνα 16: Μορφές ομαδοποίησης μέγιστου όρου και μέσο όρου [48].

1.2.5 Παραλλαγές της Συνάρτησης Συνέλιξης

Όταν αναφερόμαστε σε συνέλιξη στο πλαίσιο των νευρωνικών δικτύων, εννοούμε συνήθως μια λειτουργία που αποτελείται από πολλές εφαρμογές της συνέλιξης σε παραλληλία. Αυτό συμβαίνει επειδή η συνέλιξη με έναν μόνο πυρήνα μπορεί να εξαγάγει μόνο ένα είδος χαρακτηριστικού, αν και σε πολλές χωρικές τοποθεσίες. Συνήθως θέλουμε κάθε επίπεδο του δικτύου μας να εξαγει πολλά είδη χαρακτηριστικών σε πολλές τοποθεσίες.

Επιπλέον, η είσοδος συνήθως δεν είναι απλώς ένα πλέγμα πραγματικών τιμών. Αντίθετα, είναι ένα πλέγμα παρατηρήσεων που εκτιμώνται από τον φορέα (vector-valued observations). Για παράδειγμα, μια έγχρωμη εικόνα έχει ένταση (intensity) κόκκινου, πράσινου και μπλε σε κάθε εικονοστοιχείο. Σε ένα πολυεπίπεδο συνελκτικό δίκτυο, η είσοδος στο δεύτερο επίπεδο είναι η έξοδος του πρώτου επιπέδου, το οποίο συνήθως έχει την έξοδο πολλών διαφορετικών συνελίξεων σε κάθε θέση. Όταν δουλεύουμε με εικόνες, συνήθως θεωρούμε ότι η είσοδος και η έξοδος της συνέλιξης είναι τρισδιάστατοι τανυστές, με έναν δείκτη (index) στα διαφορετικά κανάλια και δύο δείκτες στις χωρικές συντεταγμένες κάθε καναλιού. Οι υλοποιήσεις λογισμικού συνήθως λειτουργούν σε παρτίδες (batchmode), οπότε χρησιμοποιούνται 4-D τανυστές, με τον τέταρτο άξονα να λειτουργεί ως δείκτης για τα διαφορετικά δείγματα της παρτίδας.

Επειδή τα συνελκτικά δίκτυα χρησιμοποιούν συνήθως πολυκαναλική συνέλιξη (multi-channel convolution), οι γραμμικές συναρτήσεις στις οποίες βασίζονται δεν είναι εγγυημένες ότι είναι αντιμεταθετικές (commutative). Αυτές οι πολυκαναλικές λειτουργίες αντιμεταθέτονται μόνο εάν κάθε συνάρτηση έχει τον ίδιο αριθμό καναλιών εξόδου και εισόδου [47].

Αν υποθέσουμε ότι έχουμε έναν 4-D πυρήνα τανυστή K με το στοιχείο $K_{i,j,k,l}$ να δίνει την ισχύ σύνδεσης μεταξύ μιας μονάδας στο κανάλι i της εξόδου και μιας μονάδας στο κανάλι j της εισόδου, με μια μετατόπιση k σειρών και l στηλών μεταξύ της μονάδας εξόδου και της μονάδας εισόδου. Αν υποθέσουμε ότι η είσοδος μας αποτελείται από δεδομένα V με το στοιχείο $V_{i,j,k}$ δίνοντας την τιμή της μονάδας εισόδου μέσα στο κανάλι i , στη σειρά j και στήλη k . Ας υποθέσουμε ότι η έξοδος μας αποτελείται από το Z με την ίδια μορφή με το V . Εάν το Z παράγεται συνελίσοντας το K με το V χωρίς αναστροφή του K , τότε:

$$Z_{i,j,k} = \sum_{l,m,n} V_{l,j+m-1,k+n-1} K_{i,l,m,n} \quad (32)$$

όπου το άθροισμα πάνω από τα l , m και n αποτελείται από όλες τις τιμές για τις οποίες είναι έγκυρες οι λειτουργίες ευρετηρίασης (indexing) τανυστών μέσα στο άθροισμα.

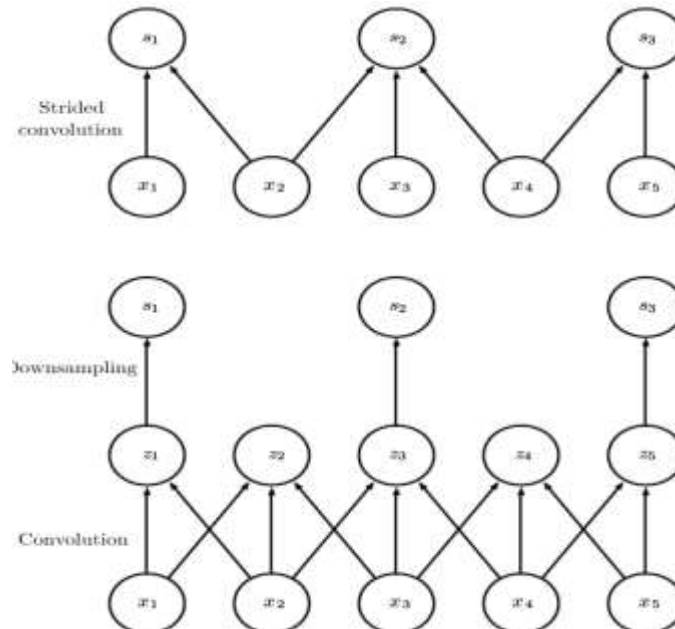
Μπορεί να χρειαστεί να παρακαμφθούν κάποιες θέσεις του πυρήνα έτσι ώστε να μειωθεί το υπολογιστικό κόστος (σε βάρος της μη εξαγωγής των χαρακτηριστικών με τη βέλτιστη ακρίβεια). Μπορούμε να θεωρηθεί ως υποδειγματοληψία (downsampling) της εξόδου της συνάρτησης πλήρους συνέλιξης. Αν θέλουμε να δοκιμάσουμε μόνο κάθε s εικονοστοιχείο σε κάθε κατεύθυνση στην έξοδο, τότε μπορούμε να ορίσουμε μια υποδειγματοληπτούμενη συνάρτηση συνέλιξης (downsampled convolution function) c έτσι ώστε:

$$Z_{i,j,k} = c(\mathbf{K}, \mathbf{V}, s)_{i,j,k} = \sum_{l,m,n} [V_{l,(j-1) \times s + m, (k-1) \times s + n} K_{i,l,m,n}] \quad (32)$$

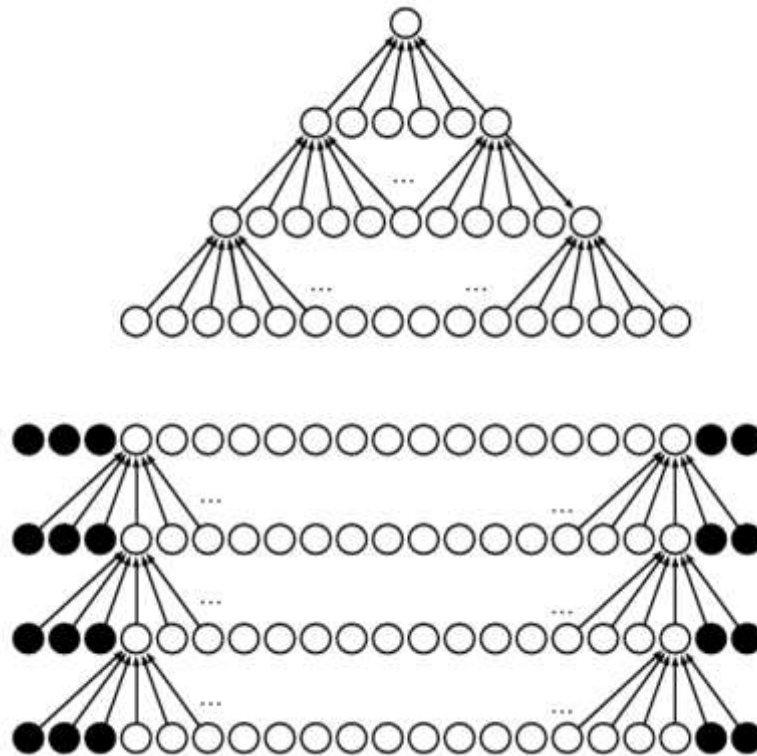
Αναφερόμαστε στο s ως ο βηματισμός (stride=δρασκελιά, διασκελισμός) της υποδειγματοληπτούμενης συνέλιξης[49]. Είναι επίσης δυνατό να οριστεί ένα ξεχωριστό βήμα για κάθε κατεύθυνση κίνησης όπως απεικονίζει η εικόνα 17.

Ένα βασικό χαρακτηριστικό οποιασδήποτε εφαρμογής του συνελκτικού δικτύου είναι η δυνατότητα μηδενικής (zero-rad) εισαγωγής στην είσοδο V για να έχει μεγαλύτερο εύρος[47]. Χωρίς αυτό το χαρακτηριστικό, το πλάτος της αναπαράστασης μειώνεται κατά ένα εικονοστοιχείο μικρότερο από το πλάτος του πυρήνα σε κάθε επίπεδο. Κάνοντας αυτό, μας επιτρέπει να ελέγξουμε το πλάτος του πυρήνα και το μέγεθος της εξόδου. Χωρίς αυτό, είμαστε αναγκασμένοι να επιλέξουμε μεταξύ της συρρίκνωσης της χωρικής έκτασης του δικτύου γρήγορα ή τη χρήση μικρών πυρήνων. Το αποτέλεσμα είναι ότι και τα δύο σενάρια περιορίζουν σημαντικά τις δυνατότητες (expressive powers) του δικτύου όπως δείχνει παραστατικά το σχήμα στην εικόνα 18.

Τα Συνελκτικά Νευρωνικά Δίκτυα CNN αποτελούν τη βάση πάνω στα οποία δημιουργήθηκαν τα σημαντικότερα και πιο γνωστά νευρωνικά δίκτυα που χρησιμοποιούνται στην αναγνώριση αντικειμένων σε εικόνες. Ανάμεσα σε αυτές είναι και το νευρωνικό δίκτυο MobileNet που χρησιμοποιήθηκε για την εκπαίδευση του νευρωνικού δικτύου για την αναγνώριση ποσότητας γάλακτος στην εργασία.



Εικόνα 17: Συνέλιξη με δρασκέλισμα[47].



Εικόνα 18: Η επίδραση της μηδενικής επένδυσης (zero padding) στο μέγεθος του δικτύου: Δείχνει ένα συνελκτικό δίκτυο με πυρήνα πλάτους έξι σε κάθε επίπεδο[47].

1.2.6 Τύποι Δεδομένων

Τα δεδομένα που χρησιμοποιούνται από ένα συνελκτικό νευρωνικό δίκτυο συνήθως αποτελούνται από διάφορα κανάλια, το καθένα εκ των οποίων αποτελεί παρατήρηση μίας διαφορετικής ποσότητας ή μεγέθους σε κάποιο σημείο του χώρου ή του χρόνου. Ο πίνακας 1 παρουσιάζει παραδείγματα με διάφορους τύπους δεδομένων με διαφορετικές διαστάσεις και αριθμό καναλιών[47].

Πίνακας 1: Παραδείγματα δεδομένων που μπορούν να χρησιμοποιηθούν με συνελκτικά νευρωνικά δίκτυα.

	Ένα Κανάλι	Πολυκαναλικό
1-D	Απλή κυματομορφή ήχου	Δεδομένα κίνησης σκελετού όπως είναι τα τρισδιάστατα κινούμενα σχέδια
2-D	Ηχητικά δεδομένα που έχουν επεξεργαστεί με τη συνάρτηση Fourier	Έγχρωμες εικόνες
3-D	Ογκομετρικά δεδομένα όπως είναι οι εικόνες από τεχνολογία σάρωσης CT στο τομέα της ιατρικής	Δεδομένα έγχρωμου βίντεο

Ένα πλεονέκτημα για τα συνελκτικά νευρωνικά δίκτυα είναι ότι μπορούν να επεξεργαστούν εισόδους με ποικίλα χωρικά μεγέθη. Αυτά τα είδη εισόδων δεν μπορούν να αναπαρασταθούν από παραδοσιακά νευρωνικά δίκτυα που βασίζονται σε πολλαπλασιασμό πινάκων. Αυτό επιτάσσει τη χρήση συνελκτικών νευρωνικών δικτύων ακόμη και όταν το υπολογιστικό κόστος και η υπερπροσαρμογή (overfitting) δεν είναι σημαντικά ζητήματα.

Η χρήση της συνέλιξης για την επεξεργασία εισόδων μεταβλητού μεγέθους έχει νόημα μόνο για εισόδους που έχουν μεταβλητό μέγεθος επειδή περιέχουν ποικίλες ποσότητες

παρατήρησης του ίδιου στοιχείου. Η συνέλιξη δεν έχει νόημα εάν η είσοδος έχει μεταβλητό μέγεθος επειδή μπορεί προαιρετικά να περιλαμβάνει διαφορετικά είδη παρατηρήσεων.

1.2.7 Ταξινόμηση Εικόνων και Ανίχνευση Αντικειμένων σε Εικόνες

Οι τεχνικές ταξινόμησης εικόνας και ανίχνευσης αντικειμένων σε εικόνες είναι σημαντικές μέθοδοι που χρησιμοποιούνται στο τομέα της υπολογιστικής όρασης (Computer Vision). Με τη χρήση αυτών των μεθόδων οι υπολογιστικές μηχανές μπορούν να κατανοήσουν και να αναγνωρίσουν αντικείμενα και περιβάλλοντα σε πραγματικό χρόνο με τη βοήθεια ψηφιακών εικόνων ως εισόδους [50].

Η μέθοδος που χρησιμοποιείται στη εκάστοτε περίπτωση έχει σχέση με το σκοπό του μοντέλου που θα εφαρμοστεί. Εάν ο σκοπός του μοντέλου είναι η ταξινόμηση εικόνων σε συγκεκριμένες κατηγορίες τότε χρησιμοποιείται μοντέλο που πραγματοποιεί ταξινόμηση εικόνων. Εάν ο σκοπός του μοντέλου είναι να εντοπιστεί αντικείμενο με συγκεκριμένα χαρακτηριστικά ή ακόμα και ο αριθμός που παρουσιάζεται η συγκεκριμένη κατηγορία αντικειμένου σε μια εικόνα τότε χρησιμοποιείται μοντέλο που πραγματοποιεί ανίχνευση αντικειμένων σε εικόνες. Εάν όμως χρειάζεται να ταξινομηθεί μια εικόνα όπου το αντικείμενο που πρόκειται να κατηγοριοποιηθεί είναι πολύ μικρό σε σχέση με την εικόνα τότε με τη μέθοδο ανίχνευσης αντικειμένων σε εικόνες θα είχαμε καλύτερα αποτελέσματα σε σχέση με μια απλή ταξινόμηση.

Στην περίπτωση της ταξινόμησης εικόνας, το δίκτυο παρουσιάζει το αποτέλεσμα της αναγνώρισης με το όνομα της κλάσης στην οποία θεωρεί ότι ανήκει η εικόνα. Επίσης, σε κάποιες περιπτώσεις δείχνει και το ποσοστό βεβαιότητας της ταξινόμησης[51].

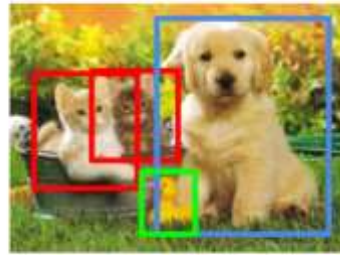
Στη περίπτωση της ανίχνευσης αντικειμένων σε εικόνες η έξοδος της αναγνώρισης είναι η εικόνα με οριοθετημένες τις περιοχές εντός της εικόνας που θεωρεί ότι βρίσκεται το εκάστοτε αντικείμενο. Οι οριοθετημένες περιοχές δημιουργούνται από ένα σύνολο προεπιλεγμένων πλαισίων(κουτιών) με διαφορετικές αναλογίες και κλίμακες οι οποίες προσαρμόζονται πάνω στη περιοχή που έχει αναγνωριστεί ότι ανήκει στη κατηγορία. Επιπλέον, το μοντέλο δημιουργεί βαθμολογίες που αφορά το ποσοστό βεβαιότητας της παρουσίας κάθε κατηγορίας αντικειμένων σε κάθε προεπιλεγμένο πλαίσιο που έχει δημιουργηθεί στην εικόνα. Μοντέλα που μπορούν να κάνουν αυτή την εργασία είναι ο Ανιχνευτής Πολλαπλών Πλαισίων Μεμονωμένης Λήψης (Single Shot Multibox Detector ή SSD Multibox) καθώς επίσης το Faster-R-CNN. Στη εικόνα 50 παρουσιάζονται οπτικά οι διαφορές μεταξύ της ταξινόμησης εικόνων και του εντοπισμού αντικειμένων σε εικόνες. Για τους σκοπούς της εργασίας χρησιμοποιήθηκε η ταξινόμηση αφού είναι πιο απλή σε σχέση με τον εντοπισμό αντικειμένων και ανταποκρίνεται στους σκοπούς της εργασίας [51,52].

Classification



CAT

Object Detection



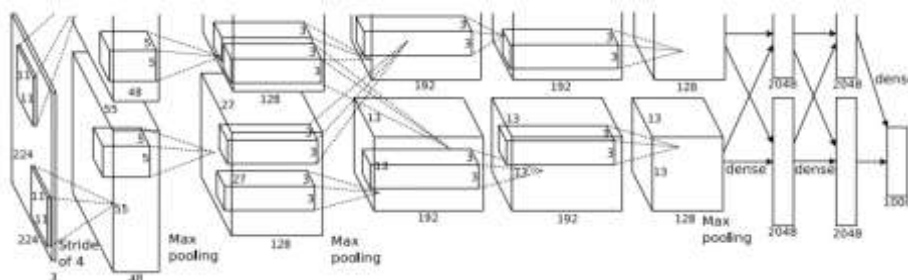
CAT, DOG, DUCK

Εικόνα 19: Αριστερά παρουσιάζεται μια εικόνα που έχει ταξινομηθεί ενώ δεξιά τα αντικείμενα της εικόνας έχουν εντοπιστεί και ταξινομηθεί [50].

1.2.8 AlexNet

Τα συνελκτικά νευρωνικά δίκτυα είναι δημοφιλείς για την ανάπτυξη μοντέλων που αφορά αναγνώριση αντικειμένων καθώς είναι εύκολο να εκπαιδευτούν και είναι δύσκολο να υπάρξει υπερπροσαρμογή όταν χρησιμοποιούν πολλαπλές εικόνες για την εκπαίδευσή τους. Έχουν το μειονέκτημα ότι υπάρχει δυσκολία να εφαρμοστούν σε εικόνες υψηλής ευκρίνειας[53].

Η αρχιτεκτονική του AlexNet αποτελείται από 650000 νευρώνες σε 8 επίπεδα εκ των οποίων 5 επίπεδα είναι συνελκτικά και 3 αποτελούνται από πλήρως συνδεδεμένους νευρώνες τα οποία καταλήγουν στο τέλος σε επίπεδο Softmax 1000 εξόδων. Χρησιμοποιεί τη ReLU ως συνάρτηση ενεργοποίησης αντί της συνάρτησής tanh που ήταν η συνηθισμένη μέχρι τότε. Έχει το χαρακτηριστικό ότι μπορεί να χρησιμοποιήσει πολλαπλές GPU για την εκπαίδευσή της αφού οι μισοί νευρώνες είναι στη μία GPU και η άλλοι μισοί στη δεύτερη GPU. Αυτό έχει ως αποτέλεσμα να μπορούν να εκπαιδευτούν μεγαλύτερα μοντέλα σε λιγότερο χρόνο. Τέλος, έχει επικαλυπτόμενη ομαδοποίηση (Overlapping Pooling) που βοηθάει στη μείωση της πιθανότητας υπερπροσαρμογής του μοντέλου[54].



Εικόνα 20: Η αρχιτεκτονική του AlexNet[54].

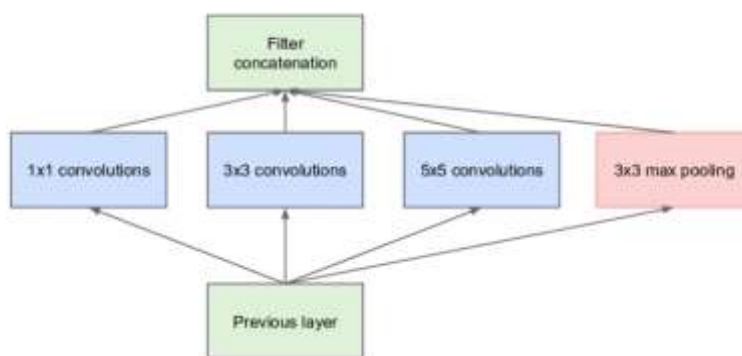
1.2.9 Google Inception V1, V2 και V3

Σκοπός δημιουργίας της αρχιτεκτονικής Inception ήταν η μείωση της υπολογιστικής ισχύος που ήταν απαραίτητη για την εκπαίδευση του νευρωνικού δικτύου ενώ ταυτόχρονα έχει πολύ υψηλή απόδοση. Καθώς συνεχίστηκε η μελέτη της αρχιτεκτονικής Inception παρουσιάστηκαν βελτιώσεις από την πρώτη έκδοση v1 στις επόμενες v2 και v3. Πιο κάτω παρουσιάζονται και αναλύονται οι αρχιτεκτονικές Inception v1, v2 και v3 [55].

1.2.9.1 Inception V1

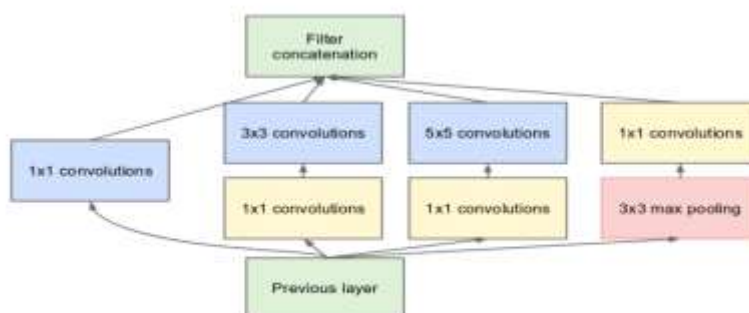
Τα αντικείμενα στις εικόνες που μας ενδιαφέρουν να αναγνωρίσουμε μπορεί να διαφέρουν ως προς το μέγεθός τους και τη θέση τους πάνω στην εικόνα. Με αυτό ως δεδομένο, η εύρεση κατάλληλου φίλτρου είναι δύσκολη καθώς φίλτρο μεγάλου μεγέθους είναι κατάλληλο για πληροφορία που διαμοιράζεται στην εικόνα ισομερώς ενώ τα μικρότερα φίλτρα είναι ιδανικά για εικόνες που τα αντικείμενά βρίσκονται σε τοπικά σημεία μέσα στην εικόνα. Επιπλέον, νευρωνικά δίκτυα με μεγάλο βάθος είναι επιρρεπή στο φαινόμενο της υπερπροσαρμογής. Τέλος, με την προσθήκη στο σύστημα συνελίξεων μεγάλου μεγέθους αυξάνει και το υπολογιστικό κόστος[56].

Για την επίλυση αυτών των προβλημάτων το Inception έχει πολλαπλά φίλτρα στο ίδιο επίπεδο που έχουν διαφορετικά μεγέθη. Το νευρωνικό δίκτυο δηλαδή, αυξάνεται σε πλάτος και όχι σε βάθος. Κάτωθι, στην εικόνα 21 παρουσιάζεται σε απλοϊκή μορφή μια βαθμίδα του Inception με 3 φίλτρα διαφορετικού μεγέθους(1x1, 3x3, 5x5) και ένα επίπεδο ομαδοποίησης μέγιστης τιμής, η κάθε μία βρίσκεται σε παραλληλία με την άλλη. Στη συνέχεια οι έξοδοι τους συγκεντρώνονται και αποστέλλονται στην επόμενη βαθμίδα του νευρωνικού δικτύου. Προκειμένου να μειωθεί το υπολογιστικό κόστος, μειώνεται ο αριθμός των καναλιών εισόδου με την πρόσθεση συνελίξεων 1x1 πριν από τις συνελίξεις 3x3 και 5x5 και ακριβώς μετά το επίπεδο ομαδοποίησης μέγιστης τιμής όπως δείχνει στη εικόνα 22 [57].



(a) Inception module, naïve version

Εικόνα 21: Απλοϊκή βαθμίδα της αρχιτεκτονικής Inception [56].

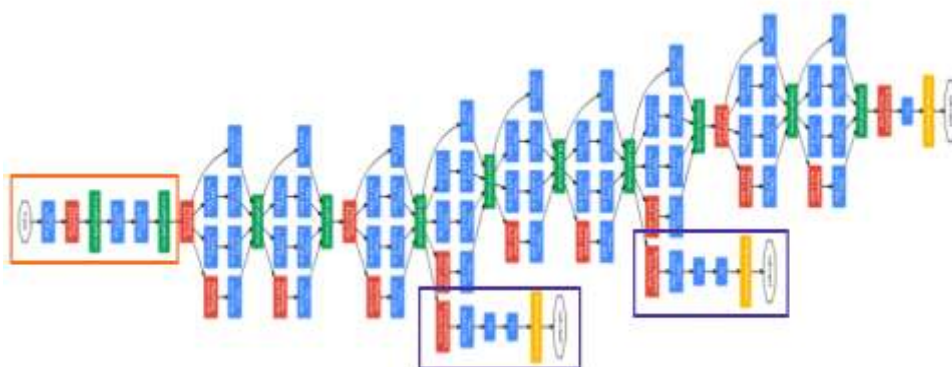


(b) Inception module with dimension reductions

Εικόνα 22: Βαθμίδα της αρχιτεκτονικής Inception με μειωμένες διαστάσεις[56].

Η βαθμίδα της εικόνας 22 αποτελεί την βάση πάνω στην οποία δημιουργήθηκε η αρχιτεκτονική του νευρωνικού δικτύου GoogLeNet ή αλλιώς Inception v1. Η αρχιτεκτονική του παρουσιάζεται στην εικόνα 23. Το GoogLeNet διαθέτει 9 βαθμίδες Inception η κάθε μια στοιβαγμένη μετά την άλλη. Έχει βάθος 22 επιπέδων και αν μετρήσουμε και τα επίπεδα ομαδοποίησης ο αριθμός των επιπέδων είναι 27. Χρησιμοποιεί ομαδοποίηση γενικού μέσου όρου μετά από την τελευταία βαθμίδα Inception[57].

Λόγω του μεγάλου βάθους του νευρωνικού δικτύου, παρουσιάζεται το πρόβλημα των εξαφανιζόμενων κλίσεων (Vanishing Gradient Problem) το οποίο σημαίνει ότι από ένα σημείο και μετά, καθώς αυξάνεται το βάθος του νευρωνικού δικτύου, παρουσιάζεται κορεσμός στην απόδοσή του. Γι' αυτό το λόγο, προστίθενται 2 βοηθητικοί ταξινομητές τα οποία παρουσιάζονται εντός των μοβ πλαισίων, στη εικόνα 23. Αυτό που κάνουν είναι ότι εφαρμόζουν τη συνάρτηση Softmax στις εξόδους 2 βαθμίδων Inception και υπολογίζουν τη βοηθητική απώλεια πάνω στις ίδιες ταμπέλες. Η συνολική συνάρτηση απώλειας είναι ο σταθμισμένος μέσος όρος της βοηθητικής απώλειας και τις πραγματικής απώλειας[56].

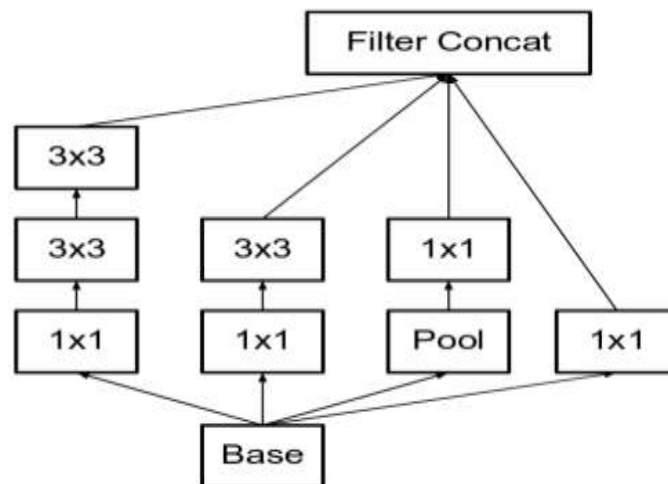


Εικόνα 23: Αρχιτεκτονική GoogLeNet[56].

