ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ

ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ
ΚΑΙ ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ

ΤΟΜΕΑΣ ΣΥΣΤΗΜΑΤΩΝ ΜΕΤΑΔΟΣΗΣ ΠΛΗΡΟΦΟΡΙΑΣ ΚΑΙ ΤΕΧΝΟΛΟΓΙΑΣ
ΥΛΙΚΩΝ

Ανάπτυξη Μοντέλου Εκτίμησης της Περιεχόμενης Ποσότητας Υδατανθράκων στα Λαμβανόμενα Γεύματα από Φωτογραφικά Στιγμιότυπα

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

Σωτήριος Α. Κούκιος-Πανόπουλος

Επιβλέπουσα : Κωνσταντίνα Σ. Νικήτα

Καθηγήτρια Ε.Μ.Π.

Αθήνα, Φεβρουάριος 2018

Η σελίδα αυτή είναι σκόπιμα λευκή.

ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ

ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ
ΚΑΙ ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ

ΤΟΜΕΑΣ ΣΥΣΤΗΜΑΤΩΝ ΜΕΤΑΔΟΣΗΣ ΠΛΗΡΟΦΟΡΙΑΣ ΚΑΙ ΤΕΧΝΟΛΟΓΙΑΣ ΥΛΙΚΩΝ

Ανάπτυξη Μοντέλου Εκτίμησης της Περιεχόμενης Ποσότητας Υδατανθράκων στα Λαμβανόμενα Γεύματα από Φωτογραφικά Στιγμιότυπα

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

Σωτήριος Α. Κούκιος-Πανόπουλος

Επιβλέπουσα: Κωνσταντίνα Σ. Νικήτα

Καθηγήτρια Ε.Μ.Π.

Εγκρίθηκε από την τριμελή εξεταστική επιτροπή την 9/2/2018.

(Υπογραφή)	(Υπογραφή)	(Υπογραφή)
.....	.....	.....
Κωνσταντίνα Σ. Νικήτα Καθηγήτρια Ε.Μ.Π.	Ανδρέας-Γεώργιος Σταφυλοπάτης Καθηγητής Ε.Μ.Π.	Γιώργος Στάμου Αναπληρωτής Καθηγητής Ε.Μ.Π.

Αθήνα, Φεβρουάριος 2018

(Υπογραφή)

.....

Κούκιος - Πανόπουλος Σωτήριος

Διπλωματούχος Ηλεκτρολόγος Μηχανικός και Μηχανικός Υπολογιστών Ε.Μ.Π.

Περιληψη

Στην παρούσα διπλωματική εργασία διερευνήθηκε η χρήση τεχνικών βαθιάς μάθησης και πιο συγκεκριμένα Συνελικτικών Νευρωνικών Δικτύων (ΣΝΔ) για την αυτόματη αναγνώριση τροφών από φωτογραφικά τους στιγμιότυπα. Για την ανάπτυξη των μοντέλων ταξινόμησης εφαρμόστηκε η αρχιτεκτονική ResNet των 50 επιπέδων, η οποία περιλαμβάνει την επανάληψη 50 δομικών μπλοκ βασιζόμενα σε φίλτρα συνέλιξης. Για την εκπαίδευσή του ΣΝΔ εφαρμόστηκαν και συγκρίθηκαν ως προς τις απαιτήσεις τους σε υπολογιστική ισχύ δύο frameworks: (i) το MatConvNet, που βασίζεται στο περιβάλλον Matlab, και το (ii) Torch, που βασίζεται στην γλώσσα σεναρίων ανοιχτού κώδικα Lua. Για την αξιολόγηση της απόδοσης και της ακρίβειας του υπό μελέτη μοντέλου χρησιμοποιήθηκε η βιβλιογραφικά διαθέσιμη βάση εικόνων γευμάτων Food-101, η οποία αποτελείται από 101000 φωτογραφίες γευμάτων που ανήκουν σε 101 κατηγορίες. Για την εκπαίδευση του μοντέλου χρησιμοποιήθηκε το 75% των εικόνων και για την αξιολόγησή του το υπόλοιπο 25% . Επίσης πραγματοποιήθηκε μια αξιολόγηση του εκπαιδευμένου ΣΝΔ χρησιμοποιώντας δειγματοληπτικά εικόνες από σύνολα δεδομένων, διαφορετικά από αυτά που χρησιμοποιήθηκαν για την εκπαίδευση του ΣΝΔ.

Το υπό μελέτη μοντέλο πέτυχε ακρίβεια ταξινόμησης $85{,}82\%$ λαμβάνοντας υπόψη μόνο την μεγαλύτερη πιθανότητα που εξάγει το ΣΝΔ (top-1 accuracy), και $97{,}24\%$ λαμβάνοντας υπόψη τις 5 μεγαλύτερες πιθανότητες (top-5 accuracy).

Λεξεις Κλειδια

Σακχαρώδης Διαβήτης, Παχυσαρκία, Σύστημα Αναγνώρισης Τροφίμων, Εκτίμηση Διατροφικής Αξίας, Υδατάνθρακες, Νευρωνικά Δίκτυα, Συνελικτικά Νευρωνικά Δίκτυα, Συνέλιξη, ResNet, Food-101, MatConvNet, Torch, ImageNet, Finetune, Image Classification, Feature Visualization

Η σελίδα αυτή είναι σκόπιμα λευκή.

ABSTRACT

In the present diploma thesis, the use of deep learning and more specifically Convolutional Neural Networks (CNNs) has been investigated in order to automatically recognize meal contents from meal screenshots. The development of the classifier has been based on the architecture ResNet-50, which includes structural blocks based on convolutional filters, stacked together in order to form a sequence of 50 similar blocks. For its training two frameworks were considered and comparatively assessed: (i) MatConvNet, based on the MatLab environment, and (ii) Torch, based on the open-source scripting language Lua. For evaluating the classifier's performance and accuracy, the publicly available dataset named "Food-101" of food images has been used, which consists of 101000 images assigned to 101 categories. 75% of the images have been used for training purposes and the rest 25% for validation. Furthermore, the classifier has been applied and evaluated on samples images from external food image datasets different than those used for its training.

The developed model achieved classification accuracy of 85.85%, taking into consideration only the first 'guess' of the CNN (top–1 accuracy), and an accuracy of 97.24% taking into consideration the first 5 guesses (top-5 accuracy).

KEYWORDS

Diabetes Mellitus, Obesity, Food Recognition System, Dietary Monitoring, Carbohydrate Estimation, Neural Networks, Convolutional Neural Networks, Convolution, ResNet, Food-101, MatConvNet, Torch, ImageNet, Finetune, Image Classification, Feature Visualization

Η σελίδα αυτή είναι σκόπιμα λευκή.

ΕΥΧΑΡΙΣΤΙΕΣ

Σε αυτό το σημείο, θα ήθελα να ευχαριστήσω θερμά όλους όσους συνέβαλαν στην επιτυχή εκπόνηση της παρούσας διπλωματικής εργασίας, ιδιαίτερα την Καθηγήτρια Κωνσταντίνα Νικήτα και την Διδάκτορα Κωνσταντία Ζαρκογιάννη, καθώς και όλα τα μέλη του Εργαστηρίου Βιοϊατρικών Προσομοιώσεων και Απεικονιστικής Τεχνολογίας (BIOSIM).

Επίσης θα ήθελα να ευχαριστήσω την οικογένειά μου και τους φίλους μου για την πολύτιμη υπομονή και υποστήριξή τους μέχρι αυτό το σημείο.

Η σελίδα αυτή είναι σκόπιμα λευκή.

ΠΕΡΙΕΧΟΜΕΝΑ

ΕΥΡΕΤΗΡΙΟ ΕΙΚΟΝΩΝ

1 ΕΙΣΑΓΩΓΗ

Η ιδέα που παρουσιάζεται στην παρούσα διπλωματική εργασία οδηγείται από τις αυξανόμενες ανησυχίες που σχετίζονται με την παχυσαρκία ή και με το βαθμό στον οποίο το άτομο χαρακτηρίζεται υπέρβαρο. Επιπροσθέτως, η ανάμειξη νέων τεχνολογιών όπως έξυπνα τηλέφωνα (smartphones) και tablets στον τομέα της υγείας, μας παροτρύνουν να βρούμε μια βοηθητική λύση που να συνενώνει την τεχνολογία με την θεραπεία προβλημάτων ή διαταραχών υγείας όπως είναι η παχυσαρκία ή ο Σακχαρώδης Διαβήτης (ΣΔ). Στην παρούσα διπλωματική εργασία, παρουσιάζεται ένα μοντέλο αυτόματης αναγνώρισης ειδών τροφίμων από φωτογραφικά στιγμιότυπα, με στόχο την εξατομικευμένη παρακολούθηση των διατροφικών συνηθειών του χρήστη.

1.1 ΠΑΧΥΣΑΡΚΙΑ

Πρόσφατα, η εξάπλωση της παχυσαρκίας και του ΣΔ σε παγκόσμιο επίπεδο έχει φτάσει αξιοσημείωτα μεγέθη και ειδικά η παχυσαρκία θεωρείται ένα από τα μείζοντα θέματα υγείας. Σύμφωνα με τον Παγκόσμιο Οργανισμό Υγείας (World Health Organisation - WHO), το 2017, το παγκόσμιο ποσοστό των παχύσαρκων ενηλίκων άγγιζε το 13%, αναλυόμενο σε 11% του αντρικού και 15% του γυναικείου πληθυσμού. [1]. Επιπλέον, είναι εξαιρετικά ανησυχητική η αύξηση 10 φορές του ποσοστού αυτού σε ενηλίκους και παιδιά μέσα σε 4 δεκαετίες, σύμφωνα με μια έρευνα του Παγκόσμιου Οργανισμού Υγείας και του Imperial College [2]. Το 2013, η American Medical Association (AMA) επίσημα χαρακτήρισε την παχυσαρκία ως ασθένεια που χρήζει ιατρικής θεραπείας και η οποία έχει επικίνδυνες ιατρικές συνέπειες. [3]

Η παχυσαρκία ορίζεται ως μια ιατρική κατάσταση κατά την οποία προκαλείται αντικανονική συσσώρευση λίπους στο ανθρώπινο σώμα. Έτσι, η παχυσαρκία και η κατάσταση στην οποία το άτομο χαρακτηρίζεται υπέρβαρο, είναι συνδεδεμένη στενά με χρόνιες παθήσεις όπως ο ΣΔ Τύπου 2, άπνοια ύπνου, υψηλή χοληστερίνη, ισχαιμικά επεισόδια, αυξημένο κίνδυνο για στεφανιαία καρδιακή νόσο, και καρκίνο νεφρών, ουροδόχου κύστης, μαστού, και παχέος εντέρου. Ισχυρά εμπειρικά στοιχεία δείχνουν πως η παχυσαρκία προκαλείται από αυξημένη πρόσληψη τροφών υψηλών σε θερμίδες που περιέχουν πολλά σάκχαρα, λίπη και αλάτι ενώ ταυτόχρονα περιέχουν μικρές ποσότητες βιταμινών, μετάλλων και άλλων μικροθρεπτικών συστατικών. Η θεραπεία της παχυσαρκίας έχει υπάρξει ως θέμα πολλών πρόσφατων ερευνών, και τα αποτελέσματα δείχνουν πως η έλλειψη ισορροπίας στην ενέργεια που καταναλώνεται με την ενέργεια που λαμβάνεται είναι ο κύριος λόγος για το αυξανόμενο ποσοστό της παχυσαρκίας. Υπάρχουν πολλές τεχνικές για την ποσοτικοποίηση και την κατηγοριοποίηση του ποσοστού λίπους στο ανθρώπινο σώμα, όπως ο Δείκτης Μάζας Σώματος (ΔΜΣ – Body Mass Index - BMI), η περιφέρεια μέσης, η αναλογία μέσης προς γοφό (waist-to-hip ratio), και το πάχος του δέρματος. Ακολουθούν επεξηγήσεις για κάθε μία από τις μεθόδους.

1.1.1 Δείκτης Μάζας Σώματος (ΔΜΣ)

Ο Δείκτης Μάζας Σώματος είναι το συνιστώμενο εργαλείο από τον Παγκόσμιο Οργανισμό Υγείας (ΠΟΥ) για μέτρηση του συνολικού σωματικού λίπους. Η τεχνική αυτή εξαρτάται από δύο τιμές, οι οποίες είναι το ύψος και το βάρος του ατόμου. Τα αποτελέσματα της μέτρησης θα είναι σε kg/m^2. Το επίπεδο της παχυσαρκίας εξαρτάται από το αποτέλεσμα της πράξης $\Delta M\Sigma = \frac{Βάρος}{Υψος^2}$, και επίσης εξαρτάται πάρα πολύ από το φύλο, την ηλικία και το σωματότυπο του ατόμου. Άτομα που αθλούνται ή έχουν γενικά αρκετούς μυς έχουν μεγαλύτερο ΔΜΣ χωρίς να έχουν περισσότερο λίπος. Άτομα τα οποία λόγω ηλικίας ή παθήσεων έχουν χάσει μυϊκή μάζα θα έχουν μικρότερο ΔΜΣ χωρίς αυτό να σημαίνει πως έχουν λιγότερο λίπος.

Παγκοσμίως έχει γίνει αποδεκτή η εξής κατηγοριοποίηση:

- Ποσοστό λίπους μικρότερο από 18,5 δείχνει ότι το άτομο είναι **ελλιποβαρές**.
- Ποσοστό λίπους μεταξύ 18,5 και 24,9 δείχνει ότι το άτομο έχει **φυσιολογικό βάρος**.
- Ποσοστό λίπους μεταξύ 25 και 29,9 δείχνει ότι το άτομο είναι **υπέρβαρο**.
- Ποσοστό λίπους 30 και μεγαλύτερο δείχνει ότι το άτομο πάσχει από **παχυσαρκία**. Εδώ υπάρχει και μια επιπλέον κατηγοριοποίηση για τη σοβαρότητα της παχυσαρκίας:
    - Μεταξύ 30 και 34,9 θεωρείται Παχυσαρκία I
    - Μεταξύ 35 και 39,9 θεωρείται Παχυσαρκία II
    - Άνω του 40 θεωρείται Παχυσαρκία III

1.1.2 Περιφέρεια Μέσης και Αναλογία Μέσης – Γοφού

Η περιφέρεια μέσης και η αναλογία μέσης – γοφού αποτελούν σημαντικές μεθόδους για τη μέτρηση του ποσοστού λίπους στο ανθρώπινο σώμα. Η τεχνική της περιφέρειας μέσης έχει επιλεγεί ως καλύτερη μέθοδος από τον Δείκτη Μάζας Σώματος για λίπος στην κοιλιακή χώρα [4]. Βασίζεται στη χρήση μιας ταινίας μεζούρας τοποθετημένη σε κατάλληλη θέση στη μέση.

Η αναλογία μέσης-γοφού χρησιμοποιείται επίσης για τη μέτρηση του κοιλιακού λίπους. Υπολογίζεται μετρώντας τη περιφέρεια μέσης και του γοφού και διαιρώντας τη μέτρηση της μέσης με την μέτρηση του γοφού.

1.1.3 Πάχος Δέρματος

Σε αυτή τη τεχνική, ειδικοί χρησιμοποιούν ένα διαβήτη σε διάφορες περιοχές του σώματος για να προσμετρήσουν το πάχος του δέρματος και του συσσωρευμένου λίπους [5]. Έπειτα, υπολογίζεται το ποσοστό του σωματικού λίπους σύμφωνα με τις μετρήσεις αυτές.

1.2 ΣΑΚΧΑΡΩΔΗΣ ΔΙΑΒΗΤΗΣ

Μια ακόμα ενδιαφέρουσα διαταραχή που αξίζει να λάβουμε υπόψη στη μελέτη μας είναι ο ΣΔ διότι η εμφάνιση, διαχείριση και εξέλιξη του συνδέονται στενά με τις διατροφικές συνήθειες του ατόμου που πάσχει από τη νόσο. Ο ΣΔ είναι μια μεταβολική ασθένεια η οποία χαρακτηρίζεται από αύξηση της συγκέντρωσης του σακχάρου στο αίμα (υπεργλυκαιμία) και από διαταραχή του μεταβολισμού της γλυκόζης, των λιπιδίων και των πρωτεϊνών, είτε ως αποτέλεσμα ελαττωμένης έκκρισης ινσουλίνης είτε λόγω ελάττωσης της ευαισθησίας των κυττάρων του σώματος στην ινσουλίνη. Η ινσουλίνη είναι μια ορμόνη που εκκρίνεται από το πάγκρεας και είναι απαραίτητη για τη μεταφορά της γλυκόζης που λαμβάνεται από τις τροφές, μέσα στα κύτταρα. Όταν το πάγκρεας δεν παράγει αρκετή ινσουλίνη ή η ινσουλίνη που παράγει δεν δρα σωστά, τότε η γλυκόζη που λαμβάνεται από τις τροφές δεν εισέρχεται στα κύτταρα ώστε να έχουν την απαραίτητη ενέργεια για τη λειτουργία τους και παραμένει στο αίμα με αποτέλεσμα την αύξηση των επιπέδων της και συνεπώς την εκδήλωση της νόσου.

Ο στόχος θεραπείας στα άτομα με ΣΔ είναι η διατήρηση των επιπέδων σακχάρου του αίματος όσο δυνατόν πλησιέστερα στο φυσιολογικό, αποφεύγοντας τις οξείες επιπλοκές της νόσου, όπως την υπογλυκαιμία ή την κετοξέωση, καθώς και τις χρόνιες επιπλοκές, όπως τη μικροαγγειοπάθεια (νεφροπάθεια, αμφιβλήστροειδοπάθεια, νευροπάθεια) και τη μακροαγγειοπάθεια (στεφανιαία νόσος, εγκεφαλικά επεισόδια και περιφερειακή αρτηριακή νόσος), έχοντας όμως και καλή ποιότητα ζωής.

Στον ΣΔ Τύπου 1, ο οποίος χαρακτηρίζεται από πλήρη ή σχεδόν πλήρη έλλειψη ενδογενούς ινσουλίνης, η ινσουλινοθεραπεία κρίνεται απαραίτητη όχι μόνο για τη ρύθμιση του σακχάρου αλλά και για την ίδια την επιβίωση του ατόμου. Όμως, προκειμένου να καθοριστεί η βέλτιστη δόση ινσουλίνης, είναι απαραίτητη η γνώση της ποσότητας καθώς και η σύνθεση της λαμβανόμενης τροφής, και ιδιαίτερα η περιεκτικότητα σε υδατάνθρακες. Η κάθε τροφή έχει διαφορετικό ποσοστό υδατανθράκων, επομένως ο οργανισμός θα αποκριθεί διαφορετικά, όσο αφορά την αύξηση της γλυκόζης στο αίμα, με τη λήψη διαφορετικής τροφής. Ο υπολογισμός αυτός όμως από τα άτομα με ΣΔ δεν είναι πάντα μια εύκολη διαδικασία.

1.2.1 Μέτρηση υδατανθράκων σε άτομα με ΣΔ

Η λήψη τροφής είναι ένα στοιχείο το οποίο δεν μπορεί να αγνοηθεί καθώς έχει καθοριστικό ρόλο στη διακύμανση των επιπέδων της γλυκόζης στο αίμα. Η τροφή αποτελείται από υδατάνθρακες, πρωτεΐνες και λίπος. Η διάσπαση της τροφής ξεκινάει στο στόμα και ολοκληρώνεται στο στομάχι. Τα επιμέρους στοιχεία απορροφούνται στο έντερο, από όπου περνάνε στο αίμα και είναι διαθέσιμα για τις διάφορες λειτουργίες και ανάγκες του οργανισμού. Συγκεκριμένα, οι υδατάνθρακες χρησιμοποιούνται από τα κύτταρα ως βασική πηγή ενέργειας. Προκειμένου να αξιοποιηθεί η γλυκόζη του αίματος είναι απαραίτητη η ύπαρξη της ινσουλίνης η οποία επιτρέπει την εισαγωγή της γλυκόζης στα κύτταρα. Η συσσώρευση γλυκόζης στο αίμα μετά από τη λήψη γευμάτων εξαρτάται ισχυρά από την περιεχόμενη ποσότητα υδατανθράκων για αυτό το λόγο η εκτίμησή της είναι σημαντική για τον προσδιορισμό κατάλληλης δόσης ινσουλίνης

Συγκεκριμένα, η μέτρηση της περιεχόμενης ποσότητας υδατανθράκων στα γεύματα είναι πολύ σημαντική για τα άτομα με ΣΔ για δύο λόγους [6]: Πρώτον, διευκολύνει το σχεδιασμό των γευμάτων και το προγραμματισμό της φυσικής δραστηριότητας. Όσον αφορά το σχεδιασμό των γευμάτων, τα άτομα με ΣΔ μπορούν να καθορίζουν τα γεύματά τους με τέτοιο τρόπο ώστε να διατηρούν τη γλυκόζη του αίματός τους σε φυσιολογικά επίπεδα. Επίσης, μπορούν να διευρύνουν τις διατροφικές του επιλογές που καλό είναι να αποφεύγονται λόγω της υψηλής τους περιεκτικότητας σε υδατάνθρακες. Δεύτερον, η μέτρηση των υδατανθράκων επιτρέπει την προσαρμογή των προγευματικών δόσεων ινσουλίνης στην ποσότητα των υδατανθράκων που πρόκειται να καταναλωθούν. Φυσιολογικά, η έκκριση της ινσουλίνης έπειτα από ένα γεύμα εξαρτάται από την ποσότητα των υδατανθράκων και γενικότερα τη σύνθεση της τροφής. Επομένως, η ποσότητα της ινσουλίνης που λαμβάνεται πριν από το γεύμα δεν μπορεί να είναι ίδια για όλα τα γεύματα αλλά να προσαρμόζεται κατάλληλα ανάλογα με το είδος και την ποσότητα της τροφής.

Ο υπολογισμός των υδατανθράκων δεν είναι μια εύκολη διαδικασία. Πολλές φορές συστήνεται από τον θεράποντα ιατρό το ζύγισμα των τροφών με μια ζυγαριά μαγειρικής. Η διαδικασία αυτή όμως είναι περιοριστική ως προς το χρόνο και το τόπο λήψης της τροφής, και συχνά επιλέγεται η εξαγωγή μιας εκτίμησης της ποσότητας της τροφής. Για την αντιστοίχιση της ποσότητας της τροφής με περιεχόμενους υδατάνθρακες υπάρχουν λίστες οι οποίες χωρίζουν τις τροφές σε κατηγορίες και δίνουν για διάφορες τροφές την ποσότητα που αντιστοιχεί σε ένα γεύμα, καθώς και την περιεκτικότητα σε υδατάνθρακες, πρωτεΐνες και λίπος για την ποσότητα αυτή.

Ωστόσο, έρευνες έχουν δείξει πως σε πολλές περιπτώσεις οι αποκλίσεις από την πραγματική ποσότητα τροφής είναι σημαντικές και ικανές να οδηγήσουν τις τιμές γλυκόζης του αίματος εκτός των φυσιολογικών ορίων. Πιο συγκεκριμένα, έρευνα σε ασθενείς που υποβάλλονται σε ινσουλινοθεραπεία έδειξε πως ανακρίβειες της τάξης των 20 γραμματίων στην εκτίμηση των υδατανθράκων έχουν αρνητική επίδραση στο μεταγευματικό προφίλ γλυκόζης [7]. Αναλυτικότερα, η έρευνα έδειξε πως αν μια δόση ινσουλίνης που έχει υπολογιστεί για επακόλουθο γεύμα που περιέχει 60 g υδατανθράκων ακολουθηθεί από γεύμα πραγματικής περιεκτικότητας 40g ή 80g, τότε οδηγεί σε μεταγευματική υπογλυκαιμία ή υπεργλυκαιμία αντίστοιχα. Επομένως, σύμφωνα με τα αποτελέσματα της έρευνας αυτής, η εκτίμηση στην περιεκτικότητα των υδατανθράκων πρέπει να έχουν μια μέγιστη απόκλιση της τάξης των 10 γραμμαρίων υδατανθράκων από την πραγματική τιμή.

Σύμφωνα με μια έρευνα που παρουσιάστηκε στην δημοσίευση [8], αξιολογήθηκε η ακρίβεια στον υπολογισμό των υδατανθράκων από εφήβους. Η έρευνα διεξήχθη σε 48 εφήβους ηλικίας 12-18 ετών με ΣΔ Τύπου 1, οι οποίοι λάμβαναν εξωγενή ινσουλίνη. Από τους συμμετέχοντες ζητήθηκε να εκτιμήσουν την περιεκτικότητα σε υδατάνθρακες σε 32 φαγητά που αποτελούν συχνή επιλογή στην συγκεκριμένη ηλικιακή κατηγορία. Η έρευνα έδειξε πως μόνο το 23% των εφήβων που συμμετείχαν στην έρευνα αυτή υπολόγισε με ακρίβεια την ποσότητα των υδατανθράκων. Στόχος ήταν ο υπολογισμός σε ένα εύρος απόκλισης 10g υδατανθράκων. Επίσης, η έρευνα έδειξε πως οι έφηβοι που υπολόγισαν με ακρίβεια την ποσότητα των υδατανθράκων είχαν λιγότερες διακυμάνσεις γλυκόζης αίματος και χαμηλότερες τιμές.

Σε μια μεταγενέστερη έρευνα [9], διερευνήθηκε το κατά πόσο τα παιδιά ή οι έφηβοι και τα άτομα που τα επιβλέπουν υπολογίζουν με ακρίβεια την περιεκτικότητα των υδατανθράκων στα γεύματα. Στην έρευνα συμμετείχαν 102 παιδιά και έφηβοι, ηλικίας 8-18 ετών, με ΣΔ Τύπου 1 τα οποία λαμβάνουν εξωγενή ινσουλίνη, καθώς και τα άτομα που ήταν υπεύθυνα για την φροντίδα τους. Σύμφωνα με τα – πιο αισιόδοξα – αποτελέσματα της έρευνας, το 73% όλων των εκτιμήσεων είχε απόκλιση 10-15 γραμμαρίων από την πραγματική τιμή ενώ οι υπόλοιπες είχαν μεγαλύτερες αποκλίσεις.

1.3 Στοχος

Από τα παραπάνω, είναι ξεκάθαρο πως είναι πολύ σημαντικό για άτομα που έχουν πρόβλημα με το βάρος τους για λόγους υγείας ή θέλουν να διατηρήσουν το βάρος τους σε υγιεινά επίπεδα, πρέπει να προσμετράται η καθημερινή πρόσληψη τροφής, για την ισορροπία της εισερχόμενης ενέργειας στον οργανισμό. Επίσης τα άτομα με ΣΔ είναι αναγκαίο να έχουν μια ακριβή μέτρηση των υδατανθράκων που περιέχονται στα γεύματά τους.

Στόχος της παρούσας διπλωματικής εργασίας είναι η προσέγγιση του προβλήματος αυτού με ένα μοντέλο εκτίμησης της διατροφικής αξίας γευμάτων μέσα από φωτογραφικά στιγμιότυπα των λαμβανόμενων γευμάτων. Πιο συγκεκριμένα, μελετάται το στάδιο της ταξινόμησης των επιμέρους τροφίμων του γεύματος της φωτογραφίας, με χρήση μεθόδων μηχανικής μάθησης. Προς αυτή την κατεύθυνση διερευνήθηκε η χρήση Συνελικτικών Νευρωνικών Δικτύων.

2 Βιβλιογραφική Επισκοπηση

2.1 Μεθοδοι Μετρησης Θρεπτικων Στοιχειων

Υπάρχουσες προσεγγίσεις για μέτρηση θρεπτικών στοιχείων σε γεύματα μπορούν να χωριστούν σε 4 γενικές κατηγορίες. Οι κατηγορίες αυτές είναι: (i) παραδοσιακές κλινικές μέθοδοι (Traditional Clinical Methods), (ii) έξυπνα περιβάλλοντα (Smart Environments), (iii) ειδικά φορητά συστήματα (Dedicated Portable Systems), και (iv) συστήματα βασισμένα σε έξυπνα τηλέφωνα (Smartphone Based Systems). Κάθε μία από αυτές τις κατηγορίες, αναλύονται ακολούθως.

2.1.1 Παραδοσιακές Κλινικές Μέθοδοι

Η κύρια κλινική μέθοδος είναι η αναφορά από τον ασθενή όλων των φαγητών που καταναλώθηκαν το τελευταίο 24ωρο, και ονομάζεται 24ωρη διαιτητική ανάκληση (24HR dietary recall). Η ανάκληση προετοιμάζεται συνήθως πρόσωπο με πρόσωπο ή μέσω τηλεφώνου, και απαιτούνται συγκεκριμένες διερευνήσεις από το πρόσωπο που θα συντάξει την αναφορά, ώστε να βοηθήσει τον ασθενή να θυμηθεί όλα τα φαγητά που κατανάλωσε μέσα στη μέρα. Σε αυτή τη μέθοδο, αναλύονται οι ημερήσιες αναφορές του ασθενή ώστε να βρεθεί ένα καλύτερο πρόγραμμα για τις επόμενες μέρες [10]. Παρά το γεγονός ότι η μέθοδος χρησιμοποιείται κυρίως για διαιτητικούς λόγους, είναι πολύ σημαντική για άτομα με ΣΔ και χρησιμοποιείται πολύ συχνά, ειδικά στους πρώτους μήνες θεραπείας. Αν και πολύ χρήσιμη, έχει ένα μεγάλο μειονέκτημα, σχετιζόμενο με την ελλιπή αναφορά. Για παράδειγμα, επισημάνθηκε πως χαρακτηριστικά όπως η παχυσαρκία, το φύλο, η εκπαίδευση, κατάσταση υγείας, ηλικία και εθνικότητα δεν αναφέρονται επαρκώς [11]. Επίσης, έχει παρατηρηθεί πως σημαντικές πληροφορίες, συμπεριλαμβανόμενου του μεγέθους μερίδων φαγητού, έχουν ελλιπή αναφορά [12], ενώ δεν αναφέρεται επαρκώς η πρόσληψη τροφής από τους ασθενείς [13] Από τις ίδιες έρευνες έχει παρατηρηθεί πως οι μερίδες έχουν αυξηθεί σημαντικά τα τελευταία 20 με 30 χρόνια και αυτό πιθανόν να συνεισφέρει στο φαινόμενο αυτό. Συνεπώς, δημιουργείται η ανάγκη για μεθόδους ακριβέστερης μέτρησης διατροφικών πληροφοριών.

2.1.2 Έξυπνα Περιβάλλοντα

Με σκοπό τη μείωση ή την εξάλειψη της υπο-αναφοράς που παρατηρείται στη μέθοδο 24ωρης διαιτητικής ανάκλησης, προτάθηκαν έξυπνα συστήματα μαγειρικής, όπως η έξυπνη κουζίνα του [14]. Σε αυτή τη προσέγγιση, σχεδιάστηκαν κουζίνες που λαμβάνουν υπόψη θερμίδες (Calorie-Aware Kitchens), οι οποίες συμπεριλαμβάνουν κάμερες που ενισχύουν την επίγνωση της επιλογής υγιεινών φαγητών και της συμπεριλαμβανόμενης ποσότητας θερμίδων στο προετοιμασμένο γεύμα. Η κουζίνα περιλαμβάνει μια κάμερα οροφής που απαθανατίζει φωτογραφίες από την διαδικασία προετοιμασίας του γεύματος, ενώ αισθητήρες συνδεδεμένοι στον πάγκο και την εστία μετρούν όλα τα συστατικά και καλύπτουν τα περισσότερα μέρη μέσα στη κουζίνα. Αυτό έχει ως αποτέλεσμα άμεση ανάδραση στον χρήστη με μια πρόταση για τη κατάλληλη ποσότητα πρόσληψης θερμίδων. Τα κύρια μειονεκτήματα της προσέγγισης αυτής αποτελούν η περιορισμένη χρήση και η αδυναμία της «μεταφοράς» της κουζίνας αυτής εκτός σπιτιού.

Άλλα συστήματα έχουν επιπλέον προταθεί αντ' αυτού. Οι Nishimura και Kuroda πρότειναν ένα σύστημα αισθητήρων που μπορεί να φορεθεί, χρησιμοποιώντας ένα μικρόφωνο [15]. Επιπλέον, ερευνητικές προσπάθειες προσανατολίστηκαν στην δημιουργία μιας τραπεζαρίας με την ικανότητα να λαμβάνει υπόψη τη διατροφική αξία των γευμάτων προς κατανάλωση [16]. Εφαρμόστηκε ταυτοποίηση μέσω ραδιοσυχνοτήτων (radio frequency identification - RFID) ως επιφανειακός αισθητήρας για εκτίμηση του τύπου του φαγητού σε συνδυασμό με ενσωματωμένες ζυγαριές πάνω στο τραπέζι για μέτρηση του βάρους του φαγητού. Ωστόσο υπήρχαν πάλι πολλά μειονεκτήματα με τη συγκεκριμένη τεχνική, εκ των οποίων βασικότερα ήταν η δυσκολία χρήσης σε περισσότερες από μία τοποθεσίες και η πολυπλοκότητα της επισύναψης της RFID ετικέτας σε κάθε σερβιρισμένο γεύμα.

2.1.3 Ειδικά Φορητά Συστήματα

Αναγνωρίζοντας τα προβλήματα που δημιουργούνται από τις έξυπνες κουζίνες λόγω αδυναμίας μετακίνησης, προτάθηκαν μερικά φορητά συστήματα. Ένα σύστημα το οποίο αυτόματα διαβάζει το πλήθος των θερμίδων βασίστηκε στην χρήση βιοεμπέδησης για τη μέτρηση του επίπεδου γλυκόζης στα κύτταρα του χρήστη [17]. Ωστόσο υπάρχουν σοβαρές ανησυχίες για την εγκυρότητα μιας τέτοιας προσέγγισης καθώς το σύστημα δεν έχει αξιολογηθεί κατάλληλα σε ευρεία κλίμακα. Ένα επιπλέον προφανές μειονέκτημα του συστήματος αποτελεί το γεγονός ότι προσμετρά τις θερμίδες του φαγητού μόνο αφού ο χρήστης έχει καταναλώσει το φαγητό.

Η εγγύς-υπέρυθρη ακτινοσκοπία (Near infrared spectroscopy – NIRS) έχει πρόσφατα προταθεί για τον καθορισμό της σύστασης των τροφών, με παραδείγματα ήδη διαθέσιμα στην αγορά [18], [19]. Τα συστήματα αυτά έχουν την δυνατότητα να ενημερώσουν τον χρήστη, για το ποσό των κορεσμένων λιπαρών ανά 100 γραμμάρια τροφής, το οποίο, σε συνδυασμό με μια υψηλής ποιότητας βάση δεδομένων, μπορεί να δώσει πληροφορίες για την διατροφική αξία της τροφής αυτής. Όμως, τα εργαλεία αυτά δεν μπορούν να μετρήσουν το βάρος και την ποσότητα του κάθε συστατικού. Έτσι, η μετάβαση από την τιμή «λιπαρά ανά 100g» σε πραγματικές θερμίδες ή υδατάνθρακες στο φαγητό δεν είναι τετριμμένη για τους χρήστες. Επίσης, διάφανα υγρά δεν μπορούν να μετρηθούν με αυτή τη τεχνική.

2.1.4 Συστήματα Βασισμένα σε Έξυπνα Τηλέφωνα

Τα τελευταία χρόνια, τεχνικές μέτρησης βασισμένη στην όραση (Vision Based Measurement - VBM [20]) επιτρέπουν την εκτίμηση των θερμίδων λαμβάνοντας υπόψη φωτογραφικά στιγμιότυπα γευμάτων χρησιμοποιώντας ένα smartphone. Η μέτρηση θερμίδων εφαρμόζοντας τεχνικές VBM αποτελεί δύσκολη περίπτωση της αναγνώρισης αντικειμένων. Η δυσκολία του προβλήματος εντοπίζεται στην ποικιλομορφία των φαγητών, καθώς οι μερίδες εμφανίζονται σε πολλά διαφορετικά μεγέθη και μορφές. Επίσης, οι μερίδες μπορεί να είναι από ένα φαγητό, ή από πολλά, αναμειγμένα μεταξύ τους, όπως για παράδειγμα οι σαλάτες, οι σούπες, κλπ. Συνεπώς, μερικές εικόνες μεικτών φαγητών είναι δύσκολο να μετρηθούν με ακρίβεια και με μεγάλο βαθμό επιτυχίας χρησιμοποιώντας τη μέθοδο αυτή. Ένα άλλο θέμα σχετίζεται με το χρόνο επεξεργασίας καθώς οι περισσότεροι αλγόριθμοι κατάτμησης και αναγνώρισης έχουν αυξημένες απαιτήσεις σε επεξεργαστή και μνήμη.

Όπως και τα υπόλοιπα συστήματα VBM, ο υπολογισμός θερμίδων φαγητών (και γενικά θρεπτικών στοιχείων όπως οι υδατάνθρακες που μελετούμε) χωρίζεται σε τέσσερα βασικά στάδια: Προεπεξεργασία, Ανάλυση εικόνας, Αναγνώριση Μετρητέου Μεγέθους, Μέτρηση.

Προ-επεξεργασία

Σε αυτό το στάδιο η αρχική εικόνα φαγητού προετοιμάζεται για τα επόμενα στάδια. Οποιοδήποτε θάμπωμα, θόρυβος, αλλοίωση κ.α. μπορούν να αφαιρεθούν σε αυτό το στάδιο. Επιπροσθέτως, εφαρμόζονται λειτουργίες όπως κανονικοποίηση (normalization), κατωφλίωση (thresholding), αποθορυβοποίηση (denoising) και χειρισμοί εικόνων όπως αλλαγή μεγέθους (resizing), περικοπή (cropping) κ.α. εφόσον χρειάζονται.

Κατάτμηση μερίδας φαγητού

Η κατάτμηση (segmentation) θα καθορίσει τα όρια των μερίδων φαγητού μέσα στο γεύμα. Η ιδανική έξοδος της λειτουργίας κατάτμησης είναι η ομαδοποίηση των εικονοστοιχείων της εικόνας τα οποία μοιράζονται ορισμένα οπτικά χαρακτηριστικά που έχουν νόημα αντιληπτικά στους ανθρώπινους παρατηρητές. Αυτό αποτελεί ένα δύσκολο πρόβλημα καθώς οι άνθρωποι χρησιμοποιούν μια πολύπλοκη και ταυτόχρονα υποσυνείδητη διαδικασία για να εκτελέσουν αυτό το έργο. Ποικίλες μέθοδοι κατάτμησης έχουν χρησιμοποιηθεί σε εφαρμογές κατάτμησης εικόνων φαγητών, όπως Κατάτμηση Χρώματος και Υφής (Color and Texture Segmentation), Συσταδοποίηση K-Μέσων (K-means Clustering), και Κατάτμηση βασισμένη σε Κοπή Γράφου (Graph Cut Based Segmentation).

Αναγνώριση Φαγητού

Στο βήμα αυτό, τα εξαγόμενα χαρακτηριστικά από κάθε μερίδα φαγητού κατηγοριοποιούνται για να αναγνωρίσουν τη μερίδα, εφαρμόζοντας διαφορετικές μεθόδους κατηγοριοποίησης. Μερικές από αυτές τις μεθόδους είναι οι Μηχανές Διανυσμάτων Υποστήριξης (Support Vector Machines – SVMs), Νευρωνικά Δίκτυα (Neural Network) και η Βαθιά Μάθηση (Deep Learning). Επιπρόσθετα, χρησιμοποιώντας το σύννεφο (cloud) για την επεξεργασία των μεθόδων αυτών, τόσο η ακρίβεια όσο και ο χρόνος απόκρισης βελτιώνονται.

Μέτρηση Θρεπτικών Στοιχείων

Αφού αναγνωριστεί το φαγητό, χρησιμοποιούνται υπάρχοντες πίνακες θρεπτικών στοιχείων για να υπολογίσουν τις θερμίδες ή τους υδατάνθρακες. Αυτοί οι πίνακες όμως απαιτούν την ποσότητα του φαγητού (σε γραμμάρια) για να δώσουν μια τελική απάντηση. Έτσι, δεν αρκεί μόνο να αναγνωρίσουμε το φαγητό, αλλά να μετρήσουμε και τη μάζα του. Εφαρμόζονται διάφορες προσεγγίσεις για την εκτίμηση της μάζας του φαγητού, όπως η η χρήση του αντίχειρα του χρήστη για βαθμονόμηση ως σημείο αναφοράς και τον υπολογισμό της επιφάνειας, του όγκου, και μετέπειτα της μάζας της μερίδας φαγητού, χρησιμοποιώντας υπάρχοντες πίνακες πυκνότητας φαγητών [21]. Άλλη προσέγγιση, χρησιμοποιεί την απόσταση μεταξύ της κάμερας του τηλεφώνου και της μερίδας φαγητού για να υπολογίσει την επιφάνεια της μερίδας φαγητού και κατά συνέπεια την μάζα του [22].

2.2 ΤΡΕΧΟΥΣΕΣ ΥΛΟΠΟΙΗΣΕΙΣ ΒΑΣΙΣΜΕΝΕΣ ΣΕ ΟΡΑΣΗ ΥΠΟΛΟΓΙΣΤΩΝ

Οι Yang και Wu δημιούργησαν μια μέθοδο για αναγνώριση πρόσληψης τροφής fast-food από βίντεο ανθρώπων [23]. Σε αυτή τη μέθοδο, ένα πλήθος από φωτογραφίες καταγεγραμμένες σε fast-food εστιατόριο συγκρίνονται με φωτογραφίες αποθηκευμένες σε μια βάση δεδομένων. Στο πλαίσιο αυτής της μελέτης, τοποθετήθηκαν κάμερες σε 3 διαφορετικές τοποθεσίες, ενώ το σύστημα εκπαιδεύτηκε σε 101 διαφορετικούς τύπους φαγητών. Σχετικά με αυτό το μοντέλο, οι Kim και Boutin στο [24] πρότειναν μια μέθοδο για αυτόματη εκτίμηση της ποσότητας μιας δεδομένης θρεπτικής αξίας ή των θερμίδων που περιέχονται σε εμπορικά φαγητά. Η μέθοδος εφαρμόζεται όταν κανένα μέρος από οποιοδήποτε συστατικό δεν αφαιρείται κατά την διαδικασία προετοιμασίας του φαγητού. Αρχικά, το σύστημα αυτόματα βρίσκει την ποσότητα κάθε συστατικού που χρησιμοποιείται για να προετοιμαστεί το φαγητό χρησιμοποιώντας τις πληροφορίες που παρέχονται στην ετικέτα, μαζί με θρεπτικές πληροφορίες για τουλάχιστον κάποια από τα συστατικά αυτά. Έπειτα, εφαρμόζεται ο αλγόριθμος Simplex για να υπολογίσει τις ποσότητες για το θρεπτικό περιεχόμενο.

Εφαρμόζοντας διαφορετικές μεθόδους για την επεξεργασία εικόνων και για τους αλγόριθμους κατάτμησης, τα συστήματα μέτρησης θερμίδων έχουν καταφέρει να αυξήσουν την ακρίβειά τους σε μεγάλες μερίδες φαγητών [25]. Παρόμοια προσέγγιση αφορά στην χρήση μιας κάρτας βαθμονόμησης που υπάρχει μέσα στην φωτογραφία ως μοτίβο μέτρησης, ώστε να υπολογίζεται το μέγεθος της μερίδας φαγητού [26]. Σε αυτή την περίπτωση, το φαγητό αναγνωρίζεται με μη αυτόματο τρόπο, με τη βοήθεια διατροφικών πληροφοριών που ανακτώνται από βάση δεδομένων. Έπειτα, οι θερμίδες υπολογίζονται για κάθε φωτογραφία, και στο τέλος, το πλήρες σύνολο των πληροφοριών αποθηκεύεται σε διαφορετικές βάσεις δεδομένων στο ερευνητικό κέντρο. Με βάση το γνωστό μέγεθος της κάρτας βαθμονόμησης, υπολογίζεται το μέγεθος των μερίδων και συνεπώς το πλήθος των θερμίδων.

Σε άλλες μελέτες, ο χρήστης λαμβάνει φωτογραφίες γευμάτων με το έξυπνο τηλέφωνό του και τις στέλνει σε ένα βήμα προ-επεξεργασίας [21]. Έπειτα, στο βήμα κατάτμησης, χρησιμοποιείται κατάτμηση χρώματος και υφής με σκοπό να εξαχθούν πληροφορίες σχετικά με τις μερίδες του φαγητού. Για κάθε μερίδα που εντοπίζεται, εξάγονται χαρακτηριστικά όπως το μέγεθος, το σχήμα, το χρώμα και η υφή. Τα εξαχθέντα χαρακτηριστικά έπειτα μεταβιβάζονται στο βήμα κατηγοριοποίησης όπου, χρησιμοποιώντας ένα SVM, αναγνωρίζεται ο τύπος του φαγητού. Εν τέλει, εκτιμώντας το εμβαδό της μερίδας του φαγητού και χρησιμοποιώντας διατροφικούς πίνακες, υπολογίζεται η θερμιδική αξία του γεύματος.

Επιπρόσθετα με τα παραπάνω, έχει προταθεί και η χρήση Νευρωνικών Δικτύων για την εκτίμηση θερμίδων από φωτογραφικά στιγμιότυπα φαγητού [27]. Σε αυτή τη προσέγγιση, καταγράφονται φωτογραφίες από πολλά πιάτα σε δίσκο πριν και μετά το γεύμα. Συγκεκριμένα, μια εικόνα ολόκληρου του δίσκου καταγράφεται στην αρχή. Έπειτα, αυτή η εικόνα μετατρέπεται σε δυαδική χρησιμοποιώντας τιμές κατωφλιού, και μια μικρή εικόνα του φαγητού θα εξαχθεί από την εικόνα του δίσκου. Χάρη στις προηγούμενες διαδικασίες, το σύστημα θα αναγνωρίσει όλες τις πληροφορίες που σχετίζονται με την εικόνα, όπως το

μήκος, το πλάτος και το σχήμα του φαγητού. Όλες οι προηγούμενες πληροφορίες μεταβιβάζονται στο Νευρωνικό Δίκτυο. Τα αποτελέσματα μεταφέρονται σε ένα πρόγραμμα προσομοίωσης που συγκρίνει τις πληροφορίες και αναλύει τα αποτελέσματα. Δυστυχώς, αυτή η μέθοδος δύσχρηστη, καθώς απαιτεί τη λήψη πολλών φωτογραφιών. Επιπλέον, η εικόνα πρέπει να αναλυθεί από υπολογιστή, κάτι που δεν είναι πρακτικό για καθημερινή χρήση.

Αναπτύχθηκε μέθοδος η οποία με αυτόματο τρόπο εντοπίζει και αναγνωρίζει τροφές σε μια ποικιλία εικόνων συνδυάζοντας δύο γενικές ιδέες [28]. Σύμφωνα με την πρώτη, ένα σύνολο από κατατμημένα αντικείμενα χωρίζονται σε κατηγορίες όμοιων αντικειμένων με βάση τα χαρακτηριστικά τους, όπως τραπεζομάντηλα και παρασκήνιο. Κατά τη δεύτερη, οι αυτόματα κατατμημένες περιοχές κατηγοριοποιούνται χρησιμοποιώντας ένα σύστημα κατηγοριοποίησης χαρακτηριστικών πολλών καναλιών (multichannel feature classification system) ως ένα κανονικοποιημένο κόψιμο γράφου (graph cut). Αυτή η μέθοδος επίσης χρησιμοποιεί SVM για την κατηγοριοποίηση. Η τελική απόφαση λαμβάνεται συνδυάζοντας αποφάσεις κατηγορίας από μεμονωμένα χαρακτηριστικά. Στο πλαίσιο άλλης μελέτης προτείνεται μια προσέγγιση ανάλυσης ενός τροφίμου σε επίπεδο εικονοστοιχείων, κατηγοριοποιώντας κάθε εικονοστοιχείο ως ένα συγκεκριμένο συστατικό και έπειτα χρησιμοποιώντας στατιστικές και τις χωρικές σχέσεις μεταξύ των ετικετών συστατικών των εικονοστοιχείων ως χαρακτηριστικά σε έναν κατηγοριοποιητή SVM [29]. Τα αποτελέσματα αναδεικνύουν την σημασία χρήσης ετικετών συστατικών εικονοστοιχείων για την αναγνώριση τροφίμων ως προς την ακρίβεια της κατηγοριοποίησης και μέτρησης, αλλά με βάρος του υψηλότερου υπολογιστικού κόστους.

Αξιοσημείωτο είναι και ένα φορητό σύστημα αναγνώρισης φαγητού στο οποίο ο χρήστης σχεδιάζει πλαίσια οριοθέτησης αλληλοεπιδρώντας με την οθόνη αρχικά, και έπειτα το σύστημα ξεκινά την αναγνώριση του τροφίμου που περιλαμβάνεται στο πλαίσιο [30], [31]. Για ακριβέστερη αναγνώριση, κάθε τρόφιμο περνάει από το στάδιο κατάτμησης χρησιμοποιώντας κόψιμο γράφου (Graph Cut), και εξάγεται ένας σάκος χαρακτηριστικών (bag of features) με χαρακτηριστικά από ιστόγραμμα χρώματος και SURF (Speeded Up Robust Features). Τελικά, χρησιμοποιείται ένα γραμμικό SVM για να κατηγοριοποιηθεί το τρόφιμο σε μία από τις 50 κατηγορίες φαγητού. Επιπροσθέτως, το σύστημα εκτιμά την κατεύθυνση των περιοχών φαγητού όπου αναμένεται να αποκτηθεί υψηλότερο σκορ εξόδου του SVM, και εμφανίζεται με ένα βέλος στην οθόνη ώστε ο χρήστης να μετακινήσει την κάμερα του κινητού προς αυτή τη περιοχή. Η διαδικασία αναγνώρισης εκτελείται επαναληπτικά, με περίοδο ενός δευτερολέπτου κατά προσέγγιση. Τα πειράματα δείχνουν μια ακρίβεια μέτρησης 81,55% για τις υψηλότερες 5 κατηγορίες φαγητού, όταν δίνονται ακριβή πλαίσια οριοθέτησης. Οι Wang κ.ά. [32] ανέπτυξαν επίσης ένα σύστημα διαιτητικής εκτίμησης το οποίο χρησιμοποιεί εικόνες τραβηγμένες από κινητό τηλέφωνο και εκμεταλλεύεται τις διατροφικές συνήθειες του χρήστη με αναδρομική Μπαγιεσιανή εκτίμηση (recursive Bayesian estimation), με αποτέλεσμα την δυνητική αύξηση της ακρίβειας κατηγοριοποίησης φαγητού κατά 11%.

Οι προαναφερθείσες μέθοδοι είναι υπολογιστικά απαιτητικές και απαιτούν υπολογιστικούς πόρους πέρα από αυτούς που μπορούν να χειριστούν τα τυπικά smartphones. Γι’ αυτό το

λόγο, αρκετά συστήματα χρησιμοποιούν υπολογιστικό νέφος (cloud computing) για να μειωθεί ο φόρτος της διαδικασίας κατάτμησης και κατηγοριοποίησης εικόνας. Το νέφος δεν επιτρέπει μόνο υψηλότερη ακρίβεια αλλά μπορεί επίσης να μειώσει το χρόνο επεξεργασίας. Ένα τέτοιο παράδειγμα αποτελεί η μελέτη των Low κ.ά. [33], όπου επέκτειναν το πλαίσιο GraphLab ώστε να υποστηρίζει δυναμικό και παράλληλο υπολογισμό γράφων στο νέφος. Το σύστημα υλοποιεί επεκτάσεις σε διοχετευμένα κλειδώματα και versioning δεδομένων για να αποφύγει τη συμφόρηση και την καθυστέρηση του δικτύου, και έχει αναπτυχθεί επιτυχώς σε ένα μεγάλο σύμπλεγμα (cluster) Amazon EC2.

Τα τελευταία χρόνια (περίπου από το 2014) τα Συνελικτικά Νευρωνικά Δίκτυα (Convolutional Neural Networks - CNN), χρησιμοποιούνται επίσης για την αναγνώριση γευμάτων, και μάλιστα με μεγαλύτερη ευκολία και ακρίβεια. Εξαιτίας της μεγάλης ποικιλίας στα είδη τροφών, η αναγνώριση εικόνων γευμάτων παρουσιάζει δυσκολίες. Ωστόσο, η Βαθεία Μάθηση έχει δείξει πως είναι μια πολύ ισχυρή τεχνική στην αναγνώριση εικόνων. Για παράδειγμα, εφαρμόζονται ΣΝΔ για τις διεργασίες του εντοπισμού και της αναγνώρισης φαγητού, μέσω βελτιστοποίησης παραμέτρων [34]. Αρχικά κατασκευάζεται ένα σύνολο δεδομένων από τα πιο συχνά τεμάχια φαγητού σε ένα δημόσια διαθέσιμο σύστημα καταχώρησης φαγητού, το οποίο εισέρχεται σε ένα ΣΝΔ για την μετέπειτα αναγνώριση εισερχόμενων αντικείμενων φαγητού με απώτερο στόχο τον προσδιορισμό των θρεπτικών συστατικών του. Φαίνεται πως ένα ΣΝΔ αποδίδει σημαντικά καλύτερα από τις παραδοσιακές μεθόδους βασισμένες σε Μηχανές Διανυσμάτων Υποστήριξης (Support Vector Machines – SVM). Ερευνητές συνέκριναν διαφορετικά οπτικά χαρακτηριστικά στο σύνολο δεδομένων Food101 και βρήκαν πως χρησιμοποιώντας το VGG19 για την εξαγωγή χαρακτηριστικών, είχαν πολύ μεγαλύτερη ακρίβεια (40.21% top-1 ακρίβεια) σε σχέση με το μοντέλο Bag-of-Words (BoW) με SIFT χαρακτηριστικά (23.96%) [35]. Συνέκριναν επίσης δύο αρχιτεκτονικές ΣΝΔ και συμπέραναν πως αρχιτεκτονικές με μεγαλύτερο βάθος είχαν μεγαλύτερη ακρίβεια σε σχέση με πιο ρηχές (αρχιτεκτονική Overfeat, 33.91%).

Ένα άλλο σύστημα που επίσης χρησιμοποιεί ΣΝΔ εφαρμόζει αρχιτεκτονική 6 επιπέδων για το νευρωνικό δίκτυο ώστε να επιτύχει κατηγοριοποίηση εικόνων τροφών [36]. Για κάθε τεμάχιο φαγητού, εξάγονται επικαλυπτόμενα τμήματα εικόνας και κατηγοριοποιούνται, ενώ επιλέγεται η κατηγορία με τη πλειοψηφία των ψήφων. Τα πειράματα έγιναν σε ένα χειροκίνητα επιλεγμένο και κατηγοριοποιημένο σύνολο δεδομένων 573 φαγητών σε 7 κατηγορίες, και επετεύχθη ακρίβεια 84.9% που δικαιολογεί την επιλογή των συνιστωσών αυτών του συστήματος.

Οι Ao και Ling [37] εξήγαγαν χαρακτηριστικά χρησιμοποιώντας ένα προεκπαιδευμένο ΣΝΔ GoogLeNet, το οποίο επανεκπαίδευσαν χρησιμοποιώντας εικόνες γευμάτων από το σύνολο δεδομένων Food-220, το οποίο αποτελεί ένα συνδυασμό των συνόλων δεδομένων Food-101 και UEC256. Εφάρμοσαν επίσης μια τεχνική για θερμή έναρξη, όπου χρησιμοποίησαν έναν αρνητικό κατηγοριοποιητή. Τους επέτρεψε να επιτύχουν καλύτερο αποτέλεσμα με λιγότερες επαναλήψεις εκπαίδευσης, και μάλιστα πετυχαίνοντας μέση ακρίβεια 91,91% (top 5) για 220 κατηγορίες φαγητού. Εφαρμόστηκε κατάτμηση Graph Cut στις εικόνες εισόδου, οι οποίες περιείχαν έναν τύπο φαγητού, και με ΣΝΔ επιτεύχθηκε 99% ακρίβεια για 30 κατηγορίες φαγητού πλαίσιο άλλης μελέτης [38].

Πρόσφατα, η Google ανέπτυξε και διέθεσε το σύστημα Im2Calories, το οποίο αναγνωρίζει το διατροφικό περιεχόμενο των γευμάτων από μία εικόνα φαγητού [39]. Στην απλούστερη μορφή του, συνδυάζει πληροφορίες τοποθεσίας για να περιορίσει το σύνολο αναζήτησης των γευμάτων σε γνωστά μενού από 23 δημοφιλή εστιατόρια, ενώ στη γενική μορφή του αναγνωρίζει την κατηγορία του φαγητού, αλλά κάνει και μια εκτίμηση του όγκου. Για τα δύο προβλήματα αυτά χρησιμοποιείται το μοντέλο GoogLeNet CNN [40], πετυχαίνοντας ακρίβεια αναγνώρισης στην γενική μορφή 79% σε ένα σύνολο δεδομένων 201 κατηγοριών φαγητού. Όλη η διαδικασία κατηγοριοποίησης γίνεται αποκλειστικά σε φορητή συσκευή.

3 ΥΠΟΒΑΘΡΟ ΚΑΙ ΕΝΝΟΙΕΣ

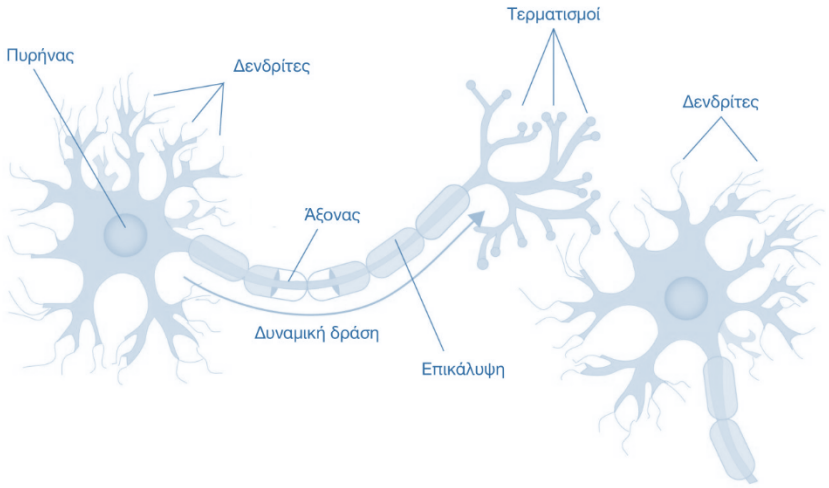
Στο κεφάλαιο αυτό, παρέχεται μια βασική περιγραφή αρχικά γενικά για τα Νευρωνικά Δίκτυα και έπειτα ειδικά για τα ΣΝΔ: πώς δουλεύει κάθε υπολογιστικό δομικό στοιχείο από τα οποία απαρτίζεται, και ποιος είναι ο αλγόριθμος που μας επιτρέπει να τα χρησιμοποιήσουμε στην κατηγοριοποίηση των εικόνων. Μετά από αυτό, ακολουθεί μια περιγραφή του αρχιτεκτονικού μοντέλου ResNet-50 [41] και μια σύντομη εισαγωγή στο MatConvNet, μια βιβλιοθήκη ανοιχτού κώδικα για το περιβάλλον αριθμητικής υπολογιστικής MatLab, αλλά και στο Torch, μια βιβλιοθήκη επίσης ανοιχτού κώδικα, βασισμένη στη γλώσσα σεναρίων Lua. Τόσο το MatLab μαζί με το MatConvNet όσο και η Lua μαζί με το Torch είναι κατάλληλα για εκπαίδευση ΣΝΔ, και παρέχουν ευελιξία στον κώδικα, μαζί με δυνατότητες αξιοποίησης GPU στους υπολογισμούς.

3.1 ΝΕΥΡΩΝΙΚΑ ΔΙΚΤΥΑ

3.1.1 Βιολογική Έμπνευση

Τα νευρωνικά δίκτυα αποτελούν μια οικογένεια υπολογιστικών αρχιτεκτονικών που αντλούν έμπνευση από τα βιολογικά νευρικά συστήματα. Ο ανθρώπινος εγκέφαλος περιέχει κατά προσέγγιση 86 δισεκατομμύρια νευρώνες, οι οποίοι συνδέονται μεταξύ τους με $10^{14} - 10^{15}$ συνάψεις.

Το νευρικό κύτταρο ή νευρώνας είναι το βασικό δομικό στοιχείο του εγκεφάλου τόσο στον άνθρωπο όσο και στα ζώα. Ο νευρώνας είναι ένα μεγάλο σε μέγεθος κύτταρο το οποίο, ανατομικά, αποτελείται από τα εξής τμήματα: το σώμα, τους δενδρίτες, τον άξονα και τις συνάψεις που συνδέουν τις διακλαδώσεις του άξονα με τους δενδρίτες άλλων νευρώνων δημιουργώντας έτσι ένα νευρωνικό δίκτυο. [42]

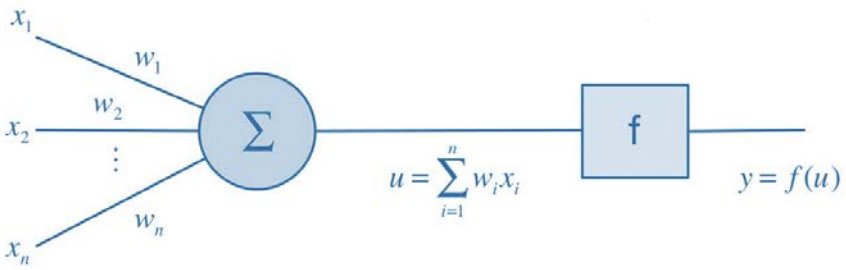


Εικόνα 1: Ένας ανθρώπινος νευρώνας

Οι δενδρίτες αποτελούν τις πύλες εισόδου του νευρώνα, ενώ ο άξονας είναι η μοναδική πύλη εξόδου του. Στέλνει σήματα προς άλλους νευρώνες υπό μορφή ηλεκτρικών παλμών

σταθερού πλάτους αλλά μεταβλητής συχνότητας. Οι συνάψεις, από την άλλη, είναι τα σημεία ένωσης μεταξύ τον άξονα ενός νευρώνα και των δενδριτών άλλων νευρώνων. Το ποσοστό της ηλεκτρικής δραστηριότητας που μεταδίδεται τελικά στον κάθε δεδνρίτη ονομάζεται συναπτικό βάρος. Οι συνάψεις επίσης χωρίζονται σε ενισχυτικές (excitatory) και ανασταλτικές (inhibitory), ανάλογα με το αν το φορτίο που εκλύεται από τη σύναψη ερεθίζει τον συνδεόμενο νευρώνα ή, αντίθετα, αν τον καταστέλλει εμποδίζοντάς τον από το να παράγει παλμούς. Στο σύνολό τους, οι δισεκατομμύρια αυτοί απλοί νευρώνες σχηματίζουν ένα εξαιρετικά πολύπλοκο διαδραστικό δίκτυο, το οποίο επιτρέπει σε εμάς ως ανθρώπινα όντα να βλέπουμε, ακούμε, κινούμαστε, επικοινωνούμε, θυμόμαστε, αναλύουμε, καταλαβαίνουμε, και ακόμα να ονειρευόμαστε [43].

Έχοντας ως έμπνευση το μοντέλο αυτό, οι Αμερικανοί επιστήμονες McCoulloch και Pitts [44] ήταν οι πρώτοι, το 1943, που περιέγραψαν ένα απλό μοντέλο δραστηριότητας του νευρώνα. Κατά το μοντέλο αυτό, η κατάσταση του νευρώνα περιγράφεται από έναν δυαδικό αριθμό y. Σε πιο εξελιγμένα μοντέλα η έξοδος λαμβάνει πραγματικές τιμές.



Εικόνα 2: Ένας τεχνητός νευρώνας

Οι συνάψεις περιγράφονται από τα συναπτικά βάρη w_i, που είναι πραγματικοί αριθμοί, θετικοί για τις ενισχυτικές συνάψεις, και αρνητικοί για τις ανασταλτικές. Τα βάρη w μπορούμε να τα δούμε σαν παραμέτρους που ορίζουν την αντίδραση του νευρώνα για ένα δεδομένο σύνολο εισόδων, και οι τιμές τους μπορούν να προσαρμοστούν με σκοπό να γίνει εκμάθηση μιας προσέγγισης του επιθυμητού σήματος εξόδου. Οι είσοδοι x_i του νευρώνα συνδυάζονται για να παράξουν το άθροισμα u του φορτίου που δέχεται ο νευρώνας:

$$u = \sum_{i=1}^{n} w_i x_i$$

Αν το άθροισμα u είναι μεγαλύτερο από ένα κατώφλι (threshold) θ τότε ο νευρώνας πυροβολεί, διαφορετικά παραμένει αδρανής. Κατά το μοντέλο των McColloch-Pitts, το αποτέλεσμα αυτό τροφοδοτείται σε έναν μη-γραμμικό μετασχηματιστή f, για να παράξει την έξοδο y του νευρώνα

$$y = f(u - \theta)$$

Η οποία συνάρτηση ενεργοποίησης f στην προκειμένη περίπτωση είναι η βηματική συνάρτηση

$$f(u) = \begin{cases} 0, & \alpha\nu\ u \leq 0 \\ 1, & \alpha\nu\ u > 0 \end{cases}$$

Το κατώφλι θ είναι ένας πραγματικός αριθμός (θετικός ή αρνητικός), όπως και τα συναπτικά βάρη. Με αυτή την έννοια, το κατώφλι θ μπορεί να θεωρηθεί πως είναι ένα επιπλέον συναπτικό βάρος, συνδεδεμένο με μια σταθερή είσοδο x_0 η οποία έχει πάντα την τιμή -1. Έτσι, θα μπορούσαμε να πούμε πως

$$u = \sum_{i=0}^{n} w_i x_i \,, \text{όπου } w_0 = \theta, x_0 = -1$$

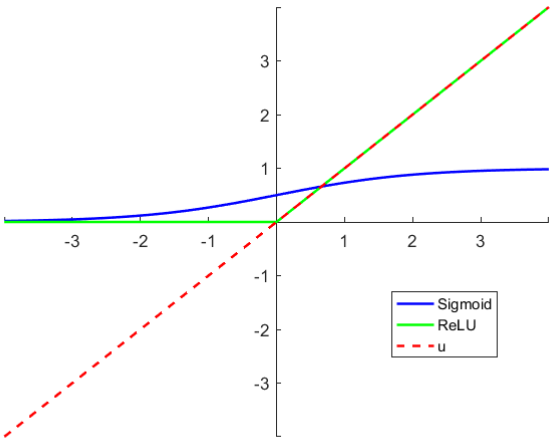
Υπάρχουν πολλές διαφορετικές μοντελοποιήσεις του νευρώνα, με τη πιο σημαντική διαφοροποίηση από το απλό μοντέλο McCulloch-Pitts να αποτελεί τη μορφή της μη γραμμικής συνάρτησης f.

Άλλες διαδεδομένες συναρτήσεις ενεργοποίησης αποτελούν οι:

- Βηματική -1/1 (ή προσήμου – step function -1/1 / sgn):
  - $f(u) = \begin{cases} -1, & \alpha\nu\ u \leq 0 \\ +1, & \alpha\nu\ u > 0 \end{cases}$
- Υπερβολική εφαπτομένη (hyperbolic tangent):
  - $f(u) = \tanh(u) = (1 - e^{-u})/(1 + e^{-u})$
- Σιγμοειδής (sigmoid):
  - $f(u) = 1/(1 + e^{-u})$
- Γραμμική Μονάδα Ανόρθωσης (Rectifier Linear Unit – ReLU):
  - $f(u) = \max\{0, u\}$

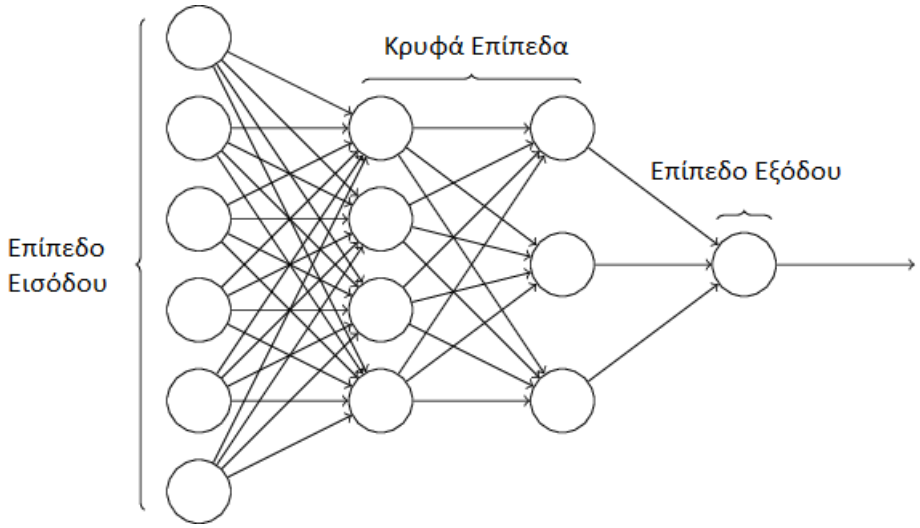
Από αυτές οι πιο διαδεδομένες πλέον στα Συνελικτικά Νευρωνικά Δίκτυα είναι η Σιγμοειδής και η ReLU, και είναι αυτές που υποστηρίζει το πακέτο εργαλείων MatConvNet που χρησιμοποιήθηκε, όπως θα δούμε και στη συνέχεια.