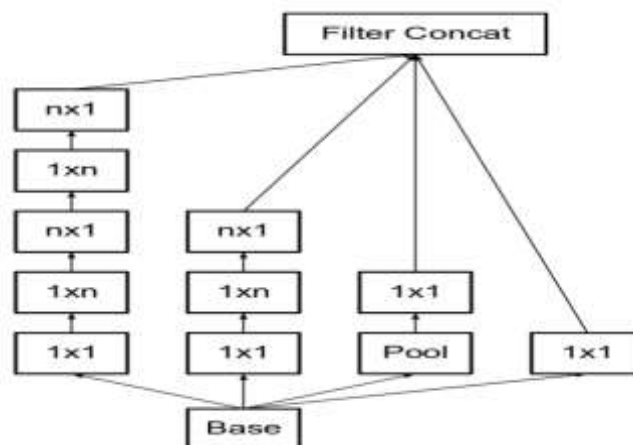
1.2.9.2 Inception V2

Τα συνελικτικά νευρωνικά δίκτυα έχουν καλύτερη απόδοση όταν δεν υπάρχει μεγάλη διαφοροποίηση των διαστάσεων της εισόδου από την συνέλιξη. Όταν μειώνονται οι διαστάσεις υπάρχει η πιθανότητα να χαθεί πληροφορία. Το φαινόμενο αυτό ονομάζεται αντιπροσωπευτική συμφόρηση (representational bottleneck). Προκειμένου να αντιμετωπιστεί αυτό, μπορούν να χρησιμοποιηθούν μέθοδοι παραγοντοποίησης έτσι ώστε να βελτιωθεί η υπολογιστική ταχύτητα του νευρωνικού δικτύου. Πιο συγκεκριμένα, η συνέλιξη 5x5 μπορεί να αντικατασταθεί ισοδύναμα από 2 συνέλιξεις 3x3, η μία μετά την άλλη, αυξάνοντας την ταχύτητα των υπολογισμών όπως δείχνει η κάτωθι εικόνα 24. Ένα φίλτρο $n \times n$ μπορεί να παραγοντοποιηθεί και να παρουσιαστεί ισοδύναμα από 2 συνέλιξεις $1 \times n$ και $n \times 1$. Πιο συγκεκριμένα, η συνέλιξη 3x3 μπορεί να αναπαρασταθεί ισοδύναμα από 2 συνέλιξεις, η μία μετά από την άλλη 1×3 και 3×1 αυξάνοντας τη ταχύτητα των υπολογισμών. Αυτό έχει ως αποτέλεσμα η βαθμίδα στην εικόνα 24 να παρουσιάζεται ισοδύναμα στην εικόνα 25. Το πρόβλημα με την αύξηση του βάθους του επιπέδου του φίλτρου είναι το φαινόμενο της αντιπροσωπευτικής συμφόρησης, δηλαδή μειώνονται οι διαστάσεις του

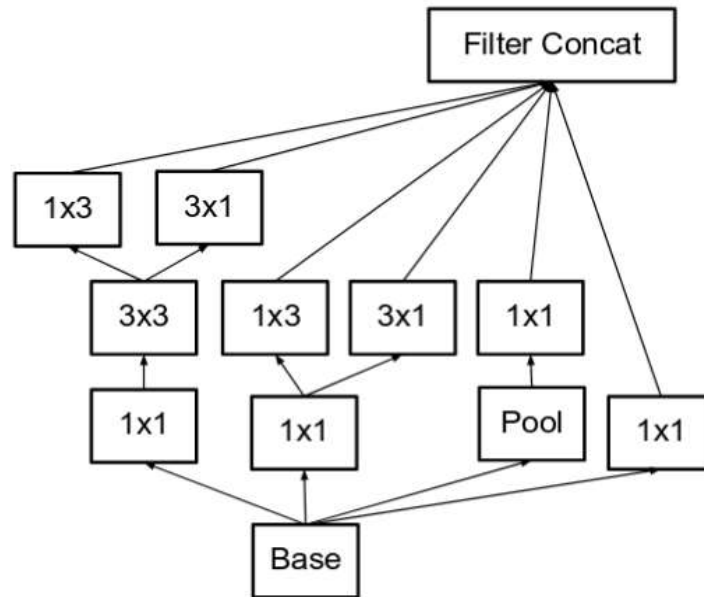
εξερχόμενου χάρτη χαρακτηριστικών σε σχέση με την είσοδο του χάνοντας έτσι πολύτιμες πληροφορίες. Για το λόγο αυτό αντί η βαθμίδα του φίλτρου να γίνει βαθύτερη, γίνεται πλατύτερη όπως δείχνει η εικόνα 26 όπου στο τελευταίο επίπεδο της βαθμίδας του φίλτρου πριν την συγκέντρωσή τους για το επόμενο επίπεδο. Εκεί, αντί για το φίλτρο μεγέθους 3×3 , έχουμε 2 φίλτρα σε παραλληλία μεγέθους 1×3 και 3×1 [58].



Εικόνα 24: Παρουσιάζεται πως η συνέλιξη 5×5 αναπαρίσταται ισοδύναμα από 2 3×3 συνελίξεις η μία μετά την άλλη [58].



Εικόνα 25: Παρουσιάζεται η αρχιτεκτονική μιας βαθμίδας Inception όταν η συνέλιξη 3×3 αντικαθίσταται ισοδύναμα από δύο συνελίξεις 1×3 και 3×1 η μία μετά την άλλη [58].



Εικόνα 26: Παρουσιάζεται η αρχιτεκτονική της εικόνας 25 όταν το τελευταίο σκέλος της βαθμίδας το οποίο είναι συνέλιξη 3x3 αντικαθίσταται ισοδύναμα από δυο συνέλιξεις 1x3 και 3x1 σε παραλληλία[58].

Οι 3 αυτές βαθμίδες που παρουσιάστηκαν πριν, εφαρμόστηκαν για τη δημιουργία του InceptionV2 όπως παρουσιάζεται στο κάτωθι πίνακα.

Πίνακας 2: Παρουσιάζονται τα χαρακτηριστικά κάθε επιπέδου του InceptionV2.

Type	Patch Size/Stride	Input Size
conv	3x3/2	229x229x3
conv	3x3/1	149x149x32
Conv padded	3x3/1	147x147x32
pool	3x3/2	147x147x64
conv	3x3/1	73x73x64
conv	3x3/2	71x71x80
conv	3x3/1	35x35x192
3x Inception	Όπως στην εικόνα 25	35x35x288
5x Inception	Όπως στην εικόνα 26	17x17x768
2x Inception	Όπως στην εικόνα 27	8x8x1280
pool	8x8	8x8x2048
linear	logits	1x1x2048
Softmax	classifier	1x1x1000

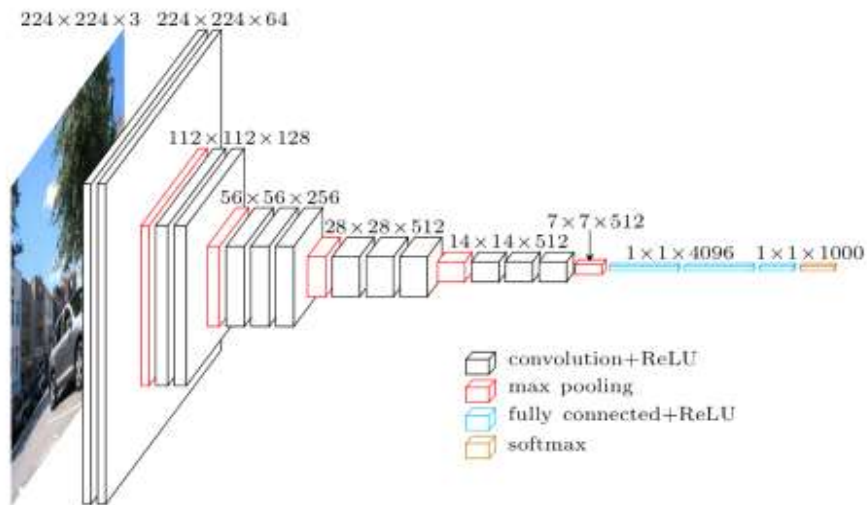
1.2.9.3 Inception V3

Με δεδομένο ότι οι βοηθητικοί ταξινομητές δεν πρόσφεραν κάτι στο σύστημα παρά μόνο προς το τέλος της εκπαίδευσης του νευρωνικού δικτύου, προτάθηκαν βελτιώσεις πάνω στη αρχιτεκτονική της InceptionV2 χωρίς όμως να υπάρξει δραστική αλλαγή στις βαθμίδες του συστήματος. Η βελτιώσεις αυτές ήταν η παραγοντοποίηση των 7x7 συνέλιξεων, η εισαγωγή του βελτιστοποιητή RPMSProp, η προσθήκη BatchNorm στους βοηθητικούς ταξινομητές και Label Smoothing το οποίο βοηθάει στην αποφυγή της υπερπροσαρμογής στο σύστημα. Η αρχιτεκτονική αυτή ονομάστηκε InceptionV3 [59].

1.2.10 VGG

Το νευρωνικό δίκτυο VGG δημιουργήθηκε από τους Karen Simonyan και Andrew Zisserman το 2015. Σκοπός της δημιουργίας του είναι η αύξηση του βάθους των νευρωνικών δικτύων έτσι ώστε να είναι πιο αποτελεσματικά στην αναγνώριση εικόνων[60].

Βασικό χαρακτηριστικό της αρχιτεκτονικής του είναι ότι σε κάθε επίπεδο που πραγματοποιείται συνέλιξη στο νευρωνικό δίκτυο το μέγεθος του φίλτρου είναι 3x3. Ο λόγος που είναι 3x3 είναι γιατί είναι το μικρότερο δυνατό μέγεθος που δίνει την ικανότητα της αντίληψης τι είναι δεξιά ή αριστερά, πάνω ή κάτω και κέντρο. Ο βηματισμός της συνέλιξης είναι μεγέθους 1 εικονοστοιχείου ενώ η μηδενική επένδυση μετά από κάθε επίπεδο είναι ίσο με 1 έτσι ώστε η έξοδος να έχει ίδιο μέγεθος με την είσοδο στο επίπεδο συνέλιξης. Πραγματοποιείται ομαδοποίηση μέγιστης τιμής 5 φορές όπως δείχνει η εικόνα 27. Η βαθμίδα των επιπέδων που πραγματοποιείται συνέλιξη ακολουθείται από 3 επίπεδα πλήρως συνδεδεμένων νευρώνων. Οι πρώτοι 2 έχουν 4096 κανάλια ο καθένας και το τρίτο επίπεδο αποτελείται από 1000 κανάλια. Το τελευταίο επίπεδο είναι το επίπεδο Softmax. Πρέπει να αναφερθεί εδώ ότι όσα αναφέρθηκαν αποτελούν κοινά χαρακτηριστικά των νευρωνικών δικτύων VGG-16 και VGG-19, τα οποία πήραν το όνομά τους από το γεγονός ότι διαθέτουν 16 και 19 επίπεδα νευρώνων αντίστοιχα. Το πλάτος του πρώτου επιπέδου συνέλιξης είναι 64 καναλιών και στη συνέχεια αυξάνεται επί 2 σε κάθε επόμενο επίπεδο έως ότου φτάσει στα 512 κανάλια. Αναπαράσταση του νευρωνικού δικτύου παρουσιάζεται κάτωθι στην εικόνα 27[60].



Εικόνα 27: Αρχιτεκτονική του νευρωνικού δικτύου VGG-16[61].

1.2.11 MobileNet



Εικόνα 28: Εφαρμογές με την αρχιτεκτονική MobileNet[62].

Το MobileNet είναι ένα μοντέλο αρχιτεκτονικής συνελικτικών νευρωνικών δικτύων CNN για την ταξινόμηση εικόνων. Αυτό που κάνει το MobileNet ξεχωριστό είναι ότι χρειάζεται πολύ λιγότερη υπολογιστική ισχύ και μπορεί εύκολα να χρησιμοποιηθεί ως βάση για άλλα μοντέλα με διαφορετικές λειτουργίες με τη χρήση βαθιάς μάθησης (Transfer Learning). Αυτό το καθιστά ιδανικό για φορητές συσκευές, ενσωματωμένα συστήματα και υπολογιστές χωρίς GPU ή με χαμηλή υπολογιστική απόδοση κάνοντας όμως συμβιβασμούς ως προς την ακρίβεια των αποτελεσμάτων. Επίσης ταιριάζει καλύτερα για προγράμματα περιήγησης στο διαδίκτυο (Web Browsers), καθώς τα προγράμματα περιήγησης έχουν περιορισμό στον υπολογισμό, την επεξεργασία εικόνων και την αποθήκευση [44].

Το μοντέλο MobileNet βασίζεται στις σε διαχωριζόμενες σε βάθος συνελίξεις (Depthwise Separable Convolutions) που με τη σειρά τους αποτελούνται όλες από μια σε βάθος συνέλιξη (Depthwise Convolution) και μια συνέλιξη 1x1 η οποία ονομάζεται pointwise συνέλιξη, μειώνοντας έτσι το υπολογιστικό κόστος σε σχέση με την τυπική συνέλιξη. Η σε βάθος συνέλιξη, η οποία είναι το πρώτο μέρος λαμβάνει την εικόνα και κάνει συνέλιξη χρησιμοποιώντας τρεις πυρήνες, ένα για κάθε κανάλι της εικόνας. Στη συνέχεια, χρησιμοποιούνται πολλαπλές συνέλιξεις pointwise οι οποίες έχουν μέγεθος 1x1 και αριθμό καναλιών ίσο με τα κανάλια της εικόνας όπως δείχνει η εικόνα 29 [63]. Αυτή η παραγοντοποίηση έχει ως αποτέλεσμα τη δραστική μείωση των υπολογισμών και του μεγέθους του μοντέλου.

Μία τυπική στοιβάδα συνέλιξης έχει ως είσοδο $D_F * D_F * M$ χάρτης χαρακτηριστικών F και παράγει την $D_G * D_G * N$ χάρτη χαρακτηριστικών G όπου D_F είναι το μήκος και πλάτος της εισαγόμενης εικόνας, M είναι ο αριθμός των καναλιών εισόδου (input channels), το D_G είναι το μήκος και πλάτος της εξαγόμενης εικόνας και N είναι ο αριθμός καναλιών εξόδου (output channels). Το τυπικό συνελικτικό επίπεδο παραμετροποιείται από το πυρήνα (kernel) συνέλιξης μεγέθους $D_K * D_K * M * N$ όπου D_K είναι οι διαστάσεις του πυρήνα που θεωρούνται τετραγωνικές.

Οι απλές συνελίξεις έχουν υπολογιστικό κόστος:

$$\text{Standard Convolutions Computational Cost} = D_K * D_K * M * N * D_F * D_F \quad (33)$$

Η σε βάθος συνέλιξη έχει υπολογιστικό κόστος:

$$\text{Depthwise Convolution Computational Cost} = D_K * D_K * M * D_F * D_F \quad (34)$$

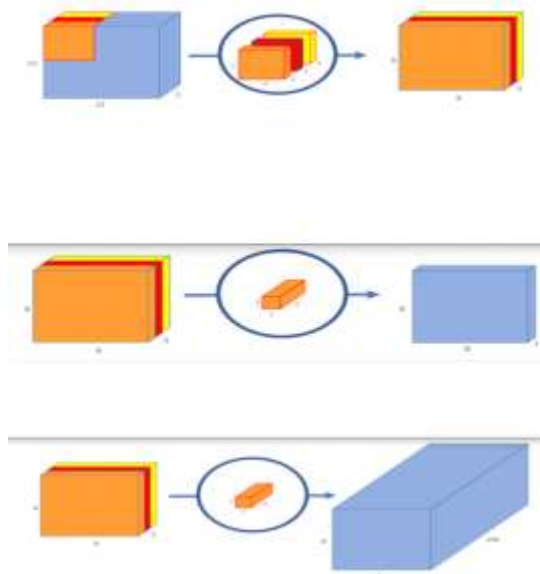
Η ίδια ενώ είναι αποδοτική σε σχέση με την απλή συνέλιξη, εν τούτοις, δεν μπορεί να συνδυάσει κανάλια εισαγωγής παρά μόνο να τα φιλτράρει. Για αυτό το λόγο έχουμε την pointwise συνέλιξη που πραγματοποιεί αυτή τη λειτουργία. Η σε βάθος διαχωριζόμενη συνέλιξη έχει υπολογιστικό κόστος:

$$\text{DepthwiseSeparableConvolutionComputationalCost} = D_K * D_K * M * D_F * D_F + M * N * D_F * D_F \quad (35)$$

Η μείωση στο υπολογιστικό κόστος της σε βάθος διαχωρίσιμης συνέλιξης σε σχέση με την απλή βγαίνει από τον υπολογισμό της διαίρεσης μεταξύ τους, δηλαδή:

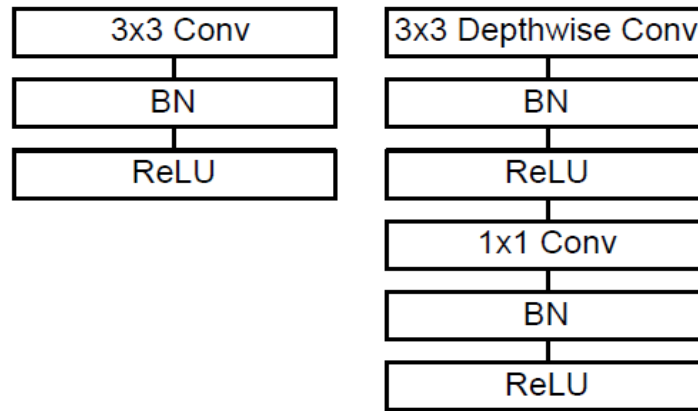
$$\frac{D_K * D_K * M * N * D_F * D_F}{D_K * D_K * M * D_F * D_F + M * N * D_F * D_F} = \frac{1}{N} + \frac{1}{D_K^2} \quad (36)$$

Το MobileNet χρησιμοποιεί 3x3 σε βάθος διαχωρίσιμες συνελίξεις που πραγματοποιούν μεταξύ 8 έως 9 φορές λιγότερους υπολογισμούς από τις τυπικές συνελίξεις με μικρή μόνο μείωση της ακρίβειας[64].



Εικόνα 29: Πάνω, γίνεται συνέλιξη σε βάθος μιας εικόνας μεγέθους 12x12 με τρία κανάλια με 3 πυρήνες μεγέθους 5x5x1 και παράγει μια νέα εικόνα με 8x8 με τρία κανάλια. Στη συνέχεια ενδιάμεσα παρουσιάζεται η συνέλιξη pointwise με ένα μόνο πυρήνα. Κάτω παρουσιάζεται το αποτέλεσμα αν γίνονταν συνέλιξη με 256 πυρήνες pointwise[63].

Στη δομή του MobileNet, η πρώτη στοιβάδα είναι η τυπική συνέλιξη ενώ στη συνέχεια οι υπόλοιπες στοιβάδες αποτελούνται από διαχωρίσιμες σε βάθος συνελίξεις. Η αρχιτεκτονική MobileNet παρουσιάζεται στον πίνακα 3. Όλα τα ενδιάμεσα επίπεδα ακολουθούνται από BatchNorm[65] και ReLU. Το τελευταίο ενδιάμεσο επίπεδο που είναι πλήρως συνδεδεμένο δεν διαθέτει μη γραμμικότητα και τροφοδοτεί το τελικό στρώμα που είναι μια Softmax[66] συνάρτηση για ταξινόμηση. Η εικόνα 30 παραθέτει αριστερά μια στοιβάδα με απλή συνέλιξη, BatchNorm και μη γραμμικότητα παραγόμενη από ReLU. Δεξιά, η εικόνα 30 δείχνει ένα παραγοντοποιημένο στρώμα με μία σε βάθος συνέλιξη, 1x1 συνέλιξη pointwise καθώς και BatchNorm και ReLU μετά από κάθε επίπεδο συνέλιξης. Η υποδειγματοληψία (downsampling) αντιμετωπίζεται με βηματική συνέλιξη(strided convolution). Στο τέλος, ο μέσος όρος των ομαδοποιήσεων μειώνει τη χωρική ανάλυση(spatial resolution) σε 1 πριν από το πλήρως συνδεδεμένο στρώμα. Εάν μετρηθούν οι σε βάθος και pointwise συνελίξεις ως ξεχωριστά επίπεδα, το MobileNet έχει 28 επίπεδα.



Εικόνα 30: Αριστερά είναι ένα επίπεδο συνέλιξης με BatchNorm και ReLU ενώ δεξιά είναι μια σε βάθος διαχωριζόμενη συνέλιξη με επίπεδα για σε βάθος συνέλιξη και συνέλιξη pointwise ακολουθούμενο από BatchNorm και ReLU[64].

Πίνακας 3: Αρχιτεκτονική MobileNet[64].

Table 1. MobileNet Body Architecture

Type / Stride	Filter Shape	Input Size
Conv / s2	$3 \times 3 \times 3 \times 32$	$224 \times 224 \times 3$
Conv dw / s1	$3 \times 3 \times 32$ dw	$112 \times 112 \times 32$
Conv / s1	$1 \times 1 \times 32 \times 64$	$112 \times 112 \times 32$
Conv dw / s2	$3 \times 3 \times 64$ dw	$112 \times 112 \times 64$
Conv / s1	$1 \times 1 \times 64 \times 128$	$56 \times 56 \times 64$
Conv dw / s1	$3 \times 3 \times 128$ dw	$56 \times 56 \times 128$
Conv / s1	$1 \times 1 \times 128 \times 128$	$56 \times 56 \times 128$
Conv dw / s2	$3 \times 3 \times 128$ dw	$56 \times 56 \times 128$
Conv / s1	$1 \times 1 \times 128 \times 256$	$28 \times 28 \times 128$
Conv dw / s1	$3 \times 3 \times 256$ dw	$28 \times 28 \times 256$
Conv / s1	$1 \times 1 \times 256 \times 256$	$28 \times 28 \times 256$
Conv dw / s2	$3 \times 3 \times 256$ dw	$28 \times 28 \times 256$
Conv / s1	$1 \times 1 \times 256 \times 512$	$14 \times 14 \times 256$
5x Conv dw / s1	$3 \times 3 \times 512$ dw	$14 \times 14 \times 512$
Conv / s1	$1 \times 1 \times 512 \times 512$	$14 \times 14 \times 512$
Conv dw / s2	$3 \times 3 \times 512$ dw	$14 \times 14 \times 512$
Conv / s1	$1 \times 1 \times 512 \times 1024$	$7 \times 7 \times 512$
Conv dw / s2	$3 \times 3 \times 1024$ dw	$7 \times 7 \times 1024$
Conv / s1	$1 \times 1 \times 1024 \times 1024$	$7 \times 7 \times 1024$
Avg Pool / s1	Pool 7×7	$7 \times 7 \times 1024$
FC / s1	1024×1000	$1 \times 1 \times 1024$
Softmax / s1	Classifier	$1 \times 1 \times 1000$

Με δεδομένο ότι το MobileNet καταναλώνει λίγη υπολογιστική ισχύ και μπορεί εύκολα να χρησιμοποιηθεί ως βάση για άλλα μοντέλα με διαφορετικές λειτουργίες (Transfer Learning), επιλέχθηκε ως το νευρωνικό δίκτυο με το οποίο θα δημιουργηθεί το αυτόματο σύστημα αναγνώρισης τροφίμων του οποίου η περιγραφή θα γίνει στο κεφάλαιο 4.

1.2.12 ResNet

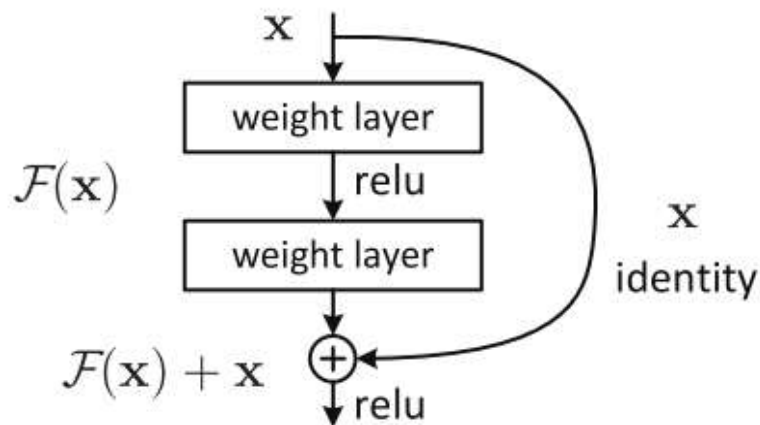
Εάν υπάρχει αρκετά μεγάλη υπολογιστική και αποθηκευτική ικανότητα, ένα νευρωνικό δίκτυο εμπρόσθιας διάδοσης μπορεί να αντιπροσωπεύσει οποιαδήποτε συνάρτηση. Ωστόσο, το επίπεδο των νευρώνων μπορεί να είναι πολύ μεγάλο και υπάρχει πάντα ο κίνδυνος της υπερπροσαρμογής. Γι' αυτό το λόγο προστίθενται επιπλέον επίπεδα νευρώνων στο δίκτυο[67].

Νευρωνικά δίκτυα με πολλαπλές στοιβάδες νευρώνων είναι δύσκολο να εκπαιδευτούν καθώς όταν γίνεται οπισθοδιάδοση από στοιβάδες πιο βαθιά στο νευρωνικό δίκτυο σε εκείνες που είναι πιο πάνω, ο επαναλαμβανόμενος πολλαπλασιασμός μεταξύ των επιπέδων μειώνει το μέγεθος της κάθε βαθμίδας. Το αποτέλεσμα είναι ότι από ένα σημείο και μετά, καθώς το δίκτυο πηγαίνει βαθύτερα, υποβαθμίζεται η ακρίβεια της εκπαίδευσης[67].

Για την επίλυση αυτού του προβλήματος το ResNet (Residual Network) παρουσιάζει ως λύση μια παράκαμψη με την οποία πηγαίνει η είσοδος (x) από προηγούμενες στοιβάδες σε ένα ή περισσότερα επίπεδα νευρώνων βαθύτερα. Το αποτέλεσμα είναι η έξοδος $F(x)$ από τη τελευταία στοιβάδα νευρώνων πριν τη διασταύρωση να προστίθεται με την είσοδο x με αποτέλεσμα το άθροισμα $F(x)+x$. Η περιγραφή που έγινε λεκτικά παρουσιάζεται στην εικόνα 31 και μαθηματικά από την κάτωθι εξίσωση 38 :

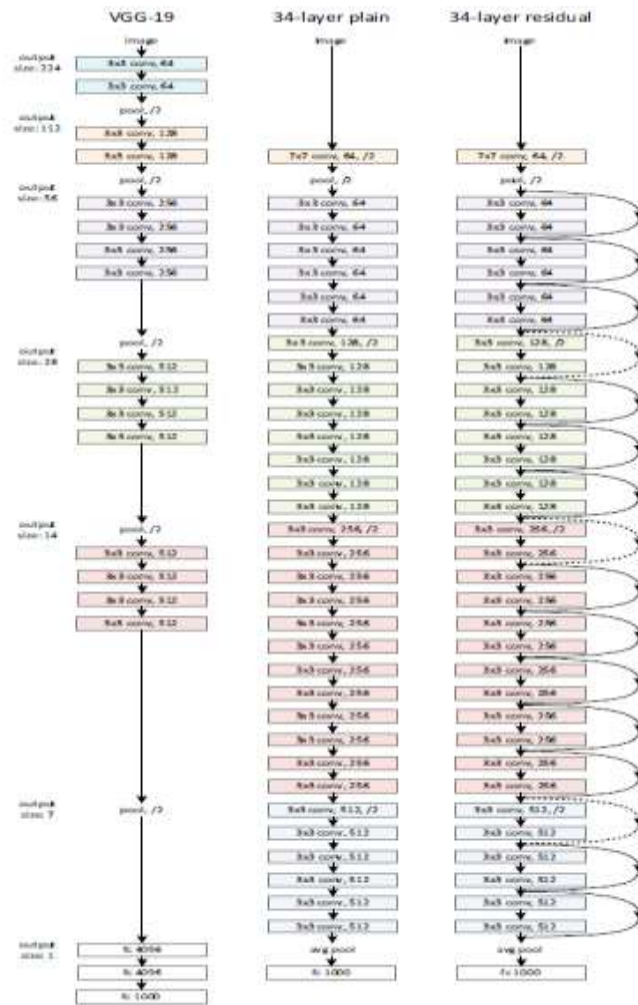
$$y=F(x,\{W_i\})+x \quad (38)$$

Το x και το y είναι η είσοδος και η τελική έξοδος της βαθμίδας που εξετάζονται αντίστοιχα. Το $F(x,\{W_i\})$ είναι η έξοδος από την τελευταία στοιβάδα νευρώνων της βαθμίδας. Το νευρωνικό δίκτυο ResNet αποτελείται από πολλαπλές τέτοιες βαθμίδες[67].



Εικόνα 31: Μια βαθμίδα της αρχιτεκτονικής του νευρωνικού δικτύου ResNet[67]

Η αποτελεσματικότητα του ResNet έχουν οδηγήσει στην ανάπτυξη παραλλαγών της όπως είναι το ResNetXt και το Συνελικτικό Νευρωνικό Δίκτυο Πυκνών συνδέσμων (Densely Connected CNN-DenseNet). Στο ResNetXt, η αρχική είσοδος x ακολουθεί πολλαπλά διαφορετικά μονοπάτια, το καθένα εκ των οποίων διαθέτει την ίδια τοπολογία νευρώνων με την άλλη. Στο τέλος οι έξοδοι από κάθε μονοπάτι προστίθεται με κάθε άλλο μονοπάτι και στη συνέχεια το αποτέλεσμά τους προστίθεται με την αρχική είσοδο x . Στο DenseNet, η κάθε βαθμίδα συνδέεται απευθείας με κάθε άλλη βαθμίδα του νευρωνικού δικτύου. Δηλαδή, η είσοδος κάθε βαθμίδας αποτελείται από το χάρτη χαρακτηριστικών όλων των προηγούμενων βαθμίδων και η έξοδος κάθε βαθμίδας αποτελεί είσοδος για όλες τις επόμενες βαθμίδες[68].



Εικόνα 32: Αριστερά παρουσιάζεται η αρχιτεκτονική του νευρωνικού δικτύου VGG με 19 επίπεδα νευρώνων, στη μέση παρουσιάζεται η αρχιτεκτονική ενός συνελκτικού νευρωνικού δικτύου με 34 επίπεδα νευρώνων και δεξιά παρουσιάζεται το ίδιο νευρωνικό δίκτυο με τη χρήση βαθμίδων ResNet[67].

Μέθοδοι Αναγνώρισης Τροφίμων και Βάσεις Δεδομένων

Μέχρι αυτό το σημείο παρουσιάστηκε η θεωρητική βάση που σχετίζεται με τη βαθιά μάθηση. Στη συνέχεια της εργασίας παρουσιάζονται μια σειρά από μεθόδους που χρησιμοποιούνται για την αναγνώριση τροφίμων. Τέλος, σε αυτό το κεφάλαιο παρουσιάζεται η βιβλιογραφική ανασκόπηση που αφορά βιβλιοθήκες που χρησιμοποιούνται για βαθιά μάθηση, τις σημαντικότερες βάσεις δεδομένων που αφορούν τροφές καθώς επίσης, βιβλιογραφικές μελέτες όπου εφαρμόστηκαν και έγινε καταγραφή των αποτελεσμάτων σε συστήματα αναγνώρισης τροφίμων.

2.1 Μέθοδοι Αναγνώρισης Τροφίμων

Για την ανάπτυξη συστήματος αναγνώριση τροφίμων μέσω της βαθιάς μάθησης υπάρχουν 3 βασικές μέθοδοι[77]:

- η ανάπτυξη νέας αρχιτεκτονικής νευρωνικού δικτύου από την αρχή,
- η χρήση ενός ήδη εκπαιδευμένου νευρωνικού δικτύου για την επανεκπαίδευσή του με νέα βάση δεδομένων μέσω της μεταφοράς μάθησης και
- η χρήση προϋπάρχουσας πλατφόρμας όπου η αναγνώριση γίνεται σε ένα είδος "μαύρου κουτιού" όπου δέχεται ως είσοδο την εικόνα προς αναγνώριση και βγάζει ως έξοδο την αναγνώριση χωρίς να έχουμε παρέμβει στην ανάπτυξη του νευρωνικού δικτύου.

Η λογική, τα προτερήματα και μειονεκτήματα της κάθε μεθόδου παρουσιάζονται στη συνέχεια.

2.1.1 Μεταφορά Μάθησης

Μία πολύ δημοφιλής μέθοδος για την ανάπτυξη νευρωνικών δικτύων που μπορούν να κάνουν αναγνώριση είναι η μεταφορά μάθησης(transfer learning).

Η μεταφορά μάθησης είναι μέθοδος της μηχανικής μάθησης με την οποία ένα μοντέλο που αναπτύχθηκε αρχικά για διαφορετική εφαρμογή, μπορεί να χρησιμοποιηθεί ως το αρχικό σημείο αναφοράς για την εκπαίδευση του δεύτερου μοντέλου το οποίο όμως είναι παρόμοιο με το αρχικό. Στο πλαίσιο της αναγνώρισης αντικειμένων και ειδικότερα εκείνης της αναγνώρισης τροφίμων, σημαίνει ότι μπορούμε να χρησιμοποιήσουμε ένα ήδη εκπαιδευμένο νευρωνικό δίκτυο προκειμένου να το εκπαιδεύσουμε με εικόνες όπου το περιεχόμενο τους είναι παρόμοιο με τις εικόνες που είχαν χρησιμοποιηθεί για την αρχική εκπαίδευση του. Το νευρωνικό δίκτυο κρατάει τα χαρακτηριστικά της αρχικής εκπαίδευσης και χρειάζεται λιγότερους υπολογισμούς για να εκπαιδευτεί επαρκώς με τις νέες εικόνες. Ο λόγος είναι γιατί το σημείο εκκίνησης της νέας εκπαίδευσης είναι πιο κοντά στο στόχο σε σχέση με το αν ξεκινούσε από τυχαίο σημείο εκκίνησης[79].

Η μέθοδος αυτή έχει το πλεονέκτημα ότι μειώνει το μέγεθος της βάσης δεδομένων με την οποία είναι απαραίτητη για να εκπαιδευτεί το νευρωνικό δίκτυο επαρκώς. Αυξάνεται η ταχύτητα εκπαίδευσης ενώ μειώνεται η υπολογιστική ισχύς που χρειάζεται για τη διαδικασία αυτή. Όλα αυτά καθιστούν τη μεταφορά μάθησης δημοφιλής στο τομέα της βαθιάς μάθησης με αποτέλεσμα να είναι λιγότερο συχνό να αναπτυχθεί νευρωνικό δίκτυο από την αρχή αφού χρειάζεται μεγάλη προσοχή για το σχεδιασμό του μοντέλου και τη συλλογή της βάσης δεδομένων[79].

2.1.2 Δημιουργία Νέας Αρχιτεκτονικής Νευρωνικού Δικτύου

Η ανάπτυξη της αρχιτεκτονικής ενός νευρωνικού δικτύου η οποία θα εφαρμοζόταν σε ένα πεδίο το οποίο δεν έχει χρησιμοποιηθεί στο παρελθόν είναι μια πολύπλοκη και χρονοβόρα προσπάθεια η οποία απαιτεί γνώσεις υψηλού επιπέδου στον τομέα του προγραμματισμού και των μαθηματικών. Από την άλλη, πλατφόρμες βαθιάς μάθησης οι οποίες είναι έτοιμες για χρήση ή μοντέλα που χρησιμοποιούν μεταφορά μάθησης χρειάζονται πολύ λιγότερο χρόνο και γνώσεις για να εφαρμοστούν, και παρέχουν πιο άμεσα αποτελέσματα[80].

2.1.3 Πλατφόρμες Βαθιάς Μάθησης

Υπάρχουν πολλές διαδικτυακές εφαρμογές οι οποίες διαθέτουν νευρωνικά δίκτυα που έχουν εκπαιδευτεί με μεγάλες βάσεις δεδομένων και οι οποίες είναι αποτελεσματικές στο να αναγνωρίσουν πολλαπλές κλάσεις με μεγάλη ακρίβεια. Ως εκ τούτου, μια νέα πλατφόρμα αναγνώρισης τροφίμων, είτε αυτή έχει δημιουργηθεί εκ του μηδενός, είτε με μεταφορά μάθησης είναι δύσκολο να ξεπεράσει τις προϋπάρχουσες πλατφόρμες ως προς την ακρίβεια αναγνώρισεων ή ακόμα και στο εύρος των κλάσεων που μπορούν να αναγνωριστούν. Στη συνέχεια παρουσιάζονται οι κύριες πλατφόρμες που χρησιμοποιούνται για την αναγνώριση τροφίμων καθώς επίσης και πλατφόρμες οι οποίες δεν επικεντρώνονται εξ ολοκλήρου στην αναγνώριση τροφίμων αλλά, εν τούτοις, μπορούν να αναγνωρίσουν τρόφιμα[81].

2.1.3.1 Vision AI

Το VisionAI είναι πλατφόρμα βαθιάς μάθησης η οποία ανήκει στη Google και υφίσταται ως υπηρεσία cloud από το 2016. Διαθέτει τη δυνατότητα να αναγνωρίσει χαρακτήρες γραμμάτων από εικόνες σε διάφορες γλώσσες ακόμα και αν εκείνες έχουν γραφτεί με το ανθρώπινο χέρι. Μπορεί να αναγνωρίσει αντικείμενα, πρόσωπα, γνωστά κτήρια και να διαχωρίσει υλικό που είναι ακατάλληλο για ανηλίκους[82].

2.1.3.2 Clarifai

Το Clarifai δημιουργήθηκε αρχικά για το διαγωνισμό οπτική αναγνώριση μεγάλης κλίμακας ImageNet (ILSVRC) το 2013. Έχει το χαρακτηριστικό ότι ο χρήστης μπορεί να επιλέξει διαφορετικά μοντέλα για την πρόβλεψη. Μπορεί να χρησιμοποιηθεί για αναγνώριση αντικειμένων, προσώπων, διαθέτει τη δυνατότητα να αναγνωρίσει και να διαχωρίσει εικόνες που είναι ακατάλληλες, αναγνώριση τροφών και χρωμάτων[83].

2.1.3.3 Amazon Rekognition

Το Amazon Rekognition είναι πλατφόρμα βαθιάς μάθησης που αναπτύχθηκε αρχικά από τη εταιρεία Obreus. Το 2016 η εταιρεία αυτή αγοράστηκε από την Amazon η οποία υποστηρίζει την πλατφόρμα από τότε. Έχει τη δυνατότητα να κάνει αναγνώριση αντικειμένων, προσώπων, γλώσσας πάνω σε έγγραφα, δραστηριότητες, και αν το περιεχόμενο μιας εικόνας είναι κατάλληλο ή όχι για ανηλίκους[84].

2.1.3.4 Azure's Computer Vision

Το Azure's Computer Vision δημιουργήθηκε από τη Microsoft και είναι πλατφόρμα που χρησιμοποιεί βαθιά μάθηση για μια σειρά από εφαρμογές. Έχει τη δυνατότητα να εξάγει κείμενο από εικόνες (Optical Character Recognition), να αναγνωρίσει εικόνες με αντικείμενα, πρόσωπα και να αντιληφθεί αν η εικόνα έχει περιεχόμενο ακατάλληλο για ανηλίκους. Μπορεί ακόμα να εντοπίσει την κίνηση των ανθρώπων σε βίντεο[85].

2.1.3.5 IBM Watson Visual Recognition

Το Watson Visual Recognition της IBM είναι πλατφόρμα αναγνώρισης αντικειμένων το οποίο μπορεί να ενσωματωθεί σε διάφορες εφαρμογές. Με τη χρήση βαθιάς μάθησης έχει τη δυνατότητα να αναγνωρίσει πρόσωπα, χρώματα, τροφές, ζώα, άψυχα αντικείμενα, τοπία και σκηνές. Δίνει τη δυνατότητα στο χρήστη να εκπαιδεύσει το δικό του νευρωνικό δίκτυο το οποίο θα διαθέτει τις κλάσεις που επιθυμεί ο ίδιος. Έχει εφαρμογές σε πεδία όπως στη εκπαίδευση, τη βιομηχανία, ασφαλιστικές υποδομές και εμπόριο[86].

2.1.3.6 Foodvisor

Το Foodvisor είναι εφαρμογή για κινητά το οποίο καταγράφει τις τροφές καθώς επίσης τις θρεπτικές τους αξίες. Σκοπός του είναι να βοηθήσει το χρήστη να χάσει κιλά, να ακολουθήσει συγκεκριμένη διαίτα και γενικά να ακολουθήσει υγιεινή διατροφή. Η εφαρμογή ενημερώνεται για το τι καταναλώνει ο χρήστης μέσω εικόνων από τα πιάτα πριν καταναλωθούν από το χρήστη. Χρησιμοποιεί βαθιά μάθηση προκειμένου να αναγνωρίσει τι πρόκειται να καταναλωθεί. Επιπλέον, η εφαρμογή υπολογίζει το βάρος του κάθε υλικού που καταναλώνεται. Προσπαθεί να υπολογίσει την απόσταση μεταξύ του κινητού και του πιάτου χρησιμοποιώντας τα δεδομένα της αυτόματης εστίασης της κάμερα του κινητού. Στη συνέχεια υπολογίζει την επιφάνεια κάθε τροφής εντός της εικόνας και από εκεί το όγκο της τροφής με βάση τον τύπο της τροφής. Εάν η εφαρμογή κάνει κάτι λάθος, τότε ο χρήστης μπορεί χειροκίνητα να το διορθώσει πριν γίνει η καταγραφή της τροφής στο ιστορικό. Αφού γίνει η καταγραφή, ο χρήστης λαμβάνει ως πληροφορία τα θρεπτικά στοιχεία της τροφής που καταναλώθηκε. Τέλος, ο χρήστης μπορεί να θέσει στόχους και να παρακολουθήσει την πρόοδο του μέσω της εφαρμογής[87].

2.1.3.7 Bitesnap

Το Bitesnap είναι πλατφόρμα που χρησιμοποιείται για να παρακολουθήσει ο χρήστης τι έχει καταναλώσει με το να αναγνωρίσει η εφαρμογή τι περιέχει το πιάτο. Έχει τη δυνατότητα να καταγράψει τις τροφές που κατανάλωσε ο χρήστης με τη χρήση βαθιάς μάθησης. Μπορεί να βοηθήσει το χρήστη να βάλει στόχους ως προς τι τροφές καταναλώνει καθώς η εφαρμογή διαθέτει τη δυνατότητα να καταμετρά θερμίδες και θρεπτικές ουσίες. Μπορεί να σαρώσει γραμμωτούς κώδικες (barcodes) συσκευασμένων τροφών και να αναλύει το περιεχόμενό τους. Οι καταγραφές μπορούν να εξαχθούν σε μορφή αρχείου csv έτσι ώστε ο χρήστης να μοιραστεί τις πληροφορίες με το γυμναστή ή το διαιτολόγο του[88].

2.1.3.8 Lose It

Το Lose It είναι πλατφόρμα που παρακολουθεί τις θερμίδες και τις θρεπτικές ουσίες που έχει καταναλώσει ο χρήστης. Σκοπός της είναι η βοήθεια του χρήστη να χάσει κιλά και να κάνει πιο υγιεινή ζωή. Μπορεί να σαρώσει γραμμωτούς κώδικες συσκευασιών τροφίμων για την ευκολότερη παρακολούθηση της κατανάλωσης τροφών. Για τον ίδιο σκοπό, μπορεί να καταγράψει σε ημερολόγιο τις τροφές που έχει καταναλώσει ο χρήστης μέσω λήψης φωτογραφιών των πιάτων που έχει καταναλώσει. Έχει τη δυνατότητα να θέσει στόχους και

να αναπτύξει δραστηριότητες έτσι ώστε να διευκολυνθεί η δίαιτα του χρήστη. Μπορεί να ενσωματωθεί σε συσκευές ευεξίας και γυμναστικής όπως το Fitbit και το Google Fit[89].

2.1.3.9 Calorie Mama

Το Calorie Mama είναι εφαρμογή για κινητά το οποίο μετρά θερμίδες από φαγητά που έχει καταναλώσει ο χρήστης μέσω φωτογραφιών που λαμβάνει από το κινητό του. Η εφαρμογή διαθέτει σύστημα βαθιάς μάθησης το οποίο μπορεί να αναλύσει εικόνες με τροφές και να βγάλει συμπεράσματα ως προς τις διατροφικές της αξίες. Μπορεί να αναγνωρίσει τροφές από διαφορετικές κουζίνες. Ο χρήστης μπορεί να επιβεβαιώσει ή να απορρίψει την αναγνώριση του συστήματος. Αν η αναγνώριση είναι σωστή τότε το σύστημα καταγράφει την αναγνώριση[90].

2.2 Βάσεις Δεδομένων

Στην εκπαίδευση ενός συνελκτικού νευρωνικού δικτύου για την αναγνώριση αντικειμένων και ειδικότερα τροφίμων απαιτούνται μεγάλες βάσεις δεδομένων που περιέχουν εικόνες. Παρόλα αυτά, βάσεις δεδομένων που επικεντρώνονται στην ανάπτυξη μοντέλων για την ταξινόμηση τροφίμων εξακολουθούν να μην καλύπτουν το εύρος των απαιτήσεων ενός τέτοιου εγχειρήματος. Η ανάπτυξη νέας βάσης δεδομένων για τροφές που να καλύπτει όλο το φάσμα των τροφών και της κατάστασής τους αποτελεί αντικείμενο ερευνών. Για την συλλογή εικόνων οι ερευνητές χρησιμοποιούν μεταξύ άλλων μηχανές αναζήτησης του διαδικτύου και μέσα κοινωνικής δικτύωσης. Οι σημαντικότερες βάσεις δεδομένων που υπάρχουν για την αναγνώριση αντικειμένων προσδίδοντας περισσότερη βαρύτητα σε εκείνες που αφορούν τροφές παρουσιάζονται στη συνέχεια[91].

2.2.1 Food-101

Το Food-101 είναι βάση δεδομένων η οποία δημιουργήθηκε το 2014 από τους Lukas Bossard, Matthieu Guillaumin, Luc Van Gool. Διαθέτει 101.000 εικόνες διαχωρισμένες σε 101 κατηγορίες τροφών. Κάθε κατηγορία διαθέτει 750 εικόνες οι οποίες προορίζονται για εκπαίδευση και 250 για το τεστ του μοντέλου. Η κάθε εικόνα έχει μέγεθος 512x512 εικονοστοιχείων. Η βάση δεδομένων έχει θόρυβο που προέρχεται κυρίως από λάθος ταυτοποιήσεις και την χρωματική απόχρωση στις εικόνες[92].

2.2.2 ImageNet

Το ImageNet είναι μια από τις σημαντικότερες βάσεις δεδομένων με εικόνες που υπάρχουν για την εκπαίδευση νευρωνικών δικτύων. Δημιουργήθηκε από τους Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, Li Fei-Fei και παρουσιάστηκε αρχικά το 2009. Διαθέτει πάνω από 14 εκατομμύρια εικόνες, οι οποίες έχουν συλλεχθεί από το διαδίκτυο, ταξινομημένες σε πάνω από 20000 κατηγορίες η κάθε μία εκ των οποίων περιέχει πάνω από 500 εικόνες. Το ImageNet συνεχίζει να αναβαθμίζεται και να προσθέτει νέες κλάσεις. Διαθέτει ένα μεγάλο εύρος κλάσεων που αφορά τροφές[93].

2.2.3 ChineseFoodNet

Το ChineseFoodNet δημιουργήθηκε από τους Xin Chen, Yu Zhu, Hua Zhou, Liang Diao, Dongyan Wang το 2017. Διαθέτει πάνω 180000 εικόνες ταξινομημένες 208 κατηγορίες. Επικεντρώνεται στις κινεζικές τροφές και οι εικόνες έχουν συλλεχθεί από το διαδίκτυο καθώς επίσης από εικόνες με συνταγές και φωτογραφίες από κινητά[94].

2.2.4 FoodX-251

Το FoodX-251 δημιουργήθηκε το 2019 από τους Parneet Kaur, Karan Sikka, Weijun Wang, Serge Belongieq, Ajay Divakaran. Πρόκειται για μια βάση δεδομένων που διαθέτει εικόνες με

τροφές οι οποίες έχουν συλλεχθεί από το διαδίκτυο. Διαθέτει 251 κατηγορίες τροφών με 158000 εικόνες στο σύνολο εκ των οποίων οι 118000 προορίζονται για εκπαίδευση του νευρωνικού δικτύου και 40000 για την επιβεβαίωση της σωστής λειτουργίας του[95].

2.2.5 UECFood100 και UECFood256

Οι βάσεις δεδομένων UECFood100 και UECFood256 δημιουργήθηκαν από τους Yuji Matsuda και Keiji Yanai το 2012 και 2014 αντίστοιχα. Περιέχουν εικόνες με τροφές που συναντώνται κυρίως στην Ιαπωνία. Το UECFood100 διαθέτει πάνω από 14000 εικόνες ταξινομημένες 100 κατηγορίες τροφών και πιάτων ενώ το UECFood256 διαθέτει πάνω από 31000 εικόνες ταξινομημένες σε 256 κατηγορίες. Το UECFood256 έχει χρησιμοποιηθεί για τη δημιουργία εφαρμογής σε κινητά που κάνει αναγνώριση Γιαπωνέζικων τροφών[96].

2.2.6 Food524DB

Το Food524DB δημιουργήθηκε από τους Raimondo Schettini, Gianluigi Ciocca, Paolo Napolitano το 2017. Είναι η μεγαλύτερη βάση δεδομένων για τροφές η οποία είναι ελεύθερη στο ευρύ κοινό. Αποτελείται από περισσότερες από 247000 εικόνες που ανήκουν σε 524 κατηγορίες. Δημιουργήθηκε από τη συνένωση των βάσεων δεδομένων VIREO, Food-101, Food50 και παραλλαγής του UECFOOD256[97].

2.2.7 EgocentricFoods

Το EgocentricFoods είναι βάση δεδομένων η οποία έχει αναπτυχθεί από τους Boloñis M. and Radeva P. του πανεπιστημίου της Βαρκελώνης το 2016. Διαθέτει εικόνες τροφών από την οπτική γωνία που τις βλέπει ο άνθρωπος καθώς οι εικόνες λήφθηκαν από κάμερες που φορούσαν άνθρωποι. Μπορεί να χρησιμοποιηθεί για αναγνώριση τροφών καθώς επίσης και για τον εντοπισμό τους εντός της εικόνας[98].

2.2.8 Recipe1M+

Το Recipe1M+ είναι βάση δεδομένων με τροφές και συνταγές μαγειρικής που δημιουργήθηκε από τους Javier Marin, Aritro Biswas, Ferda Ofli, Nicholas Hynes, Amaia Salvador, Yusuf Aytar¹, Ingmar Weber, Antonio Torralba σε συνεργασία των οργανισμών Massachusetts Institute of Technology, Qatar Computing Research Institute και HBKU Universitat Politècnica de Catalunya. Για την δημιουργία της, αρχικά, συλλέχθηκαν πολλαπλές εικόνες από συνταγές που βρίσκονταν σε γνωστές ιστοσελίδες μαγειρικής. Αυτό είχε σαν αποτέλεσμα να συλλεχθούν πάνω από 1 εκατομμύριο συνταγές μαγειρικής και πάνω από 800.000 εικόνες με τροφές. Στη συνέχεια, για κάθε συνταγή συλλέχθηκαν από το διαδίκτυο επιπλέον εικόνες οι οποίες ήταν συνολικά 13 εκατομμύρια. Πρόκειται για την μεγαλύτερη βάση δεδομένων που αφορά αποκλειστικά τροφές και συνταγές[99].

2.3 Πλατφόρμες για Βαθιά Μάθηση

Μια καλή πλατφόρμα για βαθιά μάθηση διαθέτει επιφάνεια εργασιών, βιβλιοθήκες, εργαλεία και εφαρμογές με τις οποίες μπορούν να εφαρμοστούν πολύπλοκα μοντέλα χωρίς να χρειάζεται απαραίτητα ο σχεδιασμός της αρχιτεκτονικής των δικτύων αυτών ή ακόμα και γνώση των μαθηματικών που χρησιμοποιούν. Προσφέρουν δομικά στοιχεία με τα οποία μπορεί να σχεδιαστεί, εκπαιδευτεί και να δοκιμαστεί ένα νευρωνικό δίκτυο. Οι περισσότερες πλατφόρμες βασίζονται σε κάρτες γραφικών (GPU) για να εκτελέσουν τους υπολογισμούς καθώς προσφέρουν μεγαλύτερη ταχύτητα στην εκπαίδευση του νευρωνικού δικτύου. Κάποια επιπλέον χαρακτηριστικά μιας καλής πλατφόρμας για βαθιά μάθηση είναι τα ακόλουθα:

- Βελτιστοποιημένη απόδοση.

- Ευκολία στη κατανόηση και το προγραμματισμό της.
- Είναι δημοφιλές
- Έχει υποστήριξη με τακτικές αναβαθμίσεις
- Έχει τη δυνατότητα για παράλληλες διαδικασίες έτσι ώστε να μειωθούν οι υπολογισμοί.

Με βάση τα προηγούμενα χαρακτηριστικά παρουσιάζονται οι ακόλουθες δημοφιλείς πλατφόρμες[69].

2.3.1 Tensorflow

Η Tensorflow είναι μια βιβλιοθήκη ανοιχτού κώδικα η οποία έχει ένα εύρος εφαρμογών με επίκεντρο κυρίως σε εκείνη της μηχανικής μάθησης. Αναπτύχθηκε αρχικά ως εργαλείο για εσωτερική χρήση από τη Google με σκοπό την έρευνα και ανάπτυξη εφαρμογών στο πεδίο της μηχανικής μάθησης. Έχει δημιουργηθεί με τη γλώσσα προγραμματισμού C++. Διαθέτει λειτουργίες όπως μαθηματικές συναρτήσεις και τη δυνατότητα χειραγώγησης πολυδιάστατων πινάκων. Είναι σχεδιασμένη για να εκτελεί υπολογισμούς μέσω GPU αλλά μπορεί να κάνει την ίδια εργασία σε χαμηλότερες ταχύτητες μέσω CPU καθώς επίσης σε κινητές συσκευές. Βασικά μειονεκτήματα της αποτελούν το γεγονός ότι η διεπαφή προγραμματισμού εφαρμογών είναι χαμηλού επιπέδου και η χρήση της παρουσιάζει δυσκολία για τους χρήστες που έχουν μικρή εμπειρία[70].

2.3.2 Pytorch

Η Pytorch έχει δημιουργηθεί με τις γλώσσες προγραμματισμού Python, C και CUDA. Αναπτύχθηκε από τη Facebook το 2017. Είναι ευέλικτο και έχει τη δυνατότητα ανάπτυξης πολύπλοκων αρχιτεκτονικών. Επειδή έχει τη δυνατότητα για προσαρμογές, έχει γίνει δημοφιλής για την ανάπτυξη εφαρμογών στο εμπορικό και ακαδημαϊκό τομέα. Επίσης, η Pytorch υποστηρίζει μοντέλα από το ανοιχτό ανταλλακτήριο νευρωνικού δικτύου (ONNX) το οποίο επιτρέπει την ανταλλαγή μοντέλων ανάμεσα σε διαφορετικές πλατφόρμες[71].

2.3.3 Caffe και Caffe2

Η Caffe είναι μια από τις παλαιότερες και πιο δημοφιλείς πλατφόρμες ανάπτυξης εφαρμογών βαθιάς μάθησης. Αναπτύχθηκε από τον Yangqing Jia στο Ινστιτούτο Τεχνητής Νοημοσύνης στο Μπερκλεϋ(BAIR). Υποστηρίζει την ανάπτυξη νευρωνικών δικτύων ανά επίπεδο. Κάθε στοιβάδα θα μπορούσε να αναπτυχθεί στη γλώσσα προγραμματισμού Python. Μπορεί επίσης να αναπτυχθεί σε C++CUDA. Η Caffe επίσης διαθέτει επιφάνεια εργασιών για το Matlab. Ο ίδιος ερευνητής και η ομάδα του στη Facebook ανέπτυξε το 2018 το Caffe2. Αυτή η έκδοση είναι περισσότερο εύχρηστη. Επιπλέον, μπορεί να υποστηρίξει εφαρμογές για κινητές συσκευές. Τέλος, η Caffe2 υποστηρίζει το ONNX που επιτρέπει την εύκολη ενσωμάτωση μοντέλων σε άλλες πλατφόρμες[72].