Παρατίθεται μια σύγκριση μεταξύ της γραμμικής συνάρτησης, της σιγμοειδούς και της ReLU:



Εικόνα 3: Σύγκριση μεταξύ των αποκρίσεων της γραμμικής, της σιγμοειδούς και της ReLU συνάρτησης

Ένα νευρωνικό δίκτυο σχηματίζεται διασυνδέοντας πολλούς τεχνητούς νευρώνες. Ο τρόπος με τον οποίο είναι δομημένοι οι νευρώνες ενός νευρωνικού δικτύου σχετίζεται στενά με τον αλγόριθμο μάθησης που χρησιμοποιείται για την εκπαίδευσή του δικτύου. Η πιο συνηθισμένη αρχιτεκτονική αποτελεί η διαρρύθμιση των νευρώνων σε ένα κατευθυνόμενο ακυκλικό γράφο για να σχηματίσουν ένα δίκτυο πρόσθιας τροφοδότησης. Οι νευρώνες έπειτα ομαδοποιούνται σε επίπεδα (ή στρώματα - layers), και επιτρέπονται συνδέσεις μόνο μεταξύ νευρώνων που ανήκουν σε γειτονικά επίπεδα. Σε ένα πλήρως συνδεδεμένο επίπεδο, κάθε έξοδος από το προηγούμενο επίπεδο συνδέεται με κάθε νευρώνα του τρέχοντος επιπέδου. Ένα νευρωνικό δίκτυο πρόσθιας τροφοδότησης που απαρτίζεται αποκλειστικά από πλήρως συνδεδεμένα επίπεδα ονομάζεται Perceptron Πολλών Στρωμάτων (Multilayer Perceptron – MLP).



Εικόνα 4: Multi Layer Perceptron με 3 επίπεδα και 1 έξοδο. Να σημειωθεί πως το επίπεδο εισόδου δεν θεωρείται επίπεδο της αρχιτεκτονικής.

3.1.2 Εκπαίδευση Νευρωνικών Δικτύων

Το αντικείμενο της μάθησης και της αυτοπροσαρμογής σε νέο περιβάλλον και νέες καταστάσεις αποτελεί μία από τις λειτουργίες του εγκεφάλου που οδηγούν στην νοημοσύνη, και είναι ίσως ένα από τα πιο σημαντικά χαρακτηριστικά του εγκεφάλου και γενικά των βιολογικών νευρωνικών δικτύων. Επειδή τα τεχνητά νευρωνικά δίκτυα είναι μοντέλα που μιμούνται την λειτουργία των αντίστοιχων βιολογικών νευρώνων όσο και τη δομή τους, έχουν αναπτυχθεί και μελετηθεί μαθηματικοί αλγόριθμοι που μιμούνται την αρχιτεκτονική και το πρότυπο των βιολογικών νευρωνικών δικτύων.

Οι παράμετροι σε ένα (τεχνητό) νευρωνικό δίκτυο δεν επιλέγονται χειροκίνητα – αλλά μαθαίνονται κατά τη διάρκεια της φάσης της εκπαίδευσης. Η πιο δημοφιλής προσέγγιση ονομάζεται επιβλεπόμενη μάθηση ή μάθηση με εκπαιδευτή. Μπορούμε να θεωρήσουμε ότι ο εκπαιδευτής έχει γνώση του περιβάλλοντος, η οποία αντιπροσωπεύεται από ένα σύνολο παραδειγμάτων εισόδου-εξόδου. Το περιβάλλον όμως είναι άγνωστο στο νευρωνικό δίκτυο. Κατά τη διάρκεια της εκπαίδευσης, όταν το νευρωνικό δίκτυο εκτίθεται σε ένα παράδειγμα εκπαίδευσης, ο εκπαιδευτής θα είναι σε θέση να παρέχει στο δίκτυο μια επιθυμητή έξοδο για το συγκεκριμένο παράδειγμα.

Οι παράμετροι του δικτύου λοιπόν προσαρμόζονται υπό τη συνδυασμένη επιρροή του διανύσματος (παραδείγματος) εκπαίδευσης και του σήματος σφάλματος. Το σήμα σφάλματος εκφράζει τη διαφορά μεταξύ της επιθυμητής και της πραγματικής απόκρισης του δικτύου. Αυτή η προσαρμογή εκτελείται με επαναληπτικό τρόπο, βήμα προς βήμα, με στόχο να φέρει το δίκτυο σε μια κατάσταση όπου θα προσομοιώνει τη συμπεριφορά του εκπαιδευτή. Ο στόχος μας είναι να ελαχιστοποιηθεί το σφάλμα στο σύνολο των προτύπων εκπαίδευσης, βελτιστοποιώντας τις παραμέτρους των βαρών – αυτό θα οδηγήσει στη καλύτερη προσομοίωση του εκπαιδευτή.

Η εκπαίδευση ξεκινά με αρχικοποίηση των βαρών του δικτύου σε μικρές, τυχαίες τιμές. Αυτό γίνεται διότι η αρχικοποίηση σε μια σταθερή τιμή (για παράδειγμα 0) σε όλους τους νευρώνες θα οδηγούσε στον υπολογισμό ακριβώς των ίδιων εξόδων, αποτρέποντας οποιαδήποτε μορφή μάθησης. Η Στοχαστική Κατάβαση Δυναμικού (Stochastic Gradient Descent) αποτελεί τη πιο δημοφιλή μέθοδο βελτιστοποίησης που χρησιμοποιείται για την εκπαίδευση των νευρωνικών δικτύων. Ο αλγόριθμος της κατάβασης δυναμικού υπολογίζει ένα διάνυσμα δυναμικού που περιγράφει την επιρροή κάθε βάρους στο συνολικό σφάλμα. Οι μερικές παράγωγοι μπορούν να υπολογιστούν αποδοτικά χρησιμοποιώντας τον αλγόριθμο back-propagation. Τα σφάλματα (που συμβολίζονται με το γράμμα δ) πηγάζουν από το τελευταίο στρώμα και προωθούνται προς τα πίσω, προς το πρώτο στρώμα.

Ένας πλήρης κύκλος χρήσης όλων των προτύπων ονομάζεται εποχή (epoch). Ανάλογο με το είδος και πλήθος των δεδομένων, και την χωρητικότητα του νευρωνικού δικτύου, μια πλήρης εκπαίδευση του δικτύου μπορεί να κυμαίνεται από μια έως και πολλές εκατοντάδες ή χιλιάδες εποχές. Η εκπαίδευση παίρνει επαναληπτικά από ένα πρότυπο εκπαίδευσης, υπολογίζει το τρέχον σφάλμα (φάση ανάκλησης – forward phase), υπολογίζει τα δυναμικά σε κάθε επίπεδο (φάση υπολογισμού δ – backward phase), και μεταβάλλει όλα τα βάρη κατά ένα μικρό ποσό προς την αντίθετη κατεύθυνση του σχετικού δυναμικού τους (φάση ενημέρωσης βαρών – update phase). Το μέγεθος των μεταβολών αυτών επηρεάζεται από την παράμετρο του βήματος εκπαίδευσης.

Μια εναλλακτική μέθοδος του αλγορίθμου αυτού, που ονομάζεται Κατάβαση Δυναμικού Δέσμης (Batch Gradient Descent), καθυστερεί την ενημέρωση των βαρών, και πρώτα υπολογίζει και παίρνει το μέσο όρο των δυναμικών μιας δέσμης από πρότυπα εκπαίδευσης. Αυτό επιτρέπει τον υπολογισμό να γίνει διανυσματικός και να εκτελεστεί πιο αποδοτικά σε περιβάλλοντα που υποστηρίζουν εντολές διανυσμάτων, συμπεριλαμβανομένων μονάδων επεξεργασίας γραφικών (Graphical Processing Unit – GPU), ψηφιακούς επεξεργαστές σημάτων (Digital Signal Processor – DSP) και τις περισσότερες από τις μονάδες κεντρικής επεξεργασίας (Central Processing Unit – CPU) [45].

Ενημερώνοντας επαναληπτικά τα βάρη, το νευρωνικό δίκτυο ιδανικά συγκλίνει προς μια λύση με ελάχιστο σφάλμα και ως εκ τούτου με μια καλή προσέγγιση της επιθυμητής εξόδου στο σύνολο εκπαίδευσης. Κάθε λίγες εποχές, η απόδοση του δικτύου επικυρώνεται με ένα ξεχωριστό σύνολο προτύπων επικύρωσης (validation set) τα οποία δεν είχαν χρησιμοποιηθεί κατά τη διάρκεια της εκπαίδευσης. Αποτελούν συνήθως ένα τμήμα των κατηγοριοποιημένων προτύπων εισόδου-εξόδου, που έχουν διαχωριστεί από την αρχή. Αποτελεί κοινή πρακτική να χρησιμοποιείται περίπου το $20\% - 25\%$ των προτύπων ως validation set.

Αν το σύνολο προτύπων εκπαίδευσης (training set) είναι αντιπροσωπευτικό των δεδομένων του «πραγματικού κόσμου», το δίκτυο θα αποδίδει καλές προβλέψεις για άγνωστα πρότυπα. Αν, ωστόσο, το training set είναι περιορισμένο σε μέγεθος και ποικιλομορφία, ή η χωρητικότητα του δικτύου είναι πολύ υψηλή, το νευρωνικό δίκτυο θα παρουσιάζει μια συμπεριφορά που μοιάζει με την «αποστήθιση» των προτύπων και να χάνει την ικανότητά του για γενίκευση. Το φαινόμενο αυτό της υπερ-μοντελοποίησης (overfitting) μπορεί να αντισταθμιστεί με μεγαλύτερο σύνολο εκπαίδευσης (πιθανότατα με τεχνικές αύξησης δεδομένων – data augmentation – όπως παραμορφώσεις, κατοπτρισμοί, περιστροφές) ή ακόμα και με αλλαγές στη δομή του δικτύου (όπως προσθήκη μεθόδων κανονικοποίησης).

3.2 ΣΥΝΕΛΙΚΤΙΚΑ ΔΙΚΤΥΑ

Τα Συνελικτικά Δίκτυα [46], επίσης γνωστά και ως συνελικτικά νευρωνικά δίκτυα (Convolutional Neural Networks – CNNs), αποτελούν μια ειδική κατηγορία Perceptron Πολλών Στρωμάτων (Multilayer Perceptron – MLP), τα οποία αποδεικνύονται ιδιαίτερα κατάλληλα για την ταξινόμηση προτύπων. Είναι ειδικά σχεδιασμένα ώστε να αναγνωρίζουν σχήματα σε γνωστή τοπολογία, που μοιάζει με πλέγμα, με υψηλό βαθμό μη-ευαισθησίας (ιδιότητα του αναλλοίωτου) στη μετατόπιση, την κλιμάκωση, την στρέβλωση και άλλες μορφές παραμόρφωσης. Παραδείγματα αποτελούν δεδομένα χρονοσειρών, τα οποία μπορούν να θεωρηθούν ως μονοδιάστατο πλέγμα παίρνοντας δείγματα ανά τακτά χρονικά διαστήματα, και δεδομένων εικόνας, που μπορούν να θεωρηθούν ως ένα δισδιάστατο πλέγμα από εικονοστοιχεία. Τα Συνελικτικά δίκτυα έχουν υπάρξει πολύ επιτυχημένα σε πρακτικές εφαρμογές. Το όνομά τους «Συνελικτικό Νευρωνικό Δίκτυο» δείχνει πως το δίκτυο εφαρμόζει την μαθηματική λειτουργία της *συνέλιξης*, που είναι ένα ειδικό είδος γραμμικής λειτουργίας. Τα συνελικτικά δίκτυα είναι ουσιαστικά νευρωνικά δίκτυα που χρησιμοποιούν συνέλιξη στη θέση του γενικού πολλαπλασιασμού πινάκων σε τουλάχιστον ένα από τα επίπεδά τους.

3.2.1 Έμπνευση

Το σκεπτικό στο οποίο βασίζεται η ανάπτυξη αυτών των δικτύων είναι δανεισμένο από τα βιολογικά νευρωνικά δίκτυα και ανατρέχει στην πρωτοποριακή εργασία των Hubel και Wiesel [47] πάνω στους τοπικά ευαίσθητους και επιλεκτικούς ως προς τον προσανατολισμό νευρώνες του οπτικού φλοιού της γάτας. Ο οπτικός φλοιός περιέχει μια πολύπλοκη διάταξη κυττάρων. Αυτά τα κύτταρα είναι ευαίσθητα σε μικρές υποπεριοχές του οπτικού πεδίου, τα δεκτικά πεδία. Οι υπο-περιοχές αυτές καλύπτουν ολόκληρο το πεδίο, και δρουν ως τοπικά φίλτρα πάνω στο χώρο εισόδου και είναι κατάλληλες να εκμεταλλευτούν την ισχυρή τοπική και χωρική συσχέτιση που υπάρχει στις φυσικές εικόνες. Επιπροσθέτως, έχουν αναγνωριστεί δύο βασικοί τύποι κυττάρων. Τα απλά κύτταρα ανταποκρίνονται στο μέγιστό τους σε συγκεκριμένα μοτίβα που προσεγγίζουν τις ακμές μέσα στο δεκτικό τους πεδίο. Τα πολύπλοκα κύτταρα έχουν μεγαλύτερα δεκτικά πεδία και είναι τοπικά αμετάβλητα στην ακριβή θέση του μοτίβου.

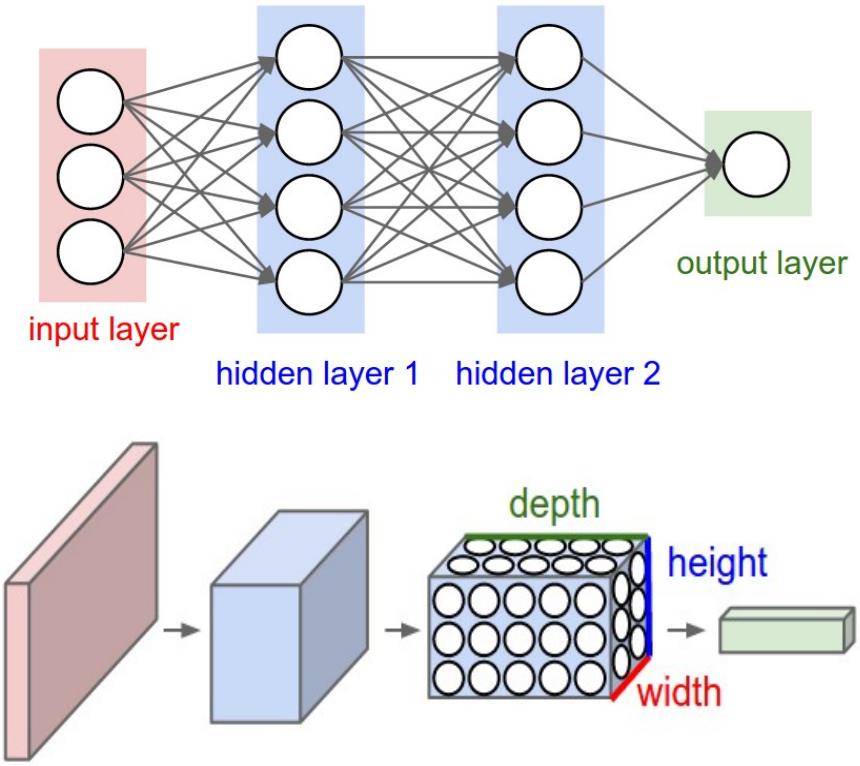
Όλα τα επίπεδα των συνελικτικών δικτύων παίρνουν έμπνευση από τη λειτουργικότητα του οπτικού φλοιού για αναγνώριση αντικειμένων, ανθρώπων, και επαναλαμβανόμενων προτύπων. Ένας από τους λόγους που οδήγησαν στην ανάπτυξη πολλών παρόμοιων μαθηματικών μοντέλων γύρω από τον οπτικό φλοιό ήταν ο Kunihiko Fukushima [48]:

> *Ο μηχανισμός της αναγώρισης προτύπων στον εγκέφαλο είναι ελάχιστα γνωστός, και φαίνεται να είναι σχεδόν αδύνατο να αποκαλυφθεί μόνο από συμβατικά φυσιολογικά πειράματα. Έτσι, επιλέγουμε μια ελαφρώς διαφορετική προσέγγιση στο πρόβλημα αυτό. Αν μπορούσαμε να κατασκευάσουμε ένα μοντέλο νευρωνικού δικτύου το οποίο έχει την ίδια ικανότητα για αναγνώριση προτύπων με το ανθρώπινο ον, θα μπορούσε να μας δώσει μια ισχυρή βοήθεια για την κατανόηση του νευρωνικού μηχανισμού στον εγκέφαλο.*

Έτσι, ένα συνελικτικό νευρωνικό δίκτυο μπορεί να αποτελεί μια ιδέα για το πώς ο εγκέφαλός μας λειτουργεί στο τμήμα της αναγνώρισης, και έχει επιτύχει πολύ καλά αποτελέσματα τα τελευταία χρόνια, ειδικά μετά το 2011, σε μερικές περιπτώσεις μάλιστα ξεπερνώντας την ανθρώπινη ακρίβεια.

Με τα όσα έχουμε δει ως τώρα για τα κανονικά νευρωνικά δίκτυα, λαμβάνουν μια είσοδο (ως ένα διάνυσμα), και τη μετασχηματίζουν μέσα από μια σειρά από κρυφά επίπεδα. Κάθε κρυφό επίπεδο αποτελείται από ένα σύνολο νευρώνων, πλήρως συνδεδεμένους με τους νευρώνες γειτονικών επιπέδων, αλλά πλήρως ανεξάρτητους στο ίδιο επίπεδο, και δεν μοιράζονται καμία σύνδεση μεταξύ τους. Το τελευταίο πλήρως συνδεδεμένο επίπεδο ονομάζεται επίπεδο εξόδου, και στο πρόβλημα της κατηγοριοποίησης αναπαριστά το σκορ για κάθε κατηγορία.

Τα κανονικά νευρωνικά δίκτυα δεν κλιμακώνουν επαρκώς για εισόδους εικόνων. Αν πάρουμε για παράδειγμα ένα πρόβλημα όπου οι είσοδοι είναι εικόνες σταθερού μεγέθους $32\ x\ 32\ x\ 3$ (32 pixel ύψος, 32 pixel πλάτος, 3 κανάλια χρωμάτων), επομένως ένας νευρώνας στο πρώτο επίπεδο θα είχε $32x32x32 = 3072$ βάρη. Μέγεθος φαινομενικά διαχειρίσιμο, όμως αν αυξήσουμε το μέγεθος σε μόλις $200\ x\ 200\ x\ 3$, θα οδηγούμασταν σε 120000 βάρη. Επιπλέον, δεν θα θέλαμε μόνο έναν νευρώνα στο δίκτυό μας, και έτσι το πλήθος των παραμέτρων θα αυξανόταν πολύ γρήγορα. Γίνεται προφανές πως η πλήρης συνδεσιμότητα είναι σπάταλη, και το τεράστιο πλήθος παραμέτρων θα οδηγούσε γρήγορα σε υπερ-μοντελοποίηση.



Εικόνα 5: Σχηματικό διάγραμμα νευρωνικού δικτύου (πανω), και συνελικτικού νευρωνικού δικτύου (κάτω)

Τα Συνελικτικά Νευρωνικά Δίκτυα αξιοποιούν το γεγονός ότι η είσοδος αποτελείται από εικόνες, και περιορίζουν την αρχιτεκτονική με ένα πιο λογικό τρόπο. Πιο συγκεκριμένα, οι νευρώνες σε ένα Συνελικτικό Δίκτυο είναι διατεταγμένοι σε 3 διαστάσεις: πλάτος, ύψος, βάθος. Στο προηγούμενο παράδειγμα, θα είχαμε ως είσοδο έναν τρισδιάστατο νευρώνα μεγέθους $32x32x3$. Οι νευρώνες κάθε επιπέδου όμως θα συνδέονται, όπως θα δούμε στη συνέχεια, με μια μικρή περιοχή του προηγούμενου επιπέδου, αντί για όλους τους νευρώνες, όπως στη πλήρως συνδεδεμένη εκδοχή. Επιπλέον, η τελική έξοδος του κατηγοριοποιητή θα είναι ένα επίπεδο με διαστάσεις $1x1xN$, όπου N το πλήθος των κατηγοριών, διότι στο τέλος του συνελικτικού δικτύου θα μετατρέψουμε την πλήρη εικόνα σε ένα μοναδικό διάνυσμα με σκορ κατηγοριών. Κάθε επίπεδο του δικτύου μετατρέπει το τρισδιάστατο όγκο εισόδου σε έναν τρισδιάστατο όγκο εξόδου που περιέχει την πληροφορία των ενεργοποιήσεων του νευρώνα. Στο παράδειγμα αυτό, το κόκκινο επίπεδο εισόδου περιέχει την εικόνα, έτσι το μήκος και το πλάτος του αντιστοιχεί στις διαστάσεις της εικόνας, και το βάθος θα είναι 3, για τα κανάλια χρώματος (Κόκκινο, Πράσινο, Μπλε – RGB).

3.2.2 Συνέλιξη

Η λειτουργία της συνέλιξης, στην γενική της μορφή, είναι μια λειτουργία πάνω σε δύο πραγματικές συναρτήσεις, και στο πεδίο του χρόνου t, εκτελεί τον υπολογισμό:

$$s(t) = \int x(\alpha)w(t-\alpha)d\alpha \text{ ,ή } s(t) = (x * w)(t)$$

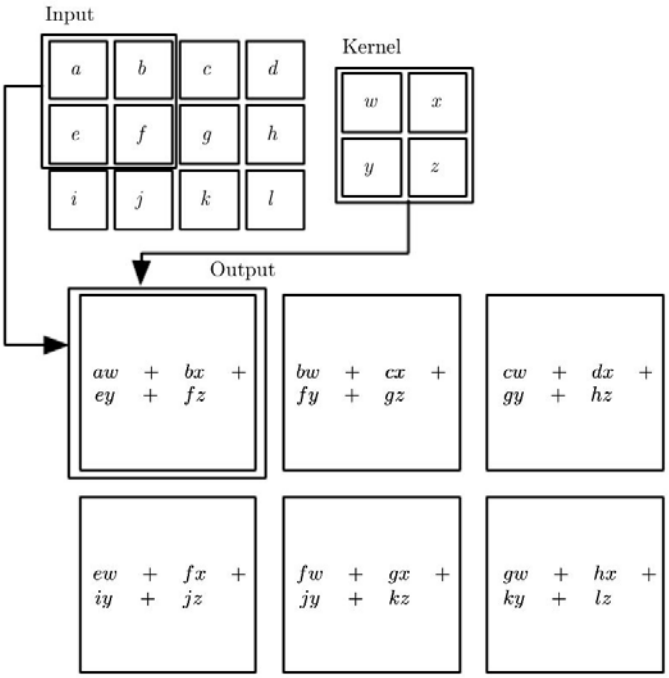
Όπου $x(t)$ η συνάρτηση ή το σήμα εισόδου, και $w(t)$ η συνάρτηση φίλτρου ή 'Πυρήνας' (Kernel), όσο αφορά την ορολογία των νευρωνικών δικτύων. Η έξοδος $s(t)$ συχνά αναφέρεται και ως χάρτης χαρακτηριστικών (feature map).

Στην πραγματικότητα, όταν εργαζόμαστε σε δεδομένα με υπολογιστή, έχουμε διακριτά μεγέθη, και χρησιμοποιούμε την διακριτή συνέλιξη, με την τιμή t να παίρνει διακριτές τιμές:

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

Σε εφαρμογές μηχανικής μάθησης, η είσοδος αποτελείται συνήθως από ένα πολυδιάστατο πίνακα δεδομένων και ο πυρήνας είναι ένας πολυδιάστατος πίνακας παραμέτρων που προσαρμόζονται με βάση τον αλγόριθμο μάθησης. Επειδή κάθε στοιχείο της εισόδου και του πυρήνα πρέπει να αποθηκεύονται ρητά και ξεχωριστά, συνήθως γίνεται η υπόθεση πως οι συναρτήσεις αυτές έχουν μηδενική τιμή παντού εκτός από το πεπερασμένο σύνολο σημείων για τα οποία αποθηκεύουμε τις τιμές. Αυτό σημαίνει πως στη πράξη, μπορούμε να υλοποιήσουμε το άπειρο άθροισμα της προηγούμενης σχέσης ως ένα άθροισμα πάνω από ένα πεπερασμένο πλήθος στοιχείων πινάκων. Επιπλέον, χειριζόμαστε συνελίξεις σε περισσότερους από έναν άξονα, και συγκεκριμένα για μια δισδιάστατη εικόνα I ως είσοδο, θα χρησιμοποιήσουμε και δισδιάστατο πυρήνα (kernel) K:

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



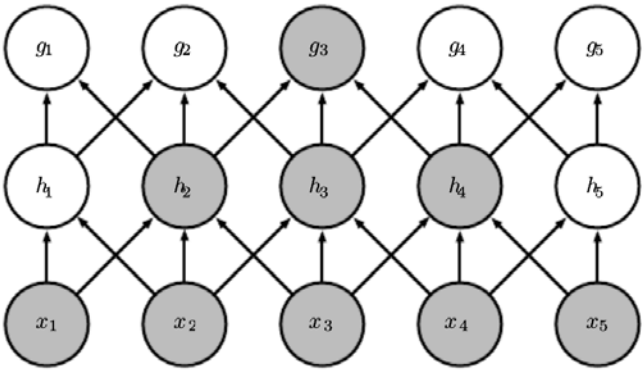
Εικόνα 6: Ένα παράδειγμα δισδιάστατης συνέλιξης

Έχοντας ως είσοδο έναν όγκο μεγέθους $h\ x\ w\ x\ d$, ένα συνελικτικό επίπεδο εφαρμόζει ένα φίλτρο (πυρήνα) μεγέθους $h_f\ x\ w_f\ x\ d\ x\ d_f$ σε κάθε χωρική θέση (h', w'), και έχει σαν έξοδο έναν όγκο μεγέθους $h_o\ x\ w_o\ x\ d_f$. Υπάρχουν δύο υπερπαράμετροι, s_h και s_w , που ονομάζονται διασκελισμός – stride, και καθορίζουν την απόσταση σε ύψος και σε πλάτος μεταξύ των σημείων (h', w') όπου θα γίνει υπολογισμός. Για τιμές $s_h = s_w = 1$, το φίλτρο εφαρμόζεται σε κάθε χωρική θέση του όγκου εισόδου [49].

Η συνέλιξη αξιοποιεί τρεις σημαντικές ιδέες που μπορούν να βοηθήσουν στην βελτίωση ενός συστήματος μηχανικής μάθησης: *αραιή αλληλεπίδραση*, *κοινή χρήση παραμέτρων*, και *ισοδύναμη αναπαράσταση*. Επιπλέον, η συνέλιξη προσφέρει ένα τρόπο να χειριζόμαστε εισόδους με μεταβλητό μέγεθος.

Τα παραδοσιακά νευρωνικά δίκτυα χρησιμοποιούν πολλαπλασιασμό με έναν πίνακα παραμέτρων, με ξεχωριστή παράμετρο που αναπαριστά τη σχέση μεταξύ κάθε νευρώνα εισόδου και εξόδου. Τα συνελικτικά δίκτυα, όμως, τυπικά έχουν αραιές αλληλεπιδράσεις (αραιή συνδεσιμότητα, αραιά βάρη), έχοντας τον πίνακα παραμέτρων μικρότερο από την είσοδο. Για παράδειγμα, μπορούμε να έχουμε μια εικόνα χιλιάδων ή εκατομμυρίων εικονοστοιχείων, αλλά μπορούμε να εντοπίζουμε μικρά και σημαντικά χαρακτηριστικά όπως ακμές με πυρήνες που καταλαμβάνουν μόνο δεκάδες ή εκατοντάδες εικονοστοιχεία. Έτσι αποθηκεύουμε λιγότερες παραμέτρους, γεγονός που μειώνει τις απαιτήσεις σε μνήμη του μοντέλου, και βελτιώνει την στατιστική αποδοτικότητά του. Επίσης χρειαζόμαστε λιγότερες πράξεις για υπολογισμό της εξόδου. Σε ένα βαθύ νευρωνικό δίκτυο, οι νευρώνες στα πιο βαθιά επίπεδα μπορούν να αλληλεπιδρούν έμμεσα με ένα μεγαλύτερο τμήμα της εισόδου, όπως φαίνεται και στο παρακάτω διάγραμμα. Αυτό επιτρέπει στο δίκτυο να περιγράψει

αποδοτικά πολύπλοκες αλληλεπιδράσεις μεταξύ πολλών μεταβλητών, κατασκευάζοντας τέτοιες αλληλεπιδράσεις από απλά δομικά στοιχεία καθένα από τα οποία περιγράφει μόνο αραιές αλληλεπιδράσεις.



Εικόνα 7: Το αυξανόμενο μέγεθος του δεκτικού πεδίου σε πιο βαθιά επίπεδα

Ο όρος *κοινή χρήση παραμέτρων* (parameter sharing) αναφέρεται στη χρήση της ίδιας παραμέτρου για περισσότερες από μια συναρτήσεις στο μοντέλο. Σε ένα παραδοσιακό νευρωνικό δίκτυο, κάθε στοιχείο του πίνακα βαρών χρησιμοποιείται ακριβώς μια φορά όταν υπολογίζεται η έξοδος του επιπέδου, καθώς πολλαπλασιάζεται με ένα στοιχείο της εισόδου και έπειτα δεν χρησιμοποιείται ξανά. Αντίθετα, σε ένα συνελικτικό νευρωνικό δίκτυο, κάθε στοιχείο του πυρήνα χρησιμοποιείται σε κάθε θέση της εισόδου, εκτός ίσως από οριακές περιπτώσεις για τα εικονοστοιχεία που βρίσκονται στο σύνορο της εικόνας. Η κοινή χρήση των παραμέτρων που χρησιμοποιείται από την λειτουργία της συνέλιξης σημαίνει ότι, αντί να μαθαίνεται ένα ξεχωριστό σύνολο παραμέτρων για κάθε θέση, εκπαιδεύεται μόνο ένα σύνολο. Αυτό μειώνει σε μεγαλύτερο βαθμό τις απαιτήσεις σε μνήμη, καθιστώντας τη συνέλιξη δραματικά αποδοτικότερη από τον απλό πολλαπλασιασμό πινάκων.

Στη συνέλιξη, αυτή η μορφή της κοινής χρήσης παραμέτρων προκαλεί τα επίπεδα του δικτύου να κατέχουν μια ιδιότητα που ονομάζεται ισοδυναμία σε μετασχηματισμούς. Λέγοντας πως μια συνάρτηση είναι ισοδύναμη σημαίνει πως αν αλλάξει η είσοδος, η έξοδος θα αλλάξει με τον ίδιο τρόπο. Στην αναγνώριση εικόνων, αν σε μια εικόνα μετακινήσουμε ένα αντικείμενο στην είσοδο, η αναπαράσταση του αντικειμένου θα μετακινηθεί το ίδιο και στην έξοδο. Αυτό είναι χρήσιμο, έχοντας τη γνώση πως κάποια συνάρτηση πάνω σε λίγα γειτονικά εικονοστοιχεία θα εφαρμοστεί σε πολλές θέσεις της εισόδου. Για παράδειγμα, όταν χειριζόμαστε εικόνες, μπορούμε να εντοπίζουμε ακμές στο πρώτο επίπεδο του συνελικτικού νευρωνικού δικτύου μας. Οι ίδιες ακμές αυτές θα επαναλαμβάνονται παντού στην εικόνα, επομένως είναι πρακτικό να μοιραζόμαστε τις παραμέτρους για ολόκληρη την εικόνα.

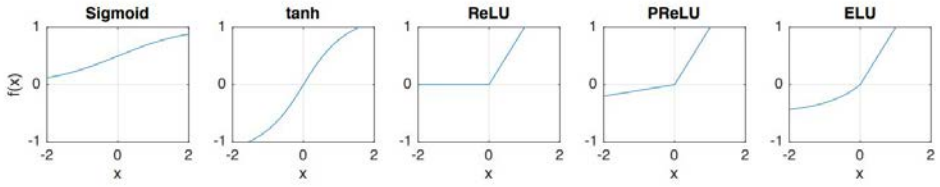
3.2.3 Επίπεδα Συνελικτικού Νευρωνικού Δικτύου

Στο πρόβλημα της κατηγοριοποίησης, η δομή ενός συνελικτικού δικτύου είναι μια σειρά από υπολογιστικά μπλοκ (ή επίπεδα) που στοιβάζονται το ένα μετά το άλλο, έχοντας ως είσοδο του τρέχοντος επιπέδου την έξοδο του προηγούμενου. Η είσοδος είναι μεγέθους $(h_{in}\ x\ w_{in}\ x\ d_{in})$ και η έξοδος είναι μεγέθους $(h_{out}\ x\ w_{out}\ x\ d_{out})$

Ένα τυπικό Συνελικτικό Νευρωνικό Δίκτυο απαρτίζεται από τα εξής επίπεδα:

- **<u>Επίπεδο Συνέλιξης</u>**: Εφαρμόζονται $(d_{in}x\ d_{out})$ φίλτρα (πυρήνες - kernels) μεγέθους $(k\ x\ k)$ για να παράξουν τους χάρτες χαρακτηριστικών εξόδου. Για φίλτρα μεγαλύτερα από $(1\ x\ 1)$, τα συνοριακά φαινόμενα θα μειώσουν τις διαστάσεις εξόδου. Για να αποφευχθεί το φαινόμενο αυτό, εφαρμόζεται τυπικά παραγέμισμα (padding) με $p = \lfloor k/2 \rfloor$ μηδενικά σε κάθε πλευρά. Τα φίλτρα μπορούν επίσης να εφαρμοστούν με βήμα s, που μειώνει επιπλέον τις διαστάσεις εξόδου σε πλάτος και ύψος κατά έναν παράγοντα s.

- **<u>Επίπεδο Μη-Γραμμικότητας:</u>** Εφαρμόζεται μια μη-γραμμική συνάρτηση ενεργοποίησης σε κάθε εικονοστοιχείο εισόδου. Η πιο διαδεδομένη συνάρτηση όπως αναφέραμε και παραπάνω είναι η ReLU, που υπολογίζει: $f(x) = \max\{0, x\}$ και αποκόπτει τις αρνητικές τιμές, περιορίζοντάς τες στο 0. Πρότερα νευρωνικά δίκτυα χρησιμοποιούσαν σιγμοειδείς συναρτήσεις, όπως την σιγμοειδή (sigmoid) και την υπερβολική εφαπτομένη (tanh), όμως δεν χρησιμοποιούνται πλέον λόγω της υπολογιστικής τους πολυπλοκότητας και της αργής σύγκλισής τους. Πρόσφατες ιδέες περιλαμβάνουν την Παραμετρική ReLU (Parametric ReLU – PreLU [50]): $f(x) = \max\{ax, x\}$, με παράμετρο α που επίσης μαθαίνεται, Maxout [51] και Εκθετικά Γραμμικά Στοιχεία (Exponential Linear Units – ELU [52]). Ακολουθεί μια σύγκριση μεταξύ τους:



Εικόνα 8: Μια σύγκριση μεταξύ διαφορετικών συναρτήσεων ενεργοποίησης. Με τη σειρά: Sigmoid, tanh, ReLU, PReLU, ELU

- **<u>Επίπεδο Συγκέντρωσης (Pooling):</u>** Στο επίπεδο αυτό μειώνονται οι χωρικές διαστάσεις της εισόδου συνοψίζοντας πολλαπλά εικονοστοιχεία εισόδου σε ένα μοναδικό εικονοστοιχείο εξόδου. Δυο δημοφιλείς επιλογές αποτελούν οι συναρτήσεις $max - pool$ και $avg - pool$, οι οποίες συνοψίζουν το τοπικό τους δεκτικό πεδίο λαμβάνοντας τη μέγιστη ή τη μέση τιμή των εικονοστοιχείων, αντίστοιχα. Εφαρμόζονται σε μικρή περιοχή, με τυπικές διαστάσεις να είναι $2x2$ ή $3x3$, με τις περιοχές να απέχουν μεταξύ τους $k = 2$ εικονοστοιχεία. Μεγαλύτερα μεγέθη περιοχών έχουν αποτέλεσμα μεγαλύτερη απώλεια πληροφορίας. Για

παράδειγμα, με ένα φίλτρο 2x2, έχουμε 1 εικονοστοιχείο εξόδου, και απορρίπτουμε 3, το 75% της εισόδου.