2.3.4 Keras

Το Keras είναι μια βιβλιοθήκη για νευρωνικά δίκτυα της οποίας οι υπολογισμοί εκτελούνται πάνω σε δημοφιλείς πλατφόρμες όπως το Tensorflow και Theano. Έχει δημιουργηθεί με τη γλώσσα προγραμματισμού Python και θεωρείται ότι η διεπαφή προγραμματισμού εφαρμογών που διαθέτει είναι υψηλού επιπέδου. Έχει φιλικό προς το χρήστη περιβάλλον. Επίσης, είναι εύκολο να προστεθούν νέα χαρακτηριστικά. Δεν απαιτεί ιδιαίτερα σημαντικές γνώσεις από το χρήστη για να χρησιμοποιηθεί[73].

2.3.5 Theano

Το Theano αναπτύχθηκε αρχικά το 2007 από το πανεπιστήμιο του Μόντρεαλ. Συντάσσει μαθηματικές εκφράσεις, μεταξύ άλλων, μέσω της βιβλιοθήκης NumPy της Python. Έχει τη δυνατότητα να τρέξει σε πολλές κάρτες γραφικών και κεντρικούς επεξεργαστές. Μετά το Νοέμβριο του 2017 το Theano έχει πάψει να αναβαθμίζεται ενεργά αλλά συνεχίζεται η συντήρησή της ενώ το Keras εκτελεί μεταξύ άλλων και το Theano[74].

2.3.6 MXNet

Το MXNet είναι μια από τις πιο δημοφιλείς βιβλιοθήκες για βαθιά μάθηση κυρίως στο εμπορικό πεδίο. Αυτό οφείλεται στη φορητότητα, την ευκολία στην ανάπτυξη εφαρμογών καθώς επίσης στη δυνατότητα της εισαγωγής νέων χαρακτηριστικών που του επιτρέπει την προσαρμογή σε ένα εύρος υπολογιστικών συστημάτων με διαφορετικές κάρτες γραφικών και κεντρικούς επεξεργαστές. Η φορητότητα του υποστηρίζει ένα μεγάλο εύρος "έξυπνων" συσκευών όπως κινητές συσκευές, συσκευές IoT και για συστήματα χωρίς διαμετακομιστή. Είναι φιλικό για εφαρμογές και υπηρεσίες που έχουν σχέση με το νέφος. Επιπλέον υποστηρίζει το ONNX που σημαίνει ότι είναι πιο εύκολη η ενσωμάτωση μοντέλων σε άλλες πλατφόρμες[75].

2.3.7 DeepLearning4j

Η DeepLearning4j δημιουργήθηκε από τους Alex D. Black, Adam Gibson, Vyacheslav Kokorin, Josh Patterson. Έχει δημιουργηθεί με τη γλώσσα προγραμματισμού Java και χρησιμοποιεί τη βιβλιοθήκη για πολυδιάστατους πίνακες ND4J. Η λειτουργία του υποστηρίζεται από κάρτες γραφικών (GPU) και κεντρικούς επεξεργαστές(CPU). Η DeepLearning4j έχει το χαρακτηριστικό ότι φορτώνει δεδομένα και εκτελεί τον αλγόριθμο εκπαίδευσης ως ξεχωριστές διεργασίες, πράγμα που το καθιστά ευέλικτο για τον χρήστη. Μπορεί να χρησιμοποιήσει διαφορετικούς τύπους δεδομένων όπως εικόνες, αρχεία .csv και απλό κείμενο[76].

2.4 Βιβλιογραφία Μελετών

Μελετώντας την βιβλιογραφία που αφορά συστήματα τα οποία έχουν την δυνατότητα αναγνώρισης τροφίμων αποδεικνύεται ότι εκείνα που χρησιμοποιούν συνελκτικά νευρωνικά δίκτυα εντός της αρχιτεκτονικής τους είναι ιδιαίτερα αποτελεσματικά. Αυτό συμβαίνει γιατί έχουν πολύ υψηλή ακρίβεια ταξινόμησης εικόνων, υψηλή ταχύτητα ενώ ταυτόχρονα χρησιμοποιούν μικρότερη υπολογιστική ισχύ σε σχέση με άλλες μεθόδους.

Στη συνέχεια παρουσιάζονται μελέτες που δείχνουν την αποτελεσματικότητα των συνελκτικών νευρωνικών δικτύων, οι οποίες πραγματοποιήθηκαν την τελευταία δεκαετία σε διάφορα επιστημονικά περιοδικά. Κάθε μελέτη χρησιμοποιεί διαφορετικές αρχιτεκτονικές συνελκτικών νευρωνικών δικτύων με διαφορετικές βάσεις δεδομένων.

Πίνακας 4: Τίτλοι Μελετών.

Αριθμός Μελέτης	Μελέτη	Τόπος Παρουσίασης Ερευνών	Συγγραφείς
Μελέτη 1	Deep learning-based food recognition.	Stanford University	Qian Yu, Dongyuan Mao, Jingfan Wang
Μελέτη 2	Application of deep learning in food: a review.	Comprehensive Reviews in Food Science and Food Safety	Lei Zhou, Chu Zhang, Fei Liu, Zhengjun Qiu, and Yong He
Μελέτη 3	Use of different food image recognition platforms in	JMIR Formative Research	Stephanie Van Asbroeck, Christophe Matthys

	dietary assessment: comparison study.		
Μελέτη 4	A Deep Learning based Food Recognition System for Lifelog Images.	7th International Conference on Pattern Recognition Applications and Methods, ICPRAM 2018, Funchal, Madeira - Portugal, January 16-18, 2018	Binh T. Nguyen, Duc-Tien Dang-Nguyen, Tien X. Dang, Thai Phat, Cathal Gurrin
Μελέτη 5	FoodAI: Food image recognition via deep learning for smart food logging.	KDD '19: The 25th ACM SIGKDD Conference on Knowledge Discovery and Data Mining Anchorage AK USA. 2019	Doyen SAHOO Singapore, Hao WANG, Ke SHU, Xiongwei WU, Hung LE
Μελέτη 6	Deep learning approaches in food recognition.	PCI 2020: 24th Pan-Hellenic Conference on Informatics Athens Greece November 20 - 22, 2020	Chairi Kiourt, George Pavlidis and Stella Markantonatou
Μελέτη 7	Food classification from images using convolutional neural networks.	IEEE Region 10 Annual International Conference, Proceedings/TENCON - USA	David J. Attokaren, Ian G. Fernandes, A. Sriram, Y.V. Srinivasa Murthy, and Shashidhar G. Koolagudi
Μελέτη 8	Where and what am i eating? image-based food menu recognition.	Computer Vision – ECCV 2018 Workshops Munich, Germany	Marc Bolanos, Marc Valdivia1, Petia Radeva

2.4.1 Μελέτη 1 – Αναγνώριση Τροφίμων με Βάση τη Βαθιά Μάθηση

Σε αυτή τη μελέτη χρησιμοποιήθηκαν για την εκπαίδευση οι βάσεις δεδομένων UPMC-FOOD-101 και ETHZ-FOOD-101 ενώ για τον έλεγχο της αποτελεσματικότητας των εκπαιδευμένων νευρωνικών δικτύων η ETHZ-FOOD-101 [100].

Εκπαιδευτήκαν και στη συνέχεια εξετάστηκαν οι αρχιτεκτονικές AlexNet, δύο διαφορετικά νευρωνικά δίκτυα αρχιτεκτονικής GoogleNet που έχουν εκπαιδευτεί με 10000 και 30000 επαναλήψεις αντίστοιχα. Επίσης εκπαιδευτήκαν και εξετάστηκαν οι αρχιτεκτονικές Inception-V3 και Inception-ResNet[100].

Αρχικά, έγινε επεξεργασία στις εικόνες εισόδου γιατί το φόντο τους διέφερε από εικόνα σε εικόνα λόγω χρώματος ή φωτισμού. Για να επιτευχθεί αυτό, έγινε επεξεργασία στις έγχρωμες εικόνες(RGB) μέσω της μεθόδου Gray World όπου ορίζει ότι ο μέσος όρος των R,G,B τιμών είναι ίσο με μια αντίστοιχη τιμή γκριζου χρώματος. Ο αλγόριθμος της μεθόδου Gray World παρουσιάζεται κάτωθι με τα α, β, γ να είναι συντελεστές κλίμακας [100]:

$$(\alpha R, \beta G, \gamma B) \rightarrow \left(\frac{\alpha R}{\sum_i R}, \frac{\beta G}{\sum_i G}, \frac{\gamma B}{\sum_i B} \right) \quad (37)$$

Στη συνέχεια, εφαρμόστηκε αλγόριθμος ισοστάθμισης ιστογράμματος (Histogram Equalization Algorithm) για να αυξηθεί η φωτεινότητα και αντίθεση στις εικόνες. Η αρχική επεξεργασία στις εικόνες βοηθάει να αντιμετωπιστεί το πρόβλημα ότι οι εικόνες έχουν διαφορετικό φόντο μειώνοντας τον χρόνο που χρειάζεται για την εκπαίδευσή τους και αυξάνοντας την ακρίβεια της εξόδου[101].

Χρησιμοποιήθηκαν τα ήδη εκπαιδευμένα νευρωνικά δίκτυα Inception-V3 και Inception-ResNet και επανεκπαιδευτήκαν αλλάζοντας το τελευταίο επίπεδο του στρώματος του νευρωνικού δικτύου μέσω της μεταφοράς μάθησης. Επειδή η απλή μεταφορά μάθησης εξήγαγε αποτελέσματα που δεν ξεπερνούσαν το 40%, έγινε εκπαίδευση με ολόκληρη τη

στοιβάδα των δικτύων για καλύτερα αποτελέσματα. Από τη βάση δεδομένων, το 80% χρησιμοποιήθηκε για την εκπαίδευση των νευρωνικών δικτύων και το υπόλοιπο 20% για την εξέταση των αποτελεσμάτων. Η διαδικασία εκπαίδευσης διαχωρίστηκε σε δύο κύριες διαδικασίες. Στη πρώτη, ο ρυθμός μάθησης για το τελευταίο επίπεδο των δικτύων ορίστηκε σε 0.01 ενώ το μέγεθος κάθε παρτίδας που κάθε φορά εκπαιδεύτηκε είναι 101 εικόνες. Στη δεύτερη που αφορά ολόκληρη τη στοιβάδα των δικτύων ορίστηκε ο ρυθμός μάθησης σε 0.001 και το μέγεθος της κάθε παρτίδας σε 7. Πέρα από τη σύγκριση που γίνεται ανάμεσα στο Inception-V3 και Inception-ResNet έγινε επίσης σύγκριση ανάμεσα σε δίκτυα που εκπαιδεύτηκαν με εικόνες που είχαν υποστεί επεξεργασία και δίκτυα που εκπαιδεύτηκαν με εικόνες που δεν υπέστησαν επεξεργασία. Ως συμπέρασμα βγήκε ότι το Inception-ResNet έχει καλύτερα αποτελέσματα σε σχέση με το Inception-V3 όπως δείχνει ο πίνακας 4 [102].

Πίνακας 5: Αποτελέσματα της Μελέτης 1.

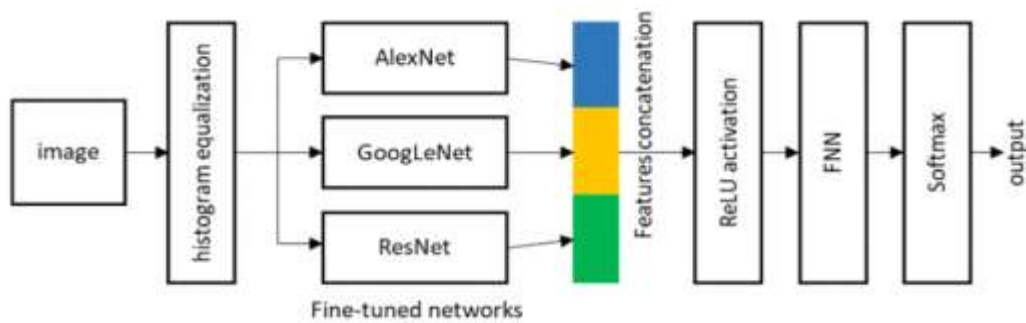
Μέθοδος	Ακρίβεια TOP-1	Ακρίβεια TOP-5
AlexNet	56.40%	NA
GoogLeNet (10,000 iterations)	70.2%	91.0%
GoogLeNet (30,000 iterations)	77.4%	93.7%
Inception V3 (last layer training)	35.32%	62.97%
Inception-ResNet (last layer training)	42.69%	72.78%
Inception V3 (full layer training)	70.60%	90.91%
Inception-ResNet (full layer training)	72.55%	91.31%

2.4.2 Μελέτη 2 – Επισκόπηση Εφαρμογών Βαθιάς Μάθησης στα Τρόφιμα

Η μελέτη αυτή έκανε επισκόπηση ως προς τη βιβλιογραφία που υπήρχε για την αναγνώριση τροφίμων και δείχνει ότι η χρήση νευρωνικών δικτύων για την αναγνώριση τροφίμων έχει καλύτερα αποτελέσματα σε σχέση με άλλες μεθόδους μηχανικής μάθησης καθώς επίσης και εκείνες που γίνεται χειροκίνητα εξαγωγή χαρακτηριστικών (manual feature extractors).

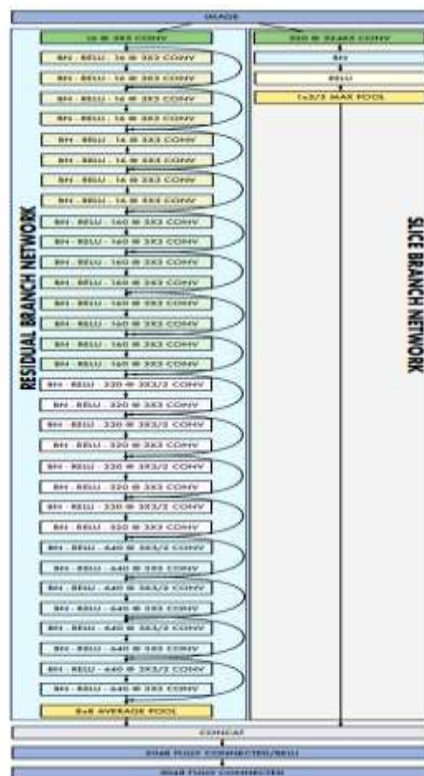
Στις εργασίες αυτές χρησιμοποιήθηκαν τρεις βάσεις δεδομένων για να εκπαιδευτούν και συνέχεια να εξεταστούν τα νευρωνικά δίκτυα, το Food-101, UECFood-256 και UECFood-100. Χρησιμοποιήθηκαν ήδη προϋπάρχουσες αρχιτεκτονικές όπως το ResNet, ResNet-50, Inception, Inception-v3+FPCNN, Inception-v3 καθώς επίσης νέες που ανέπτυξαν και παρουσίασαν οι συγγραφείς τους όπως WISer, DeepFood, FoodNet, και DeepFoodCam. Κάποιες από αυτές τις νέες αρχιτεκτονικές παρουσιάζονται στη συνέχεια.

Το FoodNet είναι συνδυασμός των αρχιτεκτονικών AlexNet, GoogleNet και ResNet. Αρχικά οι εικόνες περνάνε από τον αλγόριθμο ισοστάθμισης ιστογράμματος. Στη συνέχεια τα δεδομένα περνάνε όλα παράλληλα από τα μοντέλα AlexNet, GoogleNet και ResNet. Οι χαρακτηριστικές που έχουν εξαχθεί από κάθε αρχιτεκτονική συνενώνονται και περνάνε από ReLU όπου με τη σειρά του περνάει από ένα επίπεδο πλήρως συνδεδεμένων νευρώνων (FNN) και μετά από ένα στρώμα Softmax πριν οδηγηθεί στη έξοδο. Η αρχιτεκτονική του FoodNet παρουσιάζεται οπτικά από την κάτωθι εικόνα[103].



Εικόνα 33: Αρχιτεκτονική του νευρωνικού δικτύου FoodNet[103].

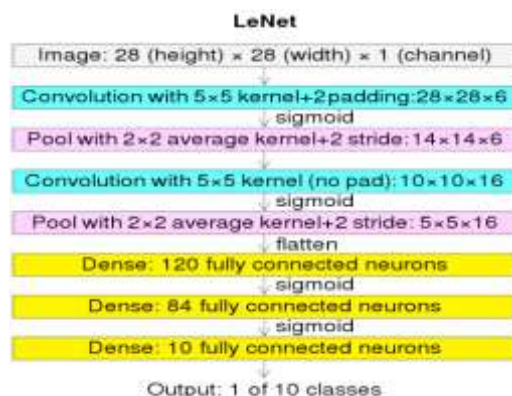
Το WISeR (Wide-Slide Residual Network) είναι νευρωνικό δίκτυο που δημιουργήθηκε με σκοπό να εντοπίσει δομικά χαρακτηριστικά στις τροφές που βρίσκονται στις εικόνες. Το δίκτυο αποτελείται από δυο διακλαδώσεις οι οποίες λειτουργούν παράλληλα η μια από την άλλη. Η μία διακλάδωση είναι ένα δίκτυο ResNet ενώ η άλλη είναι συνέλιξης φέτας (Slice) όπου το φίλτρο της συνέλιξης έχει πλάτος ίσο με εκείνο της εικόνας εισόδου. Οι χάρτες χαρακτηριστικών από τις δύο διακλαδώσεις προστίθενται μεταξύ τους. Στη συνέχεια περνάνε από ένα επίπεδο με 2048 πλήρως συνδεδεμένους νευρώνες ακολουθούμενο από ReLU και τέλος ακόμα ένα επίπεδο με πλήρως συνδεδεμένους νευρώνες πριν την έξοδο. Το σχεδιάγραμμα της αρχιτεκτονικής του νευρωνικού δικτύου WISeR παρουσιάζεται στη κάτωθι εικόνα[104]:



Εικόνα 34: Αρχιτεκτονική του νευρωνικού δικτύου WISeR[104].

Το DeepFood βασίστηκε για την δημιουργία του σε 3 προϋπάρχουσες αρχιτεκτονικές συνελικτικών νευρωνικών δικτύων. Αυτές οι αρχιτεκτονικές είναι οι LeNet-5 του οποίου η αρχιτεκτονική παρουσιάζεται στην εικόνα 35, AlexNet και GoogLeNet. Τα δεδομένα που πρόκειται να χρησιμοποιηθούν για την εκπαίδευση περνάνε αρχικά από το δίκτυο LeNet-5.

Στη συνέχεια ο χάρτη χαρακτηριστικών από το LeNet-5 περνάει στο AlexNet και τέλος, περνάει από το GoogleNet από το οποίο εξάγεται ο τελικός χάρτης χαρακτηριστικών[105].



Εικόνα 35: Αρχιτεκτονική του νευρωνικού δικτύου LeNet-5[105].

Πίνακας 6: Παρουσιάζεται η αποτελεσματικότητα μοντέλων αναγνώρισης τροφίμων με τη χρήση διαφορετικών βάσεων δεδομένων[106].

Database	Μοντέλο	Top-1%	Top-5%
Food-101	Author defined[107]	65.40	89.58
	CNN-FOOD	70.41	/
	Multitask	72.11	/
	FoodNet	72.12	91.61
	DeepFood	77.40	93.70
	Inception Module	77.00	94.00
	ResNet	78.50	94.10
	ResNet-50	82.54	95.79
	Inception-v3+FPCNN	87.96	/
UECFood-256	Inception V3 (Hassannejad et al., 2016)	88.28	96.88
	wide-slice residual networks (WiSeR) (Martinel, Foresti, & Micheloni, 2018)	90.27	98.71
	DeepFood	54.70	81.50
	Author defined [107]	60.00	85.00
	Inception Module	63.60	87.00
	DeepFoodCam	63.77	85.82
	CNN-FOOD	67.57	88.97
	ResNet	71.20	91.10
	ResNet-50	71.70	91.33
UECFood-100	Inception V3	76.17	92.58
	Inception-v3+FP-CNN	78.60	/
	WiSeR	83.15	95.45
	DeepFoodCam	72.26	92.00
	DeepFood	76.3	94.60
	Inception Module	76.30	94.60

	CNN-FOOD	78.48	94.85
	ResNet	80.60	95.90
	Inception V3	81.45	97.27
	MultiTaskCNN	82.12	97.29
	Inception-v3+FP-CNN	86.51	/
	WISeR	89.58	99.23

Από τα αποτελέσματα της μελέτης τα οποία παρουσιάζονται στο πίνακα 5 παρατηρείται ότι η αρχιτεκτονική WISeR είχε τα καλύτερα αποτελέσματα καθώς η top-1 ακρίβειά του ήταν η υψηλότερη για κάθε βάση δεδομένων που χρησιμοποιήθηκε. Το ίδιο ισχύει και όταν συγκρίνουμε τη top-5 ακρίβεια κάθε αρχιτεκτονικής ως προς και τις τρεις βάσεις δεδομένων που εξετάστηκαν[106].

2.4.3 Μελέτη 3 - Σύγκριση Διαφορετικών Πλατφορμών Αναγνώρισης Τροφίμων

Στη μελέτη αυτή έγινε σύγκριση σε δημοφιλείς πλατφόρμες αναγνώρισης τροφίμων ως προς τη δυνατότητα τους να αναγνωρίσουν σωστά τρόφιμα και ποτά υπό διαφορετικές οπτικές γωνίες και φωτισμό. Για την λήψη των φωτογραφιών χρησιμοποιήθηκε η ταμπλέτα Galaxy Tab A6 της Samsung. Το αρχικό μέγεθος κάθε εικόνας είναι 1920×1920 pixels τα οποία στη συνέχεια προσαρμόστηκαν η κάθε μια σε μέγεθος 544×544 pixels. Αρχικά επιλέχθηκαν 5 τροφές οι οποίες μπορούν να καταναλωθούν χωρίς να είναι απαραίτητη η επεξεργασία τους. Στη συνέχεια, επιλέχθηκαν 5 ακόμα επεξεργασμένες τροφές. Επιπλέον επιλέχθηκαν 7 ακόμα ποτά. Τέλος, 12 πιάτα με ανάμικτο φαγητό επιλέχθηκαν με το καθένα να διαθέτει διαφορετικά υλικά παρασκευής. Κάθε τροφή και ποτό φωτογραφήθηκε 6 φορές κάθε φορά σε διαφορετική τοποθεσία[108].

Ο κάτωθι πίνακας συγκρίνει την ακρίβεια για κάθε πλατφόρμα αναγνώρισης υπό διαφορετικές συνθήκες.

Πίνακας 7: Σύγκριση αποτελεσματικότητας για τις σημαντικότερες πλατφόρμες αναγνώρισης τροφίμων υπό διαφορετικές συνθήκες όσον αφορά το φωτισμό και τη γωνία[108].

Platform	Ideal, n (%)	Bad lighting, n (%)	Nonstandard container, n (%)	Clutter, n (%)	Unspecified angle, n (%)	Real-life, n (%)	Total, n (%)
Top 1 accuracy							
LogMeal	8 (25.8)	6 (19.4)	5 (16.1)	6 (19.4)	12 (38.7)	8 (25.8)	45 (24.2)
Clarifai	12 (38.7)	9 (29.0)	12 (38.7)	8 (25.8)	17 (54.8)	13 (41.9)	71 (38.2)
Google Vision	2 (6.5)	3 (9.7)	1 (3.2)	0 (0)	7 (22.6)	4 (12.9)	17 (9.1)
IBM Watson Recognition	10 (32.3)	5 (16.1)	5 (16.1)	7 (22.6)	8 (25.8)	12 (38.7)	47 (25.3)
Calorie Mama API	20 (64.5)	16 (51.6)	18 (58.1)	19 (61.3)	24 (77.4)	20 (64.5)	117 (62.9)

Bitesnap	16 (51.6)	11 (35.5)	13 (41.9)	15 (45.2)	21 (67.7)	16 (51.6)	91 (48.9)
Foodvisor	10 (32.3)	13 (41.9)	15 (48.4)	16 (51.6)	19 (61.3)	13 (41.9)	86 (46.2)
Total	78 (35.9)	63 (29.0)	69 (32.9)	70 (32.3)	108 (49.8)	86 (39.6)	— ^b
Top 5 accuracy							
LogMeal	14 (45.2)	13 (41.9)	10 (32.3)	14 (45.2)	17 (54.8)	14 (45.2)	82 (44.1)
Clarifai	21 (67.7)	19 (61.3)	18 (58.1)	18 (58.1)	23 (74.2)	20 (64.5)	119 (64.0)
Google Vision	7 (22.6)	7 (22.6)	2 (6.5)	3 (9.7)	16 (51.6)	10 (32.3)	45 (24.2)
IBM Watson Recognition	15 (48.4)	10 (32.3)	10 (32.3)	10 (32.3)	20 (64.5)	16 (51.6)	81 (43.5)
Calorie Mama API	27 (87.1)	28 (90.3)	25 (80.6)	26 (83.9)	29 (93.5)	28 (90.3)	163 (87.6)
Bitesnap	23 (74.2)	20 (64.5)	20 (64.5)	21 (67.7)	28 (90.3)	20 (64.5)	132 (71.0)
Foodvisor	19 (61.3)	19 (61.3)	24 (77.4)	21 (67.7)	26 (83.9)	24 (77.4)	133 (71.5)
Total	126 (58.1)	116 (53.5)	109 (51.9)	113 (52.1)	159 (73.3)	132 (60.8)	—

Επίσης, ο κάτωθι πίνακας συγκρίνει την ακρίβεια κάθε πλατφόρμας στην ικανότητα του να αναγνωρίσει διαφορετικές κατηγορίες τροφών και ποτών.

Πίνακας 8: Σύγκριση της αποτελεσματικότητας που έχουν οι σημαντικότερες πλατφόρμες αναγνώρισης για διαφορετικού τύπου τροφές και ποτά[108].

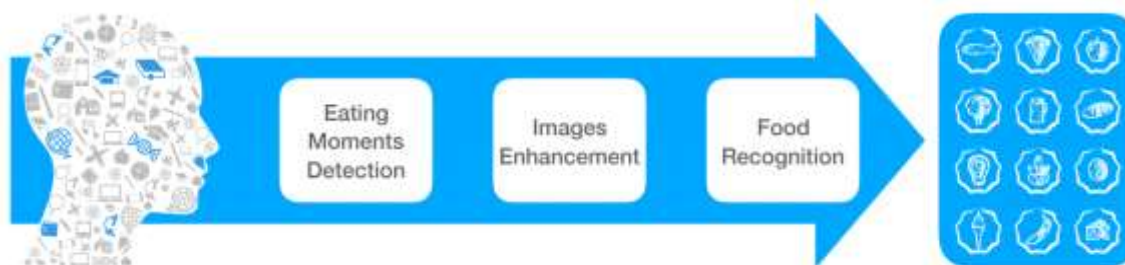
Πλατφόρμα	Simple plain		Simple processed		Beverages		Mixed dishes	
	Top 1	Top 5	Top 1	Top 5	Top 1	Top 5	Top 1	Top 5
Bitesnap	11 (36.6)	25 (83.3)	14 (46.6)	18 (60.0)	22 (53.6)	33 (80.4)	44 (61.1)	56 (77.7)
Calorie Mama API	22 (73.3)	29 (96.6)	23 (76.6)	31 (100)	19 (46.3)	29 (70.7)	50 (69.4)	70 (97.2)
Clarifai	4 (13.3)	13 (43.3)	14 (46.6)	27 (90.0)	17 (41.4)	30 (73.1)	36 (50.0)	49 (68.1)
Foodvisor	3 (10.0)	15 (50.0)	16 (53.3)	22 (73.3)	23 (56.1)	32 (78.1)	44 (61.1)	64 (88.8)
Google Vision API	0 (0)	0 (0)	12 (40.0)	17 (56.6)	1 (2.4)	9 (21.9)	4 (5.5)	19 (26.4)

IBM Watson Recognition	5 (16.6)	16 (53.3)	9 (30.0)	16 (53.3)	4 (9.7)	12 (29.2)	29 (40.2)	37 (51.3)
LogMeal	9 (30.0)	23 (76.6)	3 (10.0)	13 (43.3)	13 (31.7)	15 (36.5)	20 (27.7)	31 (43.1)
All	54 (25.7)	121 (57.6)	91 (43.3)	148 (70.5)	99 (34.5)	160 (55.7)	227 (45.1)	326 (64.6)

Τα αποτελέσματα δείχνουν ότι το Calorie Mama είχε την μεγαλύτερη ακρίβεια στην ικανότητα να αναγνωρίσει τροφές και ποτά με τις εφαρμογές Bitesnap και Foodvisor να την ακολουθούν. Επιπλέον, καμία από τις εφαρμογές δεν είχε την ικανότητα να υπολογίσει ποσότητες τροφών[108].