Η συνάρτηση pooling αντικαθιστά την του δικτύου σε μια συγκεκριμένη θέση με μια συνοπτική στατιστική των κοντινών εξόδων. Είναι μια συνάρτηση με άλλα λόγια που συνδυάζει τις συνεισφορές των γειτονικών εικονοστοιχείων εξόδου. Για παράδειγμα, η λειτουργία max pooling επιστρέφει τη μέγιστη τιμή σε μια τετραγωνική γειτονιά. Άλλες δημοφιλείς συναρτήσεις pooling περιλαμβάνουν τη μέση τιμή των γειτονικών εξόδων, η L^2 νόρμα, ή ένα σταθισμένο μέσο με βάση την απόσταση από το κεντρικό εικονοστοιχείο.

Σε κάθε περίπτωση, η διαδικασία αυτή βοηθάει στο να κάνει την αναπαράσταση προσεγγιστικά πιο αμετάβλητη σε μικρές μεταβολές στην είσοδο. Αυτό σημαίνει πως αν υπάρξει μια μεταβολή στην είσοδο κατά ένα μικρό ποσό, οι τιμές των περισσότερων εξόδων μετά το pooling θα παραμείνουν αμετάβλητες. Η αμεταβλητότητα σε μικρές τοπικές αλλαγές είναι πολύ χρήσιμη ιδιότητα αν μας ενδιαφέρει περισσότερο αν κάποιο χαρακτηριστικό είναι παρόν παρά το πού ακριβώς εντοπίζεται.

Επειδή με το pooling γίνεται μια σύνοψη των αποκρίσεων σε μια γειτονιά, είναι εφικτό να έχουμε μικρότερη έξοδο, επιστρέφοντας συνόψεις αποκρίσεων για περιοχές που απέχουν μεταξύ τους k εικονοστοιχεία αντί για 1. Βελτιώνεται έτσι η υπολογιστική αποδοτικότητα του δικτύου μας, καθώς το επόμενο επίπεδο θα έχει περίπου k φορές μικρότερη είσοδο για να επεξεργαστεί. Όταν το πλήθος των παραμέτρων στο επόμενο επίπεδο εξαρτάται από το μέγεθος της εισόδου, αυτή η μείωση μπορεί να έχει ως αποτέλεσμα βελτιωμένη στατιστική απόδοση και μειωμένες απαιτήσεις σε μνήμη για αποθήκευση των παραμέτρων.

Για πολλές εργασίες, το pooling είναι αναγκαίο για το χειρισμό εισόδων με ποικίλο μέγεθος. Για παράδειγμα, αν επιθυμούμε να κατηγοριοποιήσουμε εικόνες μεταβλητού μεγέθους, η είσοδος στον κατηγοριοποιητή πρέπει να έχει σταθερό μέγεθος. Αυτό συνήθως επιτυγχάνεται μεταβάλλοντας το μέγεθος της απόστασης μεταξύ των περιοχών στις οποίες θα εφαρμοστεί το pooling, έτσι ώστε το επίπεδο κατηγοριοποίησης πάντα να λαμβάνει στην είσοδό του τον ίδιο αριθμό χαρακτηριστικών, ανεξάρτητα από το αρχικό μέγεθος της εισόδου.

- **<u>Πλήρως Συνδεδεμένο Επίπεδο:</u>** (Fully-Connected Layer) Το επίπεδο αυτό βρίσκει χρήση στα τελευταία επίπεδα ενός Συνελικτικού Νευρωνικού Δικτύου, για να υπολογίσει τις προβλέψεις στις κατηγορίες σε εφαρμογές όπως η κατηγοριοποίση εικόνων, όπως και το αντικείμενο της παρούσας διπλωματικής εργασίας. Αποτελεί τμήμα των κλασσικών Νευρωνικών Δικτύων, και περιέχει πλήρεις συνδέσεις με κάθε ενεργοποίηση του προηγούμενου επιπέδου. Υπάρχει μια τεχνική, η οποία μετατρέπει ένα FC layer σε ένα CONV layer (Επίπεδο συνέλιξης). Αξίζει να επισημανθεί πως η μόνη διαφορά μεταξύ των δύο ειδών επιπέδων είναι πως οι νευρώνες στο επίπεδο συνέλιξης συνδέονται μόνο με μια τοπική περιοχή της εισόδου, και πως πολλοί από τους νευρώνες μοιράζονται παραμέτρους. Ωστόσο, οι νευρώνες και στα δύο επίπεδα υπολογίζουν εξίσου γινόμενα, οπότε η λειτουργικότητά τους είναι πανομοιότυπη. Για κάθε επίπεδο συνέλιξης υπάρχει ένα πλήρως συνδεδεμένο επίπεδο που υλοποιεί την

ίδια συνάρτηση. Ο πίνακας βαρών θα ήταν ένας μεγάλος πίνακας όπου θα είναι στις περισσότερες θέσεις μηδενικός εκτός από μερικά σημεία-ομάδες (εξαιτίας της τοπικής συνεκτικότητας) όπου τα βάρη σε πολλά από τις ομάδες είναι ίσα (εξαιτίας της κοινής χρήσης παραμέτρων).
Αντιστρόφως, οποιοδήποτε πλήρως συνδεδεμένο επίπεδο μπορεί να μετατραπεί σε ένα συνελικτικό επίπεδο. Για παράδειγμα, ένα πλήρες συνδεδεμένο επίπεδο με $K = 4096$ νευρώνες εξόδου, που παίρνει ως είσοδο μια εικόνα $7x7x512$, μπορεί να εκφραστεί ως ένα επίπεδο συνέλιξης με χωρικό εύρος $F = 7$, padding $P = 0$, διασκελισμό $S = 1$ και $K = 4096$ κανάλια εξόδου. Με άλλα λόγια, ορίζουμε το μέγεθος του φίλτρου να είναι ακριβώς ίδιο με τις διαστάσεις του όγκου εισόδου, και έτσι η έξοδος θα είναι ένας όγκος διαστάσεων $1x1x4096$, δίνοντας πανομοιότυπο αποτέλεσμα με το πλήρως συνδεδεμένο επίπεδο.

- **Επίπεδο Κανονικοποίησης Τοπικής Απόκρισης:** (Local Response Normalization – LRN) εφαρμόζουν ένα συναγωνισμό μεταξύ των νευρώνων γειτονικών καναλιών, κανονικοποιώντας τις αποκρίσεις τους σε σχέση με μια ορισμένη γειτονιά N καναλιών. Τα επίπεδα LRN παρουσιάστηκαν στην δημοφιλή αρχιτεκτονική AlexNet [53], όμως απαντώνται λιγότερο συχνά σε πιο πρόσφατες αρχικτεκτονικές.

- **Επίπεδο Κανονικοποίησης Δέσμης:** (Batch Normalization – BN) Εισήχθησαν το 2015 από ερευνητές της Google [54]. Εφαρμόζεται μετά από κάθε δέσμη εκπαίδευσης και κανονικοποιεί την έξοδο του κάθε επιπέδου σε κατανομή με μηδενική μέση τιμή και μοναδιαία διασπορά. Η ομοιόμορφη κατανομή εισόδου στα επόμενα επίπεδα επιτρέπει μεγαλύτερους ρυθμούς μάθησης και έτσι επιταχύνουν την διαδικασία της εκπαίδευσης και βελτιώνουν την ακρίβεια του δικτύου.

- **Επίπεδα Dropout:** Τα επίπεδα αυτά αποτελούν μια δημοφιλή μέθοδο για αντιστάθμιση του φαινομένου της υπερ-εκπαίδευσης σε μεγάλα Συνελικτικά Δίκτυα. Με τυχαίο τρόπο, αφαιρούν ένα ποσοστό από τις συνδέσεις του δικτύου κατά την εκπαίδευση, γεγονός που αποτρέπει το δίκτυο από το να μάθει πολύ συγκεκριμένες αντιστοιχίσεις μεταξύ προτύπων εισόδου και εξόδου, και επιφέρει να δημιουργηθεί μια ασάφεια και πλεονασμό πάνω στα εκπαιδεύσιμα βάρη.

- **Επίπεδα Softmax:** Αποτελούν τους πιο κοινούς κατηγοριοποιητές. Ένα επίπεδο κατηγοριοποίησης προστίθεται μετά από το τελευταίο επίπεδο πλήρους σύνδεσης ή συνέλιξης σε κάθε Συνελικτικό Νευρωνικό Δίκτυο κατηγοριοποίησης εικόνων, και έχει σαν έξοδο κανονικοποιημένες πιθανότητες κατηγοριών P_i, με είσοδο ακατέργαστα σκορ κατηγοριών z_i, σύμφωνα με τη συνάρτηση $P_i = e^{z_i} / \sum_k e^{z_k}$.

3.2.4 Πλεονεκτήματα

Συνοψίζοντας, τα συνελικτικά νευρωνικά δίκτυα παρουσιάζουν πλεονεκτήματα τα οποία τους προσδίδουν ιδιαίτερες ικανότητες αναγνώρισης και κατηγοριοποίησης εικόνων. Συγκεκριμένα, μπορούν να εξαγάγουν σχετικά χαρακτηριστικά από την εικόνα, αποτελώντας από τη φύση τους τη βέλτιστη (τη περίοδο που γράφτηκε το κείμενο αυτό) αρχιτεκτονική για όραση υπολογιστών. Αυτό συμβαίνει επειδή λειτουργούν με την ίδια λογική όπως και το ανθρώπινο σύστημα όρασης. Με ιεραρχική αφαίρεση, εξαγάγει μοναδικά χαρακτηριστικά της εισόδου. Η συνέλιξη και η υποδειγμάτωση είναι μια έμφυτη λειτουργία του ανθρώπινου οπτικού συστήματος, και η αρχιτεκτονική των συνελικτικών δικτύων έχει ως στόχο να το προσομοιώσει.

Χρησιμοποιώντας μια βαθιά αρχιτεκτονική και μειώνοντας τη χωρική ανάλυση του χάρτη χαρακτηριστικών, τα συνελικτικά δίκτυα επιτυγχάνουν ένα μεγάλο βαθμό αμεταβλητότητας σε περιστροφή, μετακίνηση και παραμόρφωση. Αυτό είναι κάτι που δεν μπορεί να επιτευχθεί από τα SVM και άλλες ‘ρηχές’ αρχιτεκτονικές. Μια ακόμα σύγκριση με τα SVMs και τις αντίστοιχες ρηχές αρχιτεκτονικές μας αφήνει να παρατηρήσουμε πως οι εξαγωγείς χαρακτηριστικών στα νευρωνικά δίκτυα προκύπτουν από αυτόματη εκπαίδευση, χωρίς να χρειάζεται ο χειροκίνητος σχεδιασμός και υλοποίησή τους. Αυτό μας παρέχει μεγαλύτερη εξοικονόμηση χρόνου και κόπου, και προσεγγίζει περισσότερο την έννοια της τεχνικής νοημοσύνης.

Χάρη στην ιδέα των κοινόχρηστων βαρών, έχουμε σημαντικά λιγότερες συνδέσεις στα επίπεδα των συνελικτικών δικτύων, και κατ’επέκταση το πλήθος των παραμέτρων όλου του δικτύου είναι αρκετές τάξεις μεγέθους μικρότερο από τα αντίστοιχα δίκτυα Perceptron Πολλαπλών Επιπέδων (Multi-Layer Perceptrons – MLP). Αυτό το γεγονός καθιστά τα συνελικτικά νευρωνικά δίκτυα εκθετικά ευκολότερα στην εκπαίδευση σε σχέση με τα MLP.

3.3 ΚΑΤΗΓΟΡΙΟΠΟΙΗΣΗ ΕΙΚΟΝΩΝ

Ένα από τα πιο ενδιαφέροντα, αλλά ταυτόχρονα ένα από τα πιο δύσκολα προβλήματα στο αντικείμενο της όρασης υπολογιστών, αποτελεί η Κατηγοριοποίηση Εικόνων (Image Classification). Η εργασία της σωστής ανάθεσης μιας από τις πολλαπλές πιθανές ετικέτες σε μια δεδομένη εικόνα. Παραδείγματα τέτοιων προβλημάτων αποτελούν αποφάσεις τύπου ΝΑΙ/ΟΧΙ (Υπάρχει άνθρωπος στην εικόνα; Είναι κακοήθης ο όγκος της εικόνας;) αλλά και προβλήματα αναγνώρισης, με μεγάλο πλήθος ετικετών (Τι ράτσα σκύλου είναι αυτή; Ποιος είναι στην εικόνα;).

Στο πλαίσιο του ετήσιο διαγωνισμού *ImageNet Large Scale Visual Recognition Challenge (ILSVRC)* [55], οι συμμετέχοντες αναπτύσσουν αλγορίθμους για να κατηγοριοποιήσουν εικόνες από ένα υποσύνολο της βάσης δεδομένων ImageNet. Η βάση δεδομένων ImageNet αποτελείται από 1000 κατηγορίες, συνολικά με περισσότερες από 14 εκατομμύρια φωτογραφίες συλλεγμένες από τον Ιστό, κάθε μία έχοντας μια ετικέτα που αντιστοιχεί στην κατηγορία της. Το σύνολο εκπαίδευσης αποτελείται από περίπου 1,2 εκατομμύρια φωτογραφίες, καλύπτοντας μια τεράστια ποικιλία από αντικείμενα (από χαρτί υγείας και φρούτα, έως διαστημικά λεωφορεία και ηφαίστεια), σκηνές (από κοιλάδες και ακτές μέχρι βιβλιοθήκες και μοναστήρια) και ζώα (120 ράτσες σκύλων αλλά και καρχαρίες). Μερικά δείγματα φαίνονται στην παρακάτω εικόνα:

Εικόνα 9: Δείγματα εικόνων από τη βάση δεδομένων ImageNet (λευκός καρχαρίας, μπανάνα, ηφαίστειο, πυροσβεστικό όχημα, πομεράνιαν, διαστημικό λεωφορείο, χαρτί υγείας)

Οι συμμετέχοντες καλούνται να εκτιμήσουν το περιεχόμενο της κάθε εικόνας, με τις εικόνες που τους παρουσιάζονται να είναι υποσύνολο της βάσης δεδομένων ImageNet, χωρίς όμως να διαθέτουν κάποια επισήμανση για την αντίστοιχη ετικέτα τους. Ο κάθε αλγόριθμος που λαμβάνει μέρος στο διαγωνισμό, και συγκεκριμένα στο τομέα της κατηγοριοποίησης παράγει μια λίστα με 5 προβλέψεις, με φθίνουσα σειρά βεβαιότητας. Η ποιότητα της αντιστοίχισης σε ετικέτα εκτιμάται βασιζόμενη στην ετικέτα που ταιριάζει καλύτερα στην πραγματική. Η ιδέα πίσω από αυτό είναι για να επιτραπεί στον αλγόριθμο να εντοπίζει πολλαπλά αντικείμενα σε μια εικόνα, και να μην τιμωρείται αν ένα από τα αντικείμενα που εντοπίστηκαν είναι παρόν στην εικόνα, αλλά δεν υπήρχε στην πραγματική κατηγορία.

Για κάθε εικόνα, ένας αλγόριθμος λοιπόν παράγει 5 ετικέτες $l_j, j = 1, \dots, 5$. Οι πραγματικές ετικέτες (ground truth label) για την εικόνα είναι $g_k, k = 1, \dots, n$. Το σφάλμα του αλγορίθμου για την εικόνα θα είναι

$$e = \frac{1}{n}\sum_{k} \min_{j} d(l_j, g_k)\,.$$

Στην σχέση αυτή έχουμε

$$d(x,y) = \begin{cases} 0, & \alpha\nu\ x = y \\ 1, & \delta\iota\alpha\varphi\rho\varepsilon\tau\iota\kappa\acute{\alpha} \end{cases}.$$

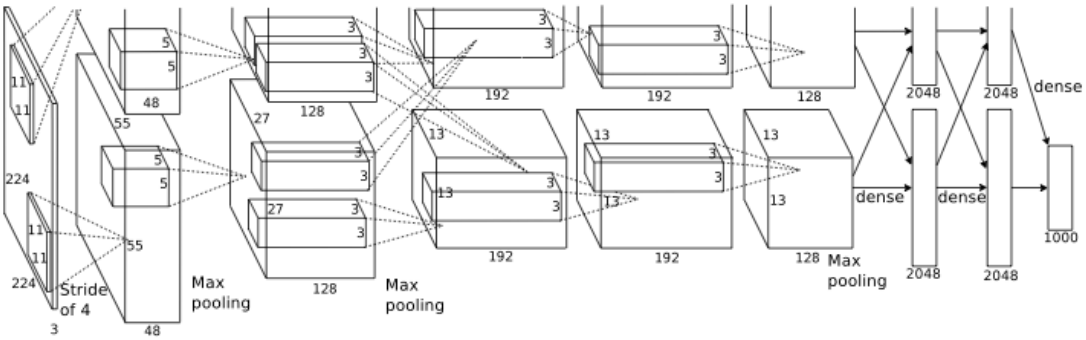
Το συνολικό σφάλμα του αλγορίθμου είναι ο μέσος όρος των σφαλμάτων, για όλες τις εικόνες δοκιμής. Στην πιο δημοφιλή έκδοση του διαγωνισμού, LSVRC2012, υπήρχε μια ετικέτα για κάθε εικόνα, δηλαδή $n = 1$.

Το σφάλμα αυτό ονομάζεται top-5 error, καθώς έχει 5 δοκιμές. Υπάρχει αντίστοιχα και το πιο αυστηρό, top-1 error, λαμβάνοντας τη πρώτη πρόβλεψη μόνο. Αξίζει να επισημανθεί πως κατά τη συλλογή των εικόνων για κάθε κατηγορία στη βάση δεδομένων, υπήρχε και ο ανθρώπινος παράγοντας του οποίου εκτιμήθηκε η ακρίβεια. Υπήρχε ανθρώπινο σφάλμα top-5 της τάξης των 5% [56].

Το πολύ μεγάλο πλήθος δειγμάτων εκπαίδευσης και η δυσκολία του προβλήματος οδήγησαν τον διαγωνισμό ImageNet να γίνει ένας ιδανικός «παιδότοπος» για αλγορίθμους μηχανικής μάθησης. Ξεκινώντας με το *AlexNet* το 2012, τα συνελικτικά νευρωνικά δίκτυα έχουν λάβει τα ηνία του διαγωνισμού ILSVRC, και τα σφάλματα top-1 και top-5 έχουν μειωθεί σημαντικά έκτοτε. Στη συνέχεια συνοψίζονται οι σημαντικότερες τοπολογίες:

AlexNet

Κατασκευασμένο από τον Alex Krizhevsky κ.ά. από το Πανεπιστήμιο του Τορόντο, ήταν το πρώτο Συνελικτικό Νευρωνικό Δίκτυο που κέρδισε τον διαγωνισμό ILSVRC το 2012. Αποτελείται από 5 συνελικτικά επίπεδα, κάποια από τα οποία ακολουθούνταν από επίπεδα max-pooling, έχει 60 εκατομμύρια παραμέτρους, 650000 νευρώνες και απαιτεί περίπου 1,1 δισεκατομμύρια πράξεις τύπου multiply-accumulate (MACC) για ένα πρόσθιο πέρασμα (forward pass). Το δίκτυο είχε πετύχει το πρωτοποριακό τότε $top - 5$ σφάλμα 15,3%, με το δεύτερο αλγόριθμο στην κατάταξη να ακολουθεί με αντίστοιχο σφάλμα 26,2% [53].



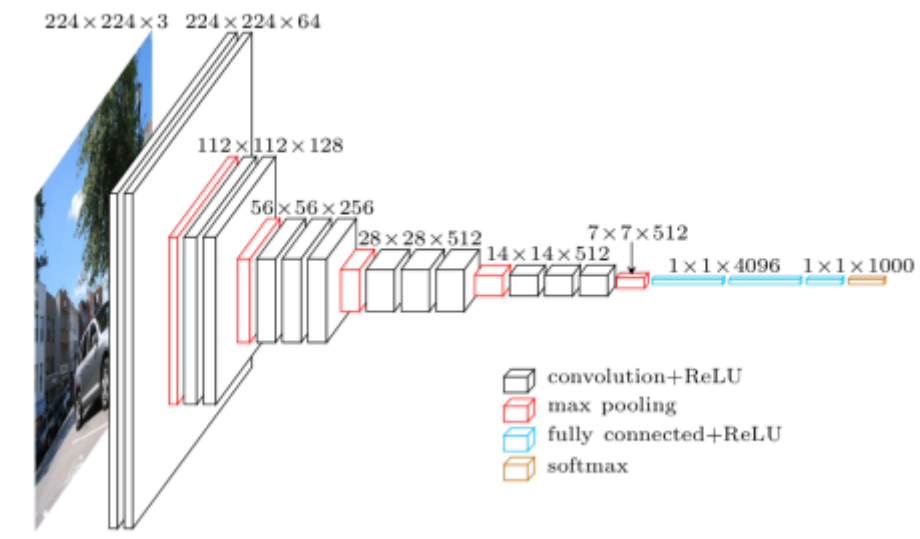
Εικόνα 10 - Το AlexNet όπως παρουσιάζεται στην δημοσίευση - το σχεδιάγραμμα είναι κομμένο στο πάνω μέρος του και στην αρχική δημοσίευση [53]

Network-in-Network (NiN)

Από τους Min Lin κ.ά., του Εθνικού Πανεπιστημίου της Σιγκαπούρης, δημοσιεύτηκε ως μια νέα αρχιτεκτονική συνελικτικού νευρωνικού δικτύου το 2013. Η αρχιτεκτονική NiN αποτελείται από μικρά, στοιβαγμένα πολυεπίπεδα perceptrons, που μετακινούνται πάνω από την αντίστοιχη είσοδο, όπως και τα συνελικτικά δίκτυα. Επιπροσθέτως, οι συγγραφείς χρησιμοποιούν επίπεδο συνέλιξης στον κατηγοριοποιητή αντί για πλήρως συνδεδεμένο επίπεδο. Αυτό κάνει το δίκτυο πολύ μικρότερο από άποψη παραμέτρων. Δεν συμμετείχε ποτέ επίσημα στον διαγωνισμό ILSVRC, όμως έχει εκπαιδευτεί κατά καιρούς πάνω στο σύνολο δεδομένων ImageNet και προσεγγίζει την ακρίβεια του AlexNet [57] [58].

VGG (Visual Geometry Group)

Ερευνητικές ομάδες από το Πανεπιστημίου της Οξφόρδης εισήγαγαν την αρχιτεκτονική VGG και κέρδισαν μέρος του διαγωνισμού ILSVRC 2014. Πειραματίστηκαν με βαθιά Συνελικτικά Νευρωνικά Δίκτυα, που αποτελούνταν από έως και 19 συνελικτικά επίπεδα. Η πιο δημοφιλής παραλλαγή $VGG-16$ έχει βάθος 16 επίπεδα, και μια πολύ κανονική δομή, αποτελούμενο αποκλειστικά από πυρήνες συνέλιξης $3x3$ και επίπεδα max-pooling $2x2$. Οι χωρικές διαστάσεις μειώνονται σταδιακά από $224x224$ εικονοστοιχεία στην είσοδο στα $7x7$, ενώ το πλήθος των καναλιών αυξανόταν ταυτόχρονα από 3 στα 4096. Το δίκτυο έφτασε top-5 σφάλμα $7{,}3\%$. Ωστόσο, το VGG-16 περιέχει περίπου 140 εκατομμύρια βάρη, και ένα πρόσθιο πέρασμα απαιτεί περίπου 16 δισεκατομμύρια MACC λειτουργίες [59].

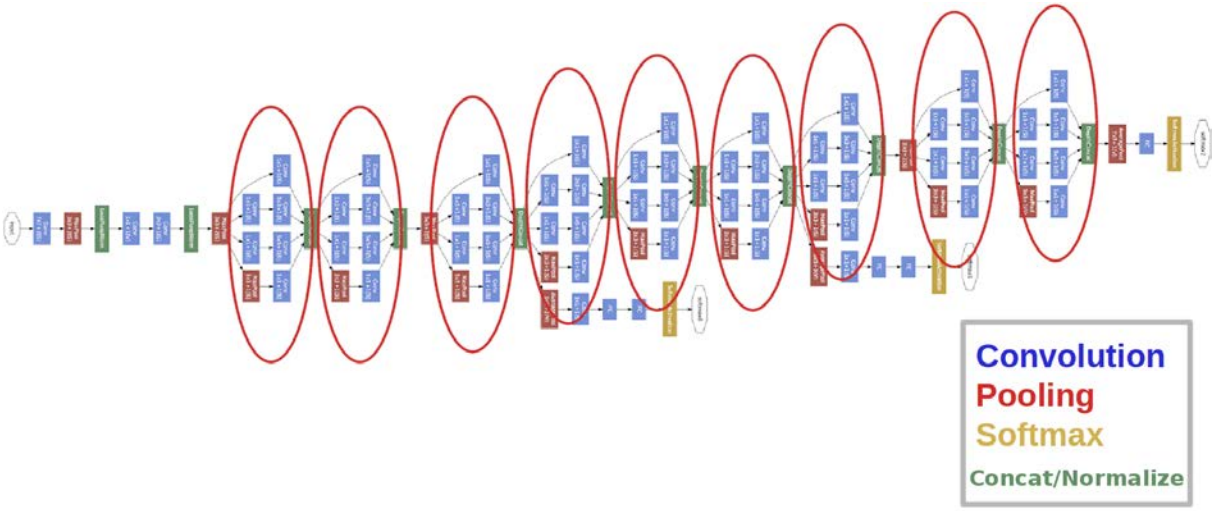


Εικόνα 11 - Δομή του VGG-16 [59]

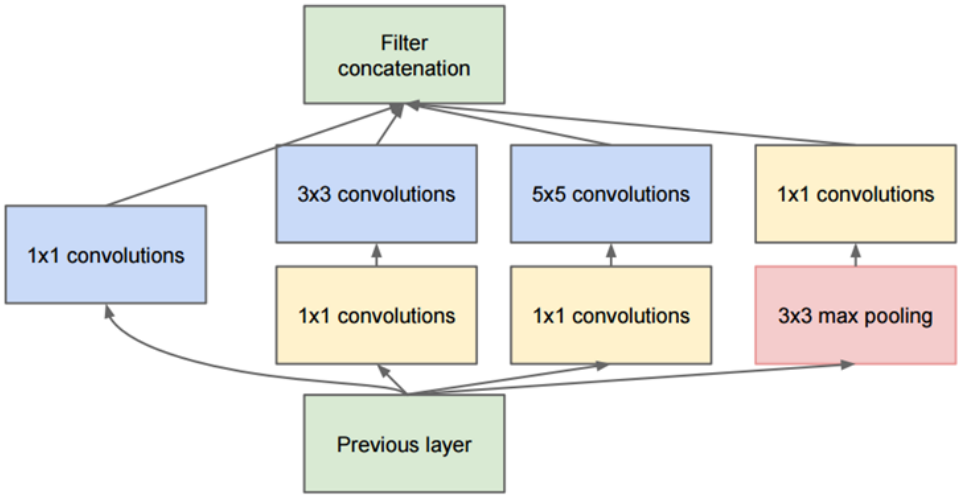
GoogLeNet

Ένα Συνελικτικό Νευρωνικό Δίκτυο από τον Christian Szegedy κ.ά. της Google, ήταν μια αρχιτεκτονική που αποτέλεσε ορόσημο, και δημοσιεύτηκε μόλις μερικές ημέρες μετά την αρχιτεκτονική VGG που παρουσιάσαμε παραπάνω. Το GoogLeNet με τα 22 συνελικτικά του

επίπεδα έθεσε νέο ρεκόρ στον διαγωνισμό ILSVRC στο έργο της κατηγοριοποίησης εικόνων, πετυχαίνοντας $top-5$ σφάλμα $6{,}67\%$, ενώ ταυτόχρονα απαιτούσε μόνο 12 εκατομμύρια παραμέτρους-βάρη. Η οικονομία αυτή στα βάρη επιτυγχάνεται με μια πιο πολύπλοκη αρχιτεκτονική, που χρησιμοποιεί τις λεγόμενες μονάδες (modules) $Inception$. Το όνομά τους είναι εμπνευσμένο από την ομώνυμη ταινία του 2010, καθώς αποτελούν ένα μικρό νευρωνικό δίκτυο μέσα στο νευρωνικό δίκτυο. Αρχικά εφαρμόζουν ένα επίπεδο συνέλιξης $1x1$ για να μειώσουν το πλήθος των καναλιών της εισόδου, προτού διαστείλουν πάλι αυτή τη συμπιεσμένη αναπαράσταση εφαρμόζοντας παράλληλα επίπεδα συνέλιξης, με φίλτρα διαστάσεων $1x1, 3x3$ και $5x5$. Η σύνοψη (reduction) στη διάσταση των καναλιών μειώνει το πλήθος των παραμέτρων και συνεπώς το πλήθος των απαιτούμενων πράξεων τύπου MACC, και η σύνθεση αυτή των πολλαπλών επιπέδων ενισχύει την μη-γραμμική εκφραστικότητα του δικτύου. Για να βελτιώσει την σύγκλιση του δικτύου, το GoogLeNet εφαρμόζει επίσης τη τεχνική κανονικοποίησης τοπικής απόκρισης (Local Response Normalization – LRN) [60].



Εικόνα 12: Η αρχιτεκτονική GoogLeNet. Σημειωμένα είναι τα Inception Modules



Εικόνα 13: Ένα μεμονωμένο Inception Module με όλα τα επιμέρους επίπεδά του [60]

ResNet

Η αρχιτεκτονική αυτή, παρουσιάστηκε από τους Kaiming He κ.ά. του Microsoft Research Asia και συμμετείχε στους διαγνωισμούς ILSVRC 2015 [61] και COCO 2015 [62], όπου απέσπασαν τη πρώτη θέση στις κατηγορίες ImageNet classification, ImageNet detection, ImageNet localization, COCO detection, και COCO segmentation.

Το πολύ βαθύ τους μοντέλο ResNet-152 πέτυχε $top-5$ σφάλμα μικρότερο του $5{,}7\%$ χρησιμοποιώντας 152 συνελικτικά επίπεδα. Μέχρι εκείνη τη στιγμή, αρχιτεκτονικές με βάθος μεγαλύτερο από 20 συνελικτικά επίπεδα ήταν πολύ δύσκολο να εκπαιδευτούν. Οι ερευνητές έλυσαν το πρόβλημα αυτό συμπεριλαμβάνοντας «παρακάμψεις» γύρω από κάθε δέσμη από 2 διαδοχικών επιπέδων συνέλιξης, αθροίζοντας τόσο την αρχική όσο και την μεταλλαγμένη απόκριση στα σημεία διασταύρωσης.

Η συγκεκριμένη τοπολογία μοιάζει με μια συνάρτηση $y = F(x) + x$, όπου το δίκτυο χρειάζεται να μάθει μόνο τη συνάρτηση-υπόλοιπο (residual function) $F(x)$, η οποία απλά προσθέτει πληροφορία, αντί να αλλάζει πλήρως την πληροφορία κάθε 2 επίπεδα.

Η μικρότερη έκδοση της αρχιτεκτονικής αυτής, η ResNet-50, χρησιμοποιεί 50 συνελικτικά επίπεδα και Κανονικοποίηση Δέσμης (Batch Normalization - BN), έχει 47 εκατομμύρια παραμέτρους και χρειάζεται $3{,}9$ δισεκατομμύρια πράξεις MACC σε κάθε πρόσθιο πέρασμα, ενώ επιτυγχάνει σφάλμα $top-5$ της τάξης του $6{,}7\%$ [63]. Αυτή η αρχιτεκτονική χρησιμοποιήθηκε στην παρούσα διπλωματική εργασία, όπως θα δούμε και στη συνέχεια.

Inception v3 και v4

Από τους δημιουργούς του GoogLeNet, υπήρξε περαιτέρω μελέτη και βελτιστοποίηση, με την πρώτη να έρχεται με τη δημοσίευση για το $InceptionV3$ [64], που περιείχε πολύτιμες συμβουλές για το σχεδιασμό και την επεξεργασία Συνελικτικών Νευρωνικών Δικτύων για καλύτερη αποδοτικότητα. Η δημοσίευση σχετικά με το InceptionV4 [65], μελετά τις θετικές επιδράσεις των residual συνδέσεων, ανάλογες με αυτές της αρχιτεκτονικής ResNet, σε αρχιτεκτονικές βασιζόμενες σε μονάδες Inception. Παρουσίασε την αρχιτεκτονική Inception-ResNet-v2, η οποία επιτυγχάνει $top-5$ σφάλμα $4{,}1\%$ στο σύνολο δεδομένων του διαγωνισμού ILSVCR. Όλες οι πρόσφατες αρχιτεκτονικές Inception χρησιμοποιούν κατά κόρον επίπεδα Κανονικοποίησης Δέσμης (Batch Normalization – BN).

3.4 FRAMEWORKS

Η εκπαίδευση ενός συνελικτικού νευρωνικού δικτύου αποτελεί χρονοβόρα διαδικασία, με μεγάλη ανάγκη για υπολογιστική ισχύ, ειδικά όταν η αρχιτεκτονική του αποκτά μεγαλύτερο βάθος, με περισσότερες παραμέτρους που πρέπει να υπολογιστούν. Η εκπαίδευση ενός μεγαλύτερου συνελικτικού δικτύου λοιπόν είναι μη πρακτική σε έναν τυπικό υπολογιστή, αξιοποιώντας απλά την υπολογιστική ισχύ της Κεντρικής Μονάδας Επεξεργασίας. Ευτυχώς, οι υπολογισμοί ενός συνελικτικού νευρωνικού δικτύου μπορούν εύκολα να αναπαρασταθούν ως πράξεις πινάκων ή τανυστών (tensors) που μπορούν να παραλληλοποιηθούν αποδοτικά. Συνεπώς, η παράλληλη υπολογιστική ισχύς των μονάδων επεξεργασίας γραφικών (GPUs) συντελεί στην ευκολότερη και πιο αποδοτική εκπαίδευση των νευρωνικών δικτύων. Γι' αυτό το λόγο και τα περισσότερα από τα δημοφιλή frameworks βαθείας μάθησης (deep learning) υποστηρίζουν επιτάχυνση με χρήση GPU.