2.4.4 Μελέτη 4- Αναγνώριση Τροφίμων για Εικόνες Lifelog

Στη μελέτη αυτή αναπτύχθηκε σύστημα βαθιάς μάθησης με το οποίο αναγνωρίζονται εικόνες με φαγητά που έχουν συλλεχθεί από στιγμές της καθημερινής ζωής μέσω κινητών τηλεφώνων ή έξυπνων συσκευών. Το σύστημα αρχικά αναγνωρίζει τις στιγμές όπου στην εικόνα δείχνει άνθρωπο να καταναλώνει τροφή. Στη συνέχεια, το σύστημα επικεντρώνεται και ενισχύει την εικόνα στα σημεία όπου δείχνει την τροφή. Τέλος, η τροφή αναγνωρίζεται μέσω ενός συνελκτικού νευρωνικού δικτύου. Στη θέση του συνελκτικού νευρωνικού δικτύου χρησιμοποιήθηκαν οι αρχιτεκτονικές AlexNet και GoogleNet προκειμένου να διαπιστωθεί ποια επιτυγχάνει τη μεγαλύτερη ακρίβεια ταξινόμησης. Το GoogleNet επιτυγχάνει τα καλύτερα αποτελέσματα αφού πετυχαίνει ακρίβεια ταξινόμησης 95.97% ενώ το AlexNet πέτυχε 91.67% αντίστοιχα[109].



Εικόνα 36: Σχεδιάγραμμα το οποίο δείχνει την αρχιτεκτονική του συστήματος αναγνώρισης[109].

2.4.5 Μελέτη 5-FoodAI

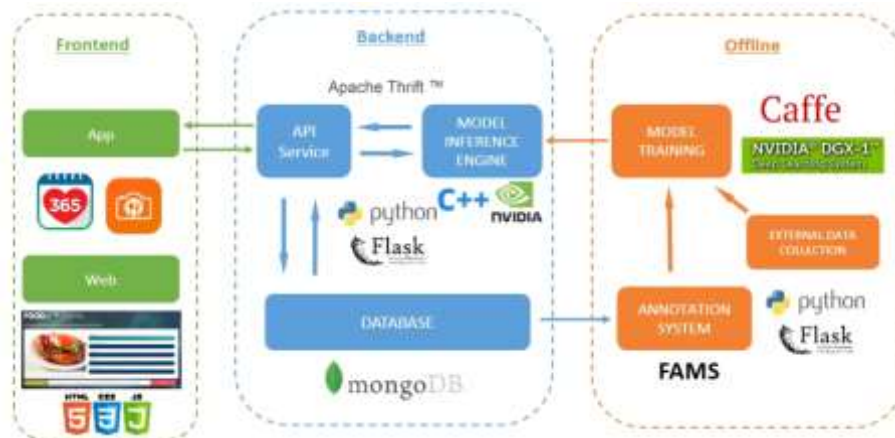
Το FoodAI δημιουργήθηκε από τους Doyen Sahoo, Wang Hao, Shu Ke, Xiongwei Wu, Hung Le, Palakorn Achananuparp, Ee-Peng Lim, Steven C. H. Hoi στο πανεπιστήμιο Διοίκησης Επιχειρήσεων της Σιγκαπούρης το 2017. Πρόκειται για σύστημα αναγνώρισης τροφίμων το οποίο επικεντρώνεται κυρίως σε τροφές που καταναλώνονται στη περιοχή της Σιγκαπούρης. Έχει την ικανότητα να αναγνωρίσει 756 κλάσεις τροφών από τις οποίες οι 100 συναντώνται κυρίως στη Σιγκαπουριανή κουζίνα. Το FoodAI έχει ενσωματωθεί σε εφαρμογές με τη μορφή RESTful υπηρεσιών τα οποία είναι προσβάσιμα από τα πρωτόκολλα επικοινωνιών HTTP/HTTPS. Αρχικά, στο μπροστινό άκρο της εφαρμογής(front end) μπορεί να είναι κάποια διαδικτυακή εφαρμογή. Στο πίσω άκρο(back end), ο διακομιστής ιστού Apache Nginx λαμβάνει τα αιτήματα (requests) και τα προωθεί στο UWSGI διακομιστή(server). Το UWSGI καλεί τη πλατφόρμα Caffe που βρίσκεται το εκπαιδευμένο νευρωνικό δίκτυο που

πραγματοποιεί τη πρόβλεψη. Τέλος, η πρόβλεψη επιστρέφει στο Apache Nginx και από εκεί λαμβάνει ο χρήστης το αποτέλεσμα. Ως έξοδο ο χρήστης λαμβάνει τις εκτιμήσεις που έχουν τη μεγαλύτερη πιθανότητα να ανταποκρίνονται στη πραγματικότητα με βάση τους υπολογισμούς του νευρωνικού δικτύου καθώς επίσης το ποσοστό βεβαιότητας για τη κάθε κλάση όπως δείχνει η εικόνα[110].



Εικόνα 37: Παρουσιάζεται εικόνα όπου γίνεται επίδειξη της εφαρμογής FoodAI[12].

Για τη βάση δεδομένων MongoDB. Για τη χρήση της εφαρμογής, ο χρήστης μπορεί να εγγραφεί στην ιστοσελίδα της FoodAI και να λάβει κλειδί API με το οποίο μπορεί να αναπτύξει τις δικές του εφαρμογές. Η εφαρμογή Healthy 365 για κινητά που έχει ως σκοπό ο χρήστης να ελέγχει τις τροφές που καταναλώνει, χρησιμοποιεί το σύστημα της FoodAI για να κάνει τις εκτιμήσεις του[111].



Εικόνα 38: Σχεδιάγραμμα λειτουργίας του συστήματος FoodAI[111].

Για τη δημιουργία του αρχικά συλλέχθηκαν από το διαδίκτυο περίπου 400000 εικόνες. Οι εικόνες αυτές κατηγοριοποιήθηκαν σε 152 υπερ-κατηγορίες τροφών που αντιπροσωπεύουν συνηθισμένες τροφές και ποτά. Για κάθε υπερ-κατηγορία αναγνωρίστηκαν και ταξινομήθηκαν διάφορα φαγώσιμα αντικείμενα. Για παράδειγμα ο ανανάς και το μήλο ανήκουν στην υπερ-κατηγορία των φρούτων. Με αυτό το τρόπο αναγνωρίστηκαν 1166 διαφορετικά ήδη φαγώσιμων αντικειμένων. Δεδομένου όμως ότι πολλές τροφές μοιάζουν μεταξύ τους ως προς την εμφάνιση, ταξινομήθηκαν σε 756 κλάσεις τροφών. Η κάθε κατηγορία τροφίμων που απαντώνται στη Σιγκαπουριανή κουζίνα αποτελείται από 500

τουλάχιστον εικόνες. Οι υπόλοιπες τροφικές κατηγορίες αποτελούνται από 174 εικόνες έως 2312 εικόνες η κάθε μία. Διαθέτει επίσης άλλη μία κατηγορία με 10000 εικόνες που περιέχουν αντικείμενα που δεν είναι φαγητά. Αυτή η βάση δεδομένων ονομάστηκε από τους δημιουργούς της FoodAI-756 και διαχωρίστηκε σε δεδομένα για εκπαίδευση, εκτίμηση και εξέτασή του μοντέλου όπως δείχνει η κάτωθι εικόνα[110].



Εικόνα 39: Ιστογράμμο το οποίο παρουσιάζει τις κλάσεις της βάσης δεδομένων που χρησιμοποιήθηκε για την εκπαίδευση των νευρωνικών δικτύων FoodAI-756[110].

Στη συνέχεια, χρησιμοποιήθηκε μεταφορά μάθησης από συνελκτικά νευρωνικά δίκτυα τα οποία είχαν εκπαιδευτεί ήδη με τη βάση δεδομένων ImageNet. Τα ήδη εκπαιδευμένα νευρωνικά δίκτυα που χρησιμοποιήθηκαν ήταν το AlexNet, VGG, GoogleNet, ResNet, ResNeXt και SENet[112]. Η αλλαγή που έγινε σε αυτά τα νευρωνικά δίκτυα ήταν να τροποποιηθεί η συνάρτηση απώλειας. Πιο συγκεκριμένα, αντί να χρησιμοποιηθεί η συνάρτηση απώλειας εντροπίας που χρησιμοποιούσαν αρχικά τα εκπαιδευμένα νευρωνικά δίκτυα, χρησιμοποιήθηκε η συνάρτηση εστιακής απώλειας (focal loss) η οποία παρουσιάζεται κάτωθι[113]:

$$FL(p_t) = -\alpha_t * (1 - p_t)^\gamma * \log(p_t) \quad (38)$$

Όπου το α είναι παράγοντας που εξισορροπεί τα βάρη των δειγμάτων από διαφορετικές κατηγορίες τροφίμων. Το $\gamma > 0$ είναι παράμετρος που ρυθμίζει τη σημασία του δείγματος με βάση την ευκολία της ταξινόμησης. Αν το δείγμα είναι εύκολο να ταξινομηθεί τότε μειώνεται η σημασία του.

Στη συνέχεια τα εκπαιδευμένα νευρωνικά δίκτυα εξετάστηκαν ως προς την επίδοσή τους να αναγνωρίσουν σωστά τις εικόνες που δέχτηκαν προς εξέταση. Για το σκοπό αυτό, εξετάστηκε αν η αναγνώριση με το μεγαλύτερο ποσοστό από το μοντέλο αντιστοιχεί με αυτό που περιέχει η εκάστοτε εικόνα (top-1 ακρίβεια) καθώς επίσης οι 5 προβλέψεις με το μεγαλύτερο ποσοστό που έκανε το μοντέλο αν αντιστοιχούν με αυτό που περιέχει η εκάστοτε εικόνα που εξετάστηκε (top-5 ακρίβεια). Στο κάτωθι πίνακα παρατίθενται τα αποτελέσματα της έρευνας.

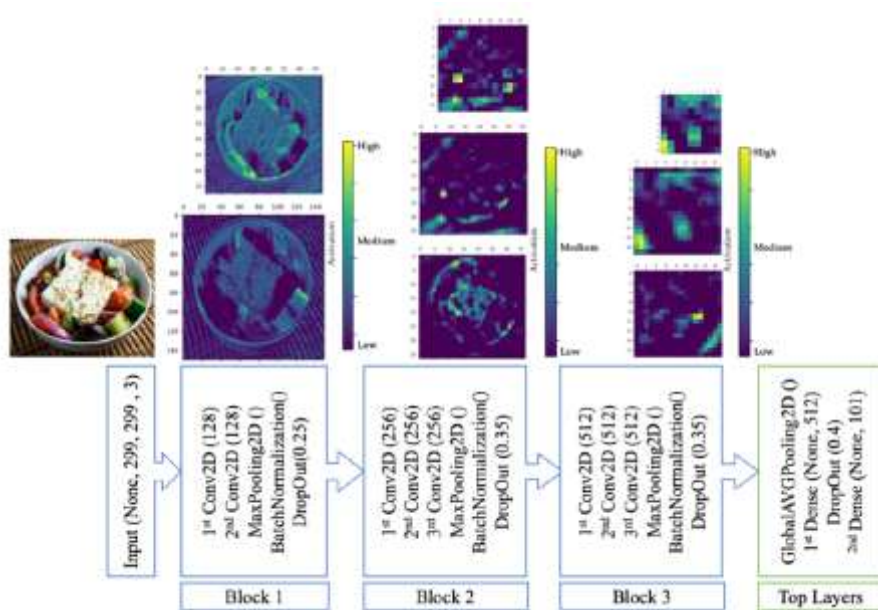
Πίνακας 9: Παρουσιάζεται η ακρίβεια για κάθε αρχιτεκτονική που εξετάστηκε στην μελέτη με τη χρήση της βάσης δεδομένων FoodAI-756[110].

Νευρωνικό Δίκτυο	Top-1 Ακρίβεια	Top-5 Ακρίβεια
ResNet-50	0.7870	0.9427
ResNet-101	0.7645	0.9366
ResNeXt-50	0.7898	0.9473
SENet ResNeXt-50	0.8086	0.9561
ResNet50 χωρίς Εστιακή Απώλεια	0.787	0.943
ResNet50 και Εστιακή Απώλεια	0.802	0.95
SENet με ResNet50 και Εστιακή Απώλεια	0.823	0.955

SENet με ResNet101 και Εστιακή Απώλεια	0.832	0.957
----------------------------------------	-------	-------

2.4.6 Μελέτη 6 - PureFoodNet

Το PureFoodNet είναι ένα μοντέλο το οποίο αναπτύχθηκε από τους Chairi Kiourt, Γιώργο Παυλίδη και Στέλλα Μαρκαντωνάτου στο Ερευνητικό Κέντρο "Αθηνά", Πανεπιστημιούπολη της Κιμμέριας, Ξάνθη το 2020. Σκοπός της ανάπτυξής του ήταν η δημιουργία ενός συστήματος αναγνώρισης τροφίμων που υπάρχουν στην ελληνική κουζίνα και η ενσωμάτωσή της σε εφαρμογές κινητής τηλεφωνίας που αφορούν το πεδίο του τουρισμού. Η αρχιτεκτονική της βασίζεται σε εκείνη της VGG. Αποτελείται από 3 βαθμίδες όπου πραγματοποιείται συνέλιξη και μία βαθμίδα όπου πραγματοποιείται ταξινόμηση. Η πρώτη βαθμίδα περιέχει 2 επίπεδα με συνελίξεις οι οποίες περιέχουν 128 φίλτρα η κάθε μία. Η δεύτερη βαθμίδα διαθέτει 3 επίπεδα με συνελίξεις που περιέχουν 256 φίλτρα η κάθε μία. Η τρίτη βαθμίδα περιέχει 3 επίπεδα συνελίξεων με 512 φίλτρα η κάθε μια. Η βαθμίδα ταξινόμησης περιέχει ένα πλήρως συνδεδεμένο επίπεδο 512 νευρώνων το οποίο με τη σειρά του οδηγεί σε ένα επίπεδο πλήρως συνδεδεμένο με 101 εξόδους, 1 για κάθε κλάση που αντιστοιχεί στη βάση δεδομένων Food101 που χρησιμοποιήθηκε για την εκπαίδευση του νευρωνικού δικτύου[114].



Εικόνα 40: Σχεδιάγραμμα της αρχιτεκτονικής του νευρωνικού δικτύου PureFoodNet[111].

Στη συνέχεια, συγκρίθηκε η αποτελεσματικότητα του συνελικτικού νευρωνικού δικτύου PureFoodNet ως προς την ικανότητά του με μια σειρά γνωστών νευρωνικών δικτύων τα οποία είναι τα VGG16, ResNet50, InceptionResNetV2, MobileNetV2, DenseNet121, NasNetMobile, NasNetLarge. Τα αποτελέσματα της μελέτης παρουσιάζονται στο κάτωθι πίνακα.

Πίνακας 10: Ποσοστά ακρίβειας Top-1 και Top-5 για κάθε βάση δεδομένων[114].

Νευρωνικό Δίκτυο	Top-1 Ακρίβεια	Top-5 Ακρίβεια
InceptionV3	0.848	0.967
VGG16	0.776	0.934
ResNet50	0.839	0.966
InceptionResNetV2	0.837	0.960

MobiliNetV2	0.817	0.956
DenseNet121	0.855	0.970
NasNetMobile	0.784	0.949
NasNetLarge	0.869	0.972
PureFoodNet	0.790	0.948

2.4.7 Μελέτη 7 -Σύστημα Ταξινόμησης Τροφίμων με τη Χρήση Μεταφοράς Μάθησης στο InceptionV3

Η μελέτη αναπτύχθηκε από τους David J. Attokaren, Ian G. Fernandes, A. Sriram, Y.V. Srinivasa Murthy, and Shashidhar G. Koolagudi του Εθνικού Ινστιτούτου Τεχνολογίας της Καρνατάκα. Το περιεχόμενο της μελέτης παρατίθεται παρακάτω. Οι σχεδιαστές της, πήραν το συνελκτικό νευρωνικό δίκτυο InceptionV3 το οποίο είχε ήδη εκπαιδευτεί με τη βάση δεδομένων ImageNet ως τη βάση του μοντέλου τους. Στη συνέχεια, πραγματοποιήθηκε μεταφορά μάθησης, εκπαιδεύοντας το μοντέλο με τη βάση δεδομένων Food-101. Στη συνέχεια παρουσιάζονται τα αποτελέσματα από την μελέτη[115].

Πίνακας 11: Αποτελέσματα ακρίβειας Top-1 για διαφορετικά νευρωνικά δίκτυα[115].

Μοντέλο	Βάση Δεδομένων	Top-1 Ακρίβεια (%)
SVM	Food-101	50.76
Μέθοδος Ταξινόμησης Random Forest	Food-101	56.76
Resnet18	Food-101	67.23
CNN	UEC-FOOD100	78.77
CNN (ILSVRC)	Food-101	79.20
CNN (Food-101)	EgocentricFood	90.90
InceptionV3 με Μεταφορά Μάθησης	Food-101	86.97

Ως προς τα συστήματα που χρησιμοποίησαν την βάση δεδομένων Food-101 το InceptionV3 παρουσιάζει τη μεγαλύτερη ακρίβεια με 86.97%. Το νευρωνικό δίκτυο CNN το οποίο χρησιμοποίησε για την εκπαίδευσή του τη βάση δεδομένων EgocentricFood, είχε το μεγαλύτερο ποσοστό ακρίβειας αποτελεσμάτων Top-1 με 90.90%.

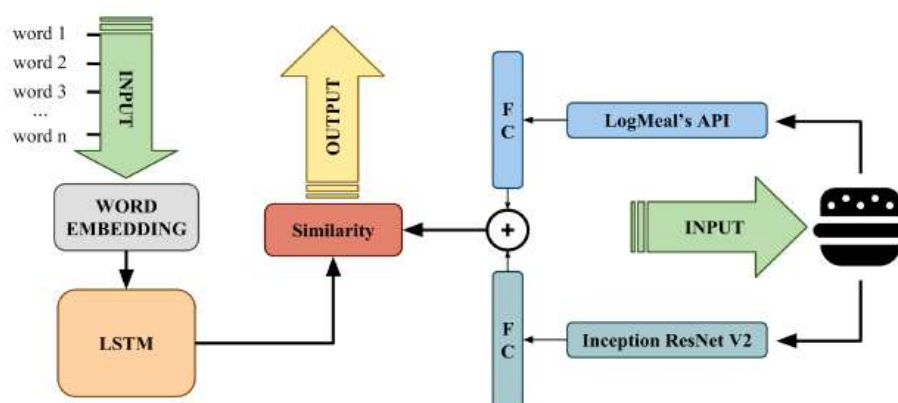
2.4.8 Μελέτη 8 Logmeal

Το Logmeal αναπτύχθηκε από ερευνητές του Τμήματος Μαθηματικών και της Επιστήμης των Υπολογιστών του Πανεπιστημίου της Βαρκελώνης. Προσφέρει λύσεις για την αναγνώριση τροφίμων με τη χρήση βαθιάς μάθησης. Πρόκειται για εφαρμογή νέφους που επιτρέπει τη δημιουργία ημερολογίου τροφών που έχουν καταναλωθεί από το χρήστη. Αυτό γίνεται μέσω επεξεργασίας που έχει γίνει από την εφαρμογή με φωτογραφίες τροφών που έχουν ληφθεί από το χρήστη. Έχει τη δυνατότητα να κάνει αναγνώριση πιάτων, να κατηγοριοποιήσει τις τροφές που αναγνωρίζει όπως και τα επιμέρους στοιχεία που αποτελούν το πιάτο που αναγνωρίζεται καθώς επίσης τη διατροφική αξία τους.

Η βάση δεδομένων για την εκπαίδευσή του συστήματος δημιουργήθηκε από εικόνες που λήφθηκαν από την πλατφόρμα Yelp. Συγκεκριμένα, οι εικόνες αυτές λήφθηκαν από εστιατόρια στη περιοχή της Καλιφόρνια. Η βάση δεδομένων διαθέτει 53877 εικόνες φαγητών με 3498 είδη πιάτων από 313 εστιατόρια. Από αυτές, 37956 χρησιμοποιήθηκαν για την

εκπαίδευση του Logmeal, 7721 για επιβεβαίωση και 10794 για να γίνει το τεστ του μοντέλου[116,117].

Το LogMeal χρησιμοποιεί ένα σύστημα πολυτροπικής μάθησης (Multimodal learning) όπου η πληροφορία της εικόνας διαχωρίζεται σε δύο διαφορετικά νευρωνικά δίκτυα που εντοπίζουν διαφορετικά χαρακτηριστικά της εικόνας. Το πρώτο νευρωνικό δίκτυο είναι το API της LogMeal. Το API ο οποίος είναι ένας ταξινομητής που έχει τη δυνατότητα να διαχωρίσει το περιεχόμενο της εικόνας σε 11 οικογένειες τροφών, 200 είδη πιάτων και 1092 συστατικά τροφών. Το δεύτερο νευρωνικό δίκτυο InceptionResNetV2 αποτελείται από διάλυμα 1536 τιμών και έχει εκπαιδευτεί με τη βάση δεδομένων ImageNet. Στη συνέχεια, η έξοδος από τα 2 μέρη πηγαίνει σε ένα επίπεδο 300 πλήρως συνδεδεμένων νευρώνων όπου μετασχηματίζει τις εξόδους από τα 2 μέρη στο ίδιο μέγεθος και μετά τα συνδυάζει. Το κείμενο που περιγράφει το όνομα της τροφής δημιουργείται από ένα δίκτυο LSTM (Long Short Term Memory). Το μέγεθος της εξόδου του LSTM είναι επίσης 300 νευρώνες για να έχει το ίδιο μέγεθος με την έξοδο από του διανύσματος της εικόνα. Πριν ακριβώς από το LSTM υπάρχει ένας ενσωματωμένος πίνακας λέξεων ο οποίος έχει αρχικοποιηθεί από το ConceptNet. Το διάγραμμα της LogMeal παρουσιάζεται στην εικόνα 41 [118,119,120].



Εικόνα 41: Αρχιτεκτονική συστήματος αναγνώρισης τροφίμων LogMeal[116].

2.5 Διαδικτυακές Εφαρμογές για αρχάριους στην ΤΝ που Μπορούν να Χρησιμοποιηθούν για Εκπαίδευση Συστήματος Αναγνώρισης Τροφίμων

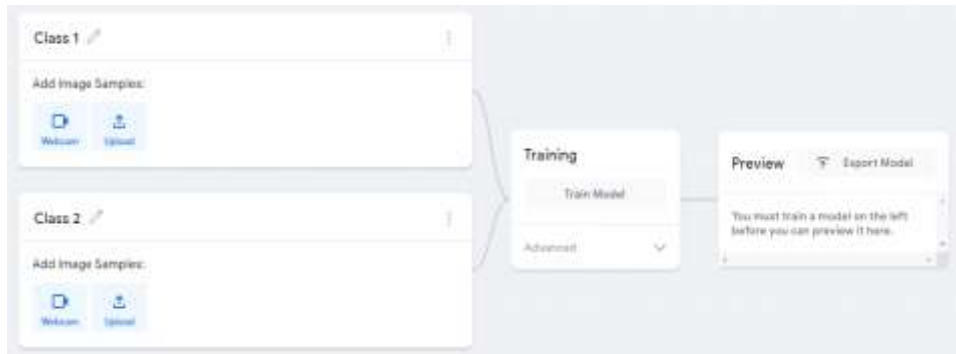
Το τελευταίο χρονικό διάστημα, έχουν αναπτυχθεί εφαρμογές και ιστοσελίδες, οι οποίες διαθέτουν εργαλεία που μπορούν να χρησιμοποιηθούν για την εύκολη εκπαίδευση νευρωνικών δικτύων για αναγνώριση αντικειμένων. Στη συνέχεια παρουσιάζονται κάποιες από τις διαδικτυακές πλατφόρμες οι οποίες έχουν την δυνατότητα εκπαίδευσης νευρωνικών δικτύων για αναγνώριση κατηγοριών τροφίμων και οι οποίες μπορούν να χρησιμοποιηθούν από αρχάριους στην ΤΝ, που δεν διαθέτουν τις τεχνικές γνώσεις για την ανάπτυξη ή την εκπαίδευση κάποιου μοντέλου.

2.5.1 Machine Learning For Kids

Το Machine Learning for Kids είναι διαδικτυακή πλατφόρμα με την οποία ο χρήστης μπορεί να αναπτύξει εφαρμογές όπου χρησιμοποιούν βαθιά μάθηση. Δημιουργήθηκε από τον Dale Lane το 2017. Λειτουργεί ως πλατφόρμα εισαγωγής στο πεδίο της μηχανική μάθησης για μαθητές. Τα εκπαιδευμένα νευρωνικά δίκτυα έχουν τη δυνατότητα να χρησιμοποιηθούν για να αναπτυχθούν εφαρμογές με τη βοήθεια των γλωσσών προγραμματισμού Python και Scratch[121].

2.5.2 Teachable Machine

Το Teachable Machine είναι διαδικτυακή πλατφόρμα όπου μπορεί κάποιος να αναπτύξει εφαρμογές που χρησιμοποιούν βαθιά μάθηση. Δημιουργήθηκε από τη Google το 2019 και αποτελεί εξέλιξη μιας παλαιότερης έκδοσης που βγήκε το 2017. Η πλατφόρμα έχει ήδη εκπαιδευμένο το συνελκτικό νευρωνικό δίκτυο MobileNet με τη βάση δεδομένων ImageNet. Ο χρήστης, με τη βοήθεια της μεταφοράς μάθησης, μπορεί να επανεκπαιδεύσει το νευρωνικό δίκτυο για διάφορες εφαρμογές στο χώρο της αναγνώρισης και εντοπισμού αντικειμένων, ήχων, ακόμα και στάσεων του σώματος. Μετά το πέρας της εκπαίδευσης, ο χρήστης μπορεί να εξαγάγει το πρόγραμμα και το κώδικα για να το ενσωματώσει σε εφαρμογές που έχει επιλέξει ο ίδιος[122].



Εικόνα 42: Επιφάνεια εργασιών όπου μπορεί να εκπαιδευτεί σύστημα αναγνώρισης μέσω της εφαρμογής Teachable Machine[121].

Στη διεθνή βιβλιογραφία έχουν προταθεί πολλαπλές βάσεις δεδομένων καθώς επίσης αρχιτεκτονικές νευρωνικών δικτύων για τη δημιουργία όλο και πιο αποτελεσματικών συστημάτων. Για το πρακτικό μέρος της εργασίας δημιουργήθηκε μια νέα βάση δεδομένων με εικόνες από κάμερες, κινητά καθώς επίσης μηχανές αναζήτησης που περιέχουν μπουκάλια με γάλα. Στο κεφάλαιο αυτό θα παρουσιαστεί η βάση δεδομένων με την οποία έχουν εκπαιδευτεί τα νευρωνικά δίκτυα. Πρόκειται για δύο νευρωνικά δίκτυα. Το ένα δημιουργήθηκε μέσω της διαδικτυακής εφαρμογή Teachable Machine. Το δεύτερο, μέσω προγράμματος που δημιουργήθηκε με την χρήση του νευρωνικού δικτύου MobileNet με μειωμένο το πλάτος κάθε στοιβάδας νευρώνων κατά 50%. Παρουσιάζονται οι ρυθμίσεις που πραγματοποιήθηκαν για την ομαλή εκπαίδευση των νευρωνικών δικτύων. Τέλος, εκθέτονται τα αποτελέσματα αξιολογώντας την αποτελεσματικότητα των εκπαιδευμένων νευρωνικών δικτύων στο να κάνουν επιτυχείς αναγνωρίσεις. Παρουσιάζονται επίσης εφαρμογές όπου χρησιμοποιήθηκαν τα δύο εκπαιδευμένα μοντέλα.