Όπως η εκπαίδευση όπως αναφέρθηκε πιο πάνω απαιτεί μεγάλη υπολογιστική ισχύ, απαιτεί άλλο τόσο κόπο στο σχεδιασμό κάθε φορά μιας νέας αρχιτεκτονικής. Ευτυχώς, υπάρχει μια πληθώρα από frameworks που σκοπό έχουν την διευκόλυνση της διαδικασίας. Κάθε κοινότητα μηχανικών ακολουθεί την τάση της σύγχρονης τεχνολογίας και αναπτύσσει τα δικά της εργαλεία που προσφέρουν ένα υψηλότερο επίπεδο αφαίρεσης, και απλοποιούν τις δυνητικά δύσκολες προγραμματιστικές διεργασίες. Έτσι, κάθε framework είναι διαφορετικό, κατασκευασμένο με διαφορετικό σκοπό, και προσφέρει ένα μοναδικό εύρος χαρακτηριστικών.

Tensorflow

Μια από τις πιο δημοφιλείς βιβλιοθήκες, το Tensorflow [69] αναπτύχθηκε από την ομάδα Google Brain και αποτελεί προϊόν ανοικτού κώδικα από το 2015. Είναι μια βιβλιοθήκη που βασίζεται στη γλώσσα Python και είναι ικανή να τρέξει σε πολλαπλές CPUs και GPUs. Είναι διαθέσιμη σε όλες τις πλατφόρμες, τόσο σταθερές όσο και κινητές. Έχει επίσης υποστήριξη για άλλες γλώσσες όπως C++ και R, και μπορεί να χρησιμοποιηθεί άμεσα για δημιουργία μοντέλων βαθείας μάθησης, ή χρησιμοποιώντας βιβλιοθήκες-ενθυλακωτές (wrapper).

Theano

Μια από τις πρώτες βιβλιοθήκες βαθείας μάθησης, το Theano [70] είναι επίσης βασισμένο σε Python και πολύ καλό όταν η διεργασία αφορά αριθμητικούς υπολογισμούς σε CPU και GPU. Ακριβώς όπως και το Tensorflow, το Theano είναι μια χαμηλού επιπέδου βιβλιοθήκη, όπου μπορεί να χρησιμοποιηθεί άμεσα ή με wrappers για να διευκολυνθεί η διαδικασία. Ωστόσο, δεν είναι πολύ καλά κλιμακώσιμη, σε αντίθεση με άλλα frameworks, και υπολείπεται από υποστήριξη πολλαπλών GPUs.

Keras

Ενώ τα Theano και Tensorflow αποτελούν πολύ καλές βιβλιοθήκες βαθείας μάθησης, η δημιουργία μοντέλων σε αυτές άμεσα μπορεί να είναι πρόκληση, καθώς είναι βιβλιοθήκες χαμηλού επιπέδου. Για να αντιμετωπίσει αυτή τη πρόκληση, το Keras [71] κατασκευάστηκε ως μια απλοποιημένη διεπαφή για δημιουργία αποδοτικών νευρωνικών δικτύων. Μπορεί να

παραμετροποιηθεί για να συνεργαστεί είτε με το Theano είτε με το Tensorflow. Είναι γραμμένο σε Python, πολύ ελαφρύ και εύκολο στη μάθηση. Παρά τη σύντομη «ζωή» του, έχει πολύ καλό documentation.

Caffe

Κατασκευασμένο με γνώμονα την έκφραση, την ταχύτητα και τον αρθρωτό σχεδιασμό, το Caffe [72] είναι μια από τις πρώτες βιβλιοθήκες βαθείας μάθησης, προγραμματισμένο κυρίως από το κέντρο μάθησης και όρασης του Berkeley (Berkeley Vision and Learning Center – BLVC). Είναι μια βιβλιοθήκη της C++, που έχει επίσης μια διεπαφή σε Python, και η κύρια εφαρμογή της είναι η μοντελοποίηση Συνελικτικών Νευρωνικών Δικτύων. Ένα από τα μεγαλύτερα πλεονεκτήματά του είναι το Caffe Model Zoo, που παρέχει ένα πλήθος από προ-εκπαιδευμένα δίκτυα, έτοιμα για άμεση χρήση.

DeepLearning4j

Το DeepLearning4j (ή DL4J για συντομία) [73] είναι ένα δημοφιλές framework βαθείας μάθησης, ανεπτυγμένο σε Java, και υποστηρίζει επίσης και άλλες γλώσσες που τρέχουν σε JVM (όπως Groovy, Scala, Kotlin, Jython κ.ά.). Βρίσκει ευρεία χρήση ως εμπορική πλατφόρμα, εστιασμένη σε βιομηχανικές λύσεις κατανεμημένης βαθείας μάθησης. Το πλεονέκτημα της χρήσης του DL4J είναι πως μπορεί να συνδυαστεί η δύναμη όλου του Java οικοσυστήματος για να εκτελεστεί αποδοτική βαθεία μάθηση, καθώς μπορεί να υλοποιηθεί πάνω από δημοφιλή εργαλεία Big Data, όπως τα Apache Hadoop και Apache Spark.

MXNet

Το MXNet [74] είναι ένα από τα frameworks που υποστηρίζει τις περισσότερες γλώσσες, με υποστήριξη για γλώσσες όπως R, Python, C++ και Julia. Αυτό προσφέρει μεγάλη βοήθεια επειδή αν κάποιος είναι ήδη γνώστης οποιασδήποτε εκ των υποστηριζόμενων γλωσσών, δεν θα χρειαστεί να μάθει νέα στοιχεία για να εκπαιδεύσει τα μοντέλα βαθείας μάθησης που επιθυμεί. Το back-end του είναι γραμμένο σε C++ και Cuda , και μπορεί να διαχειριστεί τη μνήμη του μόνο του, όπως και το Theano, με τον κατάλληλο συλλέκτη απορριμμάτων. Είναι επίσης δημοφιλές καθώς κλιμακώνει πολύ καλά και μπορεί να εκτελεστεί σε πολλαπλές GPUs και υπολογιστές, κάτι που το καθιστά ιδιαίτερα χρήσιμο σε επιχειρήσεις. Είναι ένας από τους λόγους που το Amazon το έχει ορίσει ως βασική του βιβλιοθήκη για Βαθεία Μάθηση.

Microsoft Cognitive Toolkit

Το Microsoft Cognitive Toolkit , ή όπως ήταν πρότερα γνωστό με το ακρωνύμιο CNTK, είναι ένα πακέτο εργαλείων ανοιχτού κώδικα της Microsoft για εκπαίδευση μοντέλων βαθείας μάθησης. Είναι βελτιστοποιημένο σε μεγάλο βαθμό, και έχει υποστήριξη για γλώσσες όπως Python και C++. Όντας γνωστό για την αποδοτική του χρήση πόρων, είναι εύκολο να υλοποιηθούν σε αυτό μοντέλα Ενισχυτικής Μάθησης (Reinfircement Learning) ή Γεννητικής Ανταγωνιστικής Μάθησης (Generative Adversarial Networks - GANs). Είναι σχεδιασμένο για υψηλή κλιμακωσιμότητα και επίδοση, και προσφέρει μεγαλύτερη επιτάχυνση σε σύγκριση με άλλες βιβλιοθήκες όπως το Theano και το Tensorflow, όταν εκτελείται σε πολλαπλούς υπολογιστές.

MatConvNet

Το MatConvNet είναι μια εργαλειοθήκη (Toolbox) του περιβάλλοντος MATLAB, που υλοποιεί Συνελικτικά Νευρωνικά Δίκτυα για εφαρμογές όρασης υπολογιστών. Δημιουργήθηκε από τους Andrea Vedaldi και Karel Lenc, και κυκλοφόρησε στο GitHub το Δεκέμβριο του 2014, με αντίστοιχη δημοσίευση το 2015 [75]. Η βιβλιοθήκη αυτή μπορεί να βρεθεί και να ληφθεί δωρεάν από την κεντρική ιστοσελίδα του MatConvNet [76], η οποία παρέχει και το εγχειρίδιο [77] για τη σωστή χρήση του.

Η βιβλιοθήκη MatConvNet παρέχει απλές εντολές MATLAB για δημιουργία δομικών μονάδων (blocks) όπως blocks συνέλιξης, κανονικοποίησης και pooling, οι οποίες μπορούν να συνδυαστούν για να δημιουργήσουν αρχιτεκτονικές Συνελικτικών Νευρωνικών Δικτύων. Όπως και οι περισσότερες βιβλιοθήκες για Συνελικτικά Νευρωνικά Δίκτυα, προσφέρει μια πολύ αποδοτική εφαρμογή για το πρόβλημα της εκπαίδευσης από τεράστιο όγκο δεδομένων, συχνά από εκατομμύρια εικόνες. Αυτό το πετυχαίνει μέσα από μια πληθώρα βελτιστοποιήσεων και, κατά κύριο λόγο, μέσω της υποστήριξης υπολογισμών σε επεξεργαστές γραφικών (GPUs).

Οι εντολές του MatConvNet είναι βελτιστοποιημένες τόσο για υπολογισμούς σε CPU όσο και σε GPU, με υλοποιήσεις γραμμένες σε C++ και CUDA, ενώ η εγγενής υποστήριξη του MATLAB για υπολογισμούς σε GPU επιτρέπει την συγγραφή νέων μπλοκ σε κώδικα MATLAB, διατηρώντας την υπολογιστική απόδοση. Σε σύγκριση με τη συγγραφή νέων συνιστωσών ενός Συνελικτικού Νευρωνικού Δικτύου σε γλώσσες χαμηλότερου επιπέδου, αποτελεί μια απλούστευση που επιταχύνει σημαντικά την δοκιμή νέων ιδεών.

Το MatConvNet έχει μια απλή φιλοσοφία σχεδιασμού. Αντί να ενθυλακώνει τα Συνελικτικά Νευρωνικά Δίκτυα μέσα σε πολύπλοκα στρώματα λογισμικού, εκθέτει απλές συναρτήσεις για τον υπολογισμό των βασικών δομικών μονάδων του δικτύου, όπως γραμμική συνέλιξη ή ReLU, κατευθείαν ως εντολές MATLAB. Περιέχει λοιπόν τα εξής στοιχεία:

- **<u>Υπολογιστικά μπλοκ Συνελικτικών Νευρωνικών Δικτύων</u>**. Ένα σύνολο από βελτιστοποιημένες ρουτίνες που υπολογίζουν βασικά δομικά μπλοκ. Για παράδειγμα, ένα μπλοκ συνέλιξης υλοποιείται από την εντολή `y = vl_nnconv(x, f, b)`, όπου `x` μια εικόνα, `f` ένας πυρήνας – φίλτρο, και `b` ένα διάνυσμα από πολώσεις (bias). Οι παράγωγοι υπολογίζονται ως: `[dzdx, dzdf, dzdb] = vl_nnconv(x, f, b, dzdy)` όπου `dzdy` είναι η παράγωγος της εξόδου του νευρωνικού δικτύου ως προς την έξοδο `y` της συνέλιξης.
- **<u>Ενθυλακωτές (wrappers) Συνελικτικών Νευρωνικών Δικτύων</u>**. Παρέχεται ένας απλός wrapper, που καλείται μέσω της εντολής `vl_simplenn`, που υλοποιεί ένα Συνελικτικό Δίκτυο με γραμμική τοπολογία (μια αλυσίδα από δομικά μπλοκ). Επίσης παρέχεται ένας πιο ευέλικτος wrapper, που υποστηρίζει δίκτυα με πιο αυθαίρετες τοπολογίες, έχοντας τη λογική της δομής των κατευθυνόμενων ακυκλικών γράφων (Directed Acyclic Graphs – DAG), μέσα από την MATLAB κλάση `dagnn.DagNN`.
- **<u>Παραδείγματα εφαρμογών</u>**. Παρέχονται επίσης αρκετά παραδείγματα εκμάθησης Συνελικτικών Νευρωνικών Δικτύων με στοχαστική κατάβαση δυναμικού (stochastic

gradient descent – SGD) και χρήση CPU ή GPU, πάνω σε δεδομένα MNIST, CIFAR-10 και ImageNet.

Παρέχονται επίσης προ-εκπαιδευμένα μοντέλα δημοφιλών αρχιτεκτονικών Συνελικτικών Δικτύων, για διαφορετικά έργα. Πιο συγκεκριμένα, για το έργο της Κατηγοριοποίησης εικόνων του διαγωνισμού ImageNet ILSVRC παρέχονται τα μοντέλα:

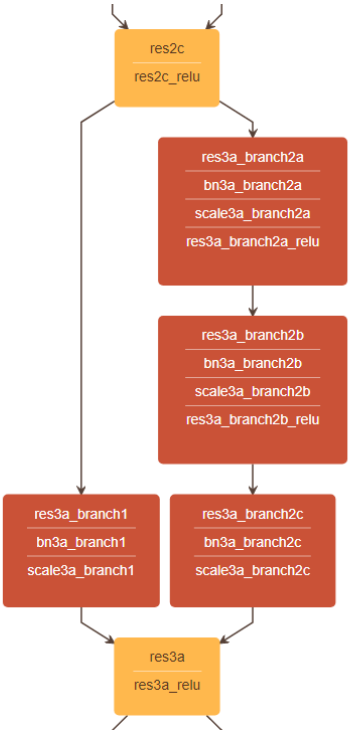
- ResNet (-50, -101, -152)
- GoogLeNet
- VGG-VD (Very Deep) (-16 , -19)
- VGG-S,M,F (-f, -m, -s, -m-2048, -m-1024, -m-128)
- Caffe Refernce
- AlexNet

Ο Πίνακας 1 αποτελεί μια σύνοψη για την απόδοση των μοντέλων που παρέχονται, για το validation σύνολο δεδομένων του ILSVRC 2012. Η ταχύτητα εκτίμησης μετρήθηκε σε έναν υπολογιστή με 12 πυρήνες, που χρησιμοποιούσε μία NVIDIA Titan X, Matlab R2015b, και τη βιβλιοθήκη CuDNN v5.1.

Πίνακας 1

Μοντέλο	Χρονολογία	Top-1 σφάλμα	Top-5 σφάλμα	Εικόνες/sec
resnet-50-dag	2015	24.6	7.7	396.3
resnet-101-dag	2015	23.4	7.0	247.3
resnet-152-dag	2015	23.0	6.7	172.5
matconvnet-vgg-verydeep-16	2014	28.3	9.5	200.9
vgg-verydeep-19	2014	28.7	9.9	166.2
vgg-verydeep-16	2014	28.5	9.9	200.2
googlenet-dag	2014	34.2	12.9	770.6
matconvnet-vgg-s	2013	37.0	15.8	586.2
matconvnet-vgg-m	2013	36.9	15.5	1212.5
matconvnet-vgg-f	2013	41.4	19.1	2482.7
vgg-s	2013	36.7	15.3	560.1
vgg-m	2013	37.3	15.9	1025.1
vgg-f	2013	41.1	18.8	1118.9
vgg-m-128	2013	40.8	18.4	1031.3
vgg-m-1024	2013	37.8	16.1	958.5
vgg-m-2048	2013	37.1	15.8	984.2
matconvnet-alex	2012	41.8	19.2	2133.3
caffe-ref	2012	42.6	19.7	1071.7
caffe-alex	2012	42.6	19.6	1379.8

Η πολύπλοκη δομή της αρχιτεκτονικής ResNet απαιτεί τον ενθυλακωτή DagNN του MatConvNet, καθώς το κύριο δομικό του μπλοκ (residual) δεν μπορεί να εκφραστεί ως απλή σύνδεση μπλοκ του ενός μετά το άλλο. Ο SimpleNN wrapper επιτρέπει μόνο παράθεση μπλοκ σε αυτή τη διάταξη, ενώ το residual block θα πρέπει να αποκτήσει την ακόλουθη δομή:



Εικόνα 14: Τμήμα της αρχιτεκτονικής ResNet-50. Το σχεδιάγραμμα παρήχθη με το εργαλείο Netscope [78] [79]

Η κύρια συνάρτηση του MatConvNet που επιτρέπει την εκπαίδευση ενός δικτύου με τη χρήση του DagNN wrapper είναι η `cnn_train_dag.m`. Αποτελεί μια πολύπλοκη συνάρτηση που δέχεται ως είσοδο τις παραμέτρους εκπαίδευσης που μπορεί να ορίσει ο χρήστης, όπως ρυθμό εκπαίδευσης, ορμή, εξασθένηση βαρών, πλήθος GPUs, την δομή του δικτύου, και σύνολο δεδομένων εικόνων για εκπαίδευση και επαλήθευση (training set και validation set αντίστοιχα).

Torch

Το Torch [80] [81] είναι μια βιβλιοθήκη μηχανικής μάθησης ανοιχτού κώδικα, που υπάρχει από το 2000. Δημιουργός του είναι ο Ronan Collobert, ο οποίος πλέον είναι Ερευνητικός Επιστήμονας στο Facebook. Έχουν κυκλοφορήσει 4 βασικές εκδόσεις, όλες σε μονό αριθμό, με την τρέχουσα έκδοση να έχει τον αριθμό 7. Είναι ανεπτυγμένη σε ποικίλες γλώσσες, αρχικά σε C, C++, πλέον σε μια γλώσσα προγραμματισμού που ονομάζεται Lua [82], διατηρώντας τις πραγματικές υλοποιήσεις των βιβλιοθηκών σε C.

Η Lua έχει ως σκοπό να χρησιμοποιηθεί ως μια ισχυρή και ελαφριά γλώσσα σεναρίων (scripting language) για κάθε πρόγραμμα που τη χρειάζεται. Είναι υλοποιημένη ως μια βιβλιοθήκη, γραμμένη σε καθαρή C. Σύμφωνα με την ιστοσελίδα της:

> *Η Lua συνδυάζει απλή διαδικαστική (procedural) σύνταξη με ισχυρές δομές περιγραφής δεδομένων που βασίζονται σε προσεταιριστικούς πίνακες (associative arrays) και επεκτάσιμη σημασιολογία (extensible semantics). Η Lua έχει δυναμικό σύστημα τύπων (dynamic typing), τρέχει ερμηνεύοντας bytecode για μια εικονική μηχανή που βασίζεται σε καταχωρητές, και έχει αυτόματη διαχείριση μνήμης με σταδιακή συλλογή απορριμμάτων*

(incremental garbage collection), που την καθιστά ιδανική για παραμετροποίηση, σχεδιασμό, και γρήγορη προτυποποίηση.

Η Lua παρέχει καλή υποστήριξη για αντικειμενοστραφή (object-oriented), συναρτησιακό (functional), ακόμα και προγραμματισμό χειρισμού γεγονότων (event-driven). Ο κύριος τύπος της είναι ο πίνακας (table), που υλοποιεί προσεταιριστικούς πίνακες (arrays) με ένα πολύ αποδοτικό τρόπο. Ένας προσεταιριστικός πίνακας είναι ένας πίνακας ο οποίος μπορεί να δεικτοδοτηθεί (indexed) όχι μόνο με ακεραίους, αλλά και με συμβολοσειρές ή οποιαδήποτε άλλη τιμή της γλώσσας. Οι πίνακες δεν έχουν προκαθορισμένο μέγεθος, μπορούν να μεταβάλουν το μέγεθός τους, και μπορούν να χρησιμοποιηθούν ως «εικονικοί πίνακες» πάνω σε έναν άλλο πίνακα, για να προσομοιώσουν διάφορες αντικειμενοστραφείς ιδεολογικές δομές. Παρά το γεγονός ότι οι πίνακες αποτελούν τη μοναδική δομή δεδομένων της Lua, είναι πολύ ισχυροί. Οι πίνακες χρησιμοποιούνται για να αναπαρασταθούν κανονικοί πίνακες, πίνακες συμβόλων, σύνολα, εγγραφές, ουρές, και άλλες δομές, με ένα απλό, ομοιόμορφο, και αποδοτικό τρόπο.

Το Torch, βασίζεται στην κλάση του «Tensor», η οποία επεκτείνει το βασικό σύνολο τύπων της Lua, παρέχοντας έναν αποδοτικό τύπο πίνακα πολλών διαστάσεων. Τα περισσότερα πακέτα (packages) του Torch ή πακέτα τρίτων που εξαρτώνται από αυτό, βασίζονται στη κλάση «Tensor» για να αναπαραστήσουν σήματα, εικόνες, βίντεο, επιτρέποντας να συνδυαστούν πολλές βιβλιοθήκες μαζί, έχοντας ένα κοινό γνώμονα. Παρέχει πολλές κλασικές λειτουργίες (όπως πράξεις γραμμικής άλγεβρας), υλοποιημένη αποδοτικά σε C, αξιοποιώντας εντολές SSE (Streaming SIMD Extensions) σε επεξεργαστές Intel, ενώ υποστηρίζει και εντολές OpenMP και υπολογισμούς σε CUDA GPU.

Χάρη στη προαναφερθείσα γλώσσα σεναρίων Lua, το Torch είναι πολύ γρήγορο και εύκολο στην επέκτασή του, με τα πακέτα που παρέχονται από τον package manager της Lua, τον `luarocks`. Στην τρέχουσα έκδοσή του, το Torch έχει 8 προεγκατεστημένα πακέτα:

- **torch:** Το βασικό πακέτο. Παρέχει τις κλάσεις των Tensors (FloatTensor, DoubleTensor, IntTensor, CudaTensor, <…>Tensor), εύκολη σειριοποίηση αρχείων και άλλες βασικές λειτουργίες
- **lab / plot:** Αυτά τα δύο πακέτα παρέχουν τυποποιημένες συναρτήσεις που μοιάζουν με τις αντίστοιχες της Matlab, για δημιουργία, μετασχηματισμό και σχεδιασμό γραφικών παραστάσεων
- **qt:** Πλήρη σύνδεση μεταξύ της βιβλιοθήκης Qt και της Lua, για εύκολη ανάπτυξη διαδραστικών demos, εκτελέσιμων σε Windows, Linux και Mac.
- **nn:** Το πακέτο nn παρέχει ένα σύνολο από τυποποιημένες ενότητες (modules) νευρωνικών δικτύων, καθώς και ένα σύνολο από ενότητες-δοχεία (container modules) που μπορούν να χρησιμοποιηθούν για να καθοριστούν αυθαίρετοι κατευθυνόμενοι (ακυκλικοί ή και μη) γράφοι. Περιγράφοντας την δομή του γράφου, συνδέοντας δομικά στοιχεία μεταξύ τους, αποφεύγουμε τη χρήση ενδιάμεσου μεταγλωττιστή ή αναλυτή, και έχουμε την ευχέρεια να δημιουργήσουμε οποιαδήποτε αρχιτεκτονική επιθυμούμε. Κάθε module παρέχει τυποποιημένες συναρτήσεις για υπολογισμό της εξόδου, και για την υλοποίηση του back propagation.

- **image:** Ένα πακέτο επεξεργασίας εικόνας. Παρέχει όλες τις τυποποιημένες συναρτήσεις επεξεργασίας: αποθήκευση/φόρτωση εικόνων, αλλαγή μεγέθους, περιστροφή, αλλαγή colorspace, συνέλιξη, φιλτράρισμα, κ.ά.
- **optim:** Ένα πακέτο που παρέχει βελτιστοποιημένες υλοποιήσεις για αλγορίθμους εκπαίδευσης, όπως πιο απότομη κατάβαση (steepest descent), συζυγείς κλίσεις (conjugate gradient), και τον αλγόριθμο BFGS (Broyden-Fletcher-Goldfarb-Shamo), με περιορισμένη χρήση μνήμης.
- **unsup:** Περιέχει ποικίλους αλγορίθμους μη-επιβλεπόμενης μάθησης, όπως K-μέσων (K-means), αραιά κωδικοποίηση (sparse coding) και αυτό-κωδικοποιητές (auto encoders)

Εκτός από τα πακέτα αυτά, είναι διαθέσιμη μια συνεχώς αναπτυσσόμενη λίστα από πακέτα τρίτων. Κάποια πακέτα, που χρησιμοποιήθηκαν στο πειραματικό μέρος της παρούσας διπλωματικής εργασίας, είναι:

- **paths:** το πακέτο αυτό παρέχει φορητές (portable) συναρτήσεις και μεταβλητές για χειρισμό του συστήματος αρχείων, με κύριες λειτουργίες να αποτελούν συναρτήσεις για χειρισμό και επεξεργασία ονομάτων αρχείων και καταλόγων, αλλά και διαδρομές για γνωστούς καταλόγους.
- **ffi:** αποτελεί συνδετικό κρίκο μεταξύ του torch και της βιβλιοθήκης FFI της Lua. Η βιβλιοθήκη αυτή επιτρέπει την απευθείας κλήση συναρτήσεων C και χρήση δομών της C από κώδικα Lua.
- **cunn:** ενώ το προ-εγκατεστημένο πακέτο nn περιέχει βασικές ενότητες για αξιοποίησή τους στη δημιουργία αρχιτεκτονικών νευρωνικών δικτύων, είναι υλοποιημένες κυρίως για να εκτελούνται στον επεξεργαστή. Με το πακέτο cunn έχουμε πρόσβαση σε υλοποιήσεις των ενοτήτων αυτών σε cuda, για επιτάχυνση της εκπαίδευσης και της δοκιμής του εκάστοτε νευρωνικού δικτύου.
- **cudnn:** το πακέτο αυτό περιέχει, σε αντιστοιχία με το cunn, υλοποιήσεις και συναρτήσεις μετατροπής των ενοτήτων από μορφή nn σε υλοποιήσεις της βιβλιοθήκης cuDNN της NVIDIA
- **gnuplot:** ένα ακόμα πακέτο για plotting, χρησιμοποιείται για να απεικονίσει αντικείμενα Tensor, και χρησιμοποιεί το gnuplot για να εμφανίσει δεδομένα.

Συγκεκριμένα για το βασικό πακέτο nn που μας ενδιαφέρει, ακολουθεί ένα παράδειγμα χρήσης του, όπου περιγράφεται μια μικρή αρχιτεκτονική, ενός multi-layer perceptron:

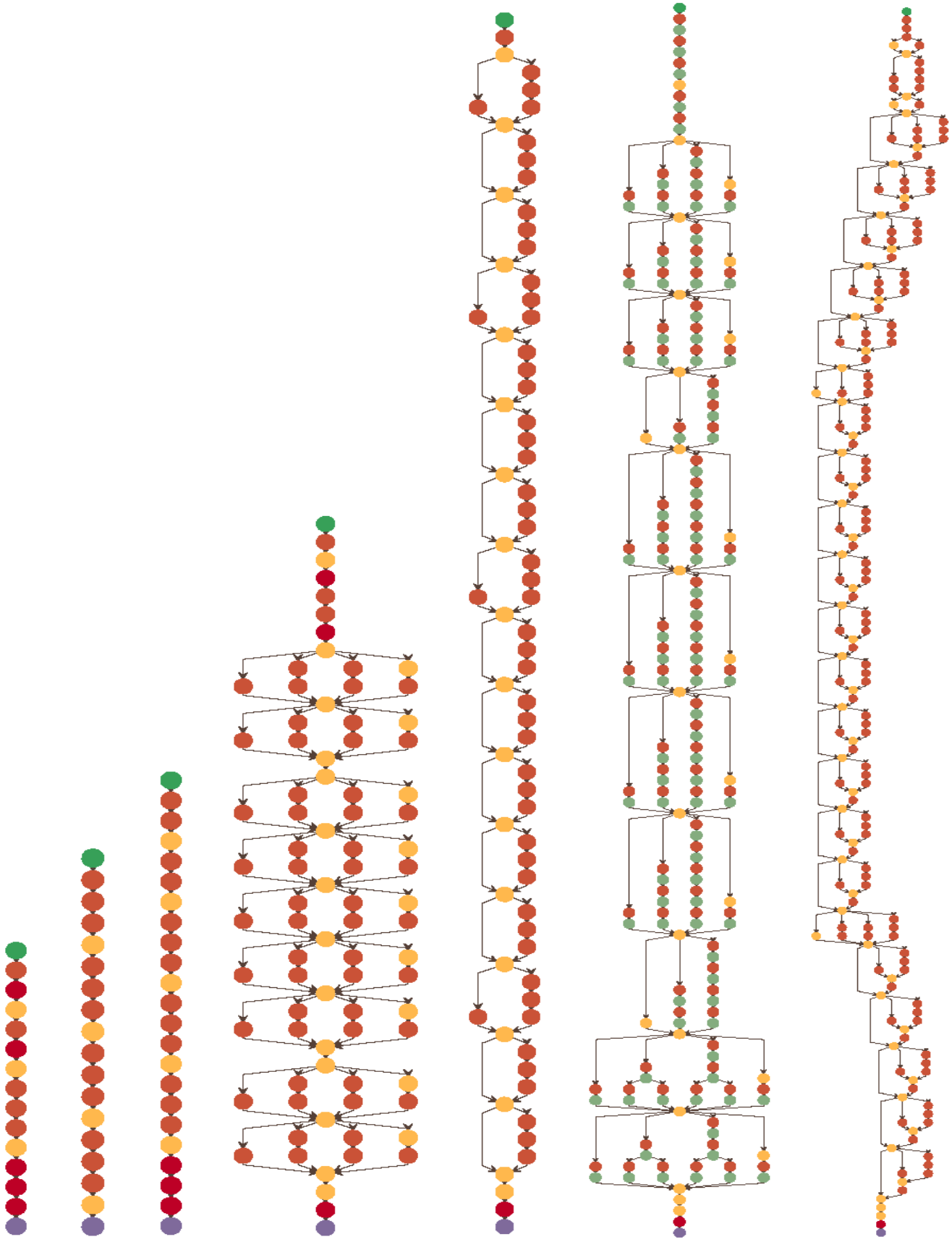
```
mlp = nn.Sequential()
mlp:add(nn.Linear(100,1000))
mlp:add(nn.Tanh())
mlp:add(nn.Linear(1000,10))
mlp:add(nn.SoftMax())
```

Ο παραπάνω κώδικας περιγράφει μια αρχιτεκτονική όπου στο πρώτο επίπεδο έχουμε ένα επίπεδο με 10 εισόδους και 1000 νευρώνες εξόδου, όπου σε κάθε νευρώνα εξόδου εφαρμόζεται μια activation function υπερβολικής εφαπτομένης (tanh). Έπειτα προσθέτει ένα νέο επίπεδο νευρώνων, με 1000 νευρώνες εισόδου και 10 νευρώνες εξόδου, των οποίων οι

έξοδοι περνάνε από μια συνάρτηση SoftMax, αλλάζοντας το πλάτος των εισόδων έτσι ώστε να βρίσκονται στο εύρος [0,1] και να αθροίζονται στο 1.

Ο κώδικας για την εκπαίδευση (ένα πέρασμα – «εποχή») είναι εξίσου απλός. Έχοντας την είσοδο X και τις σωστές κατηγορίες T, υπολογίζουμε τη παράγωγο κάποιου σφάλματος E ως προς την έξοδο Y $\left(\frac{dE}{dY}\right)$ με τον εξής τρόπο:

```
Y = mlp:forward(X)              -- υπολογισμός Y = f(X)
E = loss:forward(Y,T)           -- όπου loss μια συνάρτηση σφάλματος E = l(Y,T)
dE_dY = loss:updateGradInput(Y,T)       -- υπολογισμός του dE/dY = dl(Y,T)/dY
dE_dX = mlp:updateGradInput(X,dE_dY)    -- back-propagate τις παραγώγους,
                                           μέχρι και dE/dX
mlp:accGradParameters(X,dE_dY)        -- υπολογισμός των παραγώγων
                                          ως προς τα βάρη: dE/dW
```



Εικόνα 15: Από αριστερά προς τα δεξιά: AlexNet, Network-in-Network, Vgg-16, GoogLeNet, ResNet-50, InceptionV3, Inception-ResNet-v2

3.5 ΜΕΤΑΦΕΡΟΜΕΝΗ ΜΑΘΗΣΗ ΚΑΙ ΛΕΠΤΟΣ ΣΥΝΤΟΝΙΣΜΟΣ

Στην παρούσα διπλωματική εργασία χρησιμοποιήθηκε μια σημαντική στρατηγική μηχανικής μάθησης, η λεγόμενη μεταφερόμενη μάθηση (Transfer Learning, [83]). Οι κοινοί αλγόριθμοι μηχανικής μάθησης συνήθως αντιμετωπίζουν μεμονωμένες εργασίες. Από την άλλη, η μεταφορά μάθησης επιχειρεί να αλλάξει αυτή τη λογική, αναπτύσσοντας μεθόδους που μας επιτρέπουν να μεταφέρουμε τη γνώση που έχουμε λάβει σε ένα ή περισσότερα προβλήματα – εργασίες (πηγή - source) και να την χρησιμοποιήσουμε για να βελτιώσουμε την μάθηση σε ένα συσχετιζόμενο πρόβλημα – (στόχος – target). Ο στόχος του Transfer Learning είναι να βελτιωθεί η μάθηση σε ένα πρόβλημα-στόχο, αξιοποιώντας τη γνώση από το πρόβλημα – πηγή. Η Pratt υπήρξε η πρώτος που παρουσίασε την ιδέα της μεταφοράς γνώσης, μέσω της πρότασης του αλγορίθμου μεταφοράς με βάση τη διακριτική ευχέρεια (Discriminability-Based Transfer – DBT [84]) το 1993. Ο αλγόριθμος DBT χρησιμοποιεί ένα μέτρο πληροφορίας για να εκτιμήσει τη χρησιμότητα των υπερεπιπέδων που ορίζονται από τα πηγαία βάρη στο δίκτυο-στόχο, και μεταβάλλει τη κλίμακα των πλατών των βαρών στο νέο δίκτυο κατάλληλα. Η έρευνα πάνω στη μεταφορά γνώσης προσελκύει όλο και περισσότερη προσοχή από το 1995, με διαφορετικά ονόματα: learning to learn, knowledge transfer, inductive transfer, multi-task learning, knowledge consolidation, context-sensitive learning, knowledge-based inductive bias, meta learning και incremental/cumulative learning [85] [86] [87].

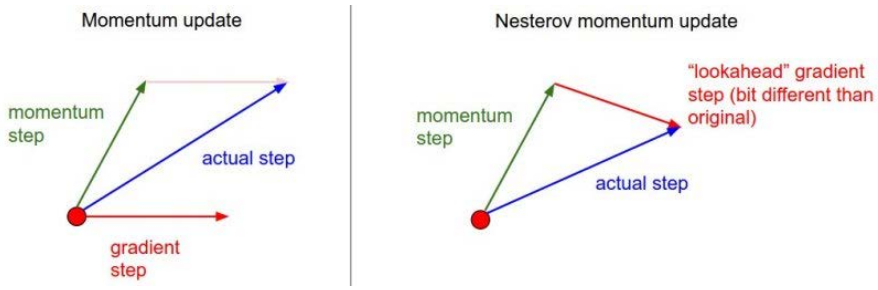
Στην πράξη, δεν γίνεται εκπαίδευση ολόκληρου Συνελικτικού Νευρωνικού Δικτύου με μεγάλο βάθος από την αρχή με τυχαία αρχικοποίηση. Αυτό συμβαίνει διότι είναι σχετικά σπάνιο να διατίθεται ένα σύνολο δεδομένων ικανοποιητικού μεγέθους που απαιτείται για το βάθος του τελικού νευρωνικού δικτύου. Για παράδειγμα, το ResNet είναι αρχικά εκπαιδευμένο στο σύνολο δεδομένων του ImageNet, ένα σύνολο δεδομένων που περιέχει πάνω από 1 εκατομμύριο εικόνες, κατηγοριοποιημένες σε 1000 κατηγορίες. Αντ' αυτού, αποτελεί κοινή τεχνική να χρησιμοποιούνται τα βάρη ενός υπάρχοντος νευρωνικού δικτύου προ-εκπαιδευμένο σε ένα μεγάλο σύνολο δεδομένων, όπως το ImageNet, είτε ως αρχικοποίηση, είτε σαν ένα σταθερό εξαγωγέα χαρακτηριστικών για το πρόβλημα που μας ενδιαφέρει. Σε περίπτωση που τα χρησιμοποιήσουμε σαν αρχικές τιμές, γίνεται λόγος για την στρατηγική Fine-Tuning (Λεπτός Συντονισμός / Τελειοποίηση) [88]. Η στρατηγική αυτή κάνει μικρο-ρυθμίσεις στα βάρη ενός ήδη εκπαιδευμένου Συνελικτικού Νευρωνικού Δικτύου, συνεχίζοντας τον αλγόριθμο BackPropagation.

Είναι εφικτό να αλλάξουν τα βάρη σε όλα τα στρώματα της αρχιτεκτονικής, ή μόνο σε ένα πιο υψηλού επιπέδου τμήμα του δικτύου, κρατώντας σταθερά κάποια από τα αρχικά στρώματα. Το γεγονός ότι τα αρχικά στρώματα ενός Συνελικτικού Δικτύου συνήθως εξαγάγουν πιο γενικά χαρακτηριστικά (ανιχνευτές ακμών ή ανιχνευτές χρωματικών κηλίδων) που μπορεί να είναι χρήσιμα σε πολλές εργασίες αποτελεί μια ώθηση στην επιλογή της δεύτερης εναλλακτικής. Τα μετέπειτα στρώματα των βαθιών αρχιτεκτονικών περιέχουν χαρακτηριστικά ολοένα και πιο σχετικά με τις λεπτομέρειες των κατηγοριών του συνόλου δεδομένων στο οποίο είναι εκπαιδευμένες.

3.6 ΟΡΜΗ NESTEROV

Σαν μέθοδος εκπαίδευσης σε μία από τις δύο εκπαιδεύσεις, όπως θα δούμε στην συνέχεια, χρησιμοποιείται η Στοχαστική Κατάβαση Δυναμικού (Stochastic Gradient Descent - SGD), με την ορμή Nesterov. Αυτό σημαίνει πως ο τρόπος της ενημέρωσης της ορμής του δικτύου γίνεται διαφορετικά, με τη μέθοδο SGD να ακολουθεί τα στοιχεία της μεθόδου NAG (Nesterov's Accelerated Gradient) [89]. Είναι μια μέθοδος που προσφέρει ισχυρότερες θεωρητικές εγγυήσεις σύγκλισης για κυρτές συναρτήσεις, και επίσης στην πράξη προσφέρει με συνέπεια ελάχιστα καλύτερα αποτελέσματα από τη κανονική ορμή.

Η βασική ιδέα πίσω από την ορμή Nesterov είναι πως, όταν το τρέχον διάνυσμα παραμέτρων είναι σε μια θέση x, τότε ξέρουμε πως θα μετακινηθεί εξαιτίας της παραμέτρου της ορμής κατά $mu * v$, όπου mu η παράμετρος της ορμής και v μια ποσότητα που εκφράζει την ταχύτητα, και υπολογίζεται κατά την εκπαίδευση, επηρεασμένη από την ορμή, την προηγούμενη ταχύτητα, και την μεταβολή των τιμών των βαρών του δικτύου. Με την μέθοδο Nesterov, γίνεται πρώτα η μετακίνηση του διανύσματος x, και έπειτα υπολογίζεται η διαφορά δυναμικού.



Εικόνα 16: Nesterov momentum

3.7 ΣΥΝΟΛΑ ΔΕΔΟΜΕΝΩΝ

Πολλές βιβλιογραφικές πηγές στην εργασία κατηγοριοποίησης εικόνων φαγητού χρησιμοποιούσαν φωτογραφίες που συγκεντρώθηκαν από την εκάστοτε ερευνητική ομάδα, είτε από το διαδίκτυο, είτε από τους ίδιους, την καθημερινότητα και το εργαστήριό τους.

Μερικές ομάδες πήραν την πρωτοβουλία και δημιούργησαν ωστόσο ολοκληρωμένα σύνολα δεδομένων, και τα άφησαν δημόσια προς χρήση από το κοινό. Στην ενότητα αυτή παρουσιάζονται μερικά από τα σύνολα δεδομένων της βιβλιογραφίας και στο τέλος μια επισκόπηση του συνόλου δεδομένων Food-101, το οποίο και χρησιμοποιήθηκε για την εκπαίδευση και αξιολόγηση του δικτύου της παρούσας διπλωματικής εργασίας.

Σύνολο δεδομένων PFID

Το σύνολο δεδομένων Pittsburgh Fast-Food Image Database (PFID) είναι ένα οπτικό σύνολο δεδομένων για 101 κατηγορίες φαγητού από 11 δημοφιλείς αλυσίδες εστιατορίων γρήγορου φαγητού. Αποτελείται από 4545 εικόνες, 606 στερεοσκοπικές εικόνες (ζεύγη εικόνων), 303 βίντεο τεμαχίων φαγητού σε 360^{o}, και 27 βίντεο από εθελοντές την ώρα που καταναλώνουν τα φαγητά. Προήλθε από έρευνα πάνω στην αναγνώριση φαγητών ταχυφαγείων για

διατροφολογική εκτίμηση, από τα ερευνητικά εργαστήρια της Intel [90], και είναι διαθέσιμο προς εξερεύνηση από την ιστοσελίδα τους [91].

Σύνολα δεδομένων UEC FOOD

Στην δημοσίευση των Matsuda, Hoashi και Yanai το 2012 [92] εισήγαγαν την εφαρμογή FoodCam, που είχε ως στόχο επίσης την αναγνώριση εικόνων φαγητού και την εκτίμηση περιεχόμενων διατροφικών στοιχείων. Η εφαρμογή ήταν σχεδιασμένη για εκτέλεση σε φορητές συσκευές (smartphones, tablets). Εισήγαγαν παράλληλα το σύνολο δεδομένων UEC FOOD 100, που αποτελούνταν από 100 κατηγορίες φαγητών, κυρίως της ιαπωνικής κουζίνας, με συνολικά 14300 εικόνες κατά προσέγγιση Οι κατηγορίες περιείχαν και πληροφορία bounding box, που όριζαν σε κάθε εικόνα πού βρίσκεται η κύρια κατηγορία τροφής, ενώ υπήρχε και έγγραφο που όριζε σε κάθε εικόνα (εφόσον υπήρχε) τις πολλαπλές κατηγορίες φαγητού που περιείχε. Με μετέπειτα δημοσιεύσεις τους το 2014 [93] [94] εισήγαγαν ένα μεγαλύτερο σύνολο δεδομένων, με 256 κατηγορίες φαγητών και συνολικά περίπου 31400 εικόνες φαγητού. Ωστόσο, οι κατηγορίες αυτές φαγητού ήταν κατά κανόνα άγνωστες και ασυνήθιστες για τα ελληνικά δεδομένα, και για το λόγο αυτό δεν χρησιμοποιήθηκε το σύνολο δεδομένων αυτό στην εκπαίδευση και αξιολόγηση του νευρωνικού δικτύου.

Σύνολο δεδομένων Food-11

Από το Πολυτεχνείο της Λωζάνης, και συγκεκριμένα από το Multimedia Signal Processing Group [95], αξιοποιήθηκε το σύνολο δεδομένων Food-11. Αποτελείται από 16643 εικόνες φαγητού, χωρισμένες σε 11 βασικές διατροφικές ομάδες: ψωμί, γαλακτοκομικά προϊόντα, επιδόρπια, αυγά, τηγανητά φαγητά, κρέας, μακαρόνια, ρύζι, θαλασσινά, σούπα, φρούτα/λαχανικά.

4 ΜΕΘΟΔΟΛΟΓΙΑ

4.1 ΠΕΡΙΒΑΛΛΟΝ ΥΛΟΠΟΙΗΣΗΣ

Για την εκπαίδευση του ΣΝΔ που παρουσιάζεται στη παρούσα διπλωματική εργασία χρησιμοποιήθηκε ένας προσωπικός υπολογιστής με τα εξής χαρακτηριστικά:

CPU	Intel Core i5-2500k @3.30GHZ
GPU	NVIDIA GTX970, 4GB GDDR5
RAM	8 GB DDR3
OS	Linux Ubuntu 16.04

Η πειραματική διαδικασία έγινε σε πρώτη φάση με χρήση της βιβλιοθήκης MatConvNet σε περιβάλλον MATLAB R2015b. Η έκδοση του MatConvNet που χρησιμοποιήθηκε ήταν 1.0-beta24. Σε δεύτερη φάση, έγινε εκπαίδευση του ίδιου δικτύου, όμως αυτή τη φορά με το πλαίσιο Torch7.

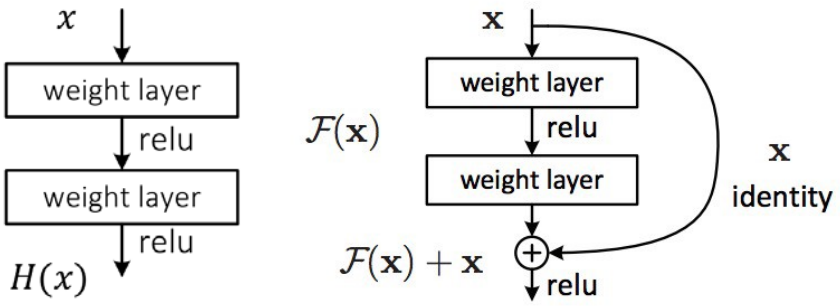
Καθώς η ανάπτυξη των μοντέλων αυτών και η εκπαίδευσή τους απαιτεί μεγάλη υπολογιστική ισχύ, ήταν σκόπιμη η αξιοποίηση της υπολογιστικής ισχύος των καρτών γραφικών της NVIDIA μέσω της αρχιτεκτονικής CUDA. Επιπλέον, η NVIDIA παρέχει μια βιβλιοθήκη από GPU-επιταχυνόμενες και βελτιστοποιημένες ρουτίνες για βαθιά νευρωνικά δίκτυα, όπως forward και back propagation, συνέλιξη, pooling, κανονικοποίηση, επίπεδα ενεργοποίησης, την cuDNN (NVIDIA CUDA Deep Neural Network Library). Και στις δύο εκπαιδεύσεις χρησιμοποιήθηκε η έκδοση 5 της βιβλιοθήκης cuDNN και η έκδοση 8.0 για την αρχιτεκτονική CUDA.

4.2 RESNET

Στην ενότητα αυτή ρίχνουμε μια πιο προσεκτική ματιά στο Συνελικτικό Νευρωνικό Δίκτυο ResNet, καθώς αυτό χρησιμοποιήθηκε στο πειραματικό μέρος της παρούσας διπλωματικής εργασίας.

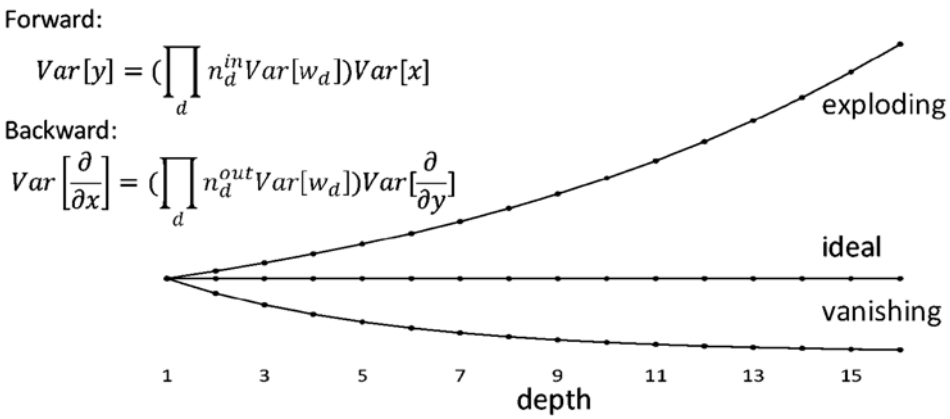
Όπως προαναφέρθηκε, η ResNet αρχιτεκτονική παρουσιάστηκε από τους Kaiming He κ.ά. από το Microsoft Research Asia (MSRA) το 2015, νικώντας κάθε έργο του διαγωνισμού ImageNet (Classification, Detection, Localization) καθώς και τις κατηγορίες Detection και Segmentation του διαγωνισμού COCO.

Η αρχιτεκτονική επιχειρεί να διευθύνει το πρόβλημα της εκπαίδευσης συνελικτικών νευρωνικών δικτύων μεγαλύτερου βάθους, αξιοποιώντας μια δομή υπολειπόμενης (residual) μάθησης. Έχει επιτύχει καλύτερα αποτελέσματα όσο αφορά την ακρίβεια, αυξάνοντας το βάθος της αρχιτεκτονικής, και ταυτόχρονα είναι εύκολο να βελτιστοποιηθεί, χάρη στη δομή αυτή.



Εικόνα 17: Residual block. Η F(x) είναι η υπολειπόμενη αντιστοίχιση (residual mapping), ενώ η x η αντιστοίχιση ταυτότητας (identity mapping). Η επιθυμητή αντιστοίχιση είναι η H(x) = F(x) + x

Το μεγαλύτερο πρόβλημα που εμφανίζεται όταν τοποθετούνται πολλά διαδοχικά επίπεδα συνέλιξης, είναι το πρόβλημα του μηδενισμού ή της εκρηκτικής αύξησης κλίσης (vanishing / exploding gradient, αντίστοιχα). Κατά το πρόσθιο (forward pass) όσο και το αντίστροφο (backward pass) πέρασμα μέσα στο νευρωνικό δίκτυο, γίνεται υπολογισμός με βάση τη κλίση των προηγούμενων επιπέδων, με πολλαπλασιαστικό τρόπο. Έτσι, αν υπάρξει ένα σφάλμα εκτός της ιδανικής τιμής, αυτό το σφάλμα θα είναι πιο έντονο σε κάθε πρόσθετο επίπεδο [66] [67] [68]:



Εικόνα 18 Το πρόβλημα του μηδενισμού και της εκρηκτικής αύξησης κλίσης

Αυτό το πρόβλημα λύθηκε εφαρμόζοντας ένα επίπεδο κανονικοποίησης δέσμης (Batch Normalization – BN) έπειτα από κάθε επίπεδο συνέλιξης, για κάθε μίνι-δέσμη (mini-batch). Αυτό επιτάχυνε σε μεγάλο βαθμό την εκπαίδευση, και καθιστά το δίκτυο λιγότερο ευαίσθητο στις τιμές αρχικοποίησης. Όμως οι ερευνητές του MSRA εστιάζουν σε ένα άλλο πρόβλημα, που αποκαλείται *υποβιβασμός*:

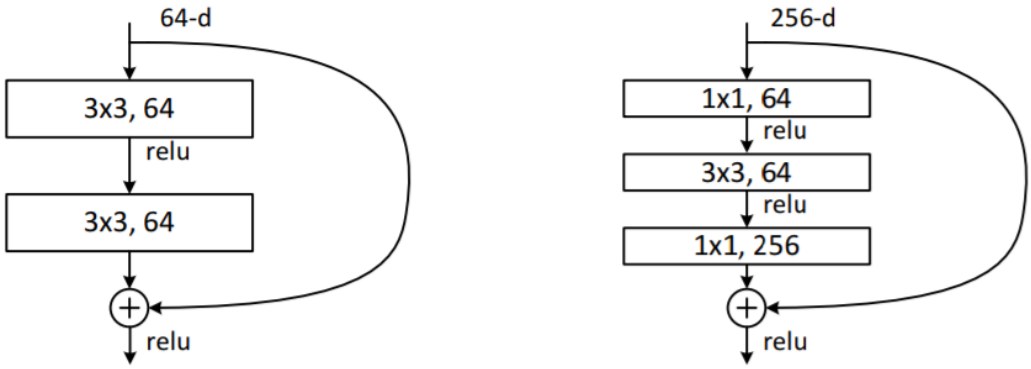
«…όταν το βάθος ενός δικτύου αυξάνεται, η ακρίβεια μπορεί να κορεστεί και έπειτα υποβιβάζεται γρήγορα. Αυτό δεν προκαλείται από υπερεκπαίδευση (overfitting), αλλά προσθέτοντας περισσότερα επίπεδα το μοντέλο πετυχαίνει υψηλότερο σφάλμα εκπαίδευσης. Ο υποβιβασμός της ακρίβειας εκπαίδευσης δείχνει πως δεν είναι το ίδιο εύκολο να βελτιστοποιηθούν όλα τα συστήματα.» [41]

Η λύση που βρήκαν για να αντιμετωπίσουν το πρόβλημα ήταν να μεταβάλλουν την αρχιτεκτονική, αφήνοντας το δίκτυο να μάθε μόνο μια υπολειπόμενη αντιστοίχιση (residual

mapping), και έπειτα να προσθέσουν την είσοδο στην έξοδο αυτή για να ανακατασκευάσουν την αρχική αντιστοίχιση.

Με αυτή τη δομή υπάρχουν δύο πιθανές περιπτώσεις κατά τη διάρκεια της μάθησης για την βελτιστοποίηση των βαρών: Αν η αντιστοίχιση ταυτότητας είναι η βέλτιστη ή κοντά στη βέλτιστη, είναι εύκολο να τεθούν τα βάρη στη τιμή 0 ή να βρεθούν μικρές μεταβολές. Αυτό οδηγεί σε μια καλή βελτιστοποίηση των βαρών, ακόμα και σε αρχιτεκτονικές με μεγάλο βάθος.

Ο σχεδιασμός των δικτύων ResNet είναι απλός, απλά πολύ βαθύς. Ακολουθεί ένα βασικό σχέδιο, όπου σχεδόν όλα τα επίπεδα συνέλιξης εφαρμόζουν φίλτρα $3x3$, και ανά τακτά σημεία υποδιπλασιάζεται το χωρικό μέγεθος, ενώ διπλασιάζεται το πλήθος των φίλτρων (σε πλήθος καναλιών), κρατώντας περίπου την ίδια πολυπλοκότητα σε κάθε επίπεδο. Υπάρχει, ωστόσο, μια διαφοροποίηση στις εκδόσεις της αρχιτεκτονικής με μεγαλύτερο βάθος. Χρησιμοποιείται ένα διαφορετικό δομικό μπλοκ, που ονομάζεται “bottleneck”, με σκοπό να βελτιωθεί ο χρόνος εκπαίδευσης. Για κάθε υπολειπόμενη συνάρτηση F, χρησιμοποιείται μια στοίβα από 3 συνελίξεις αντί για 2, και συγκεκριμένα εφαρμόζονται φίλτρα συνέλιξης $1x1, 3x3, 1x1$, όπου τα φίλτρα $1x1$ είναι υπεύθυνα για τη μείωση και έπειτα την επαναφορά των διαστάσεων, αφήνοντας το φίλτρο συνέλιξης $3x3$ με λιγότερες εισόδους και εξόδους. Στην Εικόνα 19 παρουσιάζεται ένα παράδειγμα, όπου τα δυο μπλοκ έχουν παρόμοια χρονική πολυπλοκότητα.



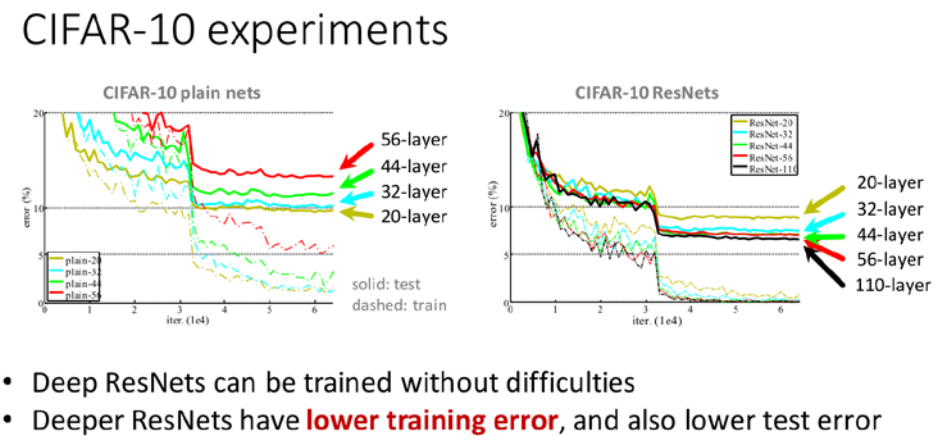
Εικόνα 19: Αριστερά, ένα απλό υπολειπόμενο μπλοκ του ResNet-34. Δεξιά, ένα bottleneck – μπλοκ των ResNet-50/101/152

Για το μοντέλο αυτό, βάζοντας μεγαλύτερο βάθος παίρνουμε πιο ακριβή αποτελέσματα. Για παράδειγμα, τα σφάλματα top-1 και top-5 για κάποιες από τις παραλλαγές του ResNet πάνω στο σύνολο δεδομένων ImageNet συνοψίζονται στον παρακάτω πίνακα:

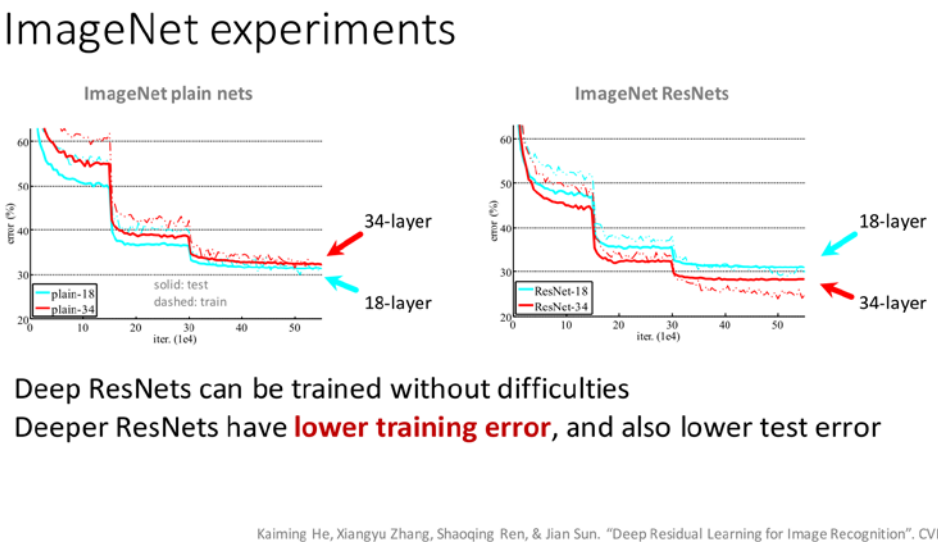
Μοντέλο	Βάθος	Σφάλμα top-1	Σφάλμα top-5
ResNet-34	34	24,19 %	7,40 %
ResNet-50	50	22,85 %	6,71 %
ResNet-101	101	21,75 %	6,05 %
ResNet-152	152	21,43 %	5,71 %

Οι ερευνητές έκαναν μια σύγκριση μεταξύ απλών συνελικτικών νευρωνικών δικτύων με ίδιο βάθος, χωρίς να αξιοποιούν τη δομή του residual block, και των αντίστοιχων ResNet, πάνω

στα σύνολα δεδομένων CIFAR-10 και ImageNet. Το σύνολο δεδομένων CIFAR-10 αποτελείται από 60000 έγχρωμες φωτογραφίες μεγέθους $32x32$, χωρισμένες σε 10 κατηγορίες.

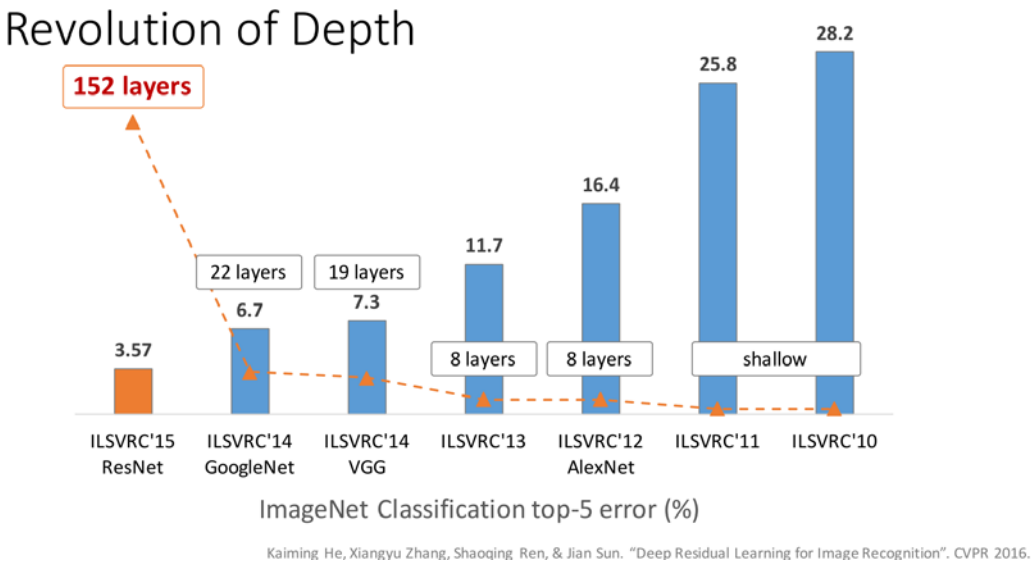


Εικόνα 20: Σύγκριση του σφάλματος εκπαίδευσης και validation στο σύνολο CIFAR-10 για απλά CNN (αριστερά) και ResNets (δεξιά)



Εικόνα 21: : Σύγκριση του σφάλματος εκπαίδευσης και validation στο σύνολο ImageNet για απλά CNN (αριστερά) και ResNets (δεξιά)

Αξίζει να σημειωθεί πως η επιλογή της αύξησης του βάθους της αρχιτεκτονικής έχει νόημα σαν επιλογή, κάτι που ενισχύεται όχι μόνο από τα σφάλματα για κάθε βάθος του πίνακα, αλλά και από ένα χρονολόγιο [41] των σφαλμάτων για την εργασία της κατηγοριοποίησης του διαγωνισμού ILSVRC, το οποίο περιέχει και δεδομένα για το βάθος της νικητήριας αρχιτεκτονικής:



Εικόνα 22: Η "Επανάσταση" του Βάθους στον διαγωνισμό ILSVRC ανά τα χρόνια, στο task της Κατηγοριοποίησης

Κατά την εκπαίδευση των μοντέλων, οι ερευνητές έκαναν τις εξής τρεις σημαντικές παρατηρήσεις:

- Μοντέλα που έχουν ως βάση βαθύτερες αρχιτεκτονικές τείνουν να παραγάγουν μικρότερα σφάλματα εκπαίδευσης, το οποίο γενικεύεται και στην φάση της επαλήθευσης. Αυτό δείχνει πως το πρόβλημα του υποβιβασμού έχει πλέον διευθετηθεί.
- Το top-1 σφάλμα ταξινόμησης μειώθηκε συγκρινόμενο με τις υπόλοιπες σύγχρονες (state of the art) αρχιτεκτονικές, γεγονός που υποδεικνύει πως η υπολειπόμενη μάθηση είναι αποδοτική σε βαθιά συστήματα.
- Για μικρότερο βάθος, τα αποτελέσματα των μοντέλων που εκπαιδεύτηκαν παρουσιάζουν παρόμοια αποτελέσματα με αυτά των απλών μοντέλων (χωρίς υπολειπόμενη μάθηση). Ωστόσο, η σύγκλιση σε ένα υπολειπόμενο (residual) δίκτυο είναι σημαντικά ταχύτερη.

Στην παρούσα διπλωματική εργασία, εκτιμάται η απόδοση του μοντέλου ResNet-50 σε σύγκριση με το μοντέλο ResNet-101 για κατηγοριοποίηση εικόνων γευμάτων, ενώ στην επόμενη σελίδα, ακολουθεί ένα σχηματικό που συγκρίνει τη δομή μερικών αρχιτεκτονικών που παρουσιάστηκαν νωρίτερα.

4.3 Σύνολο Δεδομένων Food-101

Για την αξιολόγηση των αποτελεσμάτων έγινε χρήση ένα σύνολο δεδομένων που δημιουργήθηκε στο πανεπιστήμιο ETH της Ζυρίχης το 2014, το λεγόμενο Food-101 [96]. Είναι ένα διαθέσιμο σύνολο δεδομένων για αναγνώριση φαγητών πραγματικού κόσμου, σε αντίθεση με πολλά διαφορετικά σύνολα δεδομένων που κυκλοφορούν, όπου οι φωτογραφίες των φαγητών είναι καταγραμμένες σε ελεγχόμενο περιβάλλον. Περιέχει 101,000 φωτογραφίες, χωρισμένες σε 101 κατηγορίες.

Οι φωτογραφίες αυτές συγκεντρώθηκαν από τον ιστότοπο foodspotting.com, και επιλέχθηκαν τα 101 δημοφιλέστερα είδη φαγητών, ενώ για κάθε είδος φαγητού επιλέχθηκαν 1000 φωτογραφίες. Από αυτές τις 1000 φωτογραφίες, με τυχαίο τρόπο ορίστηκαν 750 ως σύνολο εκπαίδευσης, ενώ οι υπόλοιπες 250 χρησιμοποιούνται ως σύνολο ελέγχου και επαλήθευσης, και έχουν υποστεί επεξεργασία για να καθαριστούν από θόρυβο, σε αντίθεση με τις φωτογραφίες του συνόλου εκπαίδευσης όπου ο θόρυβος έχει παραμείνει, με τη μορφή έντονων χρωμάτων ή επιπλέον περιεχομένου στην εικόνα, χωρίς να έχει σχέση με την αντίστοιχη κατηγορία. Όλες οι φωτογραφίες έχουν υποστεί μια προεπεξεργασία, έτσι ώστε να έχουν μέγιστο πλάτος τα 512 pixels. Παραδείγματα φωτογραφιών από τις κατηγορίες του Food-101 ακολουθούν:

Εικόνα 23: Στιγμιότυπα από το σύνολο δεδομένων Food-101

Στην παρούσα διπλωματική εργασία το νευρωνικό δίκτυο εκπαιδεύτηκε με τις εικόνες και τις κατηγορίες του συνόλου δεδομένων Food-101, καθώς περιείχε μεγαλύτερη ποικιλία στα δείγματα εικόνων για κάθε κατηγορία, συγκρινόμενο με τα υπόλοιπα σύνολα δεδομένων που αναφέρθηκαν, καθώς και το πλήθος των κατηγοριών, σε αντίθεση με το Food-11 που είχε μόνο 11 κατηγορίες. Ένας ακόμη λόγος ήταν η εγγύτητα των κατηγοριών φαγητών στις ελληνικές διατροφικές συνήθειες, χαρακτηριστικό που δεν συναντήθηκε στα σύνολα δεδομένων UEC. Τέλος, δεν προτιμήθηκε το σύνολο δεδομένων PFID καθώς περιείχε εικόνες τυποποιημένων φαγητών, απαθανατισμένα σε περιβάλλον εργαστηρίου, με ελεγχόμενες συνθήκες φωτισμού. Το σύνολο δεδομένων Food-101 περιέχει πραγματικές εικόνες που έχουν φωτογραφηθεί σε ποικίλα περιβάλλοντα, με διαφορετικό φωτισμό, συνδυασμούς, τρόπους προετοιμασίας και μαγειρέματος, καθώς και παραπάνω από ένα είδος τροφίμων στην ίδια εικόνα. Τα χαρακτηριστικά αυτά είναι επιθυμητά για την εκπαίδευση ενός νευρωνικού δικτύου που θα είναι έτοιμο να αξιοποιηθεί από τους χρήστες σε καθημερινή βάση.

Το πλήρες σύνολο δεδομένων περιέχει τις εξής κατηγορίες και είδη φαγητών:

Apple pie
Baby back ribs
Baklava
Beef carpaccio
Beef tartare
Beet salad
Beignets
Bibimbap
Bread pudding
Breakfast burrito
Bruschetta
Caesar salad
Cannoli
Caprese salad
Carrot cake
Ceviche
Cheesecake
Cheese plate
Chicken curry
Chicken quesadilla
Chicken wings
Chocolate cake
Chocolate mousse
Churros
Clam chowder
Club sandwich
Crab cakes
Creme brulee
Croque madame
Cup cakes
Deviled eggs
Donuts
Dumplings
Edamame
Eggs benedict
Escargots
Falafel
Filet mignon
Fish and chips
Foie gras
French fries
French onion soup
French toast
Fried calamari
Fried rice
Frozen yogurt
Garlic bread
Gnocchi
Greek salad
Grilled cheese sandwich
Grilled salmon
Guacamole
Gyoza
Hamburger
Hot and sour soup
Hot dog
Huevos rancheros
Hummus
Ice cream
Lasagna
Lobster bisque
Lobster roll sandwich
Macaroni and cheese
Macarons
Miso soup
Mussels
Nachos
Omelette
Onion rings
Oysters
Pad thai
Paella
Pancakes
Panna cotta
Peking duck
Pho
Pizza
Pork chop
Poutine
Prime rib
Pulled pork sandwich
Ramen
Ravioli
Red velvet cake
Risotto
Samosa
Sashimi
Scallops
Seaweed salad
Shrimp and grits
Spaghetti bolognese
Spaghetti carbonara
Spring rolls
Steak
Strawberry shortcake
Sushi
Tacos
Takoyaki
Tiramisu
Tuna tartare
Waffles

4.4 ΕΚΠΑΙΔΕΥΣΗ ΣΕ MATCONVNET

Για την εκπαίδευση σε MatConvNet, έγινε Fine-Tuning του ΣΝΝ ResNet-50, το οποίο είναι ήδη εκπαιδευμένο στο σύνολο δεδομένων ImageNet. Έγινε εκπαίδευση στα στιγμιότυπα του συνόλου δεδομένων εικόνων Food-101, χρησιμοποιώντας ως αλγόριθμο εκπαίδευσης τη στοχαστική κατάβαση δυναμικού (Stochastic Gradient Descent - SGD) με ορμή σε συνδυασμό με Back-Propagation.

Για την μέθοδο της μεταφερόμενης μάθησης (transfer learning), υπάρχει το ζήτημα του επιπέδου κατηγοριοποίησης. Το ήδη εκπαιδευμένο ResNet-50 που χρησιμοποιήθηκε και επανεκπαιδεύτηκε με το σύνολο δεδομένων Food-101, ήταν εκπαιδευμένο στο σύνολο δεδομένων ImageNet, το οποίο ταξινομεί τις εικόνες σε 1000 διαφορετικές κατηγορίες. Αντίθετα, το σύνολο δεδομένων μας έχει μόνο 101 κατηγορίες. Έτσι η προεπεξεργασία που πρέπει να γίνει στην αρχιτεκτονική είναι να αφαιρεθεί το τελευταίο πλήρως συνδεδεμένο στρώμα `fc1000` (Fully Connected Layer) που είχε ως έξοδο 1000 κατηγορίες, μαζί με το στρώμα `prob` (το οποίο είναι ένα στρώμα SoftMax που κανονικοποιεί τις τιμές που παράγει ο κατηγοριοποιητής του πλήρους συνδεδεμένου στρώματος).