3.1 Συλλογή Εικόνων για την Εκπαίδευση του Νευρωνικού Δικτύου

Για την εκπαίδευση του νευρωνικού δικτύου συγκεντρώθηκαν εικόνες οι οποίες διέθεταν διάφανα μπουκάλια με γάλα σε όρθια ή σε πλάγια στάση. Οι εικόνες συγκεντρώθηκαν μέσω κινητών τηλεφώνων, καμερών υπολογιστών, καθώς επίσης από μηχανές αναζήτησης του διαδικτύου. Στη συνέχεια ταξινομήθηκαν σε 4 κλάσεις με βάση το κατά πόσο τα μπουκάλια περιείχαν γάλα ανάλογα με τη ποσότητα τους. Εάν το μπουκάλι με το γάλα ήταν γεμάτο, τότε ταξινομήθηκε στη πρώτη κλάση full που αφορά γεμάτα μπουκάλια με γάλα. Εάν το μπουκάλι με το γάλα ήταν γεμάτο κατά το ήμισυ, τότε η εικόνα του ταξινομήθηκε στη δεύτερη κλάση half. Στη τρίτη κλάση half_empty ταξινομήθηκαν μπουκάλια που ήταν λιγότερο από το ήμισυ γεμάτα ή ήταν σχεδόν άδεια. Στη τέταρτη κλάση almost_full ταξινομήθηκαν μπουκάλια που ήταν γεμάτα πάνω από το ήμισυ. Δημιουργήθηκε επίσης μια πέμπτη κλάση με το όνομα random που αποτελούνταν από εικόνες που δεν διέθεταν μπουκάλια με γάλα αλλά εικόνες όπως άδεια μπουκάλια, διάφορα φαγητά και εικόνες από αντικείμενα, άδεια δωμάτια και εικόνες ανθρώπων.



Εικόνα 43:Εικόνες από την βάση δεδομένων που χρησιμοποιήθηκε προκειμένου να εκπαιδευτεί το νευρωνικό δίκτυο.

Ενδεικτικά παρουσιάζονται εικόνες από τη βάση δεδομένων στη εικόνα 43. Στη συνέχεια παρουσιάζεται πίνακας που περιγράφει το περιεχόμενο της βάσης δεδομένων που χρησιμοποιήθηκε.

Πίνακας 12: Περιεχόμενο της βάσης δεδομένων που χρησιμοποιήθηκε για την εκπαίδευση του νευρωνικού δικτύου.

ΚΛΑΣΗ	ΠΕΡΙΕΧΟΜΕΝΟ	ΑΡΙΘΜΟΣ
full	Μπουκάλια γεμάτα με γάλα	558
almost_full	Μπουκάλια σχεδόν γεμάτα με γάλα	513
half	Μπουκάλια γεμάτα με γάλα κατά το ήμισυ	540
half_empty	Μπουκάλια σχεδόν άδεια από γάλα	509
random	Εικόνες που δεν περιέχουν γάλα	515
ΣΥΝΟΛΟ ΕΙΚΟΝΩΝ		2635

3.2 Εκπαίδευση Νευρωνικού Δικτύου με Teachable Machine

Για τη δημιουργία και εκπαίδευση του πρώτου νευρωνικού δικτύου, χρησιμοποιήθηκε η πλατφόρμα Teachable Machine (<https://teachablemachine.withgoogle.com/train/image>) της Google. Η πλατφόρμα αυτή έχει ήδη εκπαιδευμένο το νευρωνικό δίκτυο MobbilleNet με τη βάση δεδομένων ImageNet. Με τη βοήθεια της μεταφοράς μάθησης και τη βάση δεδομένων που δημιουργήθηκε από τις εικόνες που συγκεντρώθηκαν, εκπαιδεύτηκε ξανά το νευρωνικό δίκτυο έτσι ώστε να μπορεί να αναγνωρίσει τις φιάλες με γάλα και τι ποσότητα γάλακτος διαθέτουν στο εσωτερικό τους.

Από το σύνολο των εικόνων που χρησιμοποιήθηκαν για την εκπαίδευση, το 85% ήταν μόνο εκείνο που χρησιμοποιήθηκε για εκπαίδευση του νευρωνικού δικτύου ενώ το υπόλοιπο 15% χρησιμοποιήθηκε για την επιβεβαίωση(validation) της ομαλής λειτουργίας του. Δηλαδή επιβεβαιώνει ότι το νευρωνικό δίκτυο εκπαιδεύεται σωστά κατά τη διάρκεια της εκπαίδευσης. Για να μπορέσει η εκπαίδευση να πραγματοποιηθεί χωρίς προβλήματα, ορίστηκε ότι το σύστημα θα λαμβάνει παρτίδες αποτελούμενες από 16 εικόνες κάθε φορά για εκπαίδευση. Ορίστηκε ότι κάθε δείγμα θα περνούσε από το μοντέλο εκπαίδευσης 350 φορές (epochs) και ο ρυθμός μάθησης ορίστηκε σε 0,0001 έτσι ώστε το νευρωνικό δίκτυο να έχει υψηλή ακρίβεια και να αποφευχθεί η υπερπροσαρμογή (Overfitting) κατά τη διάρκεια της εκπαίδευσης.

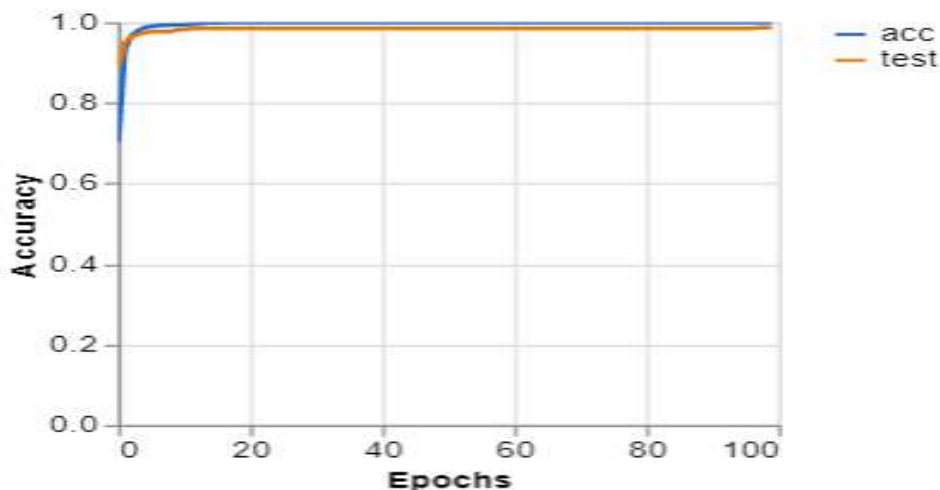
3.2.1 Αποτελέσματα Εκπαίδευσης Νευρωνικού Δικτύου με Teachable Machine

Η εικόνα 44 παρουσιάζει την ορθότητα(accuracy) ανά επανάληψη. Η ορθότητα είναι το πηλίκο των σωστών προβλεπόμενων παρατηρήσεων προς τις συνολικές παρατηρήσεις. Δίνεται από την εξίσωση 39 [123]:

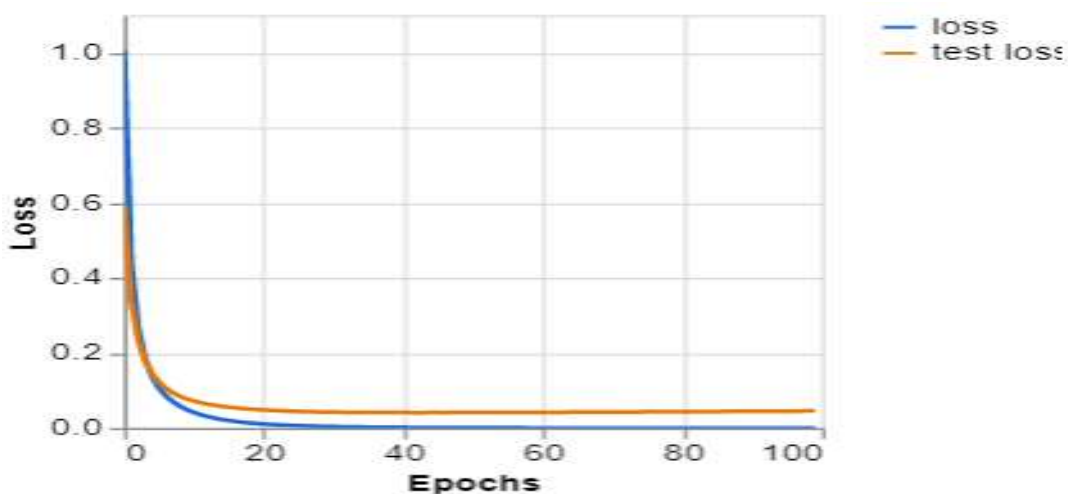
$$\text{Accuracy} = (\text{TP}+\text{TN})/((\text{TP}+\text{FP}+\text{FN}+\text{TN}))\text{(39)}$$

Η εικόνα 45 δείχνει το ρυθμό απώλειας η οποία είναι η διαφορά της προβλεπόμενης τιμής από την πραγματική τιμή ανά επανάληψη. Και στις δύο γραφικές παραστάσεις, με γαλάζιο αντιπροσωπεύεται η πρόοδος της εκπαίδευσης του νευρωνικού δικτύου ενώ με το κόκκινο επιβεβαιώνεται κατά πόσο το μοντέλο μπορεί να αναγνωρίσει τις εικόνες που είχαν οριστεί για την αξιολόγηση του μοντέλου.

Εδώ φαίνεται και από τις γραφικές παραστάσεις ότι το μοντέλο έπαψε να βελτιώνεται από την επανάληψη 50 και μετά. Οι γραφικές παραστάσεις δείχνουν ότι το μοντέλο έχει πολύ υψηλή ακρίβεια που πλησιάζει το 100%. Πιο συγκεκριμένα, με βάση το πίνακα σύγχυσης 13 που δημιουργήθηκε στο τέλος της εκπαίδευσης, υπήρξε 98.23% ακρίβεια στις προβλέψεις για τα δεδομένα επιβεβαίωσης.



Εικόνα 44: Η γραφική παράσταση παρουσιάζει την ακρίβεια ανά θήμα εκπαίδευσης(epochs).



Εικόνα 45: Στη συνάρτηση παρουσιάζεται η απώλεια (loss) σε συνάρτηση με το θήμα στο οποίο βρίσκεται η εκπαίδευση.

Πίνακας 13: Πίνακας σύγχυσης για τα δεδομένα επιβεβαίωσης της ορθής λειτουργίας του εκπαιδευμένου νευρωνικού δικτύου με Teachable Machine.

Κλάση	Πρόβλεψη					
		full	half	half_empty	almost_full	random
full		82	1	0	1	0
half		0	80	1	0	0
half_empty		0	0	77	0	0
almost_full		0	1	2	74	0
random		0	1	0	0	76

Στη συνέχεια έγινε δεύτερος έλεγχος για την αξιολόγηση της αξιοπιστίας του συστήματος, χρησιμοποιώντας ως τεστ μια ξεχωριστή βάση δεδομένων με εικόνες από μπουκάλια με γάλα που λήφθηκαν μέσω κάμερας υπολογιστή και κινητού τηλεφώνου. Κάτωθι, παρουσιάζεται ο πίνακας σύγχυσης (confusion matrix) των αποτελεσμάτων.

Πίνακας 14: Πίνακας σύγχυσης για τα δεδομένα αξιολόγησης της ορθής λειτουργίας του εκπαιδευμένου νευρωνικού δικτύου με Teachable Machine.

Κλάση	Πρόβλεψη					
		full	half	half_empty	almost_full	random
full		20	0	0	0	0
half		0	20	0	0	0
half_empty		0	1	17	2	0
almost_full		0	1	0	19	0
random		0	0	0	0	20

Από το δεύτερο πίνακα σύγχυσης παρατηρείται ότι το εκπαιδευμένο νευρωνικό δίκτυο παρουσιάζει ποσοστό επιτυχημένης αναγνώρισης top-1 96%. Στις μοναδικές περιπτώσεις όπου το μοντέλο απέτυχε να αναγνωρίσει σωστά την εικόνα ήταν για τις κλάσεις almost_full και milk_empty όπου είχαν ακρίβεια 95% και 85% αντίστοιχα.

3.3 Εκπαίδευση Νευρωνικού Δικτύου MobileNet με χρήση μεταφοράς μάθησης

Προκειμένου να γίνει η εκπαίδευση για το δεύτερο νευρωνικό δίκτυο, η αρχική βάση δεδομένων του οποίου η περιγραφή έγινε στο Πίνακα 12, διαχωρίστηκε σε δύο φακέλους με τα ονόματα train και test. Στο φάκελο train είναι οι εικόνες οι οποίες εκπαιδεύονται από το νευρωνικό δίκτυο ενώ στο φάκελο test είναι οι εικόνες που χρησιμοποιήθηκαν για την επιβεβαίωση της σωστής εκπαίδευσης του νευρωνικού δικτύου. Στο καθένα από τους 2 φακέλους υπάρχουν αντίστοιχοι φάκελοι με εικόνες που αντιπροσωπεύουν τις 5 κλάσεις από τις οποίες αποτελείται η βάση δεδομένων. Οι φάκελοι αυτοί ονομάζονται full, half, half_empty, almost_full και random. Στους φακέλους που βρίσκονται στο test ο αριθμός των εικόνων για κάθε κλάση είναι 75 που αντιστοιχεί στο 14.23% της βάσης δεδομένων. Οι υπόλοιπες εικόνες χρησιμοποιήθηκαν για την εκπαίδευση.

3.3.1 Πρόγραμμα Εκπαίδευσης Νευρωνικού Δικτύου με αρχιτεκτονική MobileNet και μεταφορά μάθησης

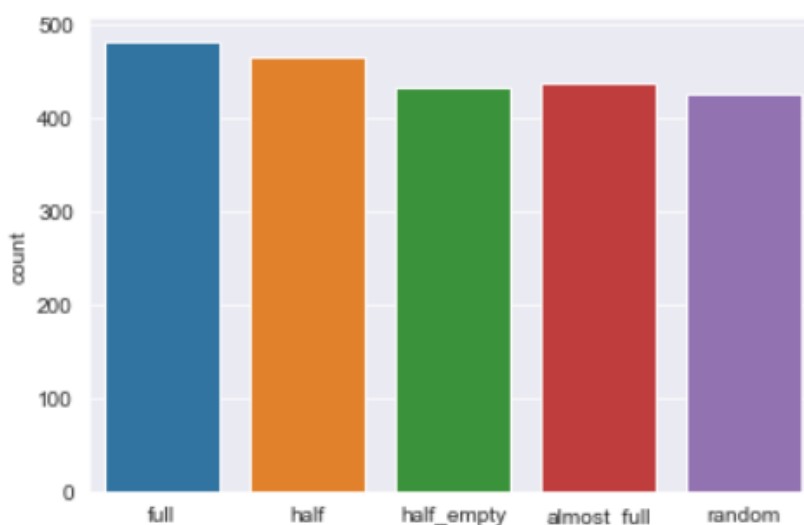
Το πρόγραμμα δημιουργήθηκε μέσω του προγραμματιστικού περιβάλλοντος Jupyter Notebook και παρουσιάζεται ο κώδικας στο Παράρτημα 1.

Για τη δημιουργία του, εισήχθησαν στο πρόγραμμα οι βιβλιοθήκες οι οποίες χρειάστηκαν να χρησιμοποιηθούν για την εκπαίδευση και αξιολόγηση του νευρωνικού δικτύου. Η βιβλιοθήκες που χρησιμοποιήθηκαν ήταν οι Keras, Tensorflow, matplotlib, seaborn, os και opencv.

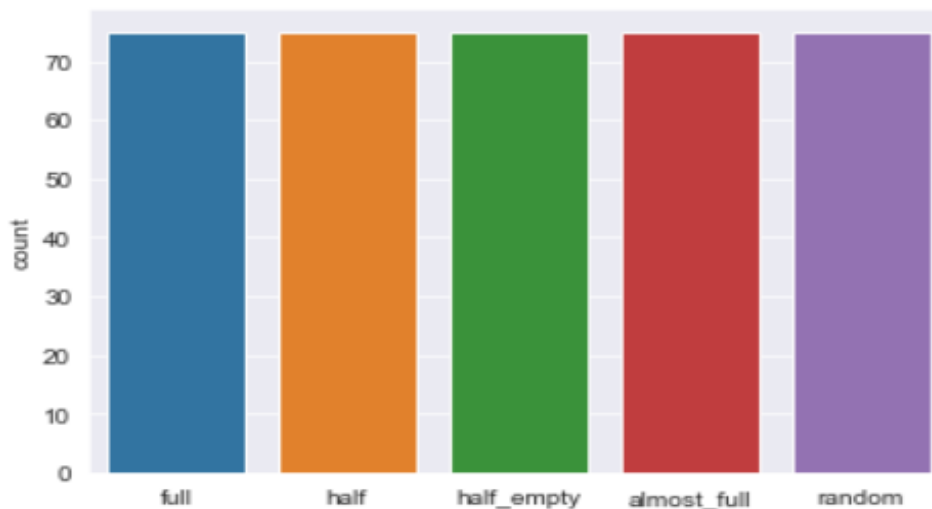
Δημιουργήθηκε μονοδιάστατος πίνακας labels 1x5 όπου σε κάθε θέση περιέχεται και το όνομα κάθε κλάσης από την βάση δεδομένων η οποία πρόκειται να εκπαιδευτεί. Στη συνέχεια δημιουργήθηκε συνάρτηση get_data με την οποία οι εικόνες στο μονοπάτι που είχε οριστεί, μετατρέπονται σε πίνακες. Αυτή η διαδικασία έχει γίνει για όλες τις εικόνες που θα χρησιμοποιηθούν για εκπαίδευση αλλά και για την επιβεβαίωση της σωστής λειτουργίας. Ο αριθμός των δεδομένων για κάθε κλάση των εικόνες που πρόκειται να εκπαιδευτούν παρουσιάζεται στην εικόνα 46. Στην εικόνα 47 παρουσιάζονται ο αριθμός των εικόνων που χρησιμοποιήθηκαν σε κάθε κλάση για επιβεβαίωση. Στη συνέχεια τα δεδομένα αυτά ομαλοποιούνται και μέσω της συνάρτησης ImageDataGenerator() προετοιμάζονται για την εκπαίδευση.

Το βασικό μοντέλο που χρησιμοποιήθηκε για την εκπαίδευση ήταν το MobileNet με μειωμένο το πλάτος κάθε στοιβάδας νευρώνων κατά 50%. Οι πίνακες εισόδου είχαν μέγεθος 224X224X3. Ορίστηκαν ως αρχικά βάρη εκείνα τα οποία είχε το αρχικό μοντέλο από την εκπαίδευση με την βάση δεδομένων ImageNet. Προκειμένου να μην αλλάξουν αυτά τα βάρη κατά την εκπαίδευση ορίστηκε ότι στο βασικό δίκτυο δεν μπορούν να ανανεωθούν. Στο τελευταίο επίπεδο ορίστηκε μια στοιβάδα με 5 νευρώνες πλήρως συνδεδεμένα. Ο κάθε νευρώνας αντιστοιχεί και σε μια κλάση που έχει εκπαιδευτεί. Επίσης ορίστηκε συνάρτηση ενεργοποίησης σε αυτό το επίπεδο η Softmax. Στο τέλος, το εκπαιδευμένο μοντέλο αποθηκεύεται σε μορφή αρχείου .h5.

Ο ρυθμός μάθησης ορίστηκε σε base_learning_rate = 0.00001. Ως συνάρτηση βελτιστοποίηση ορίστηκε η Adam. Το μέγεθος κάθε παρτίδας το οποίο εκπαιδεύεται κάθε στιγμή ορίστηκε σε 120 εικόνες ενώ η εκπαίδευση είχε 500 επαναλήψεις.



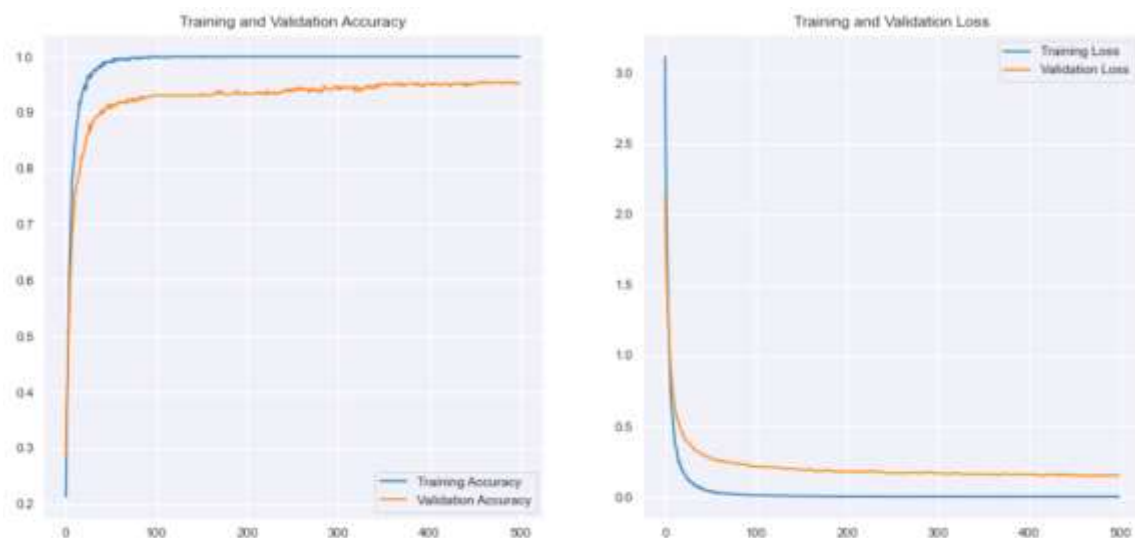
Εικόνα 46: Παρουσιάζεται ο αριθμός των εικόνων για κάθε βάση δεδομένων που είναι για εκπαίδευση.



Εικόνα 47: Παρουσιάζεται ο αριθμός των εικόνων για κάθε βάση δεδομένων που είναι για τη επιβεβαίωση της εκπαίδευσης του νευρωνικού δικτύου.

3.3.2 Αποτελέσματα Εκπαίδευσης Νευρωνικού Δικτύου Αρχιτεκτονικής MobileNet με μεταφορά μάθησης

Στην εικόνα 48 παρουσιάζεται η ακρίβεια και απώλεια ανά επανάληψη, η ακρίβεια της εκπαίδευσης έφτασε στο 100% ενώ, η ακρίβεια για τα δεδομένα επιβεβαίωσης έφτασαν στο 95.2%.



Εικόνα 48: Αριστερά παρουσιάζεται η πρόοδος της ακρίβειας κατά την διάρκεια της εκπαίδευσης ενώ δεξιά ο ρυθμός απώλειας. Με μπλε χρώμα παρουσιάζεται η πρόοδος για τα δεδομένα εκπαίδευσης ενώ με κόκκινο για τα δεδομένα που χρησιμοποιήθηκαν για επιβεβαίωση.

Στη συνέχεια, για να τελειοποιηθεί το μοντέλο έγινε ξανά εκπαίδευση στο νευρωνικό δίκτυο επιτρέποντας στα βάρη από το επίπεδο 60 και μετά να μπορούν να αλλάξουν. Ο ρυθμός μάθησης μειώθηκε σε $\text{learning_rate} = 0.000001$. Η συνάρτηση βελτιστοποίησης

άλλαξε από Adam σε RMSprop. Το μοντέλο εκπαιδεύτηκε με αυτό το τρόπο για 10 επιπλέον επαναλήψεις.

Η ακρίβεια των δεδομένων είναι 96%. Ο πίνακας σύγχυσης 15 δείχνει ότι την μεγαλύτερη ακρίβεια την παρουσιάζουν οι κλάσεις random και half ενώ οι κλάσεις με την μικρότερη ακρίβεια είναι οι half_empty και almost_full. Δεν υπήρξε κατηγορία όπου η ακρίβεια να έπεσε κάτω του 95%. Με τα δεδομένα αξιολόγησης η ακρίβεια ήταν 99% όπως δείχνει ο πίνακας 16.

Πίνακας 15: Πίνακας σύγχυσης για τα δεδομένα επιθεβαίωσης της ορθής λειτουργίας του εκπαιδευμένου νευρωνικού δικτύου MobileNet.

		Πρόβλεψη				
Κλάση		full	half	half_empty	almost_full	random
	full	72	0	1	1	1
half	0	74	0	1	0	
half_empty	0	5	70	0	0	
almost_full	3	2	0	70	0	
random	0	0	1	0	74	

Πίνακας 16: Πίνακας σύγχυσης για τα δεδομένα αξιολόγησης της ορθής λειτουργίας του εκπαιδευμένου νευρωνικού δικτύου MobileNet.

		Πρόβλεψη				
Κλάση		full	half	half_empty	almost_full	random
	full	20	0	0	0	0
half	0	20	0	0	0	
half_empty	0	0	20	0	0	
almost_full	0	1	0	19	0	
random	0	0	0	0	20	

3.4 Εφαρμογές με το εκπαιδευμένο νευρωνικό δίκτυο από την πλατφόρμα Teachable Machine

Το εκπαιδευμένο νευρωνικό δίκτυο από την πλατφόρμα Teachable Machine μπορεί να εξαχθεί και να εγκατασταθεί σε υπολογιστή. Επιπλέον, μπορεί να αποσταλεί στο νέφος και στη συνέχεια με τη χρήση συνδέσμου να χρησιμοποιηθεί σε εφαρμογές, ενσωματωμένα συστήματα ή ακόμα και στη δημιουργία εφαρμογών για κινητά. Στη συνέχεια παρουσιάζονται κάποιες από αυτές τις εφαρμογές.

3.4.1 Εφαρμογή μέσω της p5.js

Υπάρχει η δυνατότητα να γίνει εξαγωγή του εκπαιδευμένου μοντέλου και στη συνέχεια αποσταλεί σε κοινόχρηστο σύνδεσμο (shareable link) URL. Ο κοινόχρηστο σύνδεσμος URL στη συγκεκριμένη περίπτωση είναι:

<https://teachablemachine.withgoogle.com/models/BD6EpYrfM/>

Γίνεται χρήση του επεξεργαστή ιστού p5.js το οποίο διαθέτει ειδική βιβλιοθήκη για την γλώσσα προγραμματισμού Javascript. Το p5.js είναι πιο εύχρηστο για άπειρους προγραμματιστές. Ο κώδικα που χρησιμοποιήθηκε για να εκτελεστεί το εκπαιδευμένο νευρωνικό δίκτυο είναι εδώ [124]. Στο πρόγραμμα αυτό, αρχικά ορίζεται ποιος είναι ο κοινόχρηστος σύνδεσμος URL που προαναφέρθηκε. Στη συνέχεια, ορίζονται οι μεταβλητές video και flippedVideo. Με τη συνάρτηση preload() φορτώνεται το μοντέλο που έχει

αποθηκευτεί στο URL. Η συνάρτηση `setup()` ενεργοποιεί την κάμερα που χρησιμοποιείται και ορίζει το μέγεθος του πεδίου της εικόνας που παρουσιάζεται. Η συνάρτηση `draw()` ορίζει το μέγεθος και το σχήμα των γραμμάτων που παρουσιάζουν την αναγνώριση. Η συνάρτηση `classifyVideo()` πραγματοποιεί αναγνώριση στην εικόνα που τροφοδοτείται στο μοντέλο. Τέλος, η συνάρτηση `getResult()` παρουσιάζει τα αποτελέσματα.

3.4.2 Εφαρμογή με Keras.h5 και Webhook

Παρουσιάζεται εφαρμογή η οποία χρησιμοποιεί Keras.h5 και Webhook. Αρχικά, γίνεται εξαγωγή και λήψη του εκπαιδευμένου μοντέλου σε μορφή αρχείου `keras.h5`. Στη συνέχεια με τη χρήση του προγράμματος που παρουσιάζεται στο Παράρτημα 2 μπορεί να γίνει αναγνώριση εικόνων. Κάτωθι γίνεται εξήγηση ως προς το πως λειτουργεί αυτό το πρόγραμμα.