Εφόσον αφαιρέθηκαν αυτά τα στρώματα, μένει να αντικατασταθούν από τα στρώματα που χρειάζονται για την εκπαίδευση του δικτύου στις νέες, 101 κατηγορίες του συνόλου δεδομένων μας. Απαιτείται ένα πλήρως συνδεδεμένο στρώμα με μέγεθος εισόδου 4096 (όσο δηλαδή ήταν και το στρώμα που αφαιρέσαμε) και μέγεθος εξόδου ίσο με 101. Θέλουμε επίσης πάλι το στρώμα SoftMax, και για να εκτελεστεί η διαδικασία της εκπαίδευσης χρειαζόμαστε μια συνάρτηση που υπολογίζει το σφάλμα.

Στο MatConvNet αυτό γίνεται με προσθήκη κάποιων παραπάνω στρωμάτων, συγκεκριμένα τα στρώματα:

- Loss: `dagnn.Loss()` με παράμετρο `'log'`, που υπολογίζει το σφάλμα $L(X, c) = -log(X(c))$, όπου θεωρείται πως η συνάρτηση $X(c)$ είναι η προβλεπόμενη πιθανότητα της κατηγορίας c. Γι' αυτό και το διάνυσμα X πρέπει να είναι μη μηδενικό και να αθροίζεται στη μονάδα. Οι είσοδοι του στρώματος αυτού είναι το στρώμα SoftMax και το στρώμα 'label', που περιέχει τη σωστή κατηγορία της εισόδου.
- Error: `dagnn.Loss()` με παράμετρο `'classerror'`, που εκφράζει το σφάλμα κατηγοριοποίησης: $\mathrm{L(X, c) = (argmax_q X(q) \neq c)}$. Είναι μια ένδειξη για την ακρίβεια του κατηγοριοποιητή κατά την εκπαίδευση και την αξιολόγησή της, ενώ είναι ταυτόσημο με την έννοια του $top - 1\ error$. Η είσοδός του ήταν επίσης τα στρώματα 'softmax' και 'label'.
- Error5: σε αντιστοιχία με το error, είναι ένα στρώμα `dagnn.Loss()` με παράμετρο `'topkerror'`. Υπολογίζει το σφάλμα: $\mathrm{L(X, c) = (rank\ X(c)\ in\ X \leq K)}$. Για $K = 1$, είναι ταυτόσημο με το προηγούμενο στρώμα Error. Η παράμετρος K ορίζεται από την παράμετρο `topK`, η οποία ορίστηκε ίση με 5, για να έχουμε μια εικόνα του $top - 5\ error$. Όπως και με το στρώμα 'error', οι είσοδοί του ήταν τα στρώματα 'softmax' και 'label'.

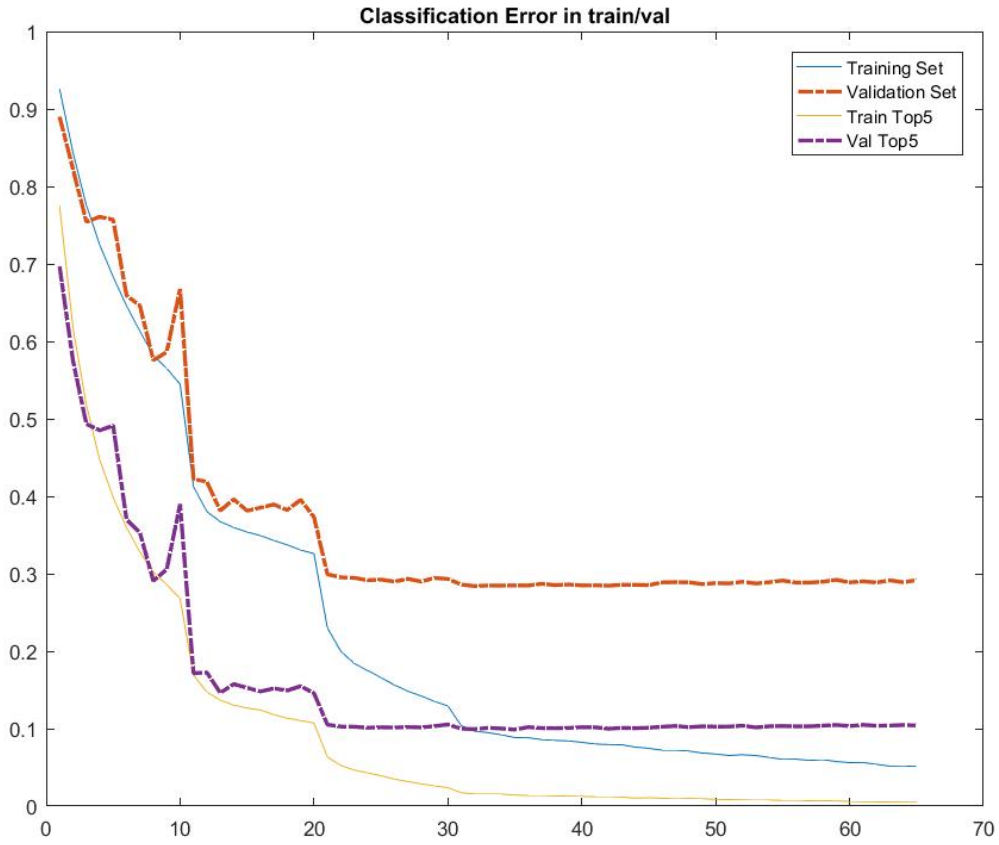
Η εκπαίδευση εκτελέστηκε για 65 εποχές. Από τις 65 αυτές εποχές, κατά τις πρώτες 10 ο ρυθμός εκπαίδευσης ορίστηκε ίσος με 0,05. Για τις επόμενες 10 ίσος με 0,01, για τις επόμενες

ίσος με 0,001 και για όλες τις υπόλοιπες ίσος με 0,0001. Η υπερπαράμετρος της ορμής δεν μεταβλήθηκε από την συνιστώμενη από το MatConvNet τιμή, και παρέμεινε στο 0,9. Το μέγεθος του Batch (δηλαδή το πόσες εικόνες μαζί θα χρησιμοποιηθούν για εκπαίδευση κάθε φορά) μειώθηκε σε 16 (από την προκαθορισμένη τιμή 32), καθώς δεν επαρκούσε η μνήμη της GPU για να επεξεργαστεί περισσότερες εικόνες ταυτόχρονα, διατηρώντας παράλληλα όλες τις παραμέτρους του δικτύου και τα επιμέρους αποτελέσματα των πράξεων στη μνήμη της.

Η μείωση του ρυθμού εκπαίδευσης αποτελεί προτεινόμενη καλή πρακτική για την εκπαίδευση μεγάλων νευρωνικών δικτύων. Έχει παρατηρηθεί πως η χρήση αυτής της πρακτικής βελτιώνει την απόδοση του δικτύου και μειώνει το χρόνο εκπαίδευσης. Με αυτή τη μέθοδο γίνονται μεγάλες μεταβολές στις παραμέτρους του δικτύου κατά τις πρώτες εποχές της εκπαίδευσης, ενώ οι μεταβολές αυτές με την πάροδο των εποχών μειώνονται. Αυτό έχει ως αποτέλεσμα την γρήγορη εύρεση «καλών» παραμέτρων νωρίς, οι οποίες προσδιορίζονται με μεγαλύτερη ακρίβεια στην πορεία.

Για να μπορέσει το MatConvNet να λάβει ως είσοδο τις εικόνες, ήταν αναγκαία η οργάνωσή τους σε μια δομή που καλείται imdb (Image DataBase). Περιέχει τις διαδρομές των αρχείων εικόνων, καθώς και την αντιστοίχιση στη κατηγορία στην οποία ανήκουν. Αναθέτει επίσης τις εικόνες σε ένα σύνολο, συγκεκριμένα στα σύνολα train και val (εκπαίδευση και επαλήθευση αντίστοιχα), για να γίνει ευκολότερη η επιλογή τους κατά τις αντίστοιχες φάσεις της εκπαίδευσης. Επίσης, διατηρούνται μεταδεδομένα όπως τα ονόματα των κατηγοριών και οι (πιο επεξηγηματικές) ετικέτες τους. Το σύνολο δεδομένων food-101 περιείχε προτεινόμενο διαχωρισμό σε σύνολα train και val (750 και 250 εικόνες από κάθε κατηγορία, αντίστοιχα), και αυτός ο διαχωρισμός διατηρήθηκε και κατά τη δημιουργία του imdb.

Κατά τη διάρκεια της εκπαίδευσης, κρατούνται στατιστικά για το error και το loss, ώστε να παραχθεί η παρακάτω γραφική παράσταση ως προς το πλήθος των εποχών.



Εικόνα 24 : Οι καμπύλες εκπαίδευσης - error ανά εποχή

Οι διακεκομμένες γραμμές αντιπροσωπεύουν το σφάλμα εκπαίδευσης στο σύνολο επαλήθευσης. Το σφάλμα αυτό είναι πιο αντιπροσωπευτικό για την πραγματική συμπεριφορά του δικτύου. Ως καλύτερη φάση του δικτύου επιλέχτηκε το δίκτυο κατά την εποχή που επετεύχθη το ελάχιστο σφάλμα στο σύνολο επαλήθευσης, που στην περίπτωση αυτή ήταν κατά την εποχή 23. Σε αυτή την εποχή η επίδοση του δικτύου ήταν ως εξής:

	Loss	Top-1 error	Top-5 error
Train	0.6790	18.43 %	4.66 %
Validation	1,1607	29.46 %	10.25 %

4.5 ΕΚΠΑΙΔΕΥΣΗ ΣΕ TORCH

Η διαδικασία του fine-tuning του νευρωνικού δικτύου μας στο περιβάλλον Torch αποδείχθηκε πιο απλή, έχοντας τη βοήθεια της υλοποίησης της αρχιτεκτονικής ResNet σε Torch7 από την ερευνητική ομάδα του Facebook [97]. Στην υλοποίηση αυτή παρέχονται προ-εκπαιδευμένα μοντέλα ResNet βάθους 18, 34, 50, 101, 152 και 200 στρωμάτων. Τα μοντέλα αυτά εκπαιδεύτηκαν στο σύνολο δεδομένων ImageNet, και έπειτα από μικρές παραμετροποιήσεις και βελτιστοποιήσεις πάνω στο το αρχικό μοντέλο ResNet πέτυχαν καλύτερη ακρίβεια, μειώνοντας τα αντίστοιχα σφάλματα:

Δίκτυο	Αρχικό top-1 error	Facebook top-1 error	Αρχικό top-5 error	Facebook top-5 error
ResNet-18	N/A	30.43 %	N/A	10.76 %
ResNet-34	N/A	26.73 %	N/A	8.74 %
ResNet-50	24.7 %	24.01 %	7.8 %	7.02 %
ResNet-101	23.6 %	22.44 %	7.1 %	6.21 %
ResNet-152	23.0 %	22.16 %	6.7 %	6.16 %
ResNet-200	N/A	21.66 %	N/A	5.79 %

Η υλοποίηση αυτή διέφερε από την πρωτότυπη υλοποίηση της αρχιτεκτονικής ResNet κατά τους εξής τρόπους:

Αύξηση κλίμακας (Scale augmentation): Χρησιμοποιείται η μέθοδος προσαύξησης δεδομένων μέσω κλίμακας και αναλογίας εικόνας (scale and aspect ratio augmentation) που χρησιμοποιήθηκε από τη δημοσίευση "Going Deeper with Convolutions" για το GoogLeNet νευρωνικό δίκτυο [40]. Στην πρωτότυπη υλοποίηση, χρησιμοποιείται μόνο προσαύξηση δεδομένων μέσω κλίμακας, ενώ με τη νέα υλοποίηση, εμφανίζεται μικρότερο σφάλμα επαλήθευσης. Η προσαύξηση δεδομένων είναι μια τεχνική κατά την οποία λαμβάνονται περισσότερα τυχαία δείγματα από το αρχικό σύνολο δεδομένων, για καλύτερη εκπαίδευση του δικτύου με περισσότερα δεδομένα.

Πιο συγκεκριμένα, με την παλιά υλοποίηση, επέστρεφε μια μετασχηματισμένη σε κλίμακα, με την μικρότερη πλευρά της να παίρνει μια τυχαία τιμή εντός κάποιων ορίων, διατηρώντας την αρχική αναλογία. Με την τεχνική του δικτύου GoogLeNet, λαμβάνεται ένα τυχαίο απόκομμα (crop) από την αρχική εικόνα, σε τυχαίο μέγεθος μεταξύ 8% και 100% του αρχικού, και τυχαία αναλογία μεταξύ 3/4 και 4/3.

Αύξηση χρώματος (Color augmentation): Χρησιμοποιήθηκαν φωτομετρικές παραμορφώσεις που παρουσιάστηκαν στην δημοσίευση του Andrew Howard [98], εκτός από τις χρωματικές προσαυξήσεις που παρουσιάστηκαν στο AlexNet [53]. Με τον ίδιο τρόπο, επέφεραν καλύτερη ακρίβεια στην τελική υλοποίηση.

Εξασθένηση βαρών (Weight Decay): Η εξασθένηση των βαρών αποτελεί προτεινόμενη πρακτική κατά τη διάρκεια της εκπαίδευσης, και είναι μια μορφή ομαλοποίησης της πολυπλοκότητας. Σε αντίθεση με την αρχική υλοποίηση, όπου η εξασθένηση βαρών

εφαρμόζεται μόνο στα βάρη των συνελικτικών επιπέδων, πλέον εφαρμόζεται και σε όλα τα βάρη και τις πολώσεις.

Συνέλιξη με Διασκελισμό (Strided convolution): Όταν χρησιμοποιείται η αρχιτεκτονική bottleneck στο δίκτυο ResNet, χρησιμοποιείται διασκελισμός με βήμα 2 στην συνέλιξη με φίλτρο 3x3, αντί για την πρώτη συνέλιξη με φίλτρο 1x1.

Για προετοιμασία των δεδομένων, δεν υπήρχε ανάγκη για κάποια εξεζητημένη δομή, σε αντίθεση με το MatConvNet toolbox όπως αναφέρθηκε νωρίτερα. Αντ'αυτού, απαιτείται μόνο οι εικόνες να είναι σε φακέλους των οποίων το όνομα αντιστοιχεί στην κατηγορία στην οποία ανήκουν. Επιπλέον, αν υπάρχει διαχωρισμός μεταξύ συνόλων εκπαίδευσης και επαλήθευσης, πρέπει να υπάρχουν ξεχωριστοί φάκελοι «train» και «val» αντίστοιχα. Έτσι η δομή των φακέλων αποκτά τη μορφή:

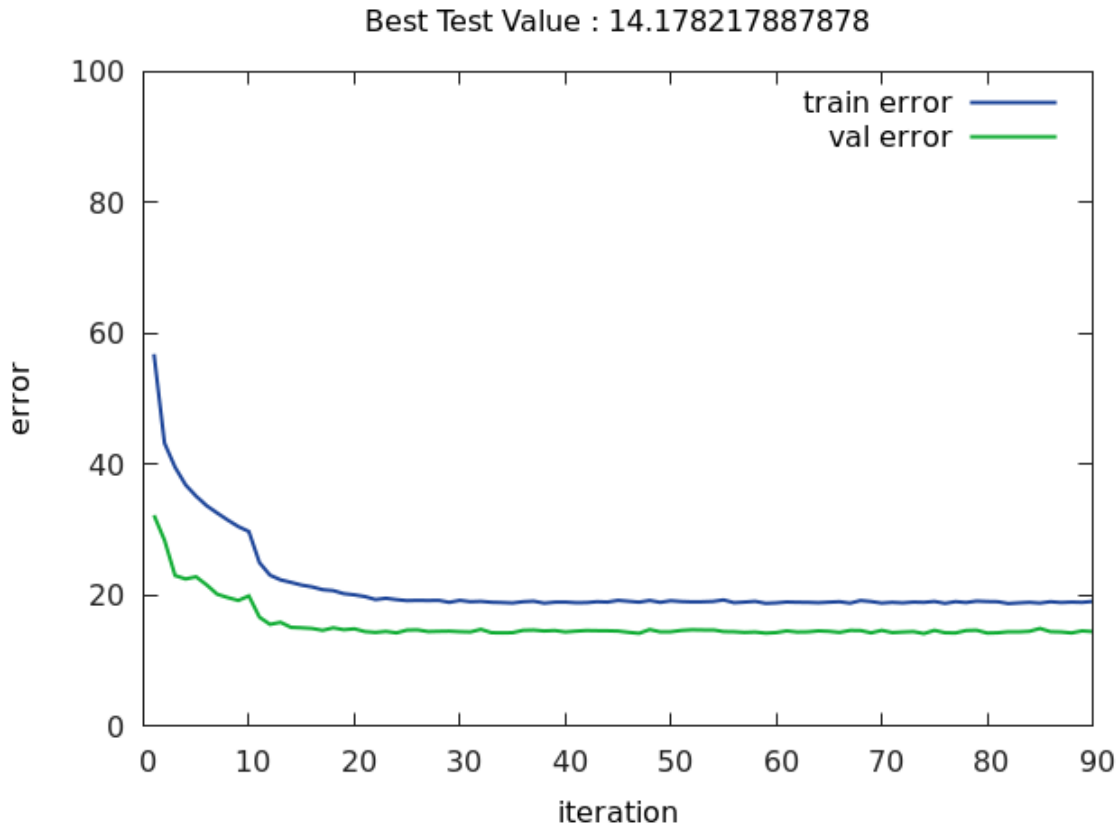
```
train/<label1>/<image.jpg>
train/<label2>/<image.jpg>
val/<label1>/<image.jpg>
val/<label2>/<image.jpg>
```

Η παρεχόμενη υλοποίηση δεν διατηρούσε την επιμέρους κατάσταση του δικτύου και τα στατιστικά της εκπαίδευσης. Για τον λόγο αυτό επεκτάθηκε ώστε να διατηρείται η ενδιάμεση κατάσταση μεταξύ των εποχών, σε περίπτωση που χρειαζόταν να διακοπεί η διαδικασία της εκπαίδευσης, και να συνεχίσει από την τελευταία πλήρως εκπαιδευμένη εποχή. Επίσης στο τέλος κάθε εποχής σχεδιάζεται μια γραφική παράσταση του σφάλματος και της τιμής της συνάρτησης σφάλματος, με τη βοήθεια του `gnuplot`.

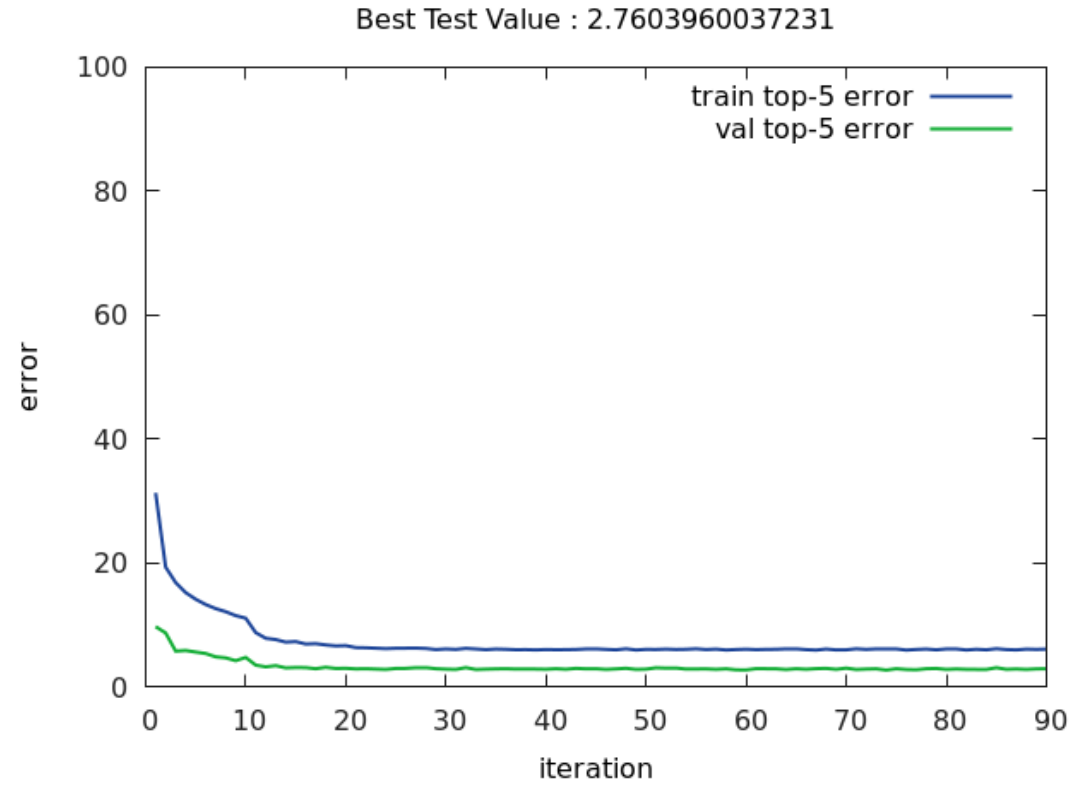
Είδαμε στην προηγούμενη ενότητα πώς λύνεται το ζήτημα του μεγέθους του κατηγοριοποιητή για την υλοποίηση της μεθόδου της μεταφερόμενης μάθησης (transfer learning). Στο Torch η διαδικασία είναι λίγο διαφορετική, κρατώντας την αρχιτεκτονική «καθαρή» χωρίς να προστίθενται επιπλέον στρώματα μόνο για την εκπαίδευση. Στο δίκτυο έχουμε απλή αντικατάσταση του τελευταίου στρώματος (το οποίο στην αρχιτεκτονική του torch είναι μόνο το πλήρες συνδεδεμένο στρώμα, χωρίς επιπλέον SoftMax) με ένα πλήρως συνδεδεμένο στρώμα 101 εξόδων. Χρησιμοποιείται μια νέα δομή, το ονομαζόμενο criterion (κριτήριο). Χρησιμοποιώντας το criterion αυτό, υπολογίζεται η συνάρτηση σφάλματος και ενημερώνονται οι παράμετροι του δικτύου. Συγκεκριμένα για την εκπαίδευση χρησιμοποιήθηκε το criterion CrossEntropyCriterion. Το CrossEntropyCriterion συνδυάζει δύο υπολογισμούς. Αρχικά εφαρμόζει την συνάρτηση SoftMax, και έπειτα εφαρμόζει ένα ήδη υπάρχον criterion, το ClassNLLCriterion, την αρνητική λογαριθμική πιθανότητα (Negative Log-Likelihood - NLL). Είναι η ίδια συνάρτηση που χρησιμοποιεί το MatConvNet με παράμετρο 'log'. Επομένως έχουμε το ίδιο κριτήριο για υπολογισμό του σφάλματος και στα δύο frameworks.

Η εκπαίδευση του δικτύου εκτελέστηκε για 90 εποχές, με μέγεθος Batch ίσο με 14, με αρχικό ρυθμό εκπαίδευσης ορισμένο στη τιμή $0{,}001$, ο οποίος υποδεκαπλασιάζεται ($lr' = lr * 0.1$) με την πάροδο προκαθορισμένου πλήθους εποχών, που ονομάζεται βήμα εξασθένησης ρυθμού εκπαίδευσης (Learning Rate Decay Step). Η παράμετρος του βήματος εξασθένησης ορίστηκε στις 10 εποχές, ενώ η παράμετρος της ορμής παρέμεινε στην προκαθορισμένη τιμή της, ίση με $0{,}9$.

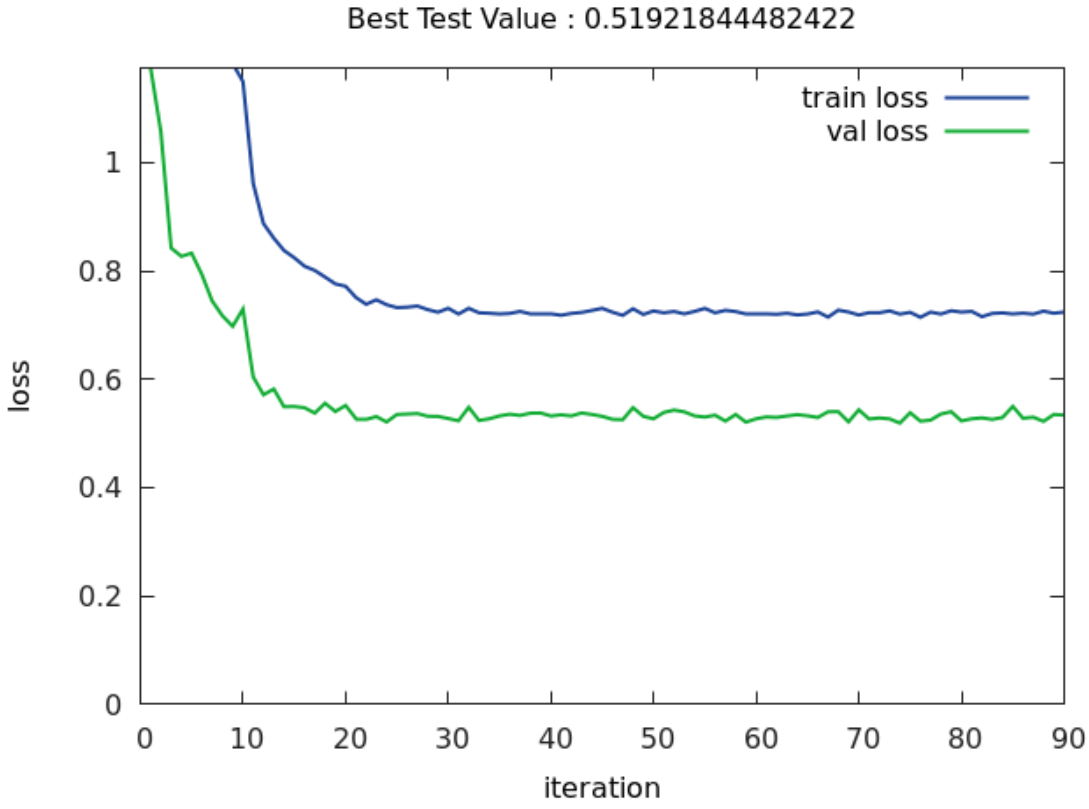
Ακολουθούν τα διαγράμματα εκπαίδευσης για `training` και `validation`, για τα μεγέθη `Loss`, `Error`, και `Error5`, σε συνάρτηση με το πλήθος εποχών. Το μέγεθος `Loss` εκφράζει την απόλυτη τιμή της εξόδου της συνάρτησης `SoftMax`, και συγκεκριμένα τη μέση τιμή της για όλες τις κατηγορίες. Το `Error` είναι το top-1 classification σφάλμα, και το `Error5` το top-5 classification σφάλμα.



Εικόνα 25: Top-1 error για Train/Validation



Εικόνα 26: : Top-5 error για Train/Validation



Εικόνα 27: Απόλυτο Loss για Train/Validation

Στα παραπάνω γραφήματα, παρατηρείται το (αρχικά) παράδοξο αποτέλεσμα του μικρότερου σφάλματος επαλήθευσης σε σχέση με το αντίστοιχο σφάλμα εκπαίδευσης, σε όλα τα μεγέθη. Αυτό συμβαίνει λόγω της υλοποίησης της συνάρτησης εκπαίδευσης του Facebook. Κατά την φάση εκπαίδευσης του δικτύου, χρησιμοποιούνται οι εικόνες που ανήκουν στο σύνολο εκπαίδευσης. Η διαφορά είναι πως δεν χρησιμοποιούνται αυτούσιες. Περνούν από προ-επεξεργασία, και παράγονται περισσότερες εικόνες έπειτα από τους μετασχηματισμούς αυτούς. Αυτή η τεχνική ονομάζεται προσαύξηση δεδομένων (Data Augmentation), και βοηθάει στην «εύρεση» νέων εικόνων που δεν υπήρχαν στο αρχικό σύνολο εκπαίδευσης, με σκοπό να αποκτήσει το εκπαιδευμένο νευρωνικό δίκτυο μια αμεταβλητότητα σε παραμορφώσεις.

Πιο συγκεκριμένα, κατά την εκπαίδευση οι εικόνες υπόκεινται μια σειρά από παραμορφώσεις (transforms) που περιλαμβάνουν μεταβολή μεγέθους (Scale), τυχαία αποκοπή (RandomSizedCrop), μικρές παραμορφώσεις χρώματος (ColorJitter), περιστροφή (Rotation), κανονικοποίηση χρώματος (ColorNormalize) και κατοπτρισμός (HorizontalFlip). Αντίθετα, κατά την επαλήθευση, οι εικόνες του συνόλου επαλήθευσης περνούν μόνο από επεξεργασίες: Scale, ColorNormalize και Crop. Επομένως η εμφάνιση μεγαλύτερου σφάλματος εκπαίδευσης είναι λογική και δεν θα πρέπει να μας παραξενεύει.

Κατά τη καλύτερη εκδοχή του μοντέλου (η οποία αποφασίζεται από την ελάχιστη συνάρτηση σφάλματος και ήταν η 80στη εποχή), είχαμε τα εξής:

	Loss	Top-1 error	Top-5 error
Train	0.720	18.947 %	6.055 %
Validation	0.519	14.178 %	2.760 %

4.6 ΟΠΤΙΚΟΠΟΙΗΣΗ ΘΕΣΗΣ ΑΝΤΙΚΕΙΜΕΝΟΥ ΦΑΓΗΤΟΥ

Ως επιπλέον εποπτική πληροφορία, έγινε μια προσέγγιση για υλοποίηση της μεθόδου που παρουσιάζεται στη δημοσίευση [99], και μπορεί να εντοπίσει τις περιοχές της εικόνας που είχαν την ισχυρότερη συνεισφορά για να παραχθεί η πρώτη εκτίμηση του δικτύου. Έτσι εμφανίζουμε ένα χάρτη θερμοκρασίας (heatmap) πάνω από την εικόνα, με τα θερμότερα χρώματα να δείχνουν τη περιοχή με την ισχυρότερη συνεισφορά. Αυτή η μέθοδος θα μπορούσε να μας βοηθήσει μετέπειτα σε κατάτμηση της φωτογραφίας στα επιμέρους συστατικά της, όμως αυτό το έργο δεν καλύπτεται στην παρούσα διπλωματική εργασία.

Το script αυτό είναι γραμμένο σε Lua, και εκτελείται με το framework Torch. Έχοντας ως βάση το εκπαιδευμένο δίκτυό μας, εξάγει την πρώτη εκτίμηση και έπειτα αφαιρεί τα 3 τελευταία επίπεδα. Δηλαδή αφαιρούνται τα επίπεδα: `cudnn.SpatialAveragePooling(7x7, 1,1)`, `nn.View(2048)` και `nn.Linear(2048 -> 101)`.

Χωρίς αυτά τα επίπεδα του δικτύου, η έξοδος του δικτύου είναι ένας τανυστής (tensor) μεγέθους ανάλογου της εισόδου του δικτύου, με 2048 κανάλια. Τα κανάλια αυτά εκφράζουν τα χαρκατηριστικά που υπολογίζονται από το δίκτυο. Τα επίπεδα που αφαιρέθηκαν παίρνουν το μέσο όρο των ενεργοποιήσεων των νευρώνων αυτών (με το επίπεδο SpatialAveragePooling), μετατρέπουν το τανυστή του αποτελέσματος σε ένα διάνυσμα 2048x1 (με το επίπεδο View(2048)) και έπειτα το πλήρως συνδεδεμένο δίκτυο nn.Linear(2048 -> 101) αναλαμβάνει να αντιστοιχίσει τα χαρακτηριστικά στις κατηγορίες εξόδου.

Έπειτα, προστίθεται ένα στάδιο συνέλιξης από τις 2048 «εικόνες» που έχουμε αυτή τη στιγμή, που θα παράξει 101 εικόνες. Αντιπροσωπεύει ουσιαστικά την αντιστοίχιση των 2048 εικόνων στις 101 κατηγορίες, όπως κάνει και το τελευταίο γραμμικό στάδιο που αφαιρέσαμε. Χρησιμοποιώντας τα βάρη του παλιού τελικού σταδίου στο στάδιο συνέλιξης που προσθέσαμε, παίρνουμε την αντιστοίχιση των 2048 εικόνων – χαρακτηριστικών (features) πάνω σε κάθε κατηγορία. Λαμβάνοντας την εικόνα εξόδου αυτή (η οποία είναι σε διαβαθμίσεις του γκρι), και μετατρέποντάς την με τη συνάρτηση `y2jet` του πακέτου `image` του Torch σε heatmap, παίρνουμε τα εξής αποτελέσματα:

Εικόνα 28: Σωστή κατηγοριοποίηση της εικόνας ως Spaghetti Carbonara με ακρίβεια 99,51% και το αντίστοιχο heatmap

Εικόνα 29: Σωστή κατηγοριοποίηση εικόνας ως Baby back ribs με ακρίβεια 98,88% και το αντίστοιχο heatmap

5 ΑΠΟΤΕΛΕΣΜΑΤΑ – ΣΥΜΠΕΡΑΣΜΑΤΑ

Με σκοπό την εκπαίδευση ενός νευρωνικού δικτύου για την κατηγοριοποίηση εικόνων που περιέχουν γεύματα στις αντίστοιχες κατηγορίες φαγητού, εφαρμόστηκε η τεχνική της μεταφερόμενης μάθησης. Αξιοποιήθηκε η αρχιτεκτονική υπολειπόμενης μάθησης συνελικτικών νευρωνικών δικτύων ResNet-50, της οποίας τα βάρη είχαν προϋπολογιστεί μετά από εκπαίδευση του δικτύου στο σύνολο δεδομένων ImageNet. Με την τεχνική της μεταφερόμενης μάθησης, έγινε επανεκπαίδευση στο σύνολο δεδομένων Food-101 με τις νέες κατηγορίες. Η εκπαίδευση έγινε σε δύο διαφορετικά περιβάλλοντα και frameworks.

Συγκρίνοντας τους πίνακες των σφαλμάτων για τις εκπαιδεύσεις που πραγματοποιήθηκαν στα δύο Frameworks, παρατηρούμε πως, ενώ η τιμή του σφάλματος είναι μικρότερη στην περίπτωση του MatConvNet, με το Torch πετύχαμε πολύ μεγαλύτερη ακρίβεια στο σύνολο επαλήθευσης. Να σημειωθεί πως ο χωρισμός των συνόλων εκπαίδευσης/επαλήθευσης ήταν ο ίδιος και στις δύο περιπτώσεις.

Αποτελέσματα MatConvNet:

	Loss	Top-1 error	Top-5 error
Train	0.6790	18.43 %	4.66 %
Validation	1.1607	29.46 %	10.25 %

Αποτελέσματα Torch:

	Loss	Top-1 error	Top-5 error
Train	0.720	18.947 %	6.055 %
Validation	0.519	14.178 %	2.760 %

Οι μεγαλύτερες τιμές που παρατηρούνται στην εκπαίδευση του δικτύου με Torch, εξηγήθηκαν νωρίτερα και οφείλονται στην προσαύξηση δεδομένων που χρησιμοποιείται, κάνοντας δυσκολότερη την κατηγοριοποίηση των εικόνων.

5.1 Έλεγχος δικτυου με εικονες απο διαφορετικα συνολα δεδομενων

Για να ελεγχθεί η ικανότητα γενίκευσης του εκπαιδευμένου δικτύου, λήφθηκαν τυχαία δείγματα εικόνων από τα σύνολα δεδομένων UECFood100/256 και Food11. Έχοντας τις εικόνες αυτές ως είσοδο, έγινε εξαγωγή των 5 κορυφαίων εκτιμήσεων του δικτύου, και έγινε σύγκριση με την αρχική κατηγορία στο αντίστοιχο σύνολο δεδομένων. Επιλέχτηκαν κατηγορίες που υπάρχουν στο σύνολο δεδομένων στο οποίο έγινε η εκπαίδευση, καθώς και μερικές περιπτώσεις όπου δεν υπήρχε αντιστοιχία, για να παρατηρηθεί η συμπεριφορά του δικτύου.

Ακολουθούν κάποια ενδεικτικά παραδείγματα εκτέλεσης του δικτύου με τις εικόνες εισόδου που λήφθηκαν τυχαία από το διαδίκτυο ή από τα σύνολα δεδομένων που αναφέραμε παραπάνω, και τις 5 καλύτερες εκτιμήσεις του δικτύου.

5.1.1 Εικόνες από σύνολα δεδομένων UECFood100 και UECFood256

94.596666 % Fried rice
0.856961 % Paella
0.621909 % Tuna tartare
0.556158 % Beet salad
0.408458 % Macaroni and cheese

74.479526 % Pizza
20.653677 % Chicken quesadilla
1.382973 % Dumplings
1.288469 % Garlic bread
0.519939 % Gyoza

68.242204 % French fries
26.010436 % Fish and chips
1.606781 % Poutine
1.599279 % Fried calamari
1.196511 % Onion rings

100.000000 % Hot dog
0.000002 % French fries
0.000001 % Hamburger
0.000000 % Grilled cheese sandwich
0.000000 % Lobster roll sandwich

97.014356 % Apple pie
1.117412 % Pancakes
0.412154 % Samosa
0.257593 % Hummus
0.252608 % Omelette

94.337910 % Sushi
4.993579 % Sashimi
0.139016 % Cheesecake
0.120216 % Grilled salmon
0.065251 % Beet salad

85.468000 % Donuts
13.109092 % Onion rings
1.054791 % Churros
0.104279 % French toast
0.064368 % Apple pie

83.699507 % Tiramisu
13.227193 % Carrot cake
0.891360 % Escargots
0.793456 % Chocolate cake
0.233896 % Baklava

Όπως παρατηρούμε σε κάποια από τα δείγματα, η ακρίβεια αναγνώρισης είναι αρκετά ικανοποιητική, δεδομένου ότι οι εικόνες αυτές δεν είχαν συναντηθεί κατά την εκπαίδευση, καθώς ανήκουν σε διαφορετικό σύνολο δεδομένων. Δεν ήταν δυνατή η αξιολόγηση σε όλες τις κατηγορίες των συνόλων UECFood100 και UECFood256, διότι δεν υπήρχαν πολλές κοινές κατηγορίες φαγητών. Τα παραπάνω αποτελέσματα ανήκουν σε κοινές κατηγορίες, και η αναγνώριση της σωστής εκάστοτε κατηγορίας είναι εφικτή.

Σε περιπτώσεις που η κατηγορία της εικόνας φαγητού δεν υπήρχε στο σύνολο δεδομένων για το οποίο εκπαιδεύσαμε το δίκτυό μας, λαμβάνουμε τα παρακάτω αποτελέσματα:

41.034126 % Caesar salad
16.057047 % Hot dog
10.179359 % Club sandwich
9.268814 % Greek salad
6.758580 % Chicken quesadilla

Σωστή κατηγορία: Tacos

72.586477 % Garlic bread
4.401102 % Chocolate cake
3.255789 % Pulled pork sandwich
2.927520 % Hamburger
1.611201 % Hot dog

Σωστή κατηγορία: Brownie

55.428243 % Pork chop
11.817037 % Foie gras
10.063053 % Steak
5.193143 % French toast
3.687693 % Grilled salmon

Σωστή κατηγορία: Baked Salmon

77.964854 % Frozen yogurt
7.145569 % Ice cream
4.048230 % Fried rice
3.415735 % Bread pudding
2.349580 % Macaroni and cheese

Σωστή κατηγορία: Popcorn

Όπως είναι φανερό, η είσοδος εικόνων στο δίκτυό μας για τις οποίες δεν έχει γνώση κάποιας κατηγοριοποίησης, το κάνει να εξάγει εσφαλμένα αποτελέσματα, προσπαθώντας να αναγνωρίσει χαρακτηριστικά τα οποία είχε ήδη μάθει.

5.1.2 Εικόνες από σύνολο δεδομένων Food-11

99.965787 % Garlic bread
0.009744 % Lasagna
0.006121 % Bruschetta
0.004680 % Beef carpaccio
0.003713 % Spaghetti Bolognese

Αρχική κατηγορία:
Ψωμί

99.994063 % Cheese plate
0.001816 % Sushi
0.001112 % Grilled salmon
0.001071 % Foie gras
0.000822 % Spring rolls

Αρχική κατηγορία:
Γαλακτοκομικά

86.401880 % Deviled eggs
7.384106 % Shrimp and grits
2.602282 % Omelette
2.084015 % Hummus
0.410091 % Breakfast burrito

Αρχική κατηγορία:
Αυγό

99.988496 % Onion rings
0.007432 % Fried calamari
0.001673 % French fries
0.000610 % Donuts
0.000607 % Churros

Αρχική κατηγορία:
Τηγανισμένα φαγητά

99.847800 % Baby back ribs
0.093895 % Steak
0.016773 % Beef carpaccio
0.009260 % Grilled salmon
0.008648 % Prime rib

Αρχική κατηγορία:
Κρεατικά

99.502426 % Spaghetti carbonara
0.490141 % Spaghetti bolognese
0.001912 % Pad thai
0.001536 % Onion rings
0.000657 % Ramen

Αρχική κατηγορία:
Ζυμαρικά

99.185151 % Sashimi
0.424212 % Sushi
0.351999 % Scallops
0.010402 % Gyoza
0.008597 % Peking duck

Αρχική κατηγορία:
Θαλασσινά

99.978620 % Cheesecake
0.012139 % Panna cotta
0.005768 % Bread pudding
0.001771 % Strawberry shortcake
0.000912 % Baby back ribs

Αρχική κατηγορία:
Επιδόρπια

99.984479 % Lobster bisque
0.009852 % Grilled cheese sandwich
0.004962 % Chicken curry
0.000400 % Spring rolls
0.000058 % Crab cakes

Αρχική κατηγορία:
Σούπες

55.765522 % Caprese salad
9.576635 % Eggs benedict
6.638539 % Spring rolls
4.031059 % Deviled eggs
3.157805 % Miso soup

Αρχική κατηγορία:
Φρούτα και Λαχανικά

Παρατηρούμε πως επί το πλείστον η κατηγοριοποίηση ήταν επιτυχής και οι πιο συγκεκριμένες κατηγορίες που βρέθηκαν ανήκουν στην αρχική κατηγορία. Μια ακόμα παρατήρηση που μπορούμε να κάνουμε είναι πως το δίκτυο αποτυγχάνει να αναγνωρίσει τις ντομάτες της τελευταίας εικόνας. Αυτό οφείλεται στο γεγονός πως το σύνολο δεδομένων Food-101 στο οποίο έγινε η εκπαίδευση δεν περιείχε κατηγορίες ούτε εικόνες από φρούτα ή λαχανικά.

5.2 Μελλοντική Έρευνα

Τα αποτελέσματα της αναγνώρισης τροφών από φωτογραφικά στιγμιότυπα χρησιμοποιώντας αλγορίθμους μηχανικής μάθησης και συγκεκριμένα Συνελικτικά Νευρωνικά Δίκτυα ήταν πολύ ενθαρρυντικά. Με βάση την επίτευξη σφάλματος ~14% , η ταξινόμηση κρίνεται ιδιαίτερα επιτυχής, ειδικά αν ληφθεί υπόψη πως η κατηγοριοποίηση γίνεται ανάμεσα σε 101 κατηγορίες φαγητού.

Ωστόσο, η τεχνική πλευρά της παρούσας διπλωματικής εργασίας επιδέχεται περαιτέρω βελτιώσεις, για να μπορεί να χρησιμοποιηθεί σε συστήματα αυτόνομης εκτίμησης ποσότητας θρεπτικής αξίας σε τρόφιμα για άτομα με Σακχαρώδη Διαβήτη.

Για αρχή, θα πρέπει το σύνολο δεδομένων να εμπλουτιστεί και οι κατηγορίες του να συγκεκριμενοποιηθούν σε είδη τροφών που είναι πιο συνηθισμένα στην ελληνική κουζίνα, καθώς πολλές από τις κατηγορίες του συνόλου δεδομένων είναι αρκετά σπάνιες να συναντηθούν. Μια επιπλέον ιδέα πάνω σε αυτό θα μπορούσε να είναι η κατηγοριοποίηση σε πιο γενικές ομάδες τροφίμων, με βάση το ποσοστό υδατανθράκων που περιέχονται σε αυτά. Η κατηγοριοποίηση αυτή θα μας επέτρεπε να επιτύχουμε μεγαλύτερη ακρίβεια στην ταξινόμηση, έχοντας λιγότερες κατηγορίες, και παράλληλα θα μας επέτρεπε να αντλήσουμε άμεσα την πληροφορία της ποσότητας υδατανθράκων, ακόμα και αν ο διαχωρισμός μεταξύ των αντικειμένων μιας κλάσης δεν είναι εύκολος.

Σχετικά με την αρχιτεκτονική του συστήματος αξιολόγησης, θα πρέπει να επισημάνουμε πως το μέγεθος του αρχείου που περιέχει όλες τις παραμέτρους του δικτύου ώστε να μπορεί να χρησιμοποιηθεί για να εκτελεστεί μια κατηγοριοποίηση, είναι αρκετά μεγάλο (για το συνελικτικό δίκτυο που βασίστηκε στην αρχιτεκτονική ResNet-50, το μέγεθος αγγίζει τα 190 MB, ενώ μια αρχιτεκτονική με μεγαλύτερο βάθος θα αύξανε ανάλογα το μέγεθος). Η χρήση του λοιπόν σε εφαρμογές φορητών συσκευών ή σε ένα αυτόνομο σύστημα δεν θα ήταν συνετή. Θα πρέπει επομένως να σχεδιαστεί και να εκπαιδευτεί από την αρχή ένα μικρότερο συνελικτικό νευρωνικό δίκτυο, που θα παρέχει την ίδια ή και καλύτερη ακρίβεια αναγνώρισης.

Είναι επίσης σημαντική η μελέτη και η υλοποίηση των επιμέρους ενοτήτων του ολοκληρωμένου συστήματος εκτίμησης περιεχόμενης ποσότητας υδατανθράκων στα γεύματα των εικόνων. Η πρώτη εργασία θα πρέπει να είναι η κατάτμηση της εικόνας για αναγνώριση πολλαπλών ειδών τροφίμων μέσα σε αυτήν. Αφού γίνει ο διαχωρισμός, μπορεί να αναγνωριστεί το κάθε είδος τροφίμου με μεθόδους κατηγοριοποίησης, όπως αυτή που παρουσιάστηκε στην παρούσα διπλωματική εργασία. Να σημειωθεί πως τα δύο αυτά στάδια μπορούν να λειτουργήσουν παράλληλα και συνεργατικά.

Το επόμενο (και δυσκολότερο) στάδιο θα πρέπει να είναι ο υπολογισμός του όγκου της τροφής που εντοπίστηκε και αναγνωρίστηκε. Έχοντας την πληροφορία του όγκου και πληροφορίες από διατροφικούς πίνακες μπορούμε να προσφέρουμε μια εκτίμηση της περιεχόμενης ποσότητας υδατανθράκων στο γεύμα του οποίου λάβαμε φωτογραφικό στιγμιότυπο.

6 ΒΙΒΛΙΟΓΡΑΦΙΑ

[1] World Health Organisation, «Obesity Study,» [Ηλεκτρονικό]. Available: http://www.who.int/mediacentre/factsheets/fs311/en/.

[2] WHO, «Tenfold increase in childhood and adolescent obesity in four decades: new study by Imperial College London and WHO,» 11 October 2017. [Ηλεκτρονικό]. Available: http://www.who.int/mediacentre/news/releases/2017/increase-childhood-obesity/en/.

[3] AMA, «Proceedings of th 2013 American Medical Association Annual Meeting,» 2013.

[4] M. Ashwell, T. J. Cole και A. K. Dixon, «Ratio of waist circumference to height is strong predictor of intra-abdominal fat,» *BMJ,* τόμ. 313, αρ. 7056, pp. 559-560, 1996.

[5] D. L. Duren, R. J. Sherwood, W. C. Chumlea, S. A. Czerwinski, M. Lee, A. Choh και R. Siervogel, «Body composition methods: comparisons and interpretation,» *Journal of Diabetes Science Technology,* τόμ. 2, αρ. 6, pp. 1139-1146, 2008.

[6] T. Kawamura, «The importance of carbohydrate counting in the treatment of children with diabetes,» *Pediatric Diabetes,* τόμ. 8, αρ. s6, pp. 57-62, 2007.

[7] C. E. Smart, B. R. King, P. McElduff και C. E. Collins, «In children using intensive insulin therapy, a 20-g variation in carbohydrate amount significantly impacts on postprandial glycaemia,» *Diabetic Medicin,* τόμ. 29, αρ. 7, pp. 21-24, 2012.

[8] F. K. Bishop, D. M. Maahs, G. Spiegel, D. Owen, G. J. Klingensmith, A. Bortsov, J. Thomas και E. J. Mayer-Davis, «The Carbohydrate Counting in Adolescents With Type 1 Diabetes (CCAT) Study,» *Diabetes Spectrum,* τόμ. 22, αρ. 1, pp. 56-62, 2009.

[9] C. Smart, K. Ross, J. Edge, B. King, P. McElduff και C. Collins, «Can children with Type 1 diabetes and their caregivers estimate the carbohydrate content of meals and snacks?,» *Diabetes Medicine,* τόμ. 27, αρ. 3, pp. 348-353, 2010.

[10] Y. C. Probst και L. C. Tapsell, «Overview of computerized computerized dietary assessment programs for research and practice in nutrition education,» *Journal of Nutrition Education and Behavior,* τόμ. 37, αρ. 1, pp. 20-26, 2005.

[11] D. H. Wand, D. H. Kogashiva και S. Kira, «Development of a new instrument for evaluating individuals' dietary intakes,» *Journal of the American Dietetic Association,* τόμ. 106, αρ. 10, pp. 1588-1593, 2006.

[12] L. Harnack, L. Steffen, D. Arnett, S. Gao και R. Luepker, «Accuracy of estimation of large food portions,» *Journal of the American Dietetic Association,* τόμ. 104, αρ. 5, pp. 804-806, 2004.

[13] R. Johnson, R. Soultanakis και D. Matthews, «Literacy and body fatness are associated with underreporting of energy intake in US low-income women using the multiple-pass 24-hour recall: a doubly labeled water study.,» *Journal of the American Dietetic Association,* τόμ. 98, αρ. 10, pp. 1136-1140, 1998.

[14] P.-Y. Chi, J.-H. Chen, H.-H. Chu και J.-L. Lo, «Enabling Calorie-Aware Cooking in a Smart Kitchen,» σε *International Conference on Persuasive Technology*, Oulu, Finland, 2008.

[15] J. Nishimura και T. Kuroda, «Human action recogintion using wireless wearable in-ear microphone,» *IEEJ Transactions on Electronics, Information and Systems,* τόμ. 131, αρ. 9, pp. 1570-1576, 2011.

[16] K.-h. Chang, S.-y. Liu, H.-h. Chu, J. Y.-j. Hsu, C. Chen, T.-y. Lin, C.-y. Chen και P. Huang, «The Diet-Aware Dining Table: Observing Dietary Behaviors over a Tabletop Surface,» σε *International Conference on Pervasive Computing*, Dublin, 2006.

[17] [Ηλεκτρονικό]. Available: http://helabe.com.

[18] [Ηλεκτρονικό]. Available: http://www.tellspecopedia.com/.

[19] «SCiO,» [Ηλεκτρονικό]. Available: https://www.consumerphysics.com/myscio/.

[20] S. Shirmohammadi και A. Ferrero, «Camera as the instrument: the rising trend of vision based measurement,» *IEEE Instrumentation and Measurement Magazine,* τόμ. 17, αρ. 3, pp. 41-47, 2014.

[21] P. Pouladzadeh, S. Shirmohammadi και R. Al-Maghrabi, «Measuring calorie and nutrition from food image,» *IEEE Transactions on Instrumentation and Measurement,* τόμ. 63, αρ. 8, pp. 1947-1956, 2014.

[22] P. Kuhad, A. Yassine και S. Shirmohammadi, «Using Distance Estimation and Deep Learning to Simplify Calibration in Food Calorie Measurement,» σε *IEEE International Conference on Computational Intelligence and Virtual Environments for Measurement Systems and Applications*, Shenzhen, China, 2015.

[23] J. Yang και W. Wu, «Fast food recognition from videos of eating for calorie estimation,» σε *IEEE International Conference on Multimedia and Expo*, New York, 2009.

[24] J. Kim και M. Boutin, «Estimating the Nutrient Content of Commercial Foods from their Label Using Numerical Optimization,» σε *New Trends in Image Analysis and Processing -- ICIAP 2015 Workshops*, Genoa, Italy, 2015.

[25] L. R. Young και M. Nestle, «The Contribution of Expanding Portion Sizes to the US Obesity Epidemic,» *American Journal of Public Health,* τόμ. 92, αρ. 2, pp. 246-249, 2002.

[26] S. M. Rebro, R. E. Patterson, A. R. Kristal και C. L. Cheney, «The effect of keeping food records on eating patterns,» *Journal of The American Dietetic Association,* τόμ. 98, αρ. 10, pp. 1163-1165, 1998.

[27] F. Takeda, K. Kumada και M. Takara, «Dish extraction method with neural network for food intake measuring system on medical use,» σε *Computational Intelligence for Measurement Systems and Applications*, 2003.

[28] F. Zhu, M. Bosch, N. Khanna, C. J. Boushey και E. J. Delp, «Multiple Hypotheses Image Segmentation and Classification With Application to Dietary Assessment,» *IEEE Journal of Biomedical and Health Informatics,* τόμ. 19, αρ. 1, pp. 377-389, 2015.

[29] J. Baxter, «Food Recognition using Ingredient-Level Features,» [Ηλεκτρονικό]. Available: http://jaybaxter.net/6869_food_project.pdf.

[30] Y. Kawano και K. Yanai, «FoodCam: A Real-Time Mobile Food Recognition System Employing Fisher Vector,» σε *International Conference on Multimedia Modeling*, Dublin, Ireland, 2014.

[31] Y. Kawano και K. Yanai, «Rapid Mobile Object Recognition Using Fisher Vector,» σε *2nd IAPR Asian Conference on Pattern Recognition (ACPR)*, Okinawa, Japan, 2013.

[32] Y. Wang, Y. He, F. Zhu, C. Boushey και E. Delp, «The Use of Temporal Information in Food Image Analysis,» σε *ICIAP 2015: New Trends in Image Analysis and Processing -- ICIAP 2015 Workshops*, Genoa, Italy, 2015.

[33] Y. Low, J. Gonzalez, A. Kyrola, D. Bickson, C. Guestrin και J. M. Hellerstein, «Distributed GraphLab: A Framework for Machine Learning in the Cloud,» σε *Proceedings of the VLDB Endowment (PVLDB), Vol. 5, No. 8, pp. 716-727*, Istanbul, Turkey, 2012.

[34] H. Kagaya, K. Aizawa και M. Ogawa, «Food Detection and Recognition Using Convolutional Neural Network,» σε *Association for Computing Machinery Multimedia*, Orlando, Florida, 2014.

[35] X. Wang, D. Kumar, N. Thome, M. Cord και F. Precioso, «Recipe recognition with large multimodal food dataset,» σε *IEEE International Conference on Multimedia & Expo Workshops (ICMEW)*, Torino, Italy, 2015.

[36] S. Christodoulidis, M. Anthimopoulos και S. Mougiakakou, «Food Recognition for Dietary Assessment Using Deep Convolutional Neural Networks,» σε *New Trends in Image Analysis and Processing -- ICIAP 2015 Workshops*, Genoa, Italy, 2015.

[37] S. Ao και C. X. Ling, «Adapting New Categories for Food Recognition with Deep Representation,» σε *IEEE Internation Conference on Data Mining Workshops*, Atlantinc City, NJ, USA, 2015.

[38] P. Pouladzadeh, P. Kuhad, S. V. B. Peddi, A. Yassine και S. Shirmohammadi, «Food Calorie Measurement Using Deep Learning Neural Network,» σε *IEEE International Instrumentation and Measurement Technology*, Taipei, Taiwan, 2016.

[39] A. Myers, N. Johnston, V. Rathod και K. Murphy, «Im2Calories: Towards an Automated Mobile Vision Food Diary,» σε *IEEE International Conference on Computer Vision (ICCV)*, Santiago, Chile, 2015.

[40] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke και A. Rabinovich, «Going deeper with convolutions,» σε *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Boston, 2015.

[41] K. He, X. Zhang, S. Ren και J. Sun, «Deep Residual Learning for Image Recognition,» 2015.

[42] Δ. Κ, Τεχνητά Νευρωνικά Δίκτυα, Κλειδάριθμος, 2017.

[43] S. Herculano-Houzel, «The human brain in numbers: A linearly scaled-up primate brain,» *Frontiers in Human Neuroscience,* τόμ. 3, αρ. 31, 2009.

[44] W. McCoulloch και W. Pitts, «A logical calculus and the ideas immanent in the nervous activity,» *Bulletin of Mathematical Biophysics,* τόμ. 5, pp. 115-133, 1943.

[45] I. Goodfellow, Y. Bengio και A. Courville, Deep Learning, MIT Press, 2016.

[46] Y. LeCun, «Generalization and network design strategies,» 1989.

[47] D. Hubel και T. Wiesel, «Receptive fields and functional architecture of monkey striate cortex,» *Journal of Physiology,* τόμ. 195, pp. 215-243, 1968.

[48] F. Kunihiko, «Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position,» *Biological Cybernetics,* τόμ. 36, αρ. 4, pp. 193-202, 1980.

[49] A. Karpathy και J. Johnson, «Course Notes on CS231n: Convolutional Neural Networks for Visual Recognition,» [Ηλεκτρονικό]. Available: https://cs231n.github.io/.

[50] K. He, X. Zhang, S. Ren και J. Sun, «Delving Deep into Rectifiers: Surpassing Human-Level Performance on ImageNet Classification,» σε *Proceedings of IEEE International Conference on Computer Vision*, 2015.

[51] I. J. Goodfellow, D. Warde-Farley, M. Mirza, A. Courville και Y. Bengio, «Maxout Networks,» σε *International Conference on Machine Learning*, Atlanta, USA, 2013.

[52] D.-A. Clevert, T. Unterthiner και S. Hochreiter, «Fast and Accurate Deep Network Learning by Exponential Linear Units (ELUs),» σε *International Conference on Learning Representations*, San Juan, Puerto Rico, 2016.

[53] A. Krizhevsky, I. Sutskever και G. E. Hinton, «ImageNet Classification with Deep Convolutional Neural Networks,» σε *Advances in Neural Information Processing Systems 25 (NIPS 2012)*, Lake Tahoe, Nevada, 2012.

[54] S. Ioffe και C. Szegedy, «Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift,» σε *Proceedings of the 32nd International Conference on Machine Learning (ICML-15)*, Lille, France, 2015.

[55] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A. C. Berg και L. Fei-Fei, «ImageNet Large Scale Visual Recognition Challenge,» *International Journal of Computer Vision,* τόμ. 115, αρ. 3, pp. 211-252, 2014.

[56] A. Karpathy, «What I learned from competing against a ConvNet on ImageNet,» 2 September 2014. [Ηλεκτρονικό]. Available: https://karpathy.github.io/2014/09/02/what-i-learned-from-competing-against-a-convnet-on-imagenet/.

[57] F. N. Iandola, K. Ashraf, M. W. Moskewicz και K. Keutzer, «FireCaffe: near-linear acceleration of deep neural network training on compute clusters,» 2015.

[58] M. Lin, Q. Chen και S. Yan, «Network in Network,» σε *International Conference on Learning Represenations 2014*, Banff, Canada, 2014.

[59] K. Simonyan και A. Zisserman, «Very Deep Convolutional Networks for Large-Scale Image Recognition,» 2014.

[60] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke και A. Rabinovich, «Going Deeper with Convolutions,» σε *IEEE Conference on Computer Vision and Pattern Recognition*, Boston, MA, USA, 2015.

[61] ImageNet, «ILSVRC 2015 Competition,» 2015. [Ηλεκτρονικό]. Available: http://image-net.org/challenges/LSVRC/2015/.

[62] COCO - Common Objects In Context, «COCO 2015 Detection Challenge,» 2015. [Ηλεκτρονικό]. Available: http://mscoco.org/dataset/#detections-challenge2015.

[63] K. He, X. Zhang, S. Ren και J. Sun, «Deep Residual Learning for Image Recognition,» 2015.

[64] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens και Z. Wojna, «Rethinking the Inception Architecture for Computer Vision,» 2015.

[65] C. Szegedy, S. Ioffe, V. Vanhoucke και A. Alemi, «Inception-v4, Inception-ResNet and the impact of residual connections on learning,» 2016.

[66] Google, «TensorFlow,» Google, [Ηλεκτρονικό]. Available: https://www.tensorflow.org/.

[67] «Theano,» [Ηλεκτρονικό]. Available: http://deeplearning.net/software/theano/.

[68] «Keras,» [Ηλεκτρονικό]. Available: https://keras.io/.

[69] «Caffe,» [Ηλεκτρονικό]. Available: http://caffe.berkeleyvision.org/.

[70] «DeepLearning for Java,» [Ηλεκτρονικό]. Available: https://deeplearning4j.org/.

[71] «MXNet,» [Ηλεκτρονικό]. Available: https://mxnet.incubator.apache.org/.

[72] A. Vidaldi και K. Lenc, «MatConvNet - Convolutional Neural Networks for MATLAB,» σε *Proc. of the ACM Int. Conf. on Multimedia*, 2015.

[73] A. Vidaldi και K. Lenc, «MatConvNet: CNNs for MATLAB,» The MatConvNet Team, 2014. [Ηλεκτρονικό]. Available: http://www.vlfeat.org/matconvnet/.

[74] A. Vedaldi, K. Lenc και A. Gupta, «MatConvNet - Manual,» [Ηλεκτρονικό]. Available: http://www.vlfeat.org/matconvnet/matconvnet-manual.pdf.

[75] «NetScope,» [Ηλεκτρονικό]. Available: https://ethereon.github.io/netscope/quickstart.html.

[76] «ResNet-50 NetScope,» [Ηλεκτρονικό]. Available: https://ethereon.github.io/netscope/#/gist/db945b393d40bfa26006.

[77] R. Collobert, C. Farabet, K. Kanukcuoglu και S. Chintala, «Torch | Scientific Computing for LuaJIT,» [Ηλεκτρονικό]. Available: http://torch.ch/.

[78] R. Collobert, K. Kavukcuoglu και C. Farabet, «Torch7: A Matlab-like Environment for Machine Learning,» σε *BigLearn, NIPS Workshop*, 2011.

[79] R. Ierusalimschy, W. Celes και L. Henrique de Figueiredo. [Ηλεκτρονικό]. Available: https://www.lua.org/about.html.

[80] L. Torrey και J. Shavlik, «Transfer Learning,» σε *Handbook of Research on Machine Learning Applications*, IGI Global, 2009.

[81] L. Y. Pratt, «Discriminability-based transfer between neural networks,» σε *Advances in Neyral Information Processing Systems 5*, San Francisco, CA, USA, 1993.

[82] S. Thrun και L. Pratt, Learning to Learn, 1998.

[83] R. Curana, «Multitask Learning,» *Machine Learning,* τόμ. 28, αρ. 1, pp. 41-75, 1997.

[84] R. Caruana, «Multitask Learning,» σε *Learning to Learn*, 1998, pp. 95-133.

[85] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li και L. Fei-Fei, «ImageNet: A Large-Scale Hierarchical Image Database,» σε *IEEE Computer Vision and Pattern Recognition (CVPR)*, 2009.

[86] I. Sutskever, J. Martens, G. Dahl και G. Hinton, «On the importance of initialization and momentum in deep learning,» σε *30th International Conference on Machine Learning (ICML 2013)*, Atlanta, USA, 2013.

[87] M. Chen, K. Dhingra, W. Wu, L. Yang, R. Sukthankar και J. Yang, «PFID: Pittsburgh Fast-Food Image Dataset,» σε *16th IEEE International Conference on Image Processing (ICIP 2009)*, Cairo, Egypt, 2009.

[88] M. Cheng, R. Sukthankar, D. Pomerleau, C. Helfrich, J. Yang, W. Wu, L. Yang, F. Kraus, A. Wang και K. D. Dhingra, «PFID,» [Ηλεκτρονικό]. Available: http://pfid.rit.albany.edu/.

[89] H. Hoashi, Y. Matsuda και K. Yanai, «Recognition of Multiple-Food Images by Detecting Candidate Regions,» σε *Proc. of IEEE International Conference on Multimedia and Expo (ICME)*, 2012.

[90] Y. Kawano και K. Yanai, «FoodCam: A Real-time Food Recognition System on a Smartphone,» *Multimedia Tools and Applications,* τόμ. 74, αρ. 14, pp. 5263-5287, 2015.

[91] Y. Kawano και K. Yanai, «Food Image Recognition with Deep Convolutional Features,» σε *Proc. of ACM UbiComp Workshop on Cooking and Eating Activities (CEA)*, 2014.

[92] EPFL - MMSPG, «Food Image Dataset,» [Ηλεκτρονικό]. Available: http://mmspg.epfl.ch/food-image-datasets.

[93] K. He, «Deep Residual Networks: Deep Learning Gets Way Deeper - ICML Tutorials 2016,» 19 June 2016. [Ηλεκτρονικό]. Available: http://icml.cc/2016/tutorials/icml2016_tutorial_deep_residual_networks_kaiminghe.pdf.

[94] Y. LeCun, L. Bottu, G. B. Orr και K.-R. Muller, «Efficient backprop,» σε *Neural Networks: Tricks of the Trade*, 1998, pp. 9-50.

[95] X. Glorot και Y. Bengio, «Understanding the difficulty of training deep feedforward neural networks,» σε *Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics*, Sardinia, Italy, 2010.

[96] L. Bossard, M. Guillaumin και L. Van Gool, «Food-101 -- Mining Descriminative Components with Random Forests,» σε *European Conference on Computer Vision*, Zurich, 2014.

[97] Facebook, «ResNet training in Torch,» [Ηλεκτρονικό]. Available: https://github.com/facebook/fb.resnet.torch.

[98] A. Howard, «Some Improvements on Deep Convolutional Neural Network Based Image Classification,» 2013.

[99] B. Zhou, A. Khosla, A. Lapedriza, A. Oliva και A. Torrabla, «Learning Deep Features for Discriminative Localization,» arXiv, 2015.

[100] L. Lab, «Convolutional Neural Networks (leNet) - DeepLearning 0.1 documentation,» LISA Lab, 05 Month 2017. [Ηλεκτρονικό]. Available: http://deeplearning.net/tutorial/lenet.html.

[101] P. Pouladzadeh, S. Shirmohammadi και S. Yassine, «Using graph cut segmentation for food calorie measurement,» σε *IEEE International Symposium on Medical Measurements and Applications*, Lisbon, Portugal, 2014.

[102] «CudaConvNet,» [Ηλεκτρονικό]. Available: https://code.google.com/archive/p/cuda-convnet/.

[103] «OverFeat,» [Ηλεκτρονικό]. Available: http://cilvr.nyu.edu/doku.php?id=code:start.

[104] «Torch,» [Ηλεκτρονικό]. Available: http://torch.ch/.

[105] Microsoft, «Microsoft Cognitive Toolkit,» [Ηλεκτρονικό]. Available: https://www.microsoft.com/en-us/cognitive-toolkit/.

[106] Κ. Ζαρκογιάννη, *Ευφυή Συστήματα Υποστήριξης Εξατομικευμένων Ιατρικών Αποφάσεων για τη Διαχείριση του Σακχαρώδους Διαβήτη,* 2011.

[107] K. Zarkogianni, M. Athanasiou, A. Thanopoulou και K. Nikita, «Comparison of machine learning approaches towards assessing the risk of developing Cardiovascular disease as a long-term diabetes complication,» *IEEE Journal of Biomedical and Health Informatics,* 2017.

[108] K. Dalakleidi, K. Zarkogianni, A. Thanopoulou και K. Nikita, «Comparative Assessment of Statistical and Machine Learning Techniques Towards Estimating the Risk of Developing Type 2 Diabetes and Cardiovascular Complications,» *Expert Systems,* 2017.

[109] K. Zarkogianni, E. Litsa, K. Mitsis, P. Wu, C. Kaddi, C. Cheng, M. Wang και K. Nikita, «A Review of Emerging Technologies for the Management of Diabetes Mellitus,» *IEEE Transactions on Biomedical Engineering,* τόμ. 62, αρ. 12, pp. 2735-2749, 2015.

[110] K. Zarkogianni, A. Vazeou, S. Mougiakakou, A. Prountzou και K. Nikita, «An insuling infusion advisory system based on autotuning nonlinear model-predictive control,» *IEEE Transactions on Biomedical Engineering,* τόμ. 58, αρ. 9, pp. 2467-77, 2011.

[111] S. Mougiakakou, C. Bartsocas, E. Bozas, N. Chaniotakis, D. Iliopoulou, I. Kouris, S. Pavlopoulos, S. Pavlopoulos, A. Prountzou, M. Skevofylakas, A. Tsoukalis, A. Vazeou, K. Zarkogianni και K. Nikita, «SMARTDIAB: A Communication and Information Technology Approach for the Intelligent Monitoring, Management and Follow-up of Type 1 Diabetes Patients,» *IEEE Transactions on Information Technology in Biomedicine, Special Issue: New and Emerging Trends in Bioinformatics and Bioengineering,* τόμ. 14, αρ. 3, pp. 622-633, 2010.

[112] K. Zarkogianni, K. Mitsis, M. Arredondo, G. Fico, A. Fioravanti και K. Nikita, «Neuro-Fuzzy Based Glucose Prediction Model for Patients with Type 1 Diabetes Mellitus,» σε *IEEE-EMBS International Conferences on Biomedical and Health Informatics*, 2014.

[113] K. Zarkogianni, E. Litsa, A. Vazeou και K. Nikita, «Personalized glucose-insulin metabolism model based on self-organizing maps for patients with type 1 diabetes mellitus,» σε *13th IEEE International Conference on BioInformatics and BioEngineering (BIBE 2013)*, 2013.