Αρχικά για να λειτουργήσει το πρόγραμμα χρησιμοποιήθηκαν 3 βιβλιοθήκες, οι `keras`, `pillow` και `numpy`. Η βιβλιοθήκη `pillow` προσθέτει στο πρόγραμμα δυνατότητες επεξεργασίας εικόνας ενώ η βιβλιοθήκη `numpy` χρησιμοποιείται για την επεξεργασία αριθμητικών δεδομένων.

Στη συνέχεια φορτώνεται το μοντέλο `keras_model.h5` μέσω της εντολής `load_model()`. Από εκεί και πέρα ορίζεται ο αριθμός των εικόνων που πρόκειται να αναγνωριστούν μέσω της εντολής `data = np.ndarray(shape=(1, 224, 224, 3), dtype=np.float32)` και στη συνέχεια φορτώνεται η εικόνα που πρόκειται να γίνει αναγνώριση από το μονοπάτι που έχει οριστεί.

Το `image = ImageOps.fit(image, size, Image.ANTIALIAS)` ορίζει το μέγεθος και τα χαρακτηριστικά της εικόνας ενώ στη συνέχεια η εικόνα μετατρέπεται σε πίνακα και κανονικοποιείται. Τέλος, φορτώνεται η εικόνα στο μοντέλο για να γίνει η αναγνώριση. Με βάση το αποτέλεσμα που θα βγάλει η πρόβλεψη παρουσιάζεται και διαφορετικό μήνυμα το οποίο αναφέρεται στο περιεχόμενο σε γάλα που δείχνει η εικόνα.

Με την προσθήκη της βιβλιοθήκης `requests` ο χρήστης έχει την δυνατότητα να χρησιμοποιήσει αιτήματα HTTP. Χρησιμοποιώντας την ιστοσελίδα (<https://webhook.site>) ο χρήστης δημιουργεί `webhook` η οποία είναι μέθοδος με την οποία αλλάζει η συμπεριφορά μιας ιστοσελίδας ή εφαρμογής. Στη συγκεκριμένη περίπτωση, η έξοδος, ανάλογα με την αρχική αναγνώριση του προγράμματος, περνάει ως έξοδος και εισάγεται στην ιστοσελίδα `webhook.site`.

3.4.3 Εφαρμογή με πρόγραμμα περιήγησης ιστού και Javascript

Στην περίπτωση όπου ο χρήστης είναι πιο έμπειρος, μπορεί να χρησιμοποιήσει τον σύνδεσμο <https://teachablemachine.withgoogle.com/models/BD6EpYrfM/> σε προγράμματα μορφής αρχείου `.html` όπου υπάρχει κώδικας `JavaScript`. Με αυτό το τρόπο, το σύστημα αναγνώρισης τροφίμων μπορεί να ενσωματωθεί σε ιστοσελίδες. Ο κώδικας του παρουσιάζεται στο Παράρτημα 3. Όταν εφαρμόζεται, η εμφάνισή του παρουσιάζεται στην εικόνα 49. Για να το θέσει σε λειτουργία και να κάνει αναγνώριση ο χρήστης το μόνο που χρειάζεται είναι κάποιο πρόγραμμα περιήγησης όπως είναι το Chrome ή το Mozilla και κάμερα. Με αυτό το τρόπο μπορεί να λαμβάνει εικόνα το πρόγραμμα και να πραγματοποιεί αναγνώριση σε πραγματικό χρόνο.



almost_full: 0.00
full: 1.00
half: 0.00
half_empty: 0.00
random: 0.00

Εικόνα 49: Επιφάνεια εργασιών όπως εμφανίζεται στο χρήστη όταν συνδέεται μέσω προγράμματος περιήγησης.

3.5 Εφαρμογές με το εκπαιδευμένο νευρωνικό δίκτυο MobileNet

Το εκπαιδευμένο νευρωνικό δίκτυο το οποίο έχει αποθηκευτεί σε μορφή αρχείου .h5 μπορεί να χρησιμοποιηθεί απευθείας με την χρήση του κώδικα που παρουσιάζεται στο Παράρτημα 2. Με την χρήση του κώδικα που παρουσιάζεται στο Παράρτημα 4 ο χρήστης μπορεί να αποθηκεύσει το εκπαιδευμένο μοντέλο σε μορφή .json. Αυτό με την σειρά του μπορεί να χρησιμοποιηθεί με την χρήση της γλώσσας προγραμματισμού JavaScript σε διάφορες εφαρμογές. Ο χρήστης δηλαδή, μέσω του προγράμματος που παρουσιάζεται στο Παράρτημα 5, έχει την δυνατότητα να φορτώσει το εκπαιδευμένο μοντέλο και να δημιουργήσει εφαρμογές που λειτουργούν μέσω προγραμμάτων περιήγησης.

3.6 Παρατηρήσεις

Από τους πίνακες σύγκυσης για το εκπαιδευμένο νευρωνικό δίκτυο με την πλατφόρμα Teachable Machine παρατηρούμε ότι έχει τη δυνατότητα να ξεχωρίσει αν η εικόνα περιέχει μπουκάλια με γάλα ή όχι αφού το ποσοστό σωστής αναγνώρισης σε αυτή τη περίπτωση είναι 100%. Για το εκπαιδευμένο νευρωνικό δίκτυο MobileNet με πλάτος στοιβάδων 50% της αρχικής αρχιτεκτονικής, το ποσοστό σωστής αναγνώρισης για τα δεδομένα επιβεβαίωσης ήταν 99.73%. Το οποίο επίσης είναι πολύ υψηλό.

Για τις εικόνες που χρησιμοποιήθηκαν για την επιβεβαίωση της σωστής εκπαίδευσης του νευρωνικού δικτύου εκπαιδευμένο με την πλατφόρμα Teachable Machine, το ποσοστό επιτυχημένης ταξινόμησης ανήλθε στο 98.2%.

Για τις εικόνες που χρησιμοποιήθηκαν για την επιβεβαίωση της σωστής εκπαίδευσης του νευρωνικού δικτύου MobileNet με πλάτος στοιβάδων στο 50% της αρχικής αρχιτεκτονικής, το ποσοστό επιτυχημένης ταξινόμησης ανήλθε στο 96.0%.

Ως προς τις εικόνες οι οποίες χρησιμοποιήθηκαν για αξιολόγηση, παρατηρείται υψηλή ακρίβεια στην αναγνώριση και με τις δύο μεθόδους. Η ακρίβεια top-1 για το μοντέλο που εκπαιδεύτηκε με Teachable Machine έφτασε στο 95.2%. Η ακρίβεια top-1 για το δεύτερο μοντέλο που δημιουργήθηκε ήταν στο 99%.

ΚΕΦΑΛΑΙΟ 4

Συμπεράσματα – Προτάσεις

Στα πλαίσια της εργασία αυτής έγινε η μελέτη και ανάπτυξη συστημάτων αναγνώρισης τροφίμων με τη χρήση βαθιάς μάθησης. Η ανάπτυξη των υπολογιστικών δυνατοτήτων των ηλεκτρονικών και κυρίως των GPU καθώς επίσης η μελέτη του πεδίου των νευρωνικών δικτύων έχουν επιτρέψει την ανάπτυξη συσκευών και εφαρμογών οι οποίες χρησιμοποιούν βαθιά μάθηση προκειμένου να πραγματοποιήσουν αναγνώριση τροφίμων. Αυτό με τη σειρά του έχει οδηγήσει σε ερωτήματα ως προς το κατά πόσο τέτοιου τύπου συστήματα είναι αποτελεσματικά, ποια είναι η καλύτερη μέθοδος ανάπτυξης τέτοιων συστημάτων, τι υλικό χρειάζεται για την υποστήριξή τους, ποια είναι η δυνατότητα συνδεσιμότητας τους με διάφορες συσκευές, ποιες είναι οι βάσεις δεδομένων που θα χρησιμοποιηθούν για την εκπαίδευση των δικτύων, καθώς επίσης τι λογισμικό και ποιο νευρωνικό δίκτυο θα χρησιμοποιηθεί για την εκπαίδευση και αναγνώριση.

Με βάση αυτά τα ερωτήματα και δεδομένα, αρχικά έγινε μια παρουσίαση των βασικών χαρακτηριστικών ενός απλού νευρωνικού δικτύου και πως λειτουργεί. Στη συνέχεια παρουσιάστηκαν τα συνελκτικά νευρωνικά δίκτυα τα οποία χρησιμοποιούνται στο τομέα της αναγνώρισης εικόνων δείχνοντας επίσης μερικές από τις βασικότερες και πιο δημοφιλείς αρχιτεκτονικές που χρησιμοποιούνται για το σκοπό αυτό. Στη συνέχεια παρουσιάστηκαν οι κύριες μέθοδοι με τις οποίες κάποιος μπορεί να αναπτύξει σύστημα που κάνει αναγνώριση τροφίμων. Παρουσιάστηκαν οι κύριες και πιο δημοφιλής βάσεις δεδομένων που χρησιμοποιούνται για ανάπτυξη συστημάτων αναγνώρισης τροφίμων. Έγινε αναφορά στις πλατφόρμες που υπάρχουν και είναι ικανές να πραγματοποιήσουν αναγνώριση τροφίμων. Επίσης παρουσιάστηκαν εργασίες οι οποίες μελετούν τον τομέα αυτό. Στις μελέτες αυτές, αποδεικνύεται ότι οι αρχιτεκτονικές συνελκτικών νευρωνικών δικτύων είναι ιδιαίτερα αποτελεσματικές ως προς την δυνατότητα τους να πραγματοποιήσουν αναγνώριση και ταξινόμηση τροφίμων. Τέλος, στο πρακτικό μέρος της εργασίας, αναπτύχθηκαν νευρωνικά δίκτυα τα οποία έχουν την δυνατότητα αναγνώρισης της ποσότητας του γάλακτος σε διαυγή μπουκάλια.

Ο τρόπος που θα χρησιμοποιηθεί για την ανάπτυξη μιας τέτοιας εφαρμογής, η βάση δεδομένων, καθώς επίσης η αρχιτεκτονική του νευρωνικού δικτύου που θα εκπαιδευτεί, έχουν άμεση επίπτωση στο χρόνο που θα χρειαστεί για την εκπαίδευση και την αποτελεσματικότητα αναγνώρισης εικόνων. Εάν χρησιμοποιηθεί ελλιπής βάση δεδομένων τότε το εκπαιδευμένο νευρωνικό δίκτυο δεν θα είναι αποτελεσματικό. Το ίδιο ισχύει αν η αρχιτεκτονική δεν είναι ικανή να βγάλει αποτελέσματα με υψηλή ακρίβεια. Από την άλλη εάν η βάση δεδομένων είναι πολύ μεγάλη ή η αρχιτεκτονική του νευρωνικού δικτύου απαιτεί πολλαπλές υπολογιστικές πράξεις, τότε η ίδια η εκπαίδευση του νευρωνικού δικτύου θα κρατήσει περισσότερο χρόνο. Ως εκ τούτου, η μελέτη και ανάπτυξη μεθόδων και συστημάτων αναγνώρισης τροφίμων συνεχίζει να είναι στο επίκεντρο της έρευνας.

Από την βιβλιογραφία έχουν προταθεί πολλαπλές βάσεις δεδομένων καθώς επίσης αρχιτεκτονικές νευρωνικών δικτύων για τη δημιουργία όλο και πιο αποτελεσματικών συστημάτων. Η επιλογή ή η δημιουργία βάσης δεδομένων εξαρτάται από το σκοπό για τα οποία θα χρησιμοποιηθούν. Για παράδειγμα, για την δημιουργία ενός συστήματος που να αναγνωρίζει τροφές από την ελληνική κουζίνα, ιδεατά θα χρησιμοποιηθεί βάση δεδομένων που περιέχει ελληνικές κυρίως τροφές. Εναλλακτικά θα έπρεπε να χρησιμοποιηθεί βάση δεδομένων από παρόμοιες κουζίνες όπως είναι η Ιταλική ή η Μεσογειακή γενικά. Επιπλέον, με βάση την υπολογιστική δύναμη και το χρόνο που έχουμε, πιθανότατα να μην υπάρχει η δυνατότητα για εκπαίδευση του νευρωνικού δικτύου εξ αρχής αλλά να χρησιμοποιηθεί ένα ήδη εκπαιδευμένο δίκτυο και να εφαρμοστεί μεταφορά μάθησης σε αυτό. Τέλος, η ίδια η αρχιτεκτονική επηρεάζει την ακρίβεια των αποτελεσμάτων αφού κάποιες αρχιτεκτονικές μπορεί να είναι πιο αποτελεσματικές για την ανάπτυξη συστήματος αναγνώρισης σε σχέση με άλλες.

Στο εμπορικό τομέα, οι επιχειρήσεις ασχολούνται όλο και πιο εντατικά για την ανάπτυξη συσκευών με ενσωματωμένο σύστημα αναγνώρισης τροφίμων αλλά και εφαρμογών που κάνουν και εκείνες αναγνώριση με τη χρήση υλικού που έχει ήδη στην κατοχή του ο χρήστης όπως τα κινητά τηλέφωνα.

Ως προς το πρακτικό μέρος της εργασίας, δημιουργήθηκε νέα βάση δεδομένων με εικόνες από κάμερες, κινητά καθώς επίσης μηχανές αναζήτησης που περιέχουν μπουκάλια με γάλα. Στη συνέχεια εκπαιδεύτηκαν δυο διαφορετικά νευρωνικά δίκτυα. Η μία εκπαίδευση έγινε μέσω της πλατφόρμας Teachable Machine η οποία διαθέτει ήδη το εκπαιδευμένο νευρωνικό δίκτυο MobileNet και στη συνέχεια χρησιμοποιώντας μεταφορά μάθησης. Η δεύτερη μέθοδος αναπτύχθηκε και πάλι μέσω μεταφοράς μάθησης χρησιμοποιώντας την αρχιτεκτονική MobileNet με το πλάτος κάθε στοιβάδας νευρώνων μειωμένο κατά 50%. Τα αποτελέσματα για τα δεδομένα επιβεβαίωσης δείχνουν ότι τα εκπαιδευμένα νευρωνικά δίκτυα είναι αποτελεσματικά στην αναγνώριση ποσότητας γάλακτος αφού τα αποτελέσματα παρουσιάζουν μεγάλη ακρίβεια ταξινόμησης.

Παρουσιάζονται εφαρμογές όπου τα εκπαιδευμένα μοντέλα έχουν την δυνατότητα να χρησιμοποιηθούν σε μορφή διαδικτυακού συνδέσμου ή να χρησιμοποιηθούν σε εφαρμογές υπολογιστών ή άλλων συσκευών. Έχουν δηλαδή τη δυνατότητα να εκτελεστεί το πρόγραμμά τους σε υπολογιστές γραφείου, ενσωματωμένα συστήματα, εφαρμογές για κινητά ή να ενσωματωθεί σε προγράμματα περιήγησης.

Στη βιβλιογραφία παρατηρείται η προσπάθεια ανάπτυξης όλο και μεγαλύτερων και πιο αποτελεσματικών βάσεων δεδομένων με εικόνες οι οποίες μπορούν να χρησιμοποιηθούν σε όλο και πιο αποτελεσματικά συστήματα αναγνώρισης τροφίμων αναγνωρίζοντας όχι μόνο τον τύπο της τροφής αλλά ακόμα και την ποσότητα αλλά και την ποιότητα της. Όσο μεγαλύτερη και πιο πολύπλοκη είναι η βάση δεδομένων, τόσο γίνεται σημαντική η ανάπτυξη μεθόδων επεξεργασίας τους καθώς επίσης αρχιτεκτονικών νευρωνικών δικτύων που θα χρειαστεί για την εκπαίδευση. Οι μελέτες δείχνουν ότι με την χρήση μεταφοράς μάθησης υπάρχει μεγαλύτερη ακρίβεια στις προβλέψεις από το εκπαιδευμένο νευρωνικό δίκτυο σε σχέση με ένα άλλο, το οποίο έχει την ίδια αρχιτεκτονική αλλά εκπαιδεύεται για πρώτη φορά.

Μελλοντική έρευνα και εργασία που θα μπορούσε να αναπτυχθεί από την διπλωματική είναι μεταξύ άλλων η δημιουργία νέας βάσης δεδομένων η οποία έχει τη δυνατότητα αναγνώρισης μεγαλύτερου αριθμού κλάσεων τροφών σε διαφορετικές ποσότητες και ποιότητα. Επιπλέον, με τη χρήση μεταφοράς μάθησης στην αρχιτεκτονική MobileNet και με την ίδια βάση δεδομένων που χρησιμοποιήθηκε, μπορεί να γίνει τροποποίηση ως προς τις

παραμέτρους της εκπαίδευσης προκειμένου να μειωθεί ο χρόνος εκπαίδευσης, να αυξηθεί η ακρίβεια αναγνώρισης και να μειωθεί το υπολογιστικό κόστος. Αυτό μπορεί να γίνει μέσω της τροποποίησης των συναρτήσεων ενεργοποίησης, απώλειας και βελτιστοποίησης που χρησιμοποιείται για την εκπαίδευση. Τέλος, με την απλοποίηση της ίδιας της αρχιτεκτονικής του νευρωνικού δικτύου μπορεί να μελετηθεί κατά πόσο έχουμε τα ίδια ή παρόμοια αποτελέσματα με λιγότερους υπολογισμούς.

ΒΙΒΛΙΟΓΡΑΦΙΑ-ΠΗΓΕΣ

- [1] Obesity and Overweight Food, URL: <https://www.who.int/en/news-room/fact-sheets/detail/obesity-and-overweight>, προσπελάστηκε στις 10/06/2021.
- [2] Proceedings of the 2013 Annual Meeting of the House of Delegates, URL: <http://www.ama-assn.org/ama/pub/news/news/2013/2013-06-18-new-ama-policies-annual-meeting.page>, προσπελάστηκε στις 10/06/2021
- [3] Nutrition: Keeping a Food Diary, URL: <https://familydoctor.org/nutrition-keeping-a-food-diary/>, προσπελάστηκε στις 10/10/2021
- [4] Salim, Nareen O., et al. "Study for Food Recognition System Using Deep Learning." Journal of Physics: Conference Series, vol. 1963, no. 1, 2021, p. 012014., <https://doi.org/10.1088/1742-6596/1963/1/012014>.
- [5] Min, W., Jiang, S., Liu, L., Rui, Y., & Jain, R. (2019). A survey on Food Computing. ACM Computing Surveys, 52(5), 1–36. <https://doi.org/10.1145/3329168>
- [6] Qaraqe, M., Usman, M., Ahmad, K., Sohail, A., & Boyaci, A. (2020). Automatic Food Recognition System for middle-eastern cuisines. IET Image Processing, 14(11), 2469–2479. <https://doi.org/10.1049/iet-ipr.2019.1051>
- [7] Truly Smart Appliances: How Computer Vision Can Improve Our Daily Lives, URL: <https://www.iotforall.com/truly-smart-appliances-how-computer-vision-can-improve-our-daily-lives/>, προσπελάστηκε στις 18/08/2020.
- [8] Smarter FridgeCam Έξυπνος Αισθητήρας, URL: https://www.kotsovolos.gr/computing/smart-home/eksupnoi-aisthithiras/204911-smarter-fridgecam?gclid=Cj0KQCQjwit_8BRCoARIsAlx3Rj7p4But3pkMatPLCzJ2_vkmus4TBr7hdrE58VS6TJrpdN-u9gQu1A0aAkOMEALw_wcB, προσπελάστηκε στις 1/12/2021.
- [9] Samsung's New Fridge Uses AI To Track Food Inventory. URL: [Samsung's Fridge Uses AI To Track Food Inventory | PYMNTS.com](https://www.pymnts.com/samsung-fridge-ai/), προσπελάστηκε στις 12/09/2020.
- [10] Samsung and LG go head to head with AI-powered fridges that recognize food. URL: [Samsung and LG go head to head with AI-powered fridges that recognize food - The Verge](https://www.theverge.com/2020/9/12/21911111/samsung-lg-ai-fridges), προσπελάστηκε στις 12/09/2020.
- [11] Do More with Bixby, URL: <https://www.samsung.com/global/galaxy/apps/bixby/vision/>, προσπελάστηκε στις 12/09/2020.
- [12] Foodai Smart Food Recognition with the state-of-the-art Visual Recognition technology URL: <https://foodai.org/>, προσπελάστηκε στις 15/11/2020.
- [13] Your personal nutrition guide, URL: <https://www.foodvisor.io/en/>, προσπελάστηκε στις 7/4/2021.
- [14] Simon Haykin (1999). Neural Networks: A Comprehensive Foundation, 2nd and 3d edition. Pearson Prentice Hall. pp.1-14.
- [15] Artificial Neural Network – Basic Concepts. URL: https://www.tutorialspoint.com/artificial_neural_network/artificial_neural_network_basic_concepts.htm, προσπελάστηκε στις 08/08/2020.

- [16] A schematic diagram of a Multi-Layer Perceptron (MLP) neural network. URL: https://www.researchgate.net/figure/A-schematic-diagram-of-a-Multi-Layer-Perceptron-MLP-neural-network_fig3_257071174 , προσπελάστηκε στις 08/08/2020.
- [17] Rachid, Taleb & Meroufel, Abdelkader & Wira, Patrice. (2009). Neural Network Control of Asymmetrical Multilevel Converters. Leonardo Journal of Sciences. 8.
- [18] Elizondo, D., & Fiesler, E. (1997). A survey of partially connected neural networks. International journal of neural systems, 8(5-6), pp. 535–558.
- [19] [Bottou Léon](#), Bousquet Olivier (2012). The Tradeoffs of Large Scale Learning. In Sra, Suvrit; Nowozin, Sebastian Wright, Stephen J. (eds.). Optimization for Machine Learning. Cambridge: MIT Press. pp. 351–368
- [20] How The backpropagation algorithm works , URL: <http://neuralnetworksanddeeplearning.com/chap2.html> , προσπελάστηκε στις 21/5/2021.
- [21] Understand the Impact of Learning Rate on Neural Network Performance, URL: <https://machinelearningmastery.com/understand-the-dynamics-of-learning-rate-on-deep-learning-neural-networks/> , προσπελάστηκε στις 21/5/2021.
- [22] Fundamentals of Deep Learning – Activation Functions and When to Use Them?, URL: <https://www.analyticsvidhya.com/blog/2020/01/fundamentals-deep-learning-activation-functions-when-to-use-them/> , προσπελάστηκε στις 21/5/2021.
- [23] Activation Functions in Neural Networks, URL: <https://towardsdatascience.com/activation-functions-neural-networks-1cbd9f8d91d6> , προσπελάστηκε στις 21/5/2021.
- [24] Rectified Linear Units (ReLU) in Deep Learning, URL: <https://www.kaggle.com/dansbecker/rectified-linear-units-relu-in-deep-learning> , προσπελάστηκε στις 14/10/2020.
- [25] A Gentle Introduction to the Rectified Linear Unit(ReLU), URL: <https://machinelearningmastery.com/rectified-linear-activation-function-for-deep-learning-neural-networks/> , προσπελάστηκε στις 14/10/2020.
- [26] Han, J., & Moraga, C. (1995). The influence of the sigmoid function parameters on the speed of backpropagation learning. Lecture Notes In Computer Science, 195-201. doi: 10.1007/3-540-59497-3_175
- [27] Raschka, Sebastian (2019). Python machine learning : machine learning and deep learning with python, scikit-learn, and tensorflow 2. Birmingham: Packt Publishing, Limited. pp. 37 - 38
- [28] Mean squared error, URL: <https://peltarion.com/knowledge-center/documentation/modeling-view/build-an-ai-model/loss-functions/mean-squared-error> , προσπελάστηκε στις 10/10/2021.
- [29] Intuition behind Log-loss score, URL: <https://towardsdatascience.com/intuition-behind-log-loss-score-4e0c9979680a> , προσπελάστηκε στις 10/10/2021.
- [30] Understand the Impact of Learning Rate on Neural Network Performance, URL: <https://machinelearningmastery.com/understand-the-dynamics-of-learning-rate-on-deep-learning-neural-networks/> , προσπελάστηκε στις 14/10/2020.

- [31] Optimization Algorithms in Neural Networks, URL: <https://www.kdnuggets.com/2020/12/optimization-algorithms-neural-networks.html> , προσπελάστηκε στις 14/10/2020.
- [32] What is the optimization theory? Is it used in machine learning?, URL: <https://www.quora.com/What-is-the-optimization-theory-Is-it-used-in-machine-learning> , προσπελάστηκε στις 14/10/2020.
- [33] Intro to optimization in deep learning: Gradient Descent, URL: <https://blog.paperspace.com/intro-to-optimization-in-deep-learning-gradient-descent/> , προσπελάστηκε στις 14/10/2020.
- [34] [Bottou, Léon](#); Bousquet, Olivier (2012). "[The Tradeoffs of Large Scale Learning](#)". In Sra, Suvrit; Nowozin, Sebastian; Wright, Stephen J. (eds.). Optimization for Machine Learning. Cambridge: MIT Press. pp. 351–368
- [35] Stochastic Gradient Descent with momentum, URL: <https://towardsdatascience.com/stochastic-gradient-descent-with-momentum-a84097641a5d> , προσπελάστηκε στις 11/10/2021.
- [36] Understanding RMSprop — faster neural network learning, URL: <https://towardsdatascience.com/understanding-rmsprop-faster-neural-network-learning-62e116fcf29a> , προσπελάστηκε στις 11/4/2021.
- [37] Recurrent Neural Networks, URL: towardsdatascience.com/recurrent-neural-networks-d4642c9bc7ce, προσπελάστηκε στις 10/10/2020.
- [38] A Beginner's Guide to Convolutional Neural Networks (CNNs) URL: <https://wiki.pathmind.com/convolutional-network> , προσπελάστηκε στις 04/02/2020.
- [39] The Complete Beginner's Guide to Deep Learning: Convolutional Neural Networks and Image Classification, URL: <https://towardsdatascience.com/wtf-is-image-classification-8e78a8235acb>, προσπελάστηκε στις 13/10/2020.
- [40] Ian Goodfellow, Yoshua Bengio, Aaron Courville (2016), Deep Learning, MIT Press. pp. 341-342.
- [41] Convolutional Neural Networks (CNNs / ConvNets), URL: <https://cs231n.github.io/convolutional-networks/>, προσπελάστηκε στις 09/10/2020.
- [42] Γεώργιος Καραγιάννης. Πέτρος Α. Μαραγκός (2011). Βασικές Αρχές Σημάτων & Συστημάτων. Αθήνα: Παπασωτηρίου ΕΚΔΟΣΕΙΣ. Pp. 91-105.
- [43] Convolution , URL : <https://mathworld.wolfram.com/Convolution.html> , προσπελάστηκε στις 07/09/2020.
- [44] Ian Goodfellow, Yoshua Bengio, Aaron Courville (2016), Deep Learning, MIT Press. pp. 331-338.
- [45] Deep learning- Overview of Convolution Neural Network, URL: <https://mc.ai/deeplearning-overview-of-convolution-neural-network/> , προσπελάστηκε στις 13/10/2020.
- [46] Understanding Parameter Sharing (or weights replication) Within Convolutional Neural Networks, URL: <https://towardsdatascience.com/understanding-parameter-sharing-or-weights-replication-within-convolutional-neural-networks-cc26db7b645a> , προσπελάστηκε στις 13/10/2020.

- [47] Ian Goodfellow, Yoshua Bengio, Aaron Courville (2016), Deep Learning, MIT Press. pp. 339-362.
- [48] Convolutional Neural Network: Learn and Apply, URL: medium.com/@sdoshi579/convolutional-neural-network-learn-and-apply-3dac9acfe2b6, προσπελάστηκε στις 01/03/2020.
- [49] Stride (Machine Learning) Definition, URL: <https://deepai.org/machine-learning-glossary-and-terms/stride>, προσπελάστηκε στις 14/10/2020.
- [50] A Beginner's Guide to Object Detection, URL: <https://www.datacamp.com/community/tutorials/object-detection-guide>, προσπελάστηκε στις 20/11/2021.
- [51] What Is The Difference Between Image Classification & Object Detection Techniques?, URL: <https://analyticsindiamag.com/what-is-the-difference-between-image-classification-object-detection-techniques/>, προσπελάστηκε στις 20/11/2021.
- [52] Liu W. et al. (2016) SSD: Single Shot MultiBox Detector. In: Leibe B., Matas J., Sebe N., Welling M. (eds) Computer Vision – ECCV 2016. ECCV 2016. Lecture Notes in Computer Science, vol 9905. Springer, Cham.
- [53] AlexNet: The Architecture that Challenged CNNs, URL: <https://towardsdatascience.com/alexnet-the-architecture-that-challenged-cnns-e406d5297951>, προσπελάστηκε στις 15/10/2021.
- [54] Alex Krizhevsky, Ilya Sutskever, Geoffrey E. Hinton. (2017). ImageNet Classification with Deep Convolutional Neural Networks.
- [55] Inception, ResNet, MobileNet, URL: <https://medium.com/@fransiska26/the-differences-between-inception-resnet-and-mobilenet-e97736a709b0>, προσπελάστηκε στις 20/4/2021.
- [56] Szegedy, C., Wei Liu, Yangqing Jia, Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., & Rabinovich, A. (2015). Going deeper with convolutions. 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). <https://doi.org/10.1109/cvpr.2015.7298594>
- [57] Understanding GoogLeNet Model – CNN Architecture, URL: <https://www.geeksforgeeks.org/understanding-googlenet-model-cnn-architecture/>, προσπελάστηκε στις 20/5/2021.
- [58] Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., & Wojna, Z. (2016). Rethinking the inception architecture for computer vision. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). <https://doi.org/10.1109/cvpr.2016.308>
- [59] A Simple Guide to the Versions of the Inception Network, URL: <https://towardsdatascience.com/a-simple-guide-to-the-versions-of-the-inception-network-7fc52b863202>, προσπελάστηκε στις 19/5/2021.
- [60] Karen Simonyan, Andrew Zisserman. (2014). Very Deep Convolutional Networks for Large-Scale Image Recognition.
- [61] VGG模型学习, URL: https://blog.csdn.net/weixin_45521594/article/details/105490421, προσπελάστηκε στις 19/5/2021.

- [62] Howard, Andrew & Zhu, Menglong & Chen, Bo & Kalenichenko, Dmitry & Wang, Weijun & Weyand, Tobias & Andreetto, Marco & Adam, Hartwig. (2017). MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications.
- [63] A Basic Introduction to Separable Convolutions, URL: <https://towardsdatascience.com/a-basic-introduction-to-separable-convolutions-b99ec3102728> , προσπελάστηκε στις 15/10/2020.
- [64] Howard, Andrew & Zhu, Menglong & Chen, Bo & Kalenichenko, Dmitry & Wang, Weijun & Weyand, Tobias & Andreetto, Marco & Adam, Hartwig. (2017). MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications.
- [65] Ioffe, Sergey & Szegedy, Christian. (2015). Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift.
- [66] Understand the Softmax Function in Minutes, URL: <https://medium.com/data-science-bootcamp/understand-the-softmax-function-in-minutes-f3a59641e86d> , προσπελάστηκε στις 15/10/2020.
- [67] Kaiming He, Xiangyu Zhang, Shaoqing Ren, Jian Sun, (2015). Deep Residual Learning for Image Recognition.
- [68] An Overview of ResNet and its Variants, URL: <https://towardsdatascience.com/an-overview-of-resnet-and-its-variants-5281e2f56035> , προσπελάστηκε στις 15/10/2020.
- [69] 5 Amazing Deep Learning Frameworks Every Data Scientist Must Know! (with Illustrated Infographic), URL: <https://www.analyticsvidhya.com/blog/2019/03/deep-learning-frameworks-comparison/> , προσπελάστηκε στις 13/9/2021.
- [70] An end-to-end open source machine learning platform, URL: <https://www.tensorflow.org/> , προσπελάστηκε στις 13/9/2021.
- [71] FROM RESEARCH TO PRODUCTION, URL: <https://pytorch.org/> , προσπελάστηκε στις 13/9/2021.
- [72] What is Caffe2? , <https://caffe2.ai/docs/caffe-migration.html> , προσπελάστηκε στις 13/9/2021.
- [73] Simple. Flexible. Powerful. , <https://keras.io/> , προσπελάστηκε στις 13/9/2021.
- [74] Theano, <https://github.com/Theano/Theano> , προσπελάστηκε στις 13/9/2021.
- [75] A FLEXIBLE AND EFFICIENT LIBRARY FOR DEEP LEARNING , <https://mxnet.apache.org/versions/1.8.0/> , προσπελάστηκε στις 13/9/2021.
- [76] Deeplearning4j Suite Overview , <https://deeplearning4j.konduit.ai/> , προσπελάστηκε στις 13/9/2021.
- [77] Chairi Kiourt, George Pavlidis and Stella Markantonatou, (2020). Deep learning approaches in food recognition.
- [78] A Comprehensive Hands-on Guide to Transfer Learning with Real-World Applications in Deep Learning, <https://towardsdatascience.com/a-comprehensive-hands-on-guide-to-transfer-learning-with-real-world-applications-in-deep-learning-212bf3b2f27a> , προσπελάστηκε στις 13/5/2021.
- [79] Transfer Learning - Machine Learning's Next Frontier , <https://runder.io/transfer-learning/index.html#whatistransferlearning> , προσπελάστηκε στις 13/5/2021

- [80] Laith Alzubaidi, Jinglan Zhang, Amjad J. Humaidi, Ayad Al-Dujaili, Ye Duan, Omran Al-Shamma, J. Santamaría, Mohammed A. Fadhel, Muthana Al-Amidie and Laith Farhan,(2021), Review of deep learning: concepts, CNN architectures, challenges, applications, future directions.
- [81] 9 nutrition and diet apps for 2020 , <https://wa-health.kaiserpermanente.org/best-diet-apps/> , προσπελάστηκε στις 13/5/2021.
- [82] Vision AI , <https://cloud.google.com/vision> , προσπελάστηκε στις 13/5/2021.
- [83] The World's AI Computer Vision NLP Audio Recognition Automatic Labeling Community, <https://www.clarifai.com/> , προσπελάστηκε στις 13/5/2021.
- [84] Amazon Rekognition, <https://aws.amazon.com/rekognition/?blog-cards.sort-by=item.additionalFields.createdDate&blog-cards.sort-order=desc> , προσπελάστηκε στις 13/5/2021.
- [85] Computer Vision, <https://azure.microsoft.com/en-us/services/cognitive-services/computer-vision/> , προσπελάστηκε στις 13/5/2021.
- [86] Watson Visual Recognition, <https://www.ibm.com/dk-en/cloud/watson-visual-recognition> , προσπελάστηκε στις 13/5/2021.
- [87] Foodvisor automatically tracks what you eat using deep learning, https://techcrunch.com/2019/10/14/foodvisor-automatically-tracks-what-you-eat-using-deep-learning/?guccounter=1&guce_referrer=aHR0cHM6Ly93d3cuZ29vZ2xlLmNvbS8&guce_referrer_sig=AQAAAAJuY2nlF6tGgNxmANUKqND3N8msvsQecFCiJDPgDJLLPPP_2s7MmEVTjzm8DlSg9WzltldMX7zWDrz7_5kcB1bcgAooOpnoGrxRIO8ffCqdVCRIxVQWhKBFlajwnfDm9Fdv2C9_2xS0kcywPeiFgTvIV7B9CTKLogKfvgSa0yGn , προσπελάστηκε στις 13/5/2021.
- [88] Bitesnap: The easier way to track what you eat., <https://getbitesnap.com/> , προσπελάστηκε στις 13/5/2021.
- [89] Top Rated Weight Loss Plan, <https://www.loseit.com/> , προσπελάστηκε στις 13/5/2021.
- [90] Calorie Mama Logo Instant Food Recognition, <https://caloriemama.ai/> , προσπελάστηκε στις 13/5/2021.
- [91] Dario Allegra, Sebastiano Battiato, Alessandro Ortis, Salvatore Urso and Riccardo Polosa,(2021), A review on food recognition technology for health applications.
- [92] Food-101 – Mining Discriminative Components with Random Forests, https://data.vision.ee.ethz.ch/cvl/datasets_extra/food-101/ , προσπελάστηκε στις 13/5/2021.
- [93] An Update to the ImageNet Website and Dataset March 11, 2021, <https://imagenet.org/update-mar-11-2021.php> , προσπελάστηκε στις 13/5/2021.
- [94] Xin Chen, Yu Zhu, Hua Zhou, Liang Diao, and Dongyan Wang,(2017), ChineseFoodNet: A Large-scale Image Dataset for Chinese Food Recognition.
- [94] Xin Chen, Yu Zhu, Hua Zhou, Liang Diao, and Dongyan Wang,(2017), ChineseFoodNet: A Large-scale Image Dataset for Chinese Food Recognition.
- [95] Parneet Kaur, Karan Sikka, Weijun Wang, Serge Belongieq, Ajay Divakaran,(2019), FoodX-251: A Dataset for Fine-grained Food Classification.

- [96] "UEC FOOD 100": 100-kind food dataset (release 1.0) and "UEC FOOD 256": 256-kind food dataset (release 1.0), 2021, <http://foodcam.mobi/dataset.html> , προσπελάστηκε στις 13/5/2021.
- [97] Food 524 Database, 2021, <http://www.ivl.disco.unimib.it/activities/food524db/> , προσπελάστηκε στις 13/5/2021.
- [98] EgocentricFood, 2021, <http://www.ub.edu/cvub/egocentricfood/> , προσπελάστηκε στις 13/5/2021.
- [99] Javier Marin, Aritro Biswas, Ferda Ofli, Nicholas Hynes, Amaia Salvador, Yusuf Aytar, Ingmar Weber, Antonio Torralba,(2019), Recipe1M+: A Dataset for Learning Cross-Modal Embeddings for Cooking Recipes and Food Images.
- [100] Chen Guanghua, Zhang Xiaolong,(2015), A Method to Improve Robustness of the Gray World Algorithm.
- [101] What is Histogram Equalization and how it works?, <https://www.mygreatlearning.com/blog/histogram-equalization-explained/> , προσπελάστηκε στις 13/5/2021.
- [102] Deep Learning Based Food Recognition, <http://cs229.stanford.edu/proj2016/report/YuMaoWang-Deep%20Learning%20Based%20Food%20Recognition-report.pdf> , προσπελάστηκε στις 13/5/2021.
- [103] Pandey, P., Deepthi, A., Mandal, B., & Puhan, N. (2017). FoodNet: Recognizing Foods Using Ensemble of Deep Networks. IEEE Signal Processing Letters, 24(12), 1758-1762. doi: 10.1109/lsp.2017.2758862
- [104] Martinel, N., Foresti, G. L., & Micheloni, C. (2018). Wide-slice residual networks for food recognition. 2018 IEEE Winter Conference on Applications of Computer Vision (WACV). <https://doi.org/10.1109/wacv.2018.00068>
- [105] Liu, C., Cao, Y., Luo, Y., Chen, G., Vokkarane, V., & Ma, Y. (2021). DeepFood: Deep Learning-Based Food Image Recognition for Computer-Aided Dietary Assessment, <https://arxiv.org/pdf/1606.05675.pdf>, προσπελάστηκε στις 6/12/2021.
- [106] Zhou, L., Zhang, C., Liu, F., Qiu, Z., & He, Y. (2019). Application of Deep Learning in Food: A Review. Comprehensive Reviews In Food Science And Food Safety, 18(6), 1793-1811. doi: 10.1111/1541-4337.12492
- [107] Jahani Heravi, E., Habibi Aghdam, H., & Puig, D. (2018). An optimized convolutional neural network with bottleneck and spatial pyramid pooling layers for classification of foods. Pattern Recognition Letters, 105, 50-58. doi: 10.1016/j.patrec.2017.12.007
- [108] Van Asbroeck, S., & Matthys, C. (2020). Use of Different Food Image Recognition Platforms in Dietary Assessment: Comparison Study. JMIR Formative Research, 4(12), e15602. Doi: 10.2196/15602
- [109] Nguyen, B.T. et al., 2018. A deep learning based food recognition system for Lifelog images. Proceedings of the 7th International Conference on Pattern Recognition Applications and Methods.
- [110] Sahoo, D. et al., 2019. FoodAI: Food image recognition via deep learning for smart food logging . Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining.

- [111] Hoi, S., Food image recognition by Deep Learning - Nvidia. Available at: <https://images.nvidia.com/content/APAC/events/ai-conference/resource/ai-for-research/FoodAI-Food-Image-Recognition-with-Deep-Learning.pdf> , προσπελάστηκε στις 6/12 /2021.
- [112] Tsang, S.-H., 2019. Review: Senet-squeeze-and-excitation network, winner of ILSVRC 2017 (image classification). Medium. Available at: , προσπελάστηκε στις 6/12/2021.
- [113] Trivedi, S., 2020. Understanding focal loss-A quick read. Medium. Available at: <https://medium.com/visionwizard/understanding-focal-loss-a-quick-read-b914422913e7>, προσπελάστηκε στις 6/12/2021.
- [114] Kiourt, C., Pavlidis, G. & Markantonatou, S., 2020. Deep learning approaches in food recognition. Learning and Analytics in Intelligent Systems, pp.83–108.
- [115] Attokaren, D.J. et al., 2017. Food classification from images using Convolutional Neural Networks. TENCON 2017 - 2017 IEEE Region 10 Conference.
- [116] Bolaños, M., Valdivia, M. & Radeva, P., 2019. Where and what am I eating? image-based food menu recognition. Lecture Notes in Computer Science, pp.590–605.
- [117] Artificial Intelligence and Deep Learning Solutions for food recognition. LogMeal Food Recognition and Detection. Available at: <https://www.logmeal.es/>, προσπελάστηκε στις 6/12/2021.
- [118] Baheti, P., 2021. Introduction to multimodal deep learning. Medium. Available at: <https://heartbeat.fritz.ai/introduction-to-multimodal-deep-learning-630b259f9291>, προσπελάστηκε στις 7/12/2021.
- [119] Anon, (PDF) long short-term memory. ResearchGate. Available at: https://www.researchgate.net/publication/13853244_Long_Short-term_Memory, προσπελάστηκε στις 7/12/2021.
- [120] Anon, ConceptNet. Available at: <https://conceptnet.io/>, προσπελάστηκε στις 7/12/2021. [121] Anon, Machine learning for kids. Machine Learning for Kids. Available at: <https://machinelearningforkids.co.uk/>, προσπελάστηκε στις 7/12/2021.
- [122] Teachable Machine. Available at: <https://teachablemachine.withgoogle.com/>, προσπελάστηκε στις 7/12/2021.
- [123] Evaluating a machine learning model. (2021). . Available at: <https://www.jeremyjordan.me/evaluating-a-machine-learning-model/>, προσπελάστηκε στις 13/12/2021.
- [124] Anon, P5.js web editor. Available at: https://editor.p5js.org/ml5/sketches/ImageModel_TM, , προσπελάστηκε στις 13/12/2021.
- [125] Webhook, 2022, <https://webhook.site> , προσπελάστηκε στις 2/2/2022.
- [126] Ruder, S. (2020, March 20). An overview of gradient descent optimization algorithms. Sebastian Ruder. <https://ruder.io/optimizing-gradient-descent/index.html#adam> , προσπελάστηκε στις 22/1/2022.

ΠΑΡΑΡΤΗΜΑ 1: ΠΡΟΓΡΑΜΜΑ ΕΚΠΑΙΔΕΥΣΗΣ ΝΕΥΡΩΝΙΚΟΥ ΔΙΚΤΥΟΥ ΜΟΒΙΛΕΝΕΤ ΜΕ ΜΕΤΑΦΟΡΑ ΜΑΘΗΣΗΣ

```
import matplotlib.pyplot as plt
import seaborn as sns
import keras
from keras.models import Sequential
from keras.layers import Dense, Conv2D, MaxPool2D, Flatten, Dropout
from keras.preprocessing.image import ImageDataGenerator
from keras.optimizers import Adam
from keras.losses import CategoricalCrossentropy
from sklearn.metrics import classification_report, confusion_matrix
import tensorflow as tf
import cv2
import os
import numpy as np
labels = ['full', 'half', 'half_empty', 'almost_full', 'random']
img_size = 224
def get_data(data_dir):
    data = []
    for label in labels:
        path = os.path.join(data_dir, label)
        class_num = labels.index(label)
        for img in os.listdir(path):
            try:
                img_arr = cv2.imread(os.path.join(path, img))[...,:-1]
                resized_arr = cv2.resize(img_arr, (img_size, img_size))
                data.append([resized_arr, class_num])
            except Exception as e:
                print(e)
    return np.array(data)
#####
train = get_data('C:/Users/panto/OneDrive/Desktop/newtrain/train')
val = get_data('C:/Users/panto/OneDrive/Desktop/newtrain/test')
#####
```

```
# Δημιουργείται γραφική παράσταση που δείχνει τον αριθμό των εικόνων σε κάθε φακέλου
#για εκπαίδευση.
```

```
l = []
```

```
for i in train:
```

```
    if(i[1] == 0):
```

```
        l.append("full")
```

```
    elif(i[1]==1):
```

```
        l.append("half")
```

```
    elif(i[1]==2):
```

```
        l.append("half_empty")
```

```
    elif(i[1]==3):
```

```
        l.append("almost_full")
```

```
    else:
```

```
        l.append("random")
```

```
sns.set_style('darkgrid')
```

```
sns.countplot(l)
```

```
#####
```

```
l = []
```

```
for i in val:
```

```
    if(i[1] == 0):
```

```
        l.append("full")
```

```
    elif(i[1]==1):
```

```
        l.append("half")
```

```
    elif(i[1]==2):
```

```
        l.append("half_empty")
```

```
    elif(i[1]==3):
```

```
        l.append("almost_full")
```

```
    else:
```

```
        l.append("random")
```

```
sns.set_style('darkgrid')
```

```
sns.countplot(l)
```

```
x_train = []
```

```
y_train = []
```

```
x_val = []
```

```
y_val = []
```

```
for feature, label in train:
```

```
    x_train.append(feature)
```

```

y_train.append(label)
for feature, label in val:
    x_val.append(feature)
    y_val.append(label)
# Normalize the data
x_train = np.array(x_train) / 255
x_val = np.array(x_val) / 255
x_train.reshape(-1, img_size, img_size, 1)
y_train = np.array(y_train)
x_val.reshape(-1, img_size, img_size, 1)
y_val = np.array(y_val)
#####
datagen = ImageDataGenerator(
    featurewise_center=False, # set input mean to 0 over the dataset
    samplewise_center=False, # set each sample mean to 0
    featurewise_std_normalization=False, # divide inputs by std of the dataset
    samplewise_std_normalization=False, # divide each input by its std
    zca_whitening=False, # apply ZCA whitening
    fill_mode='nearest',
    brightness_range=[0.0,1.5],
    rotation_range = 30, # randomly rotate images in the range (degrees, 0 to 180)
    zoom_range = 0.2, # Randomly zoom image
    width_shift_range=0.15, # randomly shift images horizontally (fraction of total width)
    height_shift_range=0.15, # randomly shift images vertically (fraction of total height)
    horizontal_flip = True, # randomly flip images
    vertical_flip=True) # randomly flip images
datagen.fit(x_train)
base_model = tf.keras.applications.MobileNet(input_shape = (224, 224, 3), alpha=0.50,
include_top = False, weights = "imagenet")
base_model.trainable = False
base_model.summary()
#####
model = tf.keras.Sequential([base_model,
                             #tf.keras.layers.MaxPool2D(),
                             #tf.keras.layers.GlobalAveragePooling2D(),
                             tf.keras.layers.Dropout(0.2),
                             tf.keras.layers.Flatten(),

```

```

        #tf.keras.layers.Dense(50, activation="relu"),
        #tf.keras.layers.Dense(25, activation="relu"),
        tf.keras.layers.Dense(5, activation="softmax")
    ])
base_learning_rate = 0.00001
model.compile(optimizer= Adam(lr=base_learning_rate),
              loss= tf.keras.losses.SparseCategoricalCrossentropy(from_logits=False),
              metrics=['accuracy'])
model.summary()
#####
epochs_first=500
batch=120
history = model.fit(x_train, y_train, epochs = epochs_first ,batch_size=batch, shuffle=True,
                    validation_data = (x_val, y_val))
#####
acc = history.history['accuracy']
val_acc = history.history['val_accuracy']
loss = history.history['loss']
val_loss = history.history['val_loss']
epochs_range = range(epochs_first)
plt.figure(figsize=(15, 15))
plt.subplot(2, 2, 1)
plt.plot(epochs_range, acc, label='Training Accuracy')
plt.plot(epochs_range, val_acc, label='Validation Accuracy')
plt.legend(loc='lower right')
plt.title('Training and Validation Accuracy')
plt.subplot(2, 2, 2)
plt.plot(epochs_range, loss, label='Training Loss')
plt.plot(epochs_range, val_loss, label='Validation Loss')
plt.legend(loc='upper right')
plt.title('Training and Validation Loss')
plt.show()

```


ΠΑΡΑΡΤΗΜΑ 2: KERAS.H5

```
import requests
import json
import tensorflow as tf
from keras.models import load_model
from PIL import Image, ImageOps, ImageFont, ImageDraw
import numpy as np
url_webhook = 'https://webhook.site/b6048c00-34cc-4d5b-bfe9-5644f9c7e76a'
labels = ['full', 'half', 'half_empty', 'almost_full', 'random']
# Load the model
#model = load_model('C:/Users/panto/OneDrive/Desktop/milk2/converted_keras(1)/keras_model.h5')
#model = load_model('C:/Users/panto/OneDrive/Desktop/milk3/converted_keras(1)/keras_model.h5')
model = load_model('C:/Users/panto/OneDrive/Desktop/newtrain/tunedmodel99.h5')
# Create the array of the right shape to feed into the keras model
# The 'length' or number of images you can put into the array is
# determined by the first position in the shape tuple, in this case 1.
data = np.ndarray(shape=(1, 224, 224, 3), dtype=np.float32)
# Replace this with the path to your image
image = Image.open('C:/Users/panto/OneDrive/Desktop/images for
test/mobile_more_than_half/IMG_20200914_140528.jpg')
#resize the image to a 224x224 with the same strategy as in TM2:
#resizing the image to be at least 224x224 and then cropping from the center
image.show()
size = (224, 224)
image = ImageOps.fit(image, size, Image.ANTIALIAS)
#turn the image into a numpy array
image_array = np.asarray(image)
# Normalize the image
normalized_image_array = (image_array.astype(np.float32) / 127.0) - 1
# Load the image into the array
data[0] = normalized_image_array
# run the inference
prediction = model.predict(data)
```

```

k = prediction[0]
print(k)
j = np.unravel_index(k.argmax(), k.shape)
print(k[j])
print(j)
if j==(0,):
    print("There is a bottle with milk")
    k.show()
    data={'name': 'Class 0',
          'message': 'No need to buy a bottle of milk'}
    r = requests.post(url_webhook, data = data)
elif j==(1,):
    print("half")
    data={'name': 'Class 1',
          'message': 'No need to buy a bottle of milk'}
    r = requests.post(url_webhook, data = data)
    k.show()
elif j==(2,):
    print("a bottle of milk-Half empty")
    data={'name': 'Class 2',
          'message': 'Need to buy a bottle of milk'}
    r = requests.post(url_webhook, data = data)
    k.show()
elif j==(3,):
    print("a bottle of milk-Half full")
    data={'name': 'Class 3',
          'message': 'No need to buy a bottle of milk'}
    r = requests.post(url_webhook, data = data)
    k.show()
elif j==(4,):
    print("Image with no bottle with milk")
    data={'name': 'Class 4',
          'message': 'Need to buy a bottle of milk'}
    r = requests.post(url_webhook, data = data)
    k.show()
else:
    print('There is no image')

```

ΠΑΡΑΡΤΗΜΑΤΑ ΠΑΡΑΡΤΗΜΑ 3: ΠΡΟΓΡΑΜΜΑ .HTML ΑΡΧΕΙΟΥ ΜΕ ΕΝΣΩΜΑΤΩΜΕΝΟ ΠΡΟΓΡΑΜΜΑ JAVASCRIPT ΤΟ ΟΠΟΙΟ ΚΑΝΕΙ ΑΝΑΓΝΩΡΙΣΗ ΤΡΟΦΙΜΩΝ ΜΕΣΩ ΤΟΥ ΕΚΠΑΙΔΕΥΜΕΝΟΥ ΜΟΝΤΕΛΟΥ

```
<!DOCTYPE html>
<html>
<head>
<style>
.myDiv {
  border: 5px outset red;
  background-color: lightblue;
  text-align: center;
}
</style>
<style>
.centered {
  position: fixed;
  top: 30%;
  left: 40%;
}
</style>
<style>
.borders {
  position: fixed;
  top: 10%;
  left: 40%;
  border: 0;
  line-height: 2.5;
  padding: 0 20px;
  font-size: 1rem;
  text-align: center;
  color: #fff;
  text-shadow: 1px 1px 1px #000;
  border-radius: 10px;
  background-color: rgba(220, 0, 0, 1);
  background-image: linear-gradient(to top left,
                                   rgba(0, 0, 0, .2),
```

```

        rgba(0, 0, 0, .2) 30%,
        rgba(0, 0, 0, 0));
    box-shadow: inset 2px 2px 3px rgba(255, 255, 255, .6),
        inset -2px -2px 3px rgba(0, 0, 0, .6);
}
</style>
<style>
.lob {
    position: fixed;
    top: 58%;
    left: 40%;
    color: red;
}
</style>
</head>
<body>
<div class="myDiv">Teachable Machine Image Model</div>
<button class="borders" type="button" onclick="init()">Start</button>
<div class="centered" id="webcam-container"></div>
<div class="lob" id="label-container"></div>
<script src="https://cdn.jsdelivr.net/npm/@tensorflow/tfjs@1.3.1/dist/tf.min.js"></script>
<script
src="https://cdn.jsdelivr.net/npm/@teachablemachine/image@0.8/dist/teachablemachine-
image.min.js"></script>
<script type="text/javascript">
    const URL = "https://teachablemachine.withgoogle.com/models/BD6EpYrfM/";
    let model, webcam, labelContainer, maxPredictions;
    async function init() {
        const modelURL = URL + "model.json";
        const metadataURL = URL + "metadata.json";
        model = await tmImage.load(modelURL, metadataURL);
        maxPredictions = model.getTotalClasses();
        const flip = true;
        webcam = new tmImage.Webcam(200, 200, flip);
        await webcam.setup();
        await webcam.play();
        window.requestAnimationFrame(loop);
        document.getElementById("webcam-container").appendChild(webcam.canvas);
        labelContainer = document.getElementById("label-container");

```

```
    for (let i = 0; i < maxPredictions; i++) {
      labelContainer.appendChild(document.createElement("div"));
    }
  }
  async function loop() {
    webcam.update();
    await predict();
    window.requestAnimationFrame(loop);
  }
  async function predict() {
    const prediction = await model.predict(webcam.canvas);
    for (let i = 0; i < maxPredictions; i++) {
      const classPrediction =
        prediction[i].className + ": " + prediction[i].probability.toFixed(2);
      labelContainer.childNodes[i].innerHTML = classPrediction;
    }
  }
</script>
</body>
</html>
```

ΠΑΡΑΡΤΗΜΑ 4: ΠΡΟΓΡΑΜΜΑ ΠΟΥ ΑΛΛΑΖΕΙ ΑΡΧΕΙΟ .H5 ΣΕ .JSON

```
import matplotlib.pyplot as plt
import keras
from keras import models
import tensorflow as tf
import tensorflowjs as tfjs
new_model=keras.models.load_model('/content/drive/MyDrive/Colab
Notebooks/newtrain/newtrain/tunedmodel99.h5')
tfjs.converters.save_keras_model(new_model,'/content/drive/MyDrive/Colab
Notebooks/newtrain/newtrain/tunedmodel99.js')
```

ΠΑΡΑΡΤΗΜΑ 5: ΣΥΝΑΡΤΗΣΗ ΠΡΟΓΡΑΜΜΑΤΟΣ ΠΟΥ ΦΟΡΤΩΝΕΙ ΤΟ ΕΚΠΑΙΔΕΥΜΕΝΟ ΜΟΝΤΕΛΟ ΜΟΡΦΗΣ .JSON

```
async function load(){  
  const model= await tf.loadLayersModel('tunedmodel99.js/model.json');  
  return model;  
};
